Homologous Control of Protein Signaling Networks.

D. Napoletani †, M. Signore‡, T. Sauer†, L. Liotta§, E. Petricoin§

Abstract
In a previous paper we introduced a method called augmented sparse reconstruction (ASR) that identifies links among nodes of ordinary differential equation networks, given a small set of observed trajectories with various initial conditions. The main purpose of that technique was to reconstruct intracellular protein signaling networks.

In this paper we show that a recursive augmented sparse reconstruction generates artificial networks that are homologous to a large, reference network, in the sense that kinase inhibition of several reactions in the network alters the trajectories of a sizable number of proteins in comparable ways for reference and reconstructed networks. We show this surprising result using a large in-silico model of the epidermal growth factor receptor (EGF-R) driven signaling cascade to generate the data used in the reconstruction algorithm.

The most significant consequence of this observed homology is that a nearly optimal combinatorial dosage of kinase inhibitors can be inferred, for many nodes, from the reconstructed network, opening the way for a variety of potential applications in personalized medicine.

Keywords: sparse network reconstructions, protein network models, signaling pathways, kinase inhibitors.

1 Introduction
One of the most intriguing and promising fields in medical research is based on the assumption that the great amount of information generated by high throughput technologies would allow us to understand cancer's complexity at various levels. In recent years, the completion of the Human Genome Project and other rapid advances in genomics have led to increasing anticipation of an
era of genomic and personalized medicine, in which an individual’s health is optimized through the use of all available patient data, including data on the individual’s genome and its downstream products.

Because variations in individuals’ genetic profiles oftentimes correlate with differences in how individuals develop diseases and respond to treatment, personalized medicine supported by genetic and genomic assays has the potential to facilitate optimal risk identification, disease screening, disease diagnosis, therapy, and monitoring [Kawamoto et al. 2009], [Willard and Ginsburg 2009], [West et al. 2006]. In addition to genomic assays, proteomic and metabolomic signatures hold great potential for serving as pillars of personalized medicine in the future [Willard et al. 2005], [Beretta 2007], [Gerszten et al. 2008], [Lewis et al. 2008], [Kaddurah-Daouk et al. 2008].

While personalized medicine guided by genomics is still in early stages of development, individuals’ genetic profiles are already starting to be used to guide patient care. As some examples, clinicians can obtain gene expression profiles of breast cancer samples to guide management [Paik et al. 2006], genotypes of HIV samples to identify the optimal antiretroviral regimen [Blum et al. 2005], and genetic profiles of patients’ cytochrome P450 drug metabolizing system to guide the selection and dosing of pharmacotherapies [de Leon et al. 2006]. However, it is the proteins that form the actual cell signaling and metabolic networks within the cell. Indeed, for the new classes of molecular targeted inhibitors, it is the proteins that are the drug targets, not the genes, and the molecular networks are underpinned by protein and protein phosphorylation.

Personalized medicine could be directed towards the generation of protein-based molecular maps of cancer networks in order to target malignant cells in their specific and unique context. The usefulness of patient-tailored therapy comes from the potential ability to depict patient-specific molecular circuitries and hence translate each targeted treatment in a favorable clinical response [Petricoin et al. 2005].

On the other hand, we still lack the ability to dynamically measure and collect enough data from every protein/node within networks with current methodologies. This restriction forces us a shift in mind set, in the sense that, rather than attempting a full reconstruction and understanding of cell pathways, we should search for equivalent, indistinguishable, classes of models that project to the same network structure, in the sense that such classes should ideally give rise, for each protein/node, to trajectories that are qualitatively similar even when the details of the topology of the connections among nodes differ.

Our approach to the problem of controlling protein signaling networks starts with this broad methodological assumption, but it necessarily moves further than that, since it is not yet clear what we should consider as a measure of similarity of trajectories for general, large networks. It is likely that properties such as the shape of the trajectories, the location of their maxima, as well as the value of the maxima themselves, are among the critical factors in deciding whether or not a given signaling response is triggered.

For example, activation of EGF-R, which is an upstream node and it is governed only by the kinetics and thermodynamics of EGF/EGF-receptor in-
teraction and the biochemistry of the kinase domain, is expected to be similar across cell lines. In contrast, signals further down the cascade are modulated by many upstream proteins, many of whose concentrations and rate constants impact on the overall output.

Despite this complex network behavior, there is a strong correlation between specific cell functions and the maximum concentration of key proteins known to be involved in cell growth, proliferation, survival and death, suggesting that suppression/enhancement of the activity of specific nodes can be seen potentially as a way to achieve the final goal of disruptively interfering with the functioning of cancer cells.

The trajectories for each node are usually generated in vitro by stimulation of cell lines and subsequent relaxation to steady state, so that the extent of suppression of a node activity can be determined by looking at the maximum value of a relatively simple curve.

Even though the maximum activity of nodes is only one of many key features of signaling networks, its accurate modification is by no means an easy task. Agents directed at an individual target in the network frequently show limited efficacy, poor safety and resistance profiles, which are often due to factors such as network robustness, redundancy, crosstalk, compensatory and neutralizing actions and anti-target and counter-target activities.

The ability to predict in silico the sensitivity of cancer cells to the inhibition of multiple reactions would allow us to combine drugs in order to achieve synergy and/or potentiation of several orders of magnitude, while avoiding undesired effects on normal cells. Systems-oriented approaches has already yielded several clinical successes and drug-discovery efforts are now focused towards near optimal combinatorial treatments that target cell pathways at several sites.

Because the fundamental goal of a combinatorial approach to cancer therapy is the control of the activity of specific nodes in the network, we use it to define an operative notion of homologous networks. We select a target node $N$, and a set of reactions $P$, and we assume the following definition of homology of networks: two networks are homologous (with respect to $N$ and $P$) if the activity of node $N$ reacts in a similar way to the suppression of the given reactions $P$ performed by known kinases.

Note that this comparison can be made on very long time scales, ideally on time intervals where the networks have each relaxed to the steady state, so that the comparison of the networks can be considered global.

**Remark 1:** In Section 3 we define more formally similarity as the concordance of the relative magnitude of the maximum difference of the trajectories of the node $N$, starting from equal initial conditions, when control of the reactions in the networks, via kinase inhibition, is on, and when control is off. In this way, we have a simple, even though partial, way to determine how close two networks will react, for specific nodes, to similar control schemes.

In a previous paper [Napoletani et al. 2008] we introduced a method called augmented sparse reconstruction (ASR) that identifies links among nodes of ordinary differential equation (ODE) networks, given a small set of observed trajectories with various initial conditions. The main purpose of that technique
was to reconstruct intracellular protein signaling networks under the assumption that most nodes interact with only a small fraction of the total number of nodes in the network. We say in that case that the network is sparse and such information can greatly help in reconstructing the network itself.

In this paper we show that augmented sparse reconstruction generates artificial networks that are homologous to the initial network, in the sense that kinase inhibition of several reactions in the network alters the trajectories of a sizable number of proteins in comparable ways. We show this surprising result using an in silico model of the epidermal growth factor receptor (EGF-R) driven signaling cascade to generate the initial observed trajectories.

The most significant consequence of this observed homology is that the optimal combinatorial dosage of kinase inhibitors can be inferred in many cases from the reconstructed network \(^1\). This result could be of great value for a variety of applications in personalized medicine.

While there have been successful attempts to derive network models from a limited number of perturbation experiments (see for example the recent works [Nelander et al. 2008], [Munsky et al. 2009]), we stress that our method achieves a degree of reconstruction and dynamical control for nodes of a network whose size far exceeds those tested so far in the literature, with the exception of the very interesting work in [Chang et al. 2009], which uses sparsity in an essential way, but that builds only a static model rather than a dynamical one. A significant amount of information can be inferred by static analysis, however a full network control in non-stationary conditions can only be achieved in a dynamical setting.

In section 2 we will show how the algorithm we described in [Napoletani et al. 2008] needs to be modified to take into consideration knowledge of specific reactions that can be inhibited. Section 3 is dedicated to comparison between an initial network and partially homologous networks obtained from its trajectories by augmented sparse reconstruction.

2 Matching Pursuit for Augmented Sparse Reconstruction

Because of our very limited understanding of the changes of dynamics in large, perturbed networks (except in those cases when the parameters of the model of individual nodes are only slightly perturbed), it is daunting to set up an homologous network from first principles, given a reference network.

We believe that the right approach to generate homologous networks is to directly use the state space, in the sense that by modifying or restricting information on the trajectories, we can use reconstruction methods to give candidate homologous networks of a reference network. Effectively, this is a signal processing approach to network dynamics: optimal signal representation of the

\(^1\)A patent application has been filed for the methods described in this paper. Provisional patent application number 61/265,815, filed on 12/02/09
trajectories becomes the main tool to explore network structure.

We select a well established model of the epidermal growth factor receptor (EGF-R) signaling pathway as reference network [Schoeberl et al. 2002]. The reason of such choice is the great importance of the epidermal growth factor receptor signaling pathway in cancer biology and the fact that it is one of the most well-studied pathways that regulate growth, survival, proliferation, and differentiation in mammalian cells [Oda et al. 2005].

In the EGF-R network, upon binding of the ligand, the receptors dimerize and phosphorylate each other, thus generating docking sites for five adaptor proteins and five enzymes. Signals from ErbBs converge to molecules forming a bow-tie core and are supposed to represent a versatile and conserved group of molecules and interactions. The amplitude of EGF-R cascades reaches high levels within minutes of stimulus and an important role is played by the recycling mechanism of receptor molecules after signal transduction, so that, in the absence of EGF molecules the system relaxes back to steady state, in line with the generic description of trajectories put forward in Section 1. The four human ErbB receptors induce a wide variety of cellular responses thereby generating a complex protein interaction network [Jones et al. 2006].

Due to its properties and involvement in tumor progression, the EGF-R network inspired several experimental and mathematical modeling studies [Birtwistle et al. 2007]; [Uetz and Stagljar 2006]. Deregulation of EGF-R signaling plays a key role in numerous cancers, including glioblastomas, breast cancer, and nonsmall cell lung cancer (NSCLC) [Kumar et al. 2008].

Another reason to choose the EGF-R pathway as a reference network is that, despite the fact that various agents have been developed to target EGF-R, there is a need for improved strategies to integrate anti-EGF-R agents with conventional therapies and to explore combinations with other molecular targets [Baselga and Arteaga 2005].

In this work we use the differential equation model of EGF-R network put forward in [Schoeberl et al. 2002] and [Hornberg et al. 2005]. This model assumes only linear and quadratic terms in the representation of the derivative of the activity of each node of the network. Linear terms correspond to unimolecular interactions and quadratic terms correspond to bimolecular interactions.

From the computational point of view, an important feature of this network is its large size (103 variables and 148 distinct reactions). Most reconstruction techniques are not able to deal with the reconstruction and control of very large networks, if the experimental data are limited and noisy, and yet this is exactly the size of networks that are of interest when exploring pathways that may not be well understood.

In equation (1) we show the model of the network at a node \( n \), in the specific integral form that is used in augmented sparse reconstruction; for a complete analysis of this integral model we refer to [Napoletani et al. 2008]. Essentially, equation (1) is nothing else than the integral of a differential equation with linear and quadratic terms, and with added random terms to make sure the reconstruction algorithm is able to eliminate errors-in-variables due especially
to the presence of non-linear terms.

\[ x_n(t) - x_n(t_0) = a_{0n} + \sum_{i=1}^{N} l_{in} \int_{t_0}^{t} x_i dt + \beta_q \sum_{i=1}^{N} \sum_{j=1}^{N} q_{ijn} \int_{t_0}^{t} x_i x_j dt \]

(1)

Here the \( \beta_q \leq 1 \) represent positive attenuation coefficients for the quadratic terms. The systems parameters at node \( n \) that we need to determine are: \( a_{0n}, l_{in}, i = 1, ..., N, q_{ijn}, i, j = 1, ..., N \). The \( n_g, g = 1, ..., G \) are discrete random vectors normally distributed, scaled to have norm 1 and multiplied by suitable parameters \( w_{gn} \) to be determined together with the system parameters.

The reconstruction algorithm of [Napoletani et al. 2008] assumes sparsity of the network, i.e. we assume that each node interacts with only a small number of nodes compared with the total of possible nodes. This assumption implies that the number of terms in each equation in (1) with nonzero parameters is small compared to the total possible number of terms.

Sparsity plays a crucial role in our method, since it allows to use fast linear programming techniques in looking for the optimal model that has as few terms as possible [Chen et al. 1998], but just as important for network reconstruction is the fact that our method avoids a direct estimate of the derivative of the trajectories, and that we augment the model with random terms. Despite these adjustments, the quality of the reconstruction worsen for nodes with many links, even when the total number of nonzero terms in equation (1) is low compared to the total number of possible terms.

Though sparsity methods are very powerful, when properly adapted to networks, and they allow for significant inference of the network under very limited and noisy data, it is unlikely that they, or any other currently known methods, will fully reconstruct the network structure from very limited, coarse data.

Despite these limitations, the fundamental claim of this work is that we can have homologous control despite our inability to gain full reconstruction of the topology of a network. This claim is intrinsically related, in ways that still need to be explored, to a fundamental assumption of systems biology, i.e. the belief that biological networks are robust under variations of the strength and type of connections of the signaling pathways.

Robustness seems to be a consequence of several recurrent factors, for example the bow-tie architecture (or hourglass structure) of the EGF-R network is considered a characteristic feature for robust evolvable systems [Kitano 2004]. Another important feature of robust biological networks is the fact that they show a diverse array of molecules for input and output, that are connected to the conserved core of the network with highly redundant and extensively crosstalking pathways and feedback control loops in various places in the pathway.

If the assumption of robustness is correct for most biological networks, augmented sparse reconstruction may not recover the exact network, but it may be
sufficiently accurate to infer a network that is homologous to the original one. We will see that this possibility is realized for our EGF-R reference model.

In this work we assume that specific reactions must be present in the reconstruction of the network, since we define homologous systems with respect to the action of kinase inhibitors. In [Napoletani et al. 2008] there was no such constraint, therefore our main objective in this section is to adapt the algorithm developed in that work in such a way that it guarantees the presence of specific reactions to be targeted with available kinase inhibitors.

Signal processing sparsity methods, that are at the core of augmented sparse reconstruction, are not able to guarantee the presence of these individual reactions, since they are more concerned with global optimality of the representation of each node. We need therefore an adaptive, recursive augmented reconstruction algorithm to extract the few terms in each equation due to the chosen reactions, before we apply the augmented sparse reconstruction algorithm to the whole representation system.

To understand the details of the reconstruction algorithm, we first recall how individual reactions are put together in a modular way to generate systems of differential equations describing the network of [Schoeberl et al. 2002], [Hornberg et al. 2005].

Suppose that phosphorylated proteins $x_i$ and $x_j$ are interacting to phosphorylate protein $x_k$, and in the process they get de-phosphorylated; this specific reaction can be modelled [Schoeberl et al. 2002], [Voit 2000] as $v = ax_i x_j - bx_k$. Its effects on the differential equations of the network are as follows: if we let $\dot{x}_i = f_i(x_1, ..., x_n)$, i.e. if we model the derivative of the phosphorylated concentration of protein $x_i$ as a function of the state of (possibly) all proteins, then, because the reaction de-phosphorylate $x_i$, then $\dot{x}_i = f_i(x_1, ..., x_n) - v$, and similarly $\dot{x}_j = f_j(x_1, ..., x_n) - v$. On the contrary, since the specific reaction increases the phosphorylation of $x_k$, we will have $\dot{x}_k = f_k(x_1, ..., x_n) + v$.

The immediate consequence of this modelling assumption is that if we know that a simple quadratic reaction $v = ax_i x_j - bx_k$ is involved in a network, then we know that the representation of the derivative of $x_i, x_j, x_k$ will have a specific quadratic and a specific linear term in the representation in (1). Only the parameters of these terms will be unknown.

The more reactions we make available as targets of kinase inhibition, the more indirect information we have about the details of the terms of the model. In many models it is possible as well that the algebraic form of the reaction is $v = ax_i x_j - bx_k x_h$, this does not affect the modular building of the network, or our approach, but only the range of proteins affected.

One way to model kinase inhibition (see [Hornberg et al. 2005]) is to assume suppression of a target reaction $v$, i.e. $v$ will appear in the representation of the derivatives of the relevant proteins concentrations multiplied by a kinase suppression coefficient $\kappa < 1$.

Most kinase inhibitors discovered to date are ATP competitive and present one to three hydrogen bonds to the amino acids located in the hinge region of the target kinase, thereby mimicking the hydrogen bonds that are normally formed by the adenine ring of ATP [Zhang et al. 2009].
Often kinase inhibitors cross-react, either, with various degrees of specificity, with other kinases among the 518 encoded in the human genome, or with the abundant nucleotide-binding enzymes that are present inside cells. The degree of kinase inhibitors selectivity depends on many factors such as their concentration and cellular context. Biochemical and cellular assays are available for the dissection of the specificity range of small molecules for various kinases, but to date the evaluation of kinase inhibitor selectivity on an organismal level remains a significant research challenge [Zhang et al. 2009].

Different models of inhibition of reactions can be easily implemented in our method, for example forward rate kinase inhibition and backward rate kinase inhibition would require $\kappa$ to act only on the first or second term of the reaction respectively. Since changes in $\kappa$ in general do not affect to a large extent the activity of any given individual node, in Section 3 we use $\kappa = 0.1$. Such value of $\kappa$ allows for detectable changes of trajectories, but it may lead in real systems to unspecific inhibition.

Our discussion up to this point clarifies how the knowledge of a reaction in the system can be used to build specific building blocks in selected differential equations of the unknown model. Next, we write down the heuristic description of a modified augmented reconstruction algorithm that includes the known reactions in its structure. The details of the algorithm are given in the appendix.

**Modified ASR Algorithm**

Select a collection of potential target reactions $v_s = a_s x_i x_j - b_s x_k$, $s = 1, \ldots, S$.

Given the collection of all time measurements for each node $n$ with $n = 1, \ldots, N$:

1. **R1** Set up a representation matrix $Z$ where each column corresponds to a term of the right hand side of equation (1) (constant, linear, quadratic and random). Set up a vector $Y_n$ that corresponds to the left hand side of (1).

2. **R2** Select the columns of $Z$ that correspond to the target reactions involved in the activity of the given node $n$.

3. **R3** Perform augmented sparse reconstruction for $Y_n$ using only the columns selected in step R2 to force those terms to have large parameters in the overall representation. Subtract the contribution of the target terms from the vector $Y_n$.

4. **R4** Perform augmented sparse reconstruction for the modified $Y_n$ using the full representation matrix $Z$. Add back the the parameters of the target terms found in the previous step to the corresponding parameters found with the full matrix $Z$.

5. **R5** Choose a threshold $T_n$. The reconstructed network equation for node $n$ will have only linear and quadratic terms that correspond to parameters larger than $T_n$. 


The modified ASR algorithm \textbf{R1-R5} generalizes the algorithm in [Napoletani et al. 2008] in such a way that, for each node, a preliminary augmented sparse reconstruction is performed only on the terms that are related to the reactions we selected as potential kinase targets, if they have an impact on that node. After this preliminary step, a full augmented sparse reconstruction is performed with all potential linear and quadratic terms.

Note that we output a full model from the algorithm, rather than a list of directed links to each node. This puts us in the position of testing our conjecture that the augmented sparse reconstruction of a network can be homologous to the reference network.

\section{Homologous Control}

With the algorithm described in the previous section, we can produce a reconstructed network model of the reference EGF-R network that is very likely to include the reactions \( v_s = \alpha_s x_i x_j - b_s x_k, \ s = 1, \ldots, S \) that we want to target with kinase inhibitors with nonzero parameters \( \alpha_s \) and \( b_s \).

Throughout this whole section it is assumed that we have 19 different noisy initial conditions for the network, and that we sample each trajectory at 11 points in the interval \([0, 11]\), with time measured in minutes. This interval is acceptable because, after \( t = 11 \), most nodes relax back to their steady state and they do not contribute to the understanding of the dynamics.

Noise level for each trajectory is assumed to be at most 10\% of the maximum value of the points along the trajectory itself. This setting gives us \( 19 \times 11 = 209 \) data points for each of the 103 nodes in the network.

This number of data points is very small from the data mining perspective, but it is large from the experimental, \textit{in vitro} point of view. However, it is within the limits of current experimental practice for reverse phase protein arrays, see for example the study in [Nishizuka and Spurrier 2008] or [Nishizuka et al. 2008]\textsuperscript{2}.

If a specific experimental setting, or cell lineage, does not activate some pathways, it is likely that those pathways will not be detected, and their contribution to kinase inhibition will not be measurable. Therefore, initial conditions should ideally be able to explore as much of the dynamical range of the reference network as possible. At the same time, we want only to detect homology according to our definition in Section 1, and we argue in this section that it is possible to generate an homologous reconstructed network using modified ASR even when only very limited data are available.

The data points used for the modified ASR algorithm are generated from the reference EGF-R network in [ Schoeberl et al. 2002]\textsuperscript{3}. The choice of meaningful initial conditions is complicated by the fact that the copy numbers of individual

\footnotesize{\textsuperscript{2}We mention here as well our work in preparation on adult stem cells, where the aggregate data set generate by RPPA encompasses about 300 data points per node: Functional Protein Network Activation Mapping of Adult Mesenchymal Stem Cells Differentiation, B. McCloud, L. Liotta, E. Petricoin.}

\footnotesize{\textsuperscript{3}The values of the trajectories are known to be positive, and this is enforced, for the reconstructed network, in the discrete differential equation solver.}
proteins vary enormously, and protein concentration varies with cell type and cell cycle stage, from less than 20000 molecules per cell for the rarest types to 100 million copies for the most common ones. In the average mammalian cell some 2000 proteins are considered to be relatively abundant [Alberts et al. 1994], [Lodish et al. 2000].

On the basis of these broad considerations, the initial conditions for the reference network are chosen to be random values uniformly distributed in the interval [2000,20000]. These values are assumed to be the average number of copies of molecules per cell. The average number of EGF receptors is taken to be a random value in the interval [1000,10000]. EGF is selected to be a random value in the interval [$10^{-8}$, $10^{-7}$], as this is a compound outside the cell and we measure it in gr/ml.

**Remark 2:** Even though this range of concentrations of each protein is reasonable, the choice of initial conditions is still not necessarily biologically meaningful, since the relative distribution of the initial conditions of the nodes with respect to each other is randomly selected. Yet we believe that our choice reflects two basic assumptions that are necessary for the success of our method, and that are indeed biologically meaningful: strong variability of trajectories, and measurable dynamical changes.

**Remark 3:** Note moreover that the number of initial conditions used for the reconstruction is very small compared with the volume of potential initial conditions, so that we are severely undersampling the space of allowed initial conditions. Yet, we will see in the following that these small data sets have predictive power when used to infer the degree of inhibition of nodes with other initial conditions. Therefore it seems that homologous control has some robustness with respect to potential variations in the experimental setting, so that not every context relevant to a specific study need to be probed for the application of our methodology.

For clinical purposes, not all nodes are of interest. For example in the reference model of EGF-R, $x_{31}$ (doubly phosphorylated MEK) and $x_{59}$ (doubly phosphorylated ERK) are particularly significant targets [Schoeberl et al. 2002]. In this paper we are interested in showing the global effect of a wide choice of kinase inhibitors on the ensemble of all network nodes, to determine a global measure of homology for the reference and reconstructed networks, with limited data available for the reconstruction of the homologous network.

To check the global effect of kinase inhibitors on the network, we select 19 reactions to target with inhibitors, on the basis of their position along the EGF-R signaling cascade, in order to comprise both upstream and downstream molecular events that span the entire signal transduction cascade from top (EGF-R docking sites) to bottom (MEK or ERK phosphorylation). The reactions are: v19,v66, v20,v67, v23,v70, v27,v74, v29,v76, v41,v83, v45,v87, v47,v89, v55,v97 ,v60 [Schoeberl et al. 2002].

In Figure 1 we can see an instance of the effect of kinase inhibitors on both reference and reconstructed networks. The top plot shows the change in behavior of node 66 of the reference EGF-R network when reactions v41 and v83 are inhibited with $\kappa = 0.1$. The bottom plot shows the corresponding change of
behavior for the 66th node of the reconstructed network.

The dynamics of reference and reconstructed networks is just marginally similar for the node in Figure 1. Often, the similarity of the trajectories is even less pronounced, or the change of behavior due to control is several order of magnitudes smaller for the reconstructed network. This is not surprising, since we are using an incredibly small amount of data to build the reconstructed network and we cannot expect similarity in the actual trajectories generated by this coarse approximation.

Note also that the sign of the change due to control is the opposite in the two networks. We will see that this is a common occurrence with this method. This sign switching may potentially be lessened by cross-validating parameter estimation for the different nodes affected by the targeted reactions. We stress that our goal is not exact trajectory reconstruction, but only an estimate, using the reconstructed network, of how much the magnitude of trajectories of the real network are changed when kinase inhibitors control is switched on.

To achieve this goal, we need a quantitative way to estimate the change of the trajectories due to kinase inhibitors control. More specifically:

(a) We randomly select several initial conditions in the same wide region used to generate the trajectories of the reference system. (b) We simulate reference network and reconstructed network with each of these initial conditions for a fixed length of time, long enough for most trajectories to relax to their steady state. (c) We perform these simulations with the target kinase inhibitors switched on, and then with the kinase inhibitors off. (d) For each node, we compute the maximum pointwise distance between trajectories with same initial conditions and with control on and off respectively, over the whole time interval used in the simulations. (e) The maximum pointwise distance for each node is divided by the maximum value of the trajectory of the same node in the reference network, with control switched off. (f) Finally, we measure the median of the scaled displacement for each node variable, with respect to the set of initial conditions, as a statistically significant measure of node displacement due to control.

We call the quantity generated by the procedure (a) — (f) the median scaled maximum pointwise displacement of trajectories of a node due to the specific choice of control kinase inhibitors, sometimes we will refer to this quantity simply as median scaled displacement.

**Remark 4:** We consider the median since most of the time the scaled displacement has very similar behavior for most initial conditions, but there is sometimes a small number of initial conditions that may lead to divergent trajectories for some nodes in the reconstructed system, making the mean too sensitive to these outliers.

In Figure 2 we plot the magnitude of the median scaled pointwise displacement of node 38 with respect to an ordering of the set of 190 different kinase inhibitor coefficients where each \( \kappa_s, s = 1, \ldots, 19 \) is allowed to be either 1 (no inhibition) or 0.1 (high kinase inhibition), and only one or at most two reactions at a time are inhibited.

This combinatorial constraint is in line with current experimental protocols,
migure Ua

An instance of relative concordance of the magnitude of displacement of trajectories due to a choice of kinase inhibitors. Dashed curves are the trajectories with inhibitors on, while solid curves are the trajectories without inhibitors. The top plot shows the effect of kinase inhibitors on the 66th node of the reference network. The bottom plot shows the trajectories with identical initial conditions for the corresponding node of the reference network.

To allow only two kinase inhibitor coefficients to be different from one for each possible combination. In this scenario an important problem is to find the optimal two kinase inhibitors to choose from a possibly large collection of inhibitors.

Note that generally a kinase inhibitor is considered useful if it changes the phosphorylation of a target by a significant amount. Our choice of $\kappa_s$ equal to 0.1 for individual reactions increases the chance that we observe relative displacements of the nodes of the reference network in the order of $10 \rightarrow 20\%$.

**Remark 5:** It is implicitly assumed in the model that the kinase inhibitors are specific at these concentrations. Notwithstanding the strict 100-fold specificity criteria used during the screening that companies usually perform for kinase inhibitors selection process, many inhibitors show various degrees of off-targets, depending either on their concentration or on the cell type. Interestingly, some approved drugs (e.g. sunitinib, dasatinib) had relatively low selectivity but are nevertheless effective for clinical use. Knowledge of target profiles should allow careful evaluation as to which drug or drug combination should be used in a particular situation to better exploit each drugs full potential [Johnson 2009].

There is a striking concordance of the shape of the two median scaled displacement curves in Figure 2, even though the magnitude for each individual kinase inhibitor combination can be vastly different and indeed the magnitude of the median scaled displacement for the reconstructed network can be far lower that the one of the reference network for many nodes, even when there is very
In this plot we show a case in which the shape of the median scaled maximum pointwise displacement curve is very similar for reference and reconstructed networks. The top plot gives the absolute value of the median scaled maximum displacements for the 35th node of the reference EGF-R network. Each point on the horizontal axis corresponds to one among 190 different pairs of kinase inhibitors combinations. The bottom plot gives the absolute value of the median scaled maximum displacement curve for the same node in the reconstructed network.

This observed concordance suggests that the optimal combination of kinase inhibitors can be inferred from the median scaled displacement curve of the reconstructed network. In general, the displacement curves tend to agree only partially and only for the largest displacements values.

For example, in Figure 3 we can see concordance of the scaled displacement for many (but not all) high displacement values relative to node 38. In this Figure we do not take the absolute value of the displacement curve, to highlight the fact that often the sign of the displacement is different for the reference network and the reconstructed network, even when we have highly correlated absolute values of the displacement curves, this is potentially a problem because the sign of the displacement will determine whether the control acts as a inhibitor or a enhancer of the target node.

It is likely that the cause for the sign switching is the inability of the restricted $l_1$ optimization in step R2 of the modified ASR algorithm to detect the proper parameter in the presence of noise, even when we enforce that such parameter should be present. Regardless, the displacement curve of the reconstructed network will be able to identify combinations of kinase inhibitors that have high impact on the node.

Remark 6: Figures 2 and 3 show that there are sets of kinase inhibitors that affect both nodes in a significant way. Especially noticeable is the similar activity level of the two nodes between kinase inhibitor combinations indexed from 85 to 100. This is a general pattern that is related to the actual size of the...
In this plot we show a case in which the shape of the median scaled maximum displacement curve is very similar for reference and reconstructed networks only for very high values of displacement. The dashed starred plot gives the median scaled maximum displacements for the 38th node of the reference EGF-R network. Each point on the horizontal axis corresponds to one among 190 different pairs of kinase inhibitors combinations. The solid circled plot gives the median scaled maximum displacement curve for the same node in the reconstructed network. Curves are scaled to have absolute maximum values equal to one for comparison purposes. In this figure we do not take the absolute value of the displacement curves here. Note the uniform reversal of sign of the median displacement curves.

Figure 3: In this plot we show a case in which the shape of the median scaled maximum displacement curve is very similar for reference and reconstructed networks only for very high values of displacement. The dashed starred plot gives the median scaled maximum displacements for the 38th node of the reference EGF-R network. Each point on the horizontal axis corresponds to one among 190 different pairs of kinase inhibitors combinations. The solid circled plot gives the median scaled maximum displacement curve for the same node in the reconstructed network. Curves are scaled to have absolute maximum values equal to one for comparison purposes. In this figure we do not take the absolute value of the displacement curves here. Note the uniform reversal of sign of the median displacement curves.
parameters of individual reactions: the larger the magnitude of the parameters, the more likely the chance that they will be properly detected and therefore that they have a significant impact in the control of both the reference and reconstructed networks. We will see that this point has important consequences when we assess the performance of our method across several choices of 19 kinase inhibitors.

One difficulty summarizing the qualitative properties of displacement curves is the tremendous variability of the individual nodes. However, it is clear that sign switching and sign concordance of the displacement curves for high displacement values are both common, and not due to random effects.

For example, in Figure 4(a) we show the number of kinase inhibitor combinations, out of the total 190, that display opposite sign, as a function of a threshold on displacement curves of both reference and reconstructed networks. For each node, the threshold sets to zero all displacements that are below the percentage of the largest displacement value denoted in the ordinate axis. Red curves are for the 5 nodes that have the highest number of combinations with sign switching. Blue curves are generated in a similar fashion, but after the median scaled displacement values have been permuted within the displacement curve of each node.

Note the very distinct behavior at the 20% threshold level between true and randomized curves in Figure 4(a). Instead, Figure 5(a) shows, for each node of the reconstructed network, the ratio of nonzero median displacement values that display sign switching with respect to the reference network, in the case of a 20% threshold. Many nodes, at this threshold level, exhibit large percentages of sign switching.

Remark 7: We stress that this sign switching is by no means consistently observed for all nodes, or for specific groups of kinase inhibitor combinations. For example, the kinase inhibitor combinations corresponding to indexing from 85 to 100, shown in Figures 2 and 3 to have vigorous control on some nodes of the network, have roughly the same number of nodes that display uniform sign switch and uniform sign concordance for large displacement values. If we study sign concordance rather than sign switching, a very similar qualitative behavior as in Figures 4(a) and 5(a) is observed, as can be inferred from Figures 4(b) and 5(b).

In Figures 4(c) and 5(c) we follow essentially the same procedure as the one that we used to explore sign switching, but focusing on the concordance of nonzero displacements. In Figure 4(c) we show the number of kinase inhibitor combinations, out of the total 190, that display concordance of nonzero displacement values, as a function of the same threshold on median displacement curves of both reference and reconstructed networks used in Figures 4(a) and 4(b). Red curves are for the 5 nodes that have the highest number of combinations with corresponding nonzero median displacement values. Blue curves are generated from permuted median displacement curves.

In Figure 4(c) we see distinct behavior between real and randomized data especially around the 40% threshold level. The fact that this percentage is higher than the one observed for the analysis of signs shows that sign concordance
Figure 4: Red starred curves in Figures 4(a), 4(b), 4(c) show the number of kinase inhibitor combinations of the 5 nodes of reference and reconstructed networks that display respectively the largest: 4(a) sign switching; 4(b) sign concordance; 4(c) concordance of nonzero median displacement values. Blue starred curves are generated in a similar fashion, but after the median scaled displacement values of the reconstructed network have been permuted within the displacement curve of each node. Curves are plotted as functions of a threshold on displacement curves of both reference and reconstructed networks. For each node, the threshold sets to zero all displacements that are below the percentage (of the largest displacement value) denoted in the ordinate axis.
Figures 5(a), 5(b), 5(c) show, for each node of the reconstructed network, the ratio of nonzero median displacement values that display respectively: 5(a) sign switching at the 20% threshold level; 5(b) sign concordance at the 20% threshold level; and 5(c) concordance of nonzero displacement values at the 40% threshold level. For each node, the threshold sets to zero all displacements that are below the given percentage of the largest displacement value.
and sign switching are more statistically significant than nonzero concordance for lower threshold. Figure 5(c) shows, for each node, the ratio of nonzero displacement values for the reconstructed network that is matched by nonzero displacement values for the corresponding node of the reference network, at the 40% threshold level. Many nodes, at this threshold level, exhibit almost complete overlapping of the nonzero displacement values of reconstructed network with (some of) the nonzero displacement values of reference network.

Our analysis so far gives a sense of the distinctive homology of reconstructed and real networks. Our final goal is to use homology to obtain a nearly optimal kinase inhibitor combination, and we propose the following strategy:

Homologous Control

C1 Given trajectories in a signaling network, and a set of reactions to be inhibited, use the modified ASR algorithm in Section 2 to obtain a reconstructed, potentially homologous network.

C2 Consider a large set of kinase inhibitor combinations that satisfy some given constraint. For each kinase inhibitor combination generate trajectories of the reconstructed network for a variety of biologically meaningful initial conditions.

C3 Generate the median scaled displacement curve for a target node protein. Identify the position of the few largest values of the median scaled displacement curve for the reconstructed network. The corresponding kinase inhibitor combinations are candidates for nearly optimal suppression/enhancement of the target node in the reference network.

Several possible choices of constraints could be enforced in step C2 of this algorithm. For example, in personalized therapies we could ask for the combination of kinase inhibitors with a minimum total norm of the corresponding kinase inhibitor coefficients $\kappa$, to avoid toxicity and loss of specificity.

In the following, we continue to explore the experimental protocol used to generate Figures 1-5 in which only two reactions at the time are inhibited. Recall that since we have 19 possible kinase inhibitor targets, there are a total of 190 distinct pairs and singlets of kinase inhibitor therapies.

In our simulations there is remarkable agreement of the location of large peaks of the median displacement curves for reference and reconstructed networks, so that it seems possible to use the largest median displacement values of the reconstructed network to predict the likely location of near-optimal combinatorial kinase inhibitions.

We cannot expect full overlapping of locations of large peaks and therefore we suggest the following possible definition of near-optimality of kinase inhibitor combinations: We assume that we found near-optimal kinase inhibitor combinations if the locations of the top three maxima in the median displacement curve of a node of the reconstructed network overlaps with the location of large median scaled displacement values of the corresponding node in
the reference network. A large displacement is defined here as a value that is a large percentage, say $P = 80\%$, of the mean of the largest 3 displacement values of the given node in the reference network.

Remark 8: Note that we use, as a benchmark of success, the mean of the largest median displacement values, rather than the absolute maximum displacement. We decided for this criterion since in principle the absolute value may be so large compared to the other displacements values that there may not be other significant kinase inhibitor combinations with displacement values close to the maximum.

Since the scaled displacement curves of many nodes of the reconstructed network are not carrying any useful information on the corresponding nodes of the reference network, we need to find a way to select the most useful nodes of the reconstructed network.

Our understanding is that, if a node displays large median displacement values for a significant number of kinase inhibitor combinations, its chance of being at least partially correlated with the corresponding displacement curves of the reference network is increased. The validity of this understanding is captured in Figure 6.

In Figure 6(a), the dashed curve shows the percentage of nodes of the reconstructed network that have a number of large median displacement values above the threshold specified on the horizontal axis. We label a median displacement value as large if it is above 10% of the maximum median displacement value for the whole median displacement curve. Note the slow decay of the percentage of nodes as a function of the threshold.

The dark starred curve in Figure 6(a) shows the percentage of nodes for which we identify near-optimal kinase inhibitor combinations (as defined above), as a function of the same threshold. We reach about a 50% success rate, meaning that for 50% of the nodes with more than 22 active displacement values in the reconstructed displacement curve we can identify near-optimal kinase inhibitor combinations. The catch is of course that only a few nodes (6 of them) display such marked activity in the reconstructed median scaled displacement curve.

It could be argued that, since the cases with significant activity in the scaled displacement curve correspond often to large relative displacement curves for the reference network, they are among the few that really matter. Note that the percentage curve is not monotone, this indicates that some strongly homologous nodes are actually removed as the threshold grows. This raises the hope that even better success rates could be achieved with more sophisticated threshold processes.

The solid circled curve in the same plot 6(a) shows the average percentage of nodes for which we identify near-optimal kinase inhibitor combinations when the sequence of median scaled displacements for the reconstructed network has been randomly permuted (average computed over 100 permutations)\(^4\).

Near-optimality for the random scrambling of the data is much lower, espe-

---

\(^4\)In all plots in Figure 6 we truncate the curves when the effective number of nodes left by the threshold is below 3, to avoid fluctuations in both real and randomized data.
The plots (a), (b), (c), (d) in Figure 6 encode the degree of homologous control that can be achieved for four different selections of 19 kinase inhibitors. For each plot, the dashed curve shows the percentage of nodes of the reconstructed network that have a number of large median displacement values above the threshold specified on the horizontal axis. We label a median displacement value as large if it is above 10% of the maximum median displacement value for the whole median displacement curve. The dark starred curve shows the percentage of nodes for which we identify near-optimal kinase inhibitor combinations (as defined in the text), as a function of the same threshold. The solid circled curve shows the average percentage of nodes for which we identify near-optimal kinase inhibitor combinations when the sequence of median scaled displacements for the reconstructed network has been randomly permuted (average computed over 100 permutations).
cially for higher threshold values. Note that the way we threshold the displacement curves allows even median scaled displacement curves of nodes that are only slightly activated by the kinase combinations to retain a strong predictive value.

The mean, across all threshold values, of the percentage of nodes for which we find near-optimal combinations in the case of random permutation of the reconstructed displacement curves in Figure 6(a) is only around 6%. For nodes selected with high values of threshold, our method is about 8 times more accurate than a random selection of kinase inhibitors in finding near-optimal combinations.

To put things in perspective, with the randomly permuted scaled displacements, on average it would be necessary to select randomly 25 combinations of kinase inhibitors to make sure that we have similar success rates to those obtained with only 3 kinase inhibitor combinations by the homologous control scheme. This represents a large potential saving, due to the data-based narrowing of combinatorial possibilities, in the search for appropriate kinase inhibitor therapies.

Figure 6(b), 6(c), 6(d) show the results of the same analysis performed in Figure 6(a), for three other selections of 19 kinase inhibitors. The main criterion underlying the choice of reactions has been the availability for most of the selected reactions, of inhibitory drugs that might be used to block the downstream signaling. Several example exist that have either passed FDA approval or are currently in clinical trials, like monoclonal antibodies or small molecules inhibitors that target EGF-R (Cetuximab and Gefitinib, respectively) and small molecule inhibitors of the Raf/MEK/ERK axis (Sorafenib). The number of reactions that satisfy this experimental plausibility criterium is relatively small in the cited EGF-R network, and there is a certain degree of overlapping among these sets.

The success rate of homologous control is even higher for the set of kinase inhibitors studied in plots 6(b) and 6(c), with respect to the results in plot 6(a). Plot 6(d) is, instead, dramatically different, not so much in terms of success rates, but in terms of the number of nodes that display any median scaled displacement at all.

A careful analysis of the different selections of kinase inhibitors shows that, while their distribution in the network is very similar, the actual parameters of the reactions selected for inhibition are significantly different for 6(d). Specifically, the number of large (bigger than one) forward and backward kinetic rates is about a third of the corresponding number of large kinetic rates in the other three selections of 19 kinase inhibitors. This is an important point that both

shows the limits of our technique, and suggests that, for this method to be eventually applied in practice, care should be made to select, when possible, kinase inhibitors that act on relatively fast reactions. Most nodes that show sizable displacement curves are likely to be close to these fast reactions, when targeted by kinase inhibitors.

**Remark 9:** Another point that needs clarification is the extent to which the modified ASR method is essential to the success of the identification of near-optimal combinations. In other words, is it necessary to know the nodes that are involved in each of the 19 reactions that we target for inhibition? The answer is affirmative: without knowledge of the presence of these reactions many of the relevant parameters are overlooked by the standard ASR algorithm, in the sense that either they are not found, or their value is underestimated. We find ourself in a scenario very similar to the one of figure 6(d), where very few nodes in the reconstructed network have any response to the kinase inhibitors (not shown). On the other hand, we reiterate that at no time was the actual knowledge of the parameters of the reaction required, as the modified ASR method finds these parameters automatically. In general, we expect the performance of the method to improve with larger sets of kinase inhibitors and to worsen with smaller sets. However, large sets of kinase inhibitors are likely to arise in the experimental setting and this is the scenario where the method should work best.

## 4 Conclusion

Our method for homologous control is an attempt to develop a signal processing approach to network dynamics and it has the potential of greatly reducing the experimental load necessary to find near-optimal combinations of kinase inhibitors for a list of potential target reactions. Its strength is in the ability to work with very limited, noisy data and with networks that have a large number of nodes, comparable with realistic protein microarray data. In this last section we would like to point out several broad areas for further development that are intimately related to the limitations of current experimental protocols for measuring node activity of signaling networks.

Our approach is dependent on the choice of a region \( \mathcal{R} \) where we select the initial conditions, and on the duration \( T \) of the time series, so that we could say that the notion of homologous systems and signal processing of networks is transient based, i.e. it depends on the choice of the cylinder \( \mathcal{R} \times [0, T] \). This raises some issue on the stability of the homology, if we run the reference system for a time \( T \), how long should we run the reconstructed system?

One characteristic of the reconstructed network is that its dynamics displays slower changes when compared to the reference network, probably because the parameters of each term in the equations are not as large as the true parameters, so that the rate of change of nodes will generally be different for the reference network and the reconstructed one. On the other hand the type of curves that we observe are usually transients with eventual relaxation, so that \( T \) can be chosen for both networks as the time such that either the trajectories of the
networks have relaxed to their steady state, or they show consistent divergence.
A possible strategy to improve parameter estimation for large networks is to run the reconstruction algorithm several times with different choices of the random terms; collect for each node the terms that display significant activity for at least one repetition of the reconstruction algorithm; repeat one last time the reconstruction for each node, only using the nodes previously selected as significant. This bootstrap version of the reconstruction algorithm may improve homology with the reference system.
An important question is to determine how infrequently we can sample the trajectories of a network and infer a well behaved reconstructed network that is homologous to the reference network. In some sense, we need a sampling theory of networks; note that the sampling is done for the trajectories, and it is a signal processing operation, but the notion of well behaved system is essential a dynamical one.
To gain a sense of the difficulty of this endeavor, consider that the trajectories’ sampling rate used in this paper clearly do not allow for high true positive rates and low false positive rates of identification of parameters in the reconstructed network. Significant noise is observed in the network parameters’ estimation, even in the parameters of the reactions selected for kinase inhibition.
Indeed, the actual trajectories generated by the reconstructed network do not need to show any strong resemblance to those of the reference network. The very notion of homologous networks is designed to be useful exactly when we undersample the network trajectories so severely that we do not hope for a proper network reconstruction. As we showed in this paper, as little as 209 data points per node are sufficient to obtain partial homology, possibly because the sensitivity of the initial dynamics of the trajectories to kinase control may be predictive of the sensitivity over the full transient leading to relaxation to the steady state. The theoretical determination of the relation between number of samples per node and network homology will require extensive study and comparison of several biologically meaningful models of pathways.
It is also crucial to understand the performance of our method when dealing with incomplete networks where only a portion of the nodes is measured. We tested, for example, the nearly-optimal kinase prediction algorithm C1-C3 on a module of the EGF-R model comprising only 16 variables, with only two reactions selected for kinase inhibition, and we were indeed able to observe homology for several nodes even though many nodes were subject to large feedbacks from unmeasured nodes.
It is yet to be seen whether a random choice of a subset of nodes belonging to a pathway are sufficient to achieve homologous control. Of course, for our method to make sense, at least all nodes involved in reactions to be inhibited and the target node must be measured. We need to perform a detailed study in which we identify the minimum number of variables (and their distribution) that need to be measured to achieve homologous control.

\(^6\)In this simulation we used 500 data points, initial condition for each node were kept very low for this simulation, while EGF was very large.
Finally, we would like to stress that there is no technical reason to restrict ourselves to the analysis of the EGF-R signaling network, or even to protein networks. A more general choice of terms for the model in equation (1) would allow our method to be used for multiscale systems, where genomic, proteomic and metabolic compounds are related in a single network.

Appendix: Recursive Augmented Sparse Reconstruction with selected target reactions

In this appendix we give details of the recursive augmented sparse reconstruction algorithm to be used in the presence of target reactions. Suppose we are given \( N \) node variables from a network and that for each variable it is possible to generate \( R \) trajectories \( X_{n,r} \), \( r = 1, ..., R \) with different initial conditions, uniformly sampled at \( L \) points. We build now the left hand side of (1) and the individual terms in the right hand side.

Call \( \dot{X}_{n,r} \) the vector \( X_{n,r}(t) - X_{n,r}(t_0) \) where \( t \) takes all \( L \) sampled values.

For a given vector \( g(t), t = t_0, ..., t_L \), let \( I(g) \) be the vector whose \( l \)-th component is the sum \( \sum_{r=0} L g(t_r) \).

Write \( Y_n = [\dot{X}_{n,1}, ..., \dot{X}_{n,R}] \), \( G_n = [I(\dot{X}_{n,1}), ..., I(\dot{X}_{n,R})] \), \( n = 1, ..., N \), \( G_{ij} = [I(\dot{X}_{i,1}X_{j,1}), ..., I(\dot{X}_{i,R}X_{j,R})] \). Finally, let \( J \) denote the unit vector with same length as \( Y_n \).

Select a collection of potential target reactions \( v_s = a_s x_i, x_j, - b_s x_k, \) \( s = 1, ..., S \). The basic process to identify the links among the nodes is the following. For each node \( n \) with \( n = 1, ..., N \):

R1 Choose an attenuation coefficient \( \beta_q \) for the quadratic terms \( G_{ij} \). Let \( n_g, g = 1, ..., G \) be discrete random vectors normally distributed scaled to have norm 1. Denote by \( || \) the 2-norm of a vector and let \( \hat{G}_t \) be the matrix whose columns are all the vectors \( \frac{\partial}{\partial x_i} \). \( \hat{G}_q \) be the matrix whose columns are all possible vectors \( \frac{\partial}{\partial x_i} \). Let \( N_G \) be the matrix whose columns are the random vectors \( n_g \) scaled to have norm 1. Choose \( G \) large enough to have the matrix \( Z = [J\hat{G}_t, \beta_q \hat{G}_q, N_G] \) with small condition number (say less that \( 10^2 \)).

R2 Set a temporary representation matrix \( M \), for each \( s = 1, ..., S \), if the node \( n \) belong to the set \( \{i_s, j_s, k_s\} \), add the vectors \( \frac{\partial}{\partial x_{i_s}} \) and \( \frac{\partial}{\partial x_{j_s}} \) as columns to the matrix \( M \).

R3 Let \( Z_M = [M N_G] \). Find the minimal \( l_1 \) solution to the underdetermined system \( Y_n = Z_M \alpha_M \). Let \( \alpha_M \) be the restriction of \( \alpha \) to the columns of \( M \), set \( Y_n = Y_n - M \alpha_M \).

R4 Find the minimal \( l_1 \) solution to the underdetermined system \( Y_n = Z \alpha \). If in part R2 we generated a nonzero matrix \( M \), then add to the components of \( \alpha \) associated to the columns of \( M \) the corresponding components of \( \alpha_M \).
R5 Choose a threshold $T_n$ and let $\alpha_{T_n}$ be the coefficients in $\alpha$ larger than $T_n$. The reconstructed network equation for $x_n$ will have only linear and quadratic terms that correspond to coefficients in $\alpha_{T_n}$, and their coefficients will be the coefficients of $\alpha_{T_n}$ divided by the norm of the corresponding $|G_i|$, if a linear term, and $|G_{ij}|$ if a quadratic term.

In [Napoletni et al. 2008] we showed that there is considerable flexibility in the choice of the number $G$ of random terms and in the choice of the attenuation coefficient $\beta_q$; in this work we use $G = 1500$ and $\beta_q = 0.8$. The threshold $T_n$ that selects the parameters to be used in the reconstructed network is taken to be a very low 2% of the maximum magnitude parameter for the corresponding node. This choice makes sure that most inferred node directed links among nodes are retained.

References


26
[Jones et al. 2006] Richard B. Jones, Andrew Gordus, Jordan A. Krall and
Gavin MacBeath. A quantitative protein interaction network for the ErbB

[Birtwistle et al. 2007] Marc R Birtwistle, Mariko Hatakeyama, Noriko Yumoto,
Babatunde A Ogummaike, Jan B Hoek and Boris N Kholodenko. Ligand-
dependent responses of the ErbB signaling network: experimental and mod-

[Uetz and Stagljar 2006] Peter Uetz and Igor Stagljar. The interactome of hu-
man EGF/ErbB receptors. Molecular Systems Biology 2:2006.0006

[Kumar et al. 2008] Amit Kumar, Edward T. Petri, Balazs Halmos, and Titus
J. Boggon. Structure and Clinical Relevance of the Epidermal Growth Fac-
tor Receptor in Human Cancer. Journal of Clinical Oncology, Vol 26, No

[Baselga and Arteaga 2005] Jose Baselga and Carlos L. Arteaga. Critical up-
date and emerging trends in Epidermal Growth Factor Receptor targeting

[Hornberg et al. 2005] Jorrit J Hornberg, Bernd Binder, Frank J Bruggeman,
Birgit Schoeberl, Reinhart Heinrich and Hans V Westerhoff. Control of
MAPK signalling: from complexity to what really matters. Oncogene
(2005) 24, 55335542.

composition by Basis Pursuit, SIAM Journal on Scientific Computing SISC,


tal validation for quantitative protein network models. Current Opinion in

[Nishizuka et al. 2008] Nishizuka S., Ramalingam S., Spurrier B., Washburn F.
L., Krishna R., Honkanen P., Young L., Shimura T., Steeg P. S. and John
Austin (2008). Quantitative Protein Network Monitoring in Response to
DNA Damage, Journal of Proteome Research, 7, 803808.

New York.

27
