Journal of Theoretical Biology 279 (2011) 29–43

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Homologous control of protein signaling networks

D. Napoletani [a,*], M. Signore [b], T. Sauer [c], L. Liotta [a], E. Petricoin [a]

[a] Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, VA 20110, USA
[b] Department of Hematology, Oncology and Molecular Medicine, Biotechnology Division, Istituto Superiore di Sanità, Viale Regina Elena 299, 00166 Rome, Italy
[c] Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA

ARTICLE INFO

Article history:
Received 14 December 2009
Received in revised form
6 March 2011
Accepted 17 March 2011
Available online 23 March 2011

Keywords:
Sparse network reconstructions
Protein network models
Signaling pathways
Kinase inhibitors

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 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, a result potentially useful for a variety of applications in personalized medicine.

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 often 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 on

* Corresponding author. Tel.: +1 703 993 9549.
E-mail address: dnapolet@gmu.edu (D. Napoletani).

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.03.020

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.

The potential therapeutic implications of such an approach are evident if we consider the great heterogeneity of cancer patients. Individuals with similar stages of the disease show diverse therapeutic responses that are oftentimes not predictable on the basis of genetic mutation analysis. Primary cancer stem cells (CSCs) from a variety of tumors have been isolated and characterized and clear evidences exist of their association with tumor's resistance to chemo- and radio-therapies, making CSCs a useful and effective target for cancer research (Ailles and Weissman, 2007; Gilbert and Ross, 2009). Moreover, an extensive proteomic, genetic and drug sensitivity profiling of the NCI60 cell line set has recently highlighted the importance of direct studies on cancer cells in order to associate drug responsiveness to specific molecular signatures (Park et al., 2010). Such studies and clinical experience, call for a multi-targeted approach to cancer treatment and underscore the importance of molecular biomarkers (Zahorowska et al., 2009). Therefore, an *in-silico* system capable of reconstructing the behavior of deregulated cancer cell signaling networks holds a significant experimental value in the prediction of therapeutic responses based on individual patients molecular characteristics.

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, and 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 interaction 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

[1] A patent application has been filed for the methods described in this paper. Patent application number 12/959,096, filed on 12/02/2010.

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 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 adapter 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 non-small 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 uni-molecular 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 Eq. (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, Eq. (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.

$$x_n(t)-x_n(t_0)=a_{0n}+\sum_{i=1}^{N} l_{in}\int_{t_0}^{t} x_i\,dt + \beta_q\sum_{i=1}^{N}\sum_{j=1}^{N} q_{ijn}\int_{t_0}^{t} x_i x_j\,dt+\sum_{g=1}^{G} w_{gn}n_g. \quad (1)$$

Here, the $\beta_q \le 1$ represent positive attenuation coefficients for the quadratic terms. The systems parameters at node n that we need to determine are: a_{0n}, l_{in}, $i=1,\ldots,N$, q_{ijn}, $i,j=1,\ldots,N$. 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 Eq. (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 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) and 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 modeled (Schoeberl et al., 2002; Voit, 2000) as $v=ax_ix_j-bx_k$. Its effects on the differential equations of the network are as follows: if we let $\dot{x}_i=f_i(x_1,\ldots,x_n)$, 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 $\dot{x}_i=f_i(x_1,\ldots,x_n)-v$, and similarly $\dot{x}_j=f_j(x_1,\ldots,x_n)-v$. On the contrary, since the specific reaction increases the phosphorylation of x_k, we will have $\dot{x}_k=f_k(x_1,\ldots,x_n)+v$.

The immediate consequence of this modeling assumption is that if we know that a simple quadratic reaction $v=ax_ix_j-bx_k$ 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 $v=ax_ix_j-bx_kx_h$, 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 $\kappa<1$.

This modeling of kinase inhibitors stems from the assumption that, whatever the current impact of a reaction on the network, a kinase inhibitor targeting that reaction can only slow down its effect on the network in relative terms. Clinical and biological evidences in cancer therapy suggest that the response of patients is often limited by both the low efficacy of drug targeting and by resistance mechanism that cancer cells evolve before and during treatments. Nonetheless, many tumors are dependent ('addicted to') on specific signaling nodes/pathways and even subtle reductions in the total amount of such nodes can impair tumor's homeostasis over time (Sharma and Settleman, 2007; Weinstein and Joe, 2008). Such evidences are the foundation of combinatorial treatments for cancer and of our assumption that even just a relative, but continuous slowing down of a reaction can have important impact on the network.

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).

Oftentimes kinase inhibitors cross-react, with various degrees of specificity, either 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 κ to act only on the first or second term of the reaction respectively. Since changes in κ 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 κ 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 knowledge of the reactors of a given set of reactions. The details of the algorithm are given in Appendix B.

2.1. Modified ASR algorithm

Select a collection of potential target reactions $v_s=a_sx_{i_s}x_{j_s}-b_sx_{k_s}$, $s=1,\ldots,S$. Given the collection of all time measurements for each node n with $n=1,\ldots,N$:

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

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

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.

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

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 **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. We stress that only the reactors in the targeted reactions need to be known, while the parameters of the reactions are found by the modified ASR method automatically.

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.

3. Partial homology of networks

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 $v_s=a_sx_{i_s}x_{j_s}-b_sx_{k_s}$, $s=1,\ldots,S$ that we want to target with kinase inhibitors with nonzero parameters a_s and b_s.

Throughout the remaining sections, it is assumed that we have sets of 20 different noisy initial conditions for time courses of nodes of the network, and that we sample each time course 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 time course is assumed to be at most 10% of the maximum value of the points along the time course itself. This setting gives us a time course microarray with $20\times 11=220$ data points for each of the 103 nodes in the reference EGF-R network.

This number of data points in the microarray is very small from the data mining perspective, but it is large from the experimental, *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).[2]

If a specific experimental setting, or cell type, 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 of protein concentrations, corresponding to distinct cell states, should ideally be able to explore as much of the dynamical range of the reference network as possible. This can be achieved experimentally by varying growth factors concentrations, e.g. EGF, across cell lines of different origins.

At the same time, we only need 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 the modified ASR algorithm even when only very limited data are available.

In this paper, the time course microarray data used in the modified ASR algorithm are simulated from the reference EGF-R network described in Schoeberl et al. (2002). The simulated initial time course microarray, and the modified ASR algorithm, provide us with a candidate reconstructed network that is then compared to the reference network.[3] Reference and reconstructed networks are simulated in the MATLAB environment available at www.mathworks.com.[4]

The choice of meaningful initial conditions for the time courses is complicated by the fact that the copy numbers of individual proteins vary enormously, and protein concentration varies with cell type and cell cycle stage, from less than 20 000 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 protein concentrations of the reference network are chosen to be random values uniformly distributed in the interval [2000, 20 000]. These values are assumed to be the average number of copies of molecules per cell, and we are assuming that different cell lines might have inherently diverse protein expression patterns. The average number of EGF receptors is taken to be a random value in the interval [1000, 10 000]. EGF is selected to be a random value in the interval $[10^{-8}, 10^{-7}]$ to simulate varying degrees of EGF stimulation. Units are different for EGF 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 for concentrations 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 time courses, and measurable dynamical changes.

Remark 3. Note moreover that the size of the simulated time course microarray used in the modified ASR 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 homology of networks 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.

Remark 4. Our analysis of the simulated EGF-R network model described in Schoeberl et al. (2002) is done assuming that the parameters of the network are constant in the interval of time during which the time course microarray is generated. This assumption seems reasonable because, even in practice, the time courses over which the microarray data are generated are limited to a time scale of days, and in the scenario of our simulations time courses are taken over only 11 min. Moreover, in case of translating the modified ASR method to a real world situation where we may seek a therapy for a disease or a pathophysiological state, the time to disease progression and hence to a potential change in the underlying network, is measurable in a time scale of months (Ernst et al., 2008). We conclude that during the time required for applying the modified ASR method to biological samples microarray dataset, the networks' parameters can be considered reasonably constant.

For clinical purposes, not all nodes are of interest. For example in the reference model of EGF-R, x_{51} (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. At the same time we focus on at least one node (node 31, or free Shc) that is particularly amenable to our techniques, and that has biological relevance.

To check the global effect of kinase inhibitors on the network, we select a set $\mathcal{S}_1$ of 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 set of reactions $\mathcal{S}_1$ composed of nine pairs of reactions for surface and internalized receptors plus a single degradation reaction.[5]

Although the function of internalization is not the same for all the receptors, it has been demonstrated that EGF-R signals through internalized receptor complexes. A clear correlation between surface and internalized molecules has not been discovered since internalization also causes receptor deactivation.

[2] We mention here as well our work in preparation on adult stem cells, where the aggregate data set generated 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.

[3] The values of the time courses are known to be positive, and this is enforced, for the simulation of the reconstructed network, in the discrete differential equation solver.

[4] Data and MATLAB codes used to generate all figures in the paper are available upon request to the corresponding author at dnapolet@gmu.edu

[5] The specific reactions that we select in $\mathcal{S}_1$ are: v19, v66; v20, v67; v23, v70; v27, v74; v29, v76; v41, v83; v45, v87; v47, v89; v55, v97; and v60. Refer to Schoeberl et al. (2002) for an actual description of these reactions.

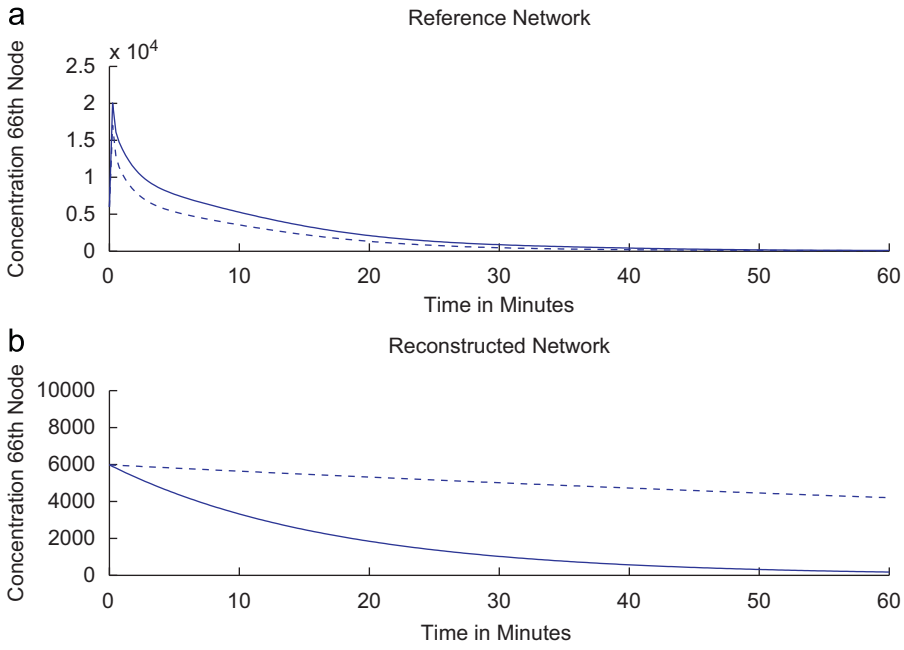

Fig. 1. An instance of relative concordance of the magnitude of displacement of trajectories for reference and reconstructed networks due to a specific pair of kinase inhibitors. Dashed curves are the trajectories with inhibitors on, while solid curves are the trajectories without inhibitors. The top plot shows the effect of inhibition of reactions v41 and v83 on the 66th node of the reference EGF-R network. The bottom plot shows the corresponding trajectories for the same node 66 of the reconstructed network. The dynamics of reference and reconstructed networks are just marginally similar for this node. The important point is that we do observe a *measurable change* of time courses due to the inhibition of the reactions for both networks.

Importantly, (Schoeberl et al., 2002) demonstrated that under certain circumstances, such as low EGF concentration, the internalization rate is a critical factor in EGF-R signaling. We took into account both the surface and internalized reactions since both contribute to the overall signaling and, at least for the EGF-R system, molecules exist already in the clinics which would allow for the selective targeting of surface receptor only (Cetuximab monoclonal antibody). This selective targeting is not feasible for many drug targets that are currently exploited for cancer therapy, but we can still treat in theory the internalized or non-internalized versions of each reaction as if they were two distinct reactions, and hence evaluate their relative impact upon inhibition. Notably, a single node that signals through both internalized and non-internalized counterparts could be inhibited by blockade of both reactions. Similarly, if the internalized version of a node is not affecting the signaling output, the reconstructed system will suggest inhibition of the non-internalized reaction.

We also note that our results do not depend essentially on selecting pairs of internalized and non-internalized reactions. In Fig. 4, we summarize results for three other sets of reactions for which there is not extensive matching of internalized and non-internalized versions of the same reaction. This scenario is the most general one, applicable also to networks other than the EGF-R signaling network, and therefore we will treat internalized and non-internalized version of reactions as if they were distinguishable.

In Fig. 1, we can see an instance of the effect of kinase inhibitors on both reference and reconstructed networks. The top plot shows the change of time course 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 time course for the 66th node of the reconstructed network. The dynamics of reference and reconstructed networks are just marginally similar for the node in Fig. 1. The important point is that we do observe a *measurable change* of time courses due to the inhibition of those two reactions. Indeed, we choose to display node 66 only because its displacements for reference and reconstructed networks are of similar order of magnitudes and amenable to a direct graphic comparison. The similarity of the time courses is, for other nodes, even less pronounced than what we observe in Fig. 1 for node 66, or it may happen that the change of time course of a node, due to control, is several order of magnitudes smaller for the reconstructed network. This is expected, 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.

Remark 5. As an aside, it is not surprising that inhibition of reactions v41 and v83 would naturally lead to the modulation of node 66 (and 35, not shown). In fact, reaction v41, [(EGF-EGFR*)2-GAP-Shc*]+[Grb2-Sos]↔[(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos], and it's internalized counterpart v83, represent the recruitment of adapter proteins to the activated receptor. These reactions are directly upstream of nodes 35 and 66, which are the canonical and internalized versions of the same node (i.e. (EGF-EGFR*)2-GAP-Shc*-Grb2-Sos). Interestingly, it has been shown that treatment of leukemia cells with a specific Grb2-SH3 (growth factor receptor-bound protein 2-SH3 domain) inhibitor that disrupts the Grb2-Sos (son of sevenless) complex, has dose-dependent cytotoxic effects, although not comparable with the efficacy of Gleevec in chronic myelogenous leukemia (Ye et al., 2008).

In Fig. 1, note also that the sign of the change due to control is the opposite in the two networks. In Appendix A we show 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 reference network are changed when kinase inhibitors control is switched on.

To achieve this goal, we propose the following estimation of the change of the trajectories due to kinase inhibitors control: (a) We randomly select several initial conditions for concentrations in the same wide region used to generate the initial time course microarray of the reference system. (b) We simulate

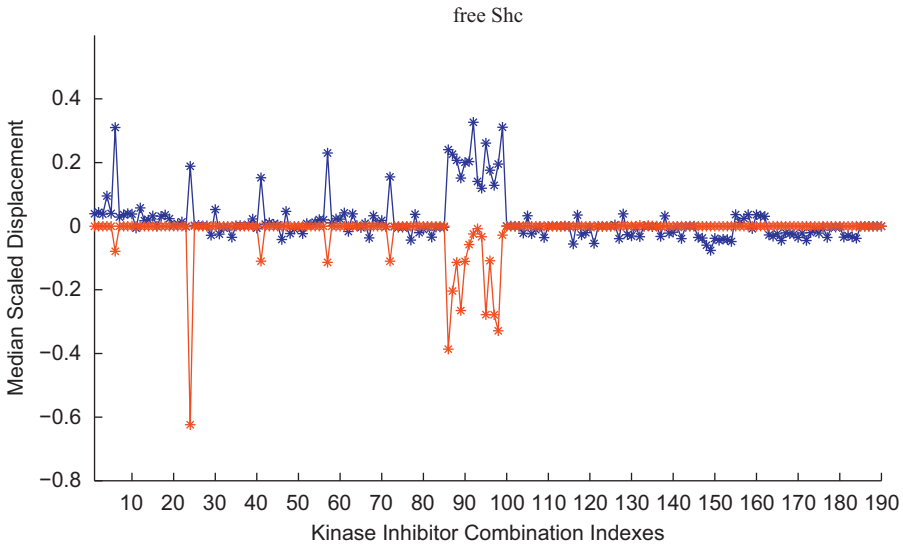

Fig. 2. In this plot we show a case in which the shape of the median scaled maximum displacement curve is very similar for reference and reconstructed networks for high values of displacement. The blue starred line gives the median scaled maximum displacements for free Shc, the 31th node of the reference EGF-R network. Each point on the horizontal axis corresponds to one among 190 different pairs of kinase inhibitor combinations. The red starred line gives the median scaled maximum displacement curve for the same node in the reconstructed network. Curves are scaled to have norm equal to one for comparison purposes. Note the reversal of sign of the median displacement curves for large displacement values.

reference network and reconstructed network with each of these initial conditions for a fixed length of time, long enough for most time courses 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 procedures (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 6. We consider the median since most of the time the scaled displacement has very similar behavior for most initial conditions, but there is rarely a small number of initial conditions that may lead to divergent time courses for some nodes in the reconstructed system, making the mean too sensitive to these outliers.

In Fig. 2, we plot the magnitude of the median scaled pointwise displacement of node 31 (free Shc) with respect to an ordering of 190 different kinase inhibitor combinations, where only one or at most two reactions of the set of 19 reactions $\mathcal{S}_1$ (described in footnote 5) are inhibited at the same time.[6] Each kinase inhibitor coefficient κ_s, $s=1,\ldots,19$ is allowed to be either 1 (no inhibition) or 0.1 (high kinase inhibition), This combinatorial constraint is in line with current experimental protocols, i.e. 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. More complex combinatorial scenarios can be envisioned as well, with several kinase inhibitors used at the same time. Note that generally a kinase inhibitor is considered useful if it changes the phosphorylation of a target by a significant amount. Our choice of κ_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 10–20%.

Remark 7. 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 drug's full potential (Johnson, 2009).

We display median scaled displacement curves for free Shc (Src homology 2 domain-containing transforming protein 1 or SHC1), since these curves exemplify a significant pattern of homology between reconstructed and reference networks. And also because Shc adapter proteins have an established and important role in transducing signals from receptor tyrosine kinases (RTKs) downstream to Ras. The biology of Shc is complex and three isoforms have been described in the literature: isoforms p46Shc and p52Shc, once phosphorylated, couple activated receptor tyrosine kinases to Ras and are implicated in the cytoplasmic propagation of mitogenic signals. Thus isoform p46Shc and isoform p52Shc may function as initiators of the Ras signaling cascade in various non-neuronal systems (Migliaccio et al., 1997). Differently from other isoforms, p66Shc does not mediate Ras activation, but is involved in signal transduction pathways that regulate the cellular response to oxidative stress and aging

[6] The total of 190 combinations is the result of taking every combination of one out of 19 reactions (19 such combinations) and every distinct combination of two reactions out of 19 (171 such combinations).

(Luzi et al., 2000; Ulivieri, 2010). Interestingly, the ratio between active (tyrosine phosphorylated) Shc and p66Shc is a strong independent prognostic indicator in breast cancer (Davol et al., 2003). Total Shc activation is dependent on interaction and subsequent phosphorylation by activated EGF-R (Alam et al., 2009), and steroids promote the proliferation of tumor cells by regulating the half life of cytosolic (free) Shc (Kumar et al., 2011). Moreover, previous studies demonstrated that even small amounts of free Shc drive most of the signal flux generated in MAPK cascade activation by EGF receptors (Gong and Zhao, 2003). Therefore we believe that, at least from a theoretical standpoint, free Shc levels are a valuable readout for the impact of combined inhibition of kinases into the EGF-R signaling. We also note that the set of endpoints for which we have strong homology may vary as we input into the modified ASR different time course microarray data, and the most useful endpoints, from a therapeutic viewpoint, may not always display homology. In Section 4 we suggest and test a criterium to automatically identify endpoints that are likely to be strongly homologous.

There is a striking concordance of the shape of the two median scaled displacement curves in Fig. 2 for large values of median scaled displacement, 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 high concordance of the shape of the median scaled displacement curves. The observed concordance suggests that the median scaled displacement curve of the reconstructed network can be used to infer the corresponding curve for the reference network.

The active combinations that display the largest absolute values of median scaled displacements in Fig. 2 are all combinations of v41 with each of the reactions in $\mathcal{S}_1$.[7] In other words, inhibition of v41 has a strong impact on free Shc and this impact is then variously modulated by inhibition of the other reactions to obtain different levels of median scaled displacement. The reaction v41, already described in Remark 5, does not directly involve free Shc as a reactor, and it is two reactions downstream of free Shc. In fact, free Shc associates with (EGF-EGFR*)2-GAP complex, this is subsequently phosphorylated by the receptor kinase to generate the (EGF-EGFR*)2-GAP-Shc* complex, and finally, in reaction v41, Grb2-Sos binds to the (EGF-EGFR*)2-GAP-Shc* complex (refer to network scheme in Hornberg et al., 2005). The distance separating v41 from free Shc confirms that modified ASR can propagate the effect of inhibition of target reactions to endpoints of the network that are not directly implicated in the reactions themselves. Essentially, by inhibiting v41, we reduce the consumption of a complex containing active Shc (Shc*), and this leads to time courses where the concentration of available free Shc in the network is larger (enhanced) than what it would be without inhibition of v41.

To gain additional useful information we would need to identify the near-optimal modulation of the inhibition of v41 by inhibition of one of the other reactions in $\mathcal{S}_1$. In Fig. 2, combination 24 (v20, v41) is wrongly identified in the reconstructed network as the one with largest absolute value of median scaled displacement, since we can see that combination 92 (v41, v67) is the one with largest absolute value of displacement for the reference network. Reaction v67 is two reactions downstream of (EGF-EGFR*)2-GAP, on a separate pathway from the one involving Shc, and therefore v67 probably indirectly affects the concentration of (EGF-EGFR*)2-GAP, which, in turn, affects the availability of free Shc. Note however that v67 is the internalized version of v20 (Hornberg et al., 2005), and therefore in this case the reconstructed network approximately identifies the biological process that best modulates the inhibition of v41 on free Shc in the reference network. Often, a more refined identification of near-optimal combinations is possible, and we further analyze this issue in Section 4 for several choices of input time course microarrays and of selections of target reactions.

In general, the displacement curves tend to agree only partially and only for the largest displacements values. We also stress the fact that often (but by no means always, see Appendix A) 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 an inhibitor or an 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 and in the following we compare only absolute values of median scaled displacement curves.

For example, we can compare Spearman's rank correlation (Gibbons and Chakraborti, 2003) of the absolute value of displacement curves for corresponding nodes in reference and reconstructed networks. A high Spearman's rank correlation would indicate a very good agreement of the pattern of increase and decrease of large displacements for a given node. In Fig. 3(a), we show three histograms of Spearman correlation of corresponding median scaled displacement curves for reference and reconstructed networks. All of them are obtained assuming that the set $\mathcal{S}_1$ of reactions (see footnote 5) is the target of kinase inhibitors, and that the modified ASR algorithm is given as input three different sets of time course microarrays, i.e. three completely different sets of initial conditions and dynamical evolutions of the EGF-R network. To avoid the impact of relatively small displacement values, all displacement values below 10% of the maximum value of each median scaled displacement curves are set to zero before computing the Spearman correlations. Note the very long right tail of large correlations, indication of excellent agreement of displacement curves for several nodes. Such tail is completely absent when, in Fig. 3(b), we compute the histograms of the Spearman correlations of the median scaled displacement curves of the nodes of the reference network, with randomly permuted versions of the median scaled displacement curves of the nodes of the reconstructed networks.

4. Homologous control of nodes

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

4.1. Homologous control

C1 Given a set of 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 give constraint. For each kinase inhibitor

[7] The separation of active combinations 6, 24, 41, 57, 72 from the main group of active combinations 86–99 in Fig. 2 is only an artifact due to the specific indexing of the pairs of reactions.

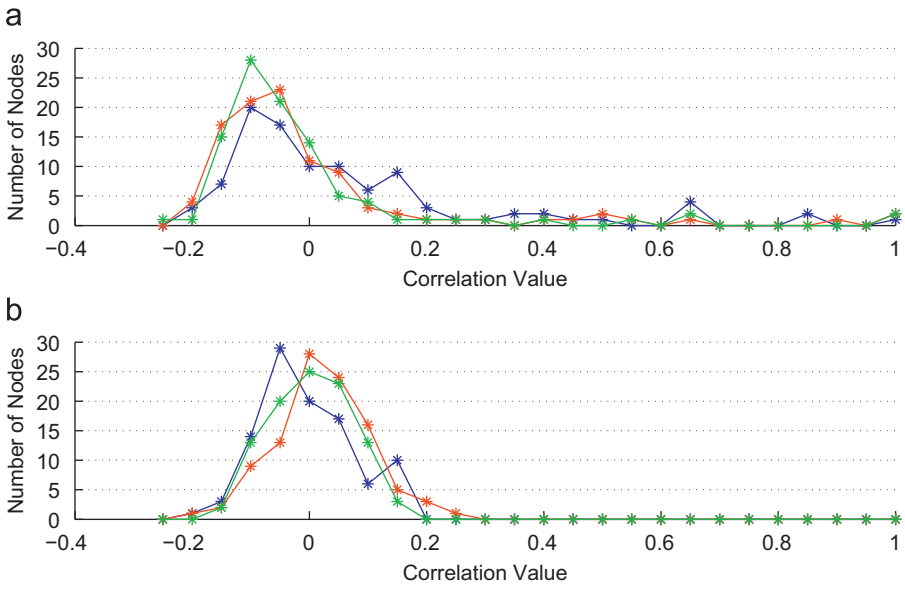

Fig. 3. In (a) we show three histograms of the Spearman correlation of the absolute value of corresponding median scaled maximum displacement curves in the reference and reconstructed network. Each histogram corresponds displacement curves generated with one of three different choices of initial time course microarrays as input for the modified ASR algorithm. In (b), we show the histograms of the Spearman correlations of the displacement curves of the nodes of the reference network, and randomly permuted versions of the displacement curves of the nodes of the reconstructed networks. All displacement values below 10% of the maximum value of each displacement curves are set to zero before computing the Spearman correlations.

combination generate time courses 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 κ, to reduce total amount of inhibitors and therefore to avoid toxicity and loss of specificity.

In the following, we continue to explore the experimental protocol used to generate Figs. 2 and 3 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 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 scaled 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 80%, of the mean of the largest three 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 three median displacement values, rather than the absolute maximum displacement. We decided for this criterium 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 filter the most useful nodes of the reconstructed network.

Our understanding is that the larger the displacements of a node in the reconstructed network, the more likely the chance that they convey some useful information about the corresponding displacements of the node of the reference network. This is the case because large displacements probably indicate an increased sensitivity of the reconstructed network to specific kinase inhibitors. The following procedure defines quantitatively a notion of large displacements: we take the median nonzero displacement value for each node in the reconstructed network, and then the mean of these medians across all the nodes in the network. We threshold to zero all displacements that are below this value, so that we retain only displacement values that are large relatively to the overall activity of the network.

After removing small displacements with the previous procedure, we filter the nodes in the network by retaining only nodes that have the mean of the remaining nonzero scaled displacements above increasing large threshold values. Note that these threshold values are dimensionless, since they relate to the scaled displacements that are by definition dimensionless.

This filtering of nodes is very stringent, in the sense that only a few nodes are left by this process, and yet it is quite effective at identifying those displacement curves in the reconstructed network that partially correlate to the corresponding displacement curves of the reference network. In particular, in Fig. 4(b), the dashed curves shows the number of nodes of the reconstructed network that have mean of the large nonzero median displacement values above the threshold specified on the horizontal axis. Again, we show our results for three different sets of time course microarrays given as input into the modified ASR algorithm.

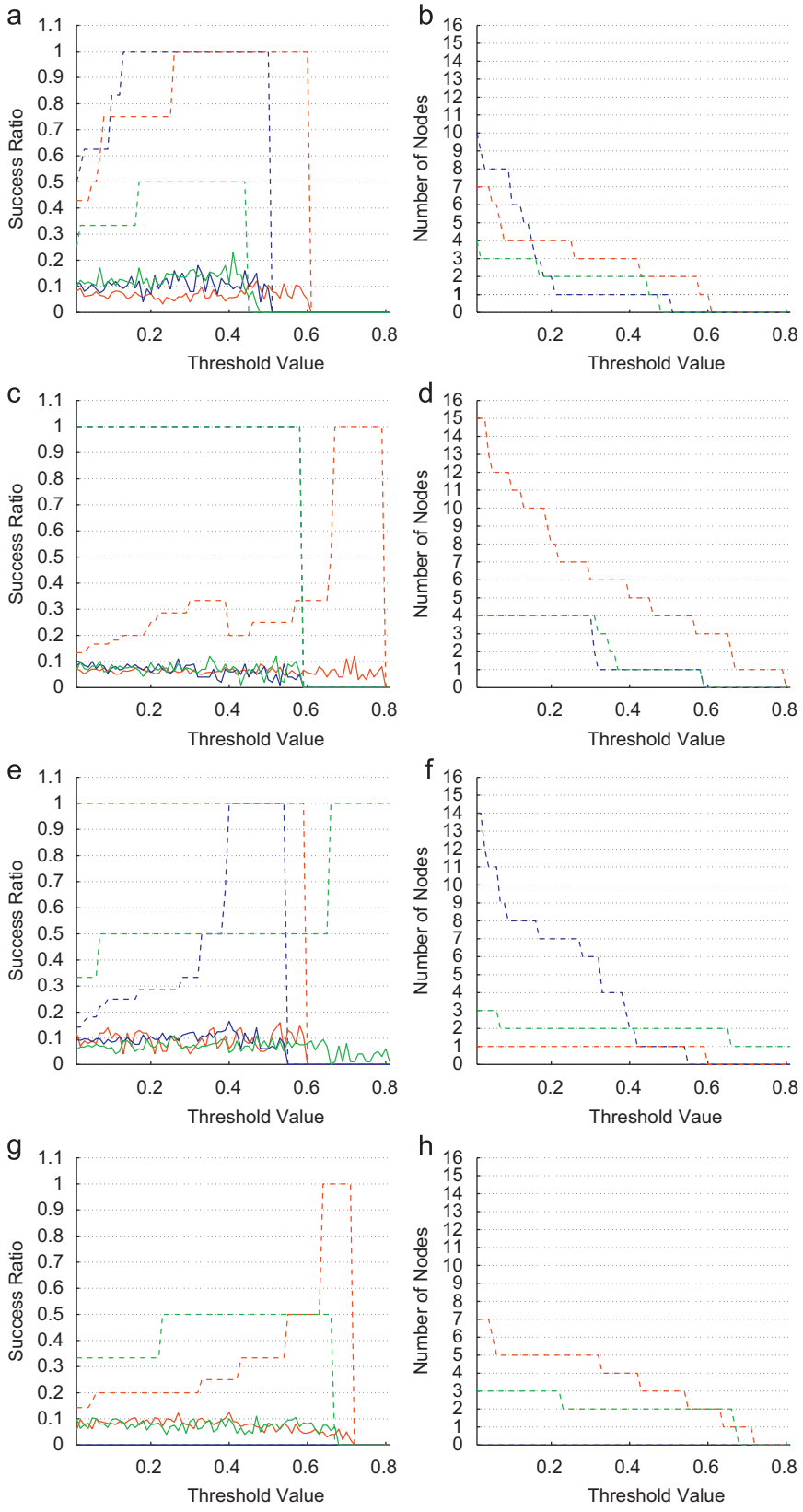

Fig. 4. Plots (a) and (b) refer to the set of 19 reactions $\mathcal{S}_1$. In plot (b) the three dashed curves show the number of nodes of the reconstructed network that have mean of large median displacement values above the threshold specified on the horizontal axis. In plot (a), the three dashed curves show the ratios of nodes for which we identify near-optimal kinase inhibitor combinations, as a function of the same threshold. The solid curves show the average ratio 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 permutated (average computed over 100 permutations). Each color corresponds to the analysis of one of three different choices of initial time course microarrays as input for the modified ASR algorithm. Plots (c), (e), (g), and (d), (f), (h) repeat the same analysis as in plots (a) and (b), respectively, for three more sets of kinase inhibitors (see footnote 8).

The dashed curve in Fig. 4(a) shows the ratio of nodes for which we identify near-optimal kinase inhibitor combinations (as defined above), as a function of the same threshold as in Fig. 4(b). We reach at least a 0.5 success ratio, meaning that, given sufficiently high threshold, for 50% of the nodes that are not filtered out by the threshold, we can identify near-optimal kinase inhibitor combinations. The catch is of course that only a few nodes display such marked homology in the reconstructed median scaled displacement curve. The 0.5 ratio mark is reached when the threshold leave respectively 10, 6, and 2 nodes for the three time course microarray inputs to the modified ASR algorithm. On the other hand, it is striking that we can eventually reach 100% success ratio (respectively, for 6 and 3 nodes) for two of the microarray inputs.

The solid curves in the same plot 4(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 permutated (average computed over 100 permutations). Near-optimality for the random scrambling of the data is much lower, especially for the nodes left by higher threshold values. The mean of the ratio of nodes for which we find near-optimal combinations in the case of random permutation of the reconstructed displacement curves in Fig. 4(a) is only around 0.1 (computed across all threshold values and all three time course microarrays inputs). For nodes selected with high values of threshold, our method is at least about five times more accurate than a random selection of kinase inhibitors in finding near-optimal combinations, and up to 10 times more accurate for some selected nodes with two microarray inputs.

To put things in perspective, with the randomly permuted scaled displacements, on average it would be necessary to select randomly between 18 and 30 kinase inhibitor combinations (depending on the microarray input instance) to make sure that we have at least a 0.5 success ratio, reachable with only three 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.

Fig. 4(c–h) show the results of the same analysis performed in Fig. 4(a and b), for three other selections of 19 kinase inhibitors,[8] and several distinct initial microarrays. 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.

Note that the success curve is not monotone for one instance (red dashed line) in Fig. 4(c), indicating that some strongly homologous nodes are actually removed as the threshold value increases. This raises the hope that even better success rates could be achieved with more sophisticated threshold processes. In particular, it may be beneficial in some cases to have an adaptive threshold that is a variable multiple of the mean of medians used to pre-condition the displacement curves. This may allow for a larger number of nodes to be retained.

The success rate of homologous control is, for high enough threshold value, at least 50% for reactions analyzed in Fig. 4(c–f). Fig. 4(g and h) is, instead, significantly different, in that, for one instance of microarray (blue dashed curves) no displacement curve of the reconstructed network passes the filtering process. 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 Fig. 4(g and h). 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 observation may justify the failure of our method for one instance of initial microarray and it is an important point that both shows the limits of our technique, and suggests that, for the 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 one of the instances of Fig. 4(g and h), where very few nodes in the reconstructed network have any response to the kinase inhibitors (not shown). 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.

5. Discussion and 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 time course 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

[8] The reactions selected for inhibition corresponding to Fig. 4(a,b,c,d,e,f,g,h) are, respectively: ($\mathcal{S}_1$) v19, v20, v23, v27, v29, v41, v45, v47, v55, v60, v66, v67, v70, v74, v76, v83, v87, v89, v97; ($\mathcal{S}_2$) v1, v2, v3, v10, v16, v28, v29, v36, v37, v45, v46, v47, v48, v64, v75, v94, v95, v126, v130; ($\mathcal{S}_3$) v1, v2, v12, v23, v27, v33, v34, v36, v45, v53, v56, v74, v76, v78, v89, v97, v111, v129, v130; ($\mathcal{S}_4$) v1, v3, v11, v18, v22, v26, v29, v33, v39, v42, v44, v52, v65, v67, v75, v88, v96, v128, v148. Refer to Schoeberl et al. (2002) for an actual description of the reactions.

rate of change of nodes will generally be different for the reference network and the reconstructed one. Nevertheless, 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 220 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.[9]

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.

The final goal of our approach is the experimental validation of homologous control over a broad range of signaling networks in defined biological contexts such as cell proliferation, survival and differentiation. The availability of hundreds chemical compounds with known specific inhibitory activity allows to test the efficacy of the predicted near-optimal kinase inhibitor combinations into cell line models. In particular, it is possible to exploit cancer cell lines obtained from diverse types of tumor and to train the modified ASR method with reverse-phase protein microarray data containing time-courses for each cell line under serum addition or hypoxic stimuli. The molecular dynamics driven by such conditions would allow a validation of the *in-silico* reconstructed network in a specific and biologically meaningful context. The combinations of inhibitors that are found by our method in this scenario could then be tested *in vitro* (and eventually *in vivo* in immunodeficient mice) to understand their ability to affect proliferation and survival mechanisms of tumor's cells. Such an approach has a great potential in the identification of novel therapeutic strategies for cancer. We stress moreover that there is no technical reason to restrict homologous control strategies to the analysis of the EGF-R signaling network, or even to protein signaling networks. A more general choice of terms for the model in Eq. (1) would allow our method to be tested for multiscale heterogeneous systems, where genomic, proteomic and metabolic compounds are related in a single network.

[9] In this simulation (data not shown) we used 550 data points, and initial condition for each node were kept very low for this simulation, while EGF was very large.

Acknowledgments

The authors acknowledge the support of the College of Science at George Mason University and the Istituto Superiore di Sanità. They also would like to thank Daniele C. Struppa for many useful discussions and the anonymous referees for their constructive remarks.

Appendix A. Sign concordance and sign switching

One difficulty summarizing the qualitative properties of displacement curves described in Section 3 is the tremendous variability of the individual nodes of the network. 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 Fig. 5(a) we show the number of kinase inhibitor combinations, out of the total 190 combinations used in the protocol of Figs. 2 and 3, 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 five 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 Fig. 5(a). Instead, Fig. 6(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 A1. 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 Fig. 2 to have vigorous control on at least a node 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,

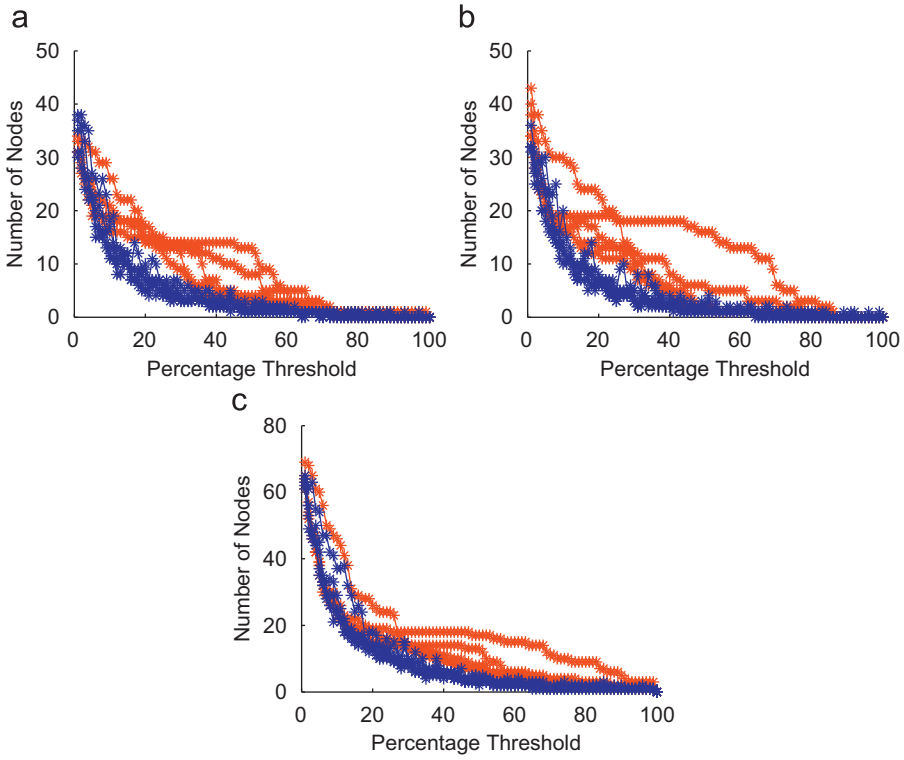

Fig. 5. Red starred curves in (a)–(c) show the number of kinase inhibitor combinations of the five nodes of reference and reconstructed networks that display, respectively, the largest: (a) sign switching; (b) sign concordance; and (c) concordance of nonzero median displacement values. Blue starred curves 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.

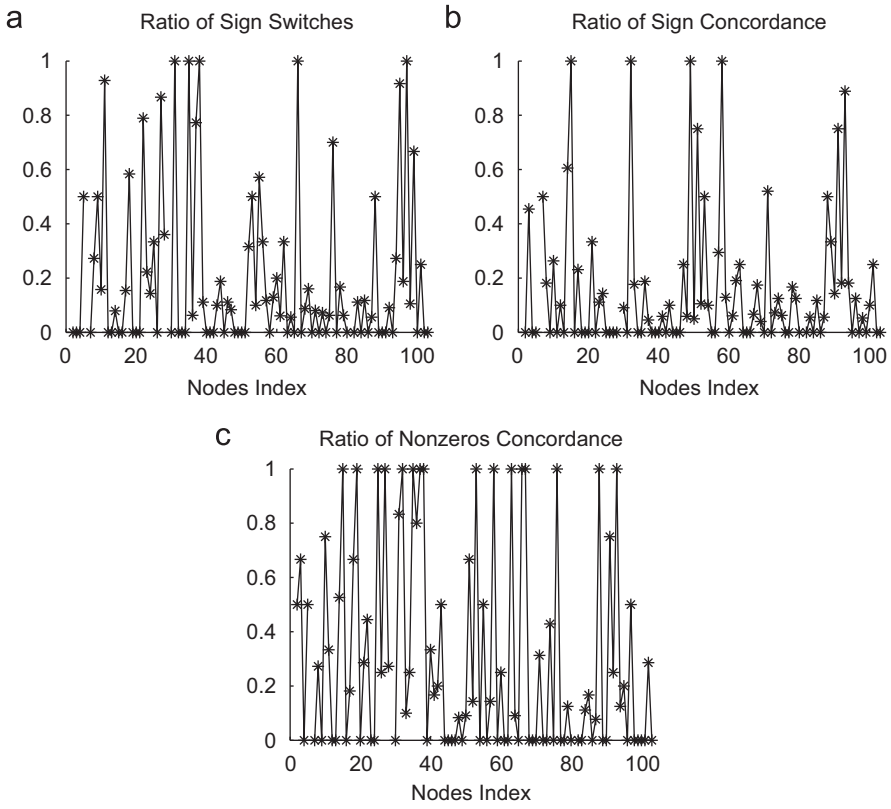

Fig. 6. (a)–(c) show, for each node of the reconstructed network, the ratio of nonzero median displacement values that display, respectively: (a) sign switching at the 20% threshold level; (b) sign concordance at the 20% threshold level; and (c) concordance of nonzero displacement values at the 40% threshold level. For each node, the threshold sets to zero all displacements that are below the given percentage of the largest displacement value.

a very similar qualitative behavior as in Figs. 5(a) and 6(a) is observed, as can be inferred from Figs. 5(b) and (6b).

In Figs. 5(c) and 6(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 Fig. 5(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 Fig. 5(a) and (b). Red curves are for the five nodes that have the highest number of combinations with corresponding nonzero median displacement values. Blue curves are generated from permuted median displacement curves.

In Fig. 5(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 and sign switching are more statistically significant than nonzero concordance for lower threshold. Fig. 6(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.

Appendix B. 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.[10] Suppose we are given N node variables from a network and that for each variable it is possible to generate R trajectories $X_{n,r}$ $r=1,\ldots,R$ 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 $\overline{X}_{n,r}$ the vector $X_{n,r}(t)-X_{n,r}(t_0)$ where t takes all L sampled values. For a given vector $g(t)$, $t=t_0,\ldots,t_L$, let $I(g)$ be the vector whose l-th component is the sum $\sum_{i=0}^{l} g(t_i)$.

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

Select a collection of potential target reactions $v_s=a_s x_{i_s}x_{j_s}-b_s x_{k_s}$, $s=1,\ldots,S$. The basic process to identify the links among the nodes is the following. For each node n with $n=1,\ldots,N$:

R1 Choose an attenuation coefficient β_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 $\hat{G}_l$ be the matrix whose columns are all the vectors $G_i/|G_i|$, $\hat{G}_q$ be the matrix whose columns are all possible vectors $G_{ij}/|G_{ij}|$. 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 $Z=[J\hat{G}_l,\beta_q\hat{G}_q,N_G]$ with small condition number (say less that 10^2).

R2 Set a temporary representation matrix M, for each $s=1,\ldots,S$, if the node n belong to the set $\{i_s,j_s,k_s\}$, add the vectors $G_{k_s}/|G_{k_s}|$ and $G_{i_sj_s}/|G_{i_sj_s}|$ as columns to the matrix M.

R3 Let $Z_M=[MN_G]$. Find the minimal l_1 solution to the underdetermined system $Y_n=Z_M\alpha_M$. Let α_M be the restriction of α 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 α associated to the columns of M the corresponding components of α_M.

R5 Choose a threshold T_n and let α_{T_n} be the coefficients in α larger than T_n. The reconstructed network equation for x_n will have only linear and quadratic terms that correspond to coefficients in α_{T_n}, and their coefficients will be the coefficients of α_{T_n} divided by the norm of the corresponding $|G_i|$, if a linear term, and $|G_{ij}|$ if a quadratic term.

In Napoletani 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 β_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.

[10] Actual code implementation in MATLAB is available upon request to dnapolet@gmu.edu.

References

Ailles, L.E., Weissman, I.L., 2007. Cancer stem cells in solid tumors. Curr. Opin. Biotechnol. 18 (5), 460–466.

Alam, S.M., Rajendran, M., Ouyang, S., Veeramani, S., Zhang, L., Lin, M.F., 2009. A novel role of Shc adaptor proteins in steroid hormone-regulated cancers. Endocr. Relat. Cancer 16 (1), 1–16 Epub 2008, November 11.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D., 1994. Molecular Biology of the Cell, third ed. Garland Publishing, New York.

Baselga, J., Arteaga, C.L., 2005. Critical update and emerging trends in epidermal growth factor receptor targeting in cancer. J. Clin. Oncol. 23 (11), 2445–2459.

Beretta, L., 2007. Proteomics from the clinical perspective: many hopes and much debate. Nat. Methods 4, 785–786.

Birtwistle, M.R., Hatakeyama, M., Yumoto, N., Ogunnaike, B.A., Hoek, J.B., Kholodenko, B.N., 2007. Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. Mol. Syst. Biol. 3, 144.

Blum, R.A., Wylie, N., England, T., French, C., 2005. HIV resistance testing in the USA and a model for the application of pharmacogenomics in the clinical setting. Pharmacogenomics 6, 169–179.

Chang, J.T., Carvalho, C., Mori, S., Bild, A.H., Gatza, M.L., Wang, Q., Lucas, J.E., Potti, A., Febbo, P.G., West, M., Nevins, J.R., 2009. A genomic strategy to elucidate modules of oncogenic pathway signaling networks. Mol. Cell 34 (1), 104–114.

Chen, S., Donoho, D., Saunders, M.A., 1998. Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. 20 (1), 33–61.

Davol, P.A., Bagdasaryan, R., Elfenbein, G.J., Maizel, A.L., Frackelton Jr., A.R., 2003. Shc proteins are strong, independent prognostic markers for both node-negative and node-positive primary breast cancer. J. Cancer Res. 63 (20), 6772–6783.

de Leon, J., Susce, M.T., Murray-Carmichael, E., 2006. The AmpliChip CYP450 genotyping test: integrating a new clinical tool. Mol. Diagn. Ther. 10, 135–151.

Ernst, T., Erben, P., Muller, M.C., Paschka, P., Schenk, T., Hoffmann, J., Kreil, S., LaRosee, P., Hehlmann, R., Hochhaus, A., 2008. Dynamics of BCR-ABL mutated clones prior to hematologic or cytogenetic resistance to imatinib. Haematologica 93 (2), 186–192.

Gerszten, R.E., Accurso, F., Bernard, G.R., Caprioli, R.M., Klee, E.W., Klee, G.G., Kullo, I., Laguna, T.A., Roth, F.P., Sabatine, M., Srinivas, P., Wang, T.J., Ware, L.B., 2008. Challenges in translating plasma proteomics from bench to bedside: update from the NHLBI clinical proteomics programs. Am. J. Physiol. Lung Cell. Mol. Physiol. 295, L16–L22.

Gibbons, J.D., Chakraborti, S., 2003. Nonparametric Statistical Inference, fourth ed. CRC Press Revised and Expanded.

Gilbert, C.A., Ross, A.H., 2009. Cancer stem cells: cell culture, markers, and targets for new therapies. J. Cell. Biochem. 108 (5), 1031–1038.

Gong, Y., Zhao, X., 2003. Shc-dependent pathway is redundant but dominant in MAPK cascade activation by EGF receptors: a modeling inference. FEBS Lett. 554 (3), 467–472.

Hornberg, J.J., Bruggeman, B.B.F.J., Schoeberl, B., Heinrich, R., Westerhoff, H.V., 2005. Control of MAPK signalling: from complexity to what really matters. Oncogene 24, 5533–5542.

Johnson, L.N., 2009. Protein inhibitors: contributions from structure to clinical compounds. Q. Rev. Biophys. 42, 1–40.

Jones, R.B., Gordus, A., Krall, J.A., MacBeath, G., 2006. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature 439, 168–174.

Kaddurah-Daouk, R., Kristal, B.S., Weinshilboum, R.M., 2008. Metabolomics: a global biochemical approach to drug response and disease. Annu. Rev. Pharmacol. Toxicol. 48, 653–683.
Kawamoto, K., Lobach, D.F., Willard, H.F., Ginsburg, G.S., 2009. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. BMC Med. Inf. Decision Making 9, 17.
Kitano, H., 2004. Biological robustness. Nat. Rev. Genet. 5, 826–837.
Kumar, A., Petri, E.T., Halmos, B., Boggon, T.J., 2008. Structure and clinical relevance of the epidermal growth factor receptor in human cancer. J. Clin. Oncol. 26 (10), 1742–1751.
Kumar, S., Kumar, S., Rajendran, M., Alam, S.M., Lin, F.F., Cheng, P.W., Lin, M.F., 2011. Steroids up-regulate p66Shc longevity protein in growth regulation by inhibiting its ubiquitination. PLoS One 6 (1), e15942.
Lewis, G.D., Asnani, A., Gerszten, R.E., 2008. Application of metabolomics to cardiovascular biomarker and pathway discovery. J. Am. Coll. Cardiol. 52, 117–123.
Lodish, H., Berk, A., Zipursky, A.L., Matsudaira, P., Baltimore, D., Darnell, J., 2000. Molecular Cell Biology, fourth ed. W.H. Freeman, New York.
Luzi, L., Confalonieri, S., Di Fiore, P.P., Pelicci, P.G., 2000. Evolution of Shc functions from nematode to human. Curr. Opin. Genet. Dev. 10 (6), 668–674.
Migliaccio, E., Mele, S., Salcini, A.E., Pelicci, G., Lai, K.M., Superti-Furga, G., Pawson, T., Di Fiore, P.P., Lanfrancone, L., Pelicci, P.G., 1997. Opposite effects of the p52shc/p46shc and p66shc splicing isoforms on the EGF receptor-MAP kinase-fos signalling pathway. EMBO J. 16 (4), 706–716.
Munsky, B., Trinh, B., Khammash, M., 2009. Listening to the noise: random fluctuations reveal gene network parameters. Mol. Syst. Biol. 5, 318.
Napoletani, D., Sauer, T., Struppa, D.C., Petricoin, E., Liotta, L., 2008. Augmented sparse reconstruction of protein signaling networks. J. Theor. Biol. 255 (1), 40–52.
Nelander, S., Wang, W., Nilsson, B., She, Q., Pratilas, C., Rosen, N., Gennemark, P., Sander, C., 2008. Models from experiments: combinatorial drug perturbations of cancer cells. Mol. Syst. Biol. 4, 216.
Nishizuka, S., Spurrier, B., 2008. Experimental validation for quantitative protein network models. Curr. Opin. Biotechnol. 19, 41–49.
Nishizuka, S., Ramalingam, S., Spurrier, B., Washburn, F.L., Krishna, R., Honkanen, P., Young, L., Shimura, T., Steeg, P.S., Austin, J., 2008. Quantitative protein network monitoring in response to DNA damage. J. Proteome Res. 7, 803–808.
Oda, K., Matsuoka, Y., Funahashi, A., Kitano, H., 2005. A comprehensive pathway map of epidermal growth factor receptor signaling. Mol. Syst. Biol. 1 2005.0010.
Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., Bryant, J., Costantino, J.P., Geyer Jr., C.E., Wickerham, D.L., Wolmark, N., 2006. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J. Clin. Oncol. 24, 3726–3734.
Park, E.S., Rabinovsky, R., Carey, M., Hennessy, B.T., Agarwal, R., Liu, W., Ju, Z., Deng, W., Lu, Y., Woo, H.G., Kim, S.B., Cheong, J.H., Garraway, L.A., Weinstein, J.N., Mills, G.B., Lee, J.S., Davies, M.A., 2010. Integrative analysis of proteomic signatures, mutations, and drug responsiveness in the NCI 60 cancer cell line set. Mol. Cancer Ther. 9 (2), 257–267.
Petricoin, E.F., Bichsel, V.E., Calvert, V.S., Espina, V., Winters, M., Young, L., Belluco, C., Trock, B.J., Lippman, M., Fishman, D.A., Sgroi, D.C., Munson, P.J., Esserman, L.J., Liotta, L.A., 2005. Mapping molecular networks using proteomics: a vision for patient-tailored combination therapy. J. Clin. Oncol. 23 (15), 3614–3621.
Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., Muller, G., 2002. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat. Biotechnol. 20, 370–375.
Sharma, S.V., Settleman, J., 2007. Oncogene addiction: setting the stage for molecularly targeted cancer therapy. Genes Dev. 21 (24), 3214–3231.
Uetz, P., Stagljar, I., 2006. The interactome of human EGF/ErbB receptors. Mol. Syst. Biol. 2 2006.0006.
Ulivieri, C., 2010. Cell death: insights into the ultrastructure of mitochondria. Tissue Cell 42 (6), 339–347.
Voit, E.O., 2000. Computational Analysis of Biochemical Systems. Cambridge University Press, Cambridge.
Weinstein, I.B., Joe, A., 2008. Oncogene addiction. Cancer Res. 68 (9), 3077–3080.
West, M., Ginsburg, G.S., Huang, A.T., Nevins, J.R., 2006. Embracing the complexity of genomic data for personalized medicine. Genome Res. 16, 559–566.
Willard, H.F., Ginsburg, G.S., 2009. Genomic and Personalized Medicine. Academic Press, San Diego, CA.
Willard, H.F., Angrist, M., Ginsburg, G.S., 2005. Genomic medicine: genetic variation and its impact on the future of health care. Philos. Trans. R. Soc. London Ser. B 360, 1543–1550.
Ye, Y.B., Lin, J.Y., Chen, Q., Liu, F., Chen, H.J., Li, J.Y., Liu, W.Q., Garbay, C., Vidal, M., 2008. The cytotoxicity of a Grb2-SH3 inhibitor in Bcr-Abl positive K562 cells. Biochem. Pharmacol. 75 (11), 2080–2091.
Zahorowska, B., Crowe, P.J., Yang, J.L., 2009. Combined therapies for cancer: a review of EGFR-targeted monotherapy and combination treatment with other drugs. J. Cancer Res. Clin. Oncol. 135 (9), 1137–1148.
Zhang, J., Yanga, P.L., Gray, N.S., 2009. Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer 9 (1), 28–39.