Published in final edited form as:
Cancer Res. 2011 August 15; 71(16): 5400–5411. doi:10.1158/0008-5472.CAN-10-4453.

COMPARING SIGNALING NETWORKS BETWEEN NORMAL AND TRANSFORMED HEPATOCYTES USING DISCRETE LOGICAL MODELS

Julio Saez-Rodriguez[1,2,3,5], **Leonidas G. Alexopoulos**[1,2,4,5], **MingSheng Zhang**[1], **Melody K. Morris**[2], **Douglas A. Lauffenburger**[2], and **Peter K. Sorger**[1,2]

[1]Dept of Systems Biology, Harvard Medical School, MA 02115

[2]Dept of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

Substantial effort in recent years has been devoted to constructing and analyzing large-scale gene and protein networks based on 'omic data and literature mining. These interaction graphs provide valuable insight into the topologies of complex biological networks, but are rarely context-specific and cannot be used to predict the responses of cell signaling proteins to specific ligands or drugs. Conversely, traditional approaches to analyzing cell signaling are narrow in scope and cannot easily make use of network-level data. Here we combine network analysis and functional experimentation using a hybrid approach in which graphs are converted into simple mathematical models that can be trained against biochemical data. Specifically, we created Boolean logic models of immediate-early signaling in liver cells by training a literature-based prior knowledge network against biochemical data obtained from primary human hepatocytes and four hepatocellular carcinoma cell lines exposed to combinations of cytokines and small-molecule kinase inhibitors. Distinct families of models were recovered for each cell type that clustered topologically into normal and diseased sets. Comparison revealed that clustering arises from systematic differences in signaling logic in three regions of the network. We also infer the existence of a new interaction involving Jak-Stat and NFκB signaling and show that it arises from the polypharmacology of an IκB kinase inhibitor rather than previously unidentified protein-protein associations. These results constitute a proof-of-principle that receptor-mediated signal transduction can be reverse engineered using biochemical data so that the immediate effects of drugs on normal and diseased cells can be studied in a systematic manner.

Keywords

liver; signal transduction; hepatocellular carcinoma; cancer; network inference; Boolean logic modeling

Quick Guide to Equations and Assumptions

We start with a Prior Knowledge Network (PKN) comprising a “signed and directed” node-edge graph that depicts interactions among proteins as arrows so that substrate-product and activation-inhibition relationships are captured: Raf→MEK→ERK for example. The 78

‡Please address correspondence to: Peter Sorger, WAB Room 438 Harvard Medical School, 200 Longwood Avenue Boston MA 02115, 617-432-6901/6902, peter_sorger@hms.harvard.edu, cc: Christopher_Bird@hms.harvard.edu.
[3]Current address: European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK
[4]Current address: Dept of Mechanical Engineering, National Technical University of Athens, 15780 Zografou, Greece
[5]These two authors contributed equally

node, 112 edge PKN in this paper came from the Ingenuity database with manual additions. A PKN cannot be compared directly to cell response data because a model is needed to specify input-output relationships. Here we use a simple Boolean formalism, in which the state of a nodes is either 0 or 1 ("off" or "on"), and nodes interact via *AND/OR/NOT* logical operators. In this formalism, if EGFR *OR* IR is active then MEK is active, etc. Conversion of a PKN into a Boolean model proceeds as follows. ***Network preprocessing*** specifies proteins to be measured or perturbed experimentally (so-called *designated nodes*) and then compresses the PKN based on two criteria (1): (a) nodes that are not directly or indirectly connected to designated nodes are eliminated (because we have no data on them) (b) cascades in which a undesignated nodes impinge on designated nodes are simplified; for example, if Raf and ERK are designated, but MEK is not, then Raf→MEK→ERK is replaced by Raf→ERK. ***Logical expansion*** computes all possible combinations of logic gates compatible with the compressed network. Each interaction in the PKN (hyperedges in graph theory) can give rise to multiple logical connections so that if two edges link into ERK (e.g. Raf →ERK and NFκB →ERK) we would generate three logic interactions: (i) Raf →ERK, (ii), NFκB →ERK and (iii) Raf *AND* NFκB →ERK. An *OR* gate (Raf *OR* NFκB →ERK) simply corresponds to (i) plus (ii). In our compressed PKN, 32 nodes and 128 logical interactions (hyperedges) give rise to $2^{128} = \sim 10^{38}$ possible models, each of which is evaluated against data. ***Training*** a family of possible models against data involves propagating the input signals along the logical network until all nodes reach inter-consistent values (a logical steady state); models are then compared to experimental measurements. Raw data are processed so that arbitrary intensity measures from xMAP sandwich immunoassays are converted into values between 0 and 1 based on various standards, a procedure that we have previously described in detail (1). For quantitative analysis of model/data match, the deviation (mean squared error; MSE) between data and a specific model is computed as $\mathrm{MSE}=\frac{1}{n_E}\sum_{k=1}^{s}\sum_{l=1}^{m}\sum_{t=1}^{n}(B^M_{k,l,t} - B^E_{k,l,t})^2$ where $B^M_{k,l,t} \in \{0, 1\}$ is predicted by the logical steady state of the model and $B^E_{k,l,t} \in [0, 1)$ is the discretized data for assay *l* recorded at time *t* under the k^{th} experimental condition. Values incompatible with a logical steady state are penalized as though they represent a mismatch between simulation and experimental data. We seek the simplest models consistent with data, using a bipartite objective function: $\Theta = \mathrm{MSE} + \alpha \Theta_S$ where Θ_S is model size and α is an adjustable parameter. We have shown that models identified using this objective function are similar in their numbers of edges and goodness of fit across a wide range of values for α (1); in the current paper we set α= 0.0001.

The training procedure consists of searching across 2^{128} models by the objective function using a standard genetic algorithm (2). Model training was iterated 50–100 times for each set of data. Because of the stochastic nature of the genetic algorithm and the non-identifiability of the models given data, different solutions were recovered each time. In common with most work on network inference, we addressed non-identifiability by analyzing families of models instead of single solutions; in the current work models within 1 % MSE of the best-fit model constituted the "consensus." For subsequent analysis, the distance between two sets of models *j* and *k* was computed as $d^2_{j,k}=\sum_{i=1}^{n}(f_i^j - f_i^k)^2$ where f_i^k is the frequency of the *ith* hyperedge in the *kth* model set. Distances were normalized with respect to the distance between hepatocyte and AvgHCC models (Table S1).

INTRODUCTION

The availability of high throughput interaction data has led to the creation of methods for summarizing and exploring networks using node-edge graphs. In these graphs, genes or proteins are represented by nodes (vertexes) and interactions by edges (3, 4)). The underlying interaction data are diverse and include manual or automated text mining of the literature (5, 6), genetic interactions obtained from gene deletion sets, and physical interactions identified by large-scale mass spectrometry or two-hybrid analysis (4, 7). Interactions in node-edge graphs can be undirected (denoting an unspecified interaction), directed but unsigned (denoting substrate-product relationships) or directed and signed (denoting both substrate-product and inhibition-activation relationships); the latter are particularly useful because they capture biochemical causality. For protein data, graphs comprising undirected edges are typically called Protein Interaction Networks (PINs) whereas those with signed directed edges are known as Protein Signaling Networks (PSNs). Most work on PINs and PSNs to date has focused on adding as much data as possible, often from more than one organism or type of experiment, so as to construct large networks with the greatest possible scope and the greatest number of interactions per node (increasing the "degree" of the network); the culmination of this effort is a proposed "Human Interactome" covering all known gene products (8).

In cancer biology, comparative analysis is the natural focus of "conventional" low-throughput studies of signal transduction with particular attention paid to differences in cellular responses to ligands or drugs in different cell types. In most cases, these differences reflect changes in the abundance or activity of signaling proteins (or of their substrates), features that could in principle be depicted by the strength of an edge in a network graph. However, existing PSNs and PINs do not encode the activities of proteins in cells that have been exposed to specific activators or inhibitors. A dearth of data on context-specific interactions makes it difficult to compare normal and diseased cells or diseased cells from different tumors. Cell- and state-specific information has been added to network graphs using gene expression data (3, 7, 9), but few attempts have been made to reconstruct comparative networks using biochemical data.

In this paper we attempt to combine concepts from global network discovery and traditional biochemistry by constructing comparative network models of signal transduction in normal and transformed liver cells. Starting with a prototypical network derived from the literature (which we will refer to as a prior knowledge network or PKN), we first constructed a set of all Boolean models compatible with the PKN, used the model "superstructure" to guide the collection of biochemical data on multiple nodes in the network across multiple cell types, and then trained the superstructure against data to uncover underlying differences in signaling logic among cell types. The net result is a computational representation of a signaling network that focuses on activity rather than literature association or physical interaction and that is explicitly comparative.

A first essential step in adding activity data to networks is to convert PKNs into models in which it is possible to compute input-output (I/O) characteristics (1). In this paper we use a two-state (Boolean) logical formalism in which each node can have only two states, *0* or *1*, but having a *1* at the output can depend on having a *1* at one of several inputs (an OR gate), all inputs (an AND gate), or *0* and *1* inputs in any combination. Boolean models have the advantage that they have no continuous free parameters and their topologies can be trained efficiently using data (1), a task that is harder with large differential equation models (10). However, we recognize that real biological systems exhibit dose response behavior that is only poorly approximated by Boolean logic. Thus, a major question at the outset of this work was whether the strengths of Boolean modeling with respect to computational

simplicity would outweigh its weaknesses. It seemed possible that the crudeness of the Boolean on/off approximation would overwhelm any differences we might measure experimentally from one cell type to the next. Conversely, success in creating comparative models would constitute a proof-of-principle for the approach.

We therefore applied Boolean modeling to distinguishing patterns of immediate early signaling in normal and transformed cells, represented here by primary human hepatocytes and HepG2, Hep3B, Focus and Huh7 liver cancer cell lines. Liver cancer (which is dominated by hepatocellular carcinoma [HCC]) is the third most common cause of cancer death in humans (11) and is known to involve alterations in the EGF-Ras-MAPK, AKT/mTOR, Jak/Stat and NFκB cascades (12). Thus, we aimed to collect multivariate data on the activities of these pathways in normal and transformed hepatocytes. We show that it is possible to assemble predictive network models that are specific to each cell type, cluster models based on topology and uncover consistent biochemical differences between transformed and normal cells. By identifying an interaction missing from the starting PKN but supported by data, we also uncover a poorly documented off-target effect of a drug being developed for asthma and inflammation (13, 14). Our findings demonstrate that discrete logical modeling can capture cell-type specific biochemical relationships, raising the possibility of constructing large comparative models of signal transduction in normal and diseased cells.

MATERIALS AND METHODS

Data generation

HepG2 and Hep3B, HuH7, and Focus cells were plated in 96-well plates coated with collagen type I (Becton Dickinson) with 100 µl phenol-free Williams' Medium E (WEM, Sigma-Aldrich) with supplements (15). Freshly pre-plated primary human hepatocytes were purchased from CellzDirect (Research Triangle Park, NC) and cultured overnight on collagen, starved for 6 hours in 95 µl of WEM, exposed to kinase inhibitors for 40 min. and then to ligands for 25 min. All cells were lysed in 90 µl of manufacturer's xMAP buffer (Bio-Rad) and intracellular signals measured using a Luminex instrument (Luminex, Austin, TX) and 16-plex phospho-protein bead sets (Bio-Rad; see Supplementary material). Different dilutions of cell extract were required for all 16 signals to be in the linear range.

HepG2 and Hep3B cell lines were obtained directly from ATCC; Huh7 and Focus cells (16), which are not available from ATCC, were obtained from Prof. Jack Wands. Cells were expanded once to create master stocks from which working cultures were periodically established; no lines were serially passage longer than three weeks. ATCC cell lines are validated by the provider (17); no testing was performed on the other lines. Freshly plated human hepatocytes were tested for contaminants as described in (15).

Reagents

To minimize experimental variability, samples were processed in parallel and common stocks of cytokines, inhibitors, and assay reagents used throughout (see Supplementary materials). Recombinant Jak2 and IKK-2 obtained from Cell Signaling Technology were preincubated with kinase inhibitors for 5 min. in manufacturer's buffer with 20µM ATP and FLT3 as Jak2 substrate or Iκb as a IKK-2 substrate. Phosphorylation was assayed using a time resolved fluorescence plate reader (PerkinElmer Victor 3) and apparent IC50 calculated from the activity A, where $A = (I + [drug] / IC50^{H})^{-1}$ and H is the apparent Hill coefficient. K_i for TPCA-1 was calculated using the Cheng-Prusoff equation, $K_i = IC50 / (1+[ATP] / K_M)$ where K_M (for ATP) was 0.43µM for Jak2 and 0.91 µM IKK-2.

Data handling

Data was processed and visualized using *DataRail* software (18) (19) with xMAP measurements normalized to a value between 0 and 1 (1).

Network construction and model calibration

The starting Protein Signaling Network (PKN) was constructed with ProMoT (20) using the database of Ingenuity Systems (21) supplemented manually. Modeling was then performed using our MATLAB toolbox *CellNetOptimizer* (*CellNOpt*; (22).

RESULTS

The dynamics of immediate-early signaling pathways was probed using a combinatorial experimental protocol (23): primary human hepatocytes and four HCC lines were treated with IL1α, IL6, TGFα, TNFα and Insulin (15), in the presence and absence of small molecule kinase inhibitors of IKK, MEK and PI3K and 16 intracellular signaling proteins were then assayed in cell extracts using multiplex sandwich immuno-assays (xMAP assays; Luminex Inc. Fig. 1 and S1; (15). Our use of kinase inhibitors and phospho-protein antibodies naturally focused the analysis on the druggable kinome (24) but similar analysis with other classes of drugs and signaling proteins is also possible. The experiments generated five sets of ligand-response data (for HepG2, Focus, Hep3B, Huh7 cell lines; “hepatocytes” refers to primary human hepatocytes); a sixth dataset (AvgHCC) was synthesized by arithmetically averaging data from the four tumor lines and attempts to capture biochemistry common to all cell lines.

All Boolean models compatible with a PKN of receptor-mediated signaling that included 78 nodes and 112 interactions was processed using freely available software of our own design, *CellNetOptimizer* (*CellNOpt*; (1). This yielded an ensemble of ~10^{38} models having 128 AND or OR gates with different connectivity or logic. The ensemble of ~10^{38} models was compared to each of six sets of experimental data (representing five cell types and the average) using a bipartite objective function that minimized the deviation between model and data while penalizing model size (Fig. 2). For any single data set optimization returned multiple models that differed slightly in topology and logic but had nearly the same value of the objective function (therefore making them indistinguishable with respect to data). Such “non-identifiability” is common in network inference and we therefore retained a family of best-fit models for each cell type differing by 1% in goodness of fit (see equations box for details).

Because Boolean models lack continuous parameters (akin to the rate constants in an differential equation network), it is not necessarily true that training will yield a model having a substantially better fit to data than the PKN, but this was the case with our data and models: the untrained ensemble containing all possible interactions and logic exhibited a poor fit (39–47 % MSE error across all data sets, Fig. 3) whereas families of trained models exhibited much better fit (9% to 13 % MSE error, see Figs. 3 and S2). We performed cross-validation and statistical tests to show that trained models were predictive of real data and were non-random (Figs. S3, S4, S5). Moreover, models trained against individual HCC cell lines were all better than the starting ensemble at predicting AvgHCC data (which was not used in training; 10–14 % MSE error, Fig. 3a). When the fit between models and data corresponding to individual ligands or biochemical assays was examined, levels of error were relatively low except in the case of p53 and IRS1s, which exhibited poor fits across all conditions (Fig. 3b and S6). This almost certainly arises because the PKN represents p53 biology in an imprecise manner and annotation of IRS1 modification is incomplete. These

are areas in which improved PKNs would clearly be useful. Nonetheless, we conclude that model training recovers substantially better network representations that the starting PKN.

Signaling network properties determined from data-trained logical models

In our procedure, training a PKN-based Boolean model against data improves the goodness of fit by removing unused edges. However, connectivity varied significantly with cell type: 85 of 128 possible gates were present in the superposition of all models in all cell types (involving ~90% of the interactions not removed in the preprocessing), but only seven gates were common to all models (Figs. 4 and S7). We therefore concluded that the primary deficiency of the literature-derived PKN with respect to our data is not the presence of true false-positive interactions (since some support can be found for most edges in data) but rather an absence of cell-type specificity.

To compare the topologies of models for all six data sets we computed pairwise distances by enumerating edges that differed between averaged best-fit models (see methods and Table S1 for details). HCC-derived models clustered together, with models built from AvgHCC data in the middle of the cluster, and well separated from models of primary hepatocytes (Fig. 5a). Models of Focus cells were farthest from primary cell models and HepG2 models were closest, in agreement with a classification of HCC lines proposed previously (25–27) (Fig. 5b). Moreover, the pattern of clustering derived from network topology was generally similar to pattern computed from transcriptional profiles. While the goal of logical modeling is not to generate cluster diagrams (being focused instead on the biochemistry of signal transduction) the similarities between clusters generated using transcript profiling and logical models suggest that the biochemical processes covered in our networks are representative of broader differences across cell lines.

Pathway differences in signaling networks among primary and transformed cells

Next we asked which interactions or logical gates were consistently present or absent when all possible models for one cell type were compared to all models of another cell type. This is a conservative approach that accounts for the inability of training to uniquely specify a model for each cell type based on available data. We observed that one interaction was absent from all models of HCC cells while being present in all models of primary hepatocytes whereas six interactions had the opposite property, being present only in models of HCC cells (Fig. 6). These differential interactions affected three regions of the signaling network. First, whereas the EGFR ligand TGFα caused ERK activation in all cell types, up-regulation of Hsp27-S78 phosphorylation (presumably by PRAK kinase, which lies downstream of ERK) was observed only in primary cells (differential interactions 1–2). In two of four HCC cell lines (Focus and Huh7) Hsp27-S78 was phosphorylated to a significant degree but it was p38-rather than ERK-dependent (Fig. S7). These findings are consistent with a reported association between low levels of Hsp27-S78 phosphorylation and tumor progression in HCC our data shows that the situation is more complex than previously thought (28) potentially involving a switch in Hsp27 kinases. A second significant difference between primary and HCC cells involved a change in the inferred logic of the IKK-NFκB pathway: in primary hepatocytes Iκb-S32/S36 phosphorylation (a signal for Iκb degradation and consequent nuclear localization of NFκB) required TNFα and an activator of PI3K such as TGFα (differential interactions 3–4). In contrast, in HCC lines only TNFα was required, implying differential control over canonical NFκB-mediated signaling.

The third significant difference involved phosphorylation of PI3K/AKT and GSK3-S9/S21 in Insulin-treated HCC cells but not in primary hepatocytes (differential interactions 5–8). AKT is a potent pro-survival kinase and its phosphorylation of GSK3 on S9/S21 is known to down-regulate GSK3 activity and promote nuclear localization of β–catenin, NFAT and

other pro-growth factors (29). Insulin receptor substrate 1 (IRS-1) is over-expressed in HCC cell lines (30) and it is possible that this shifts Insulin receptor (IR) signaling (or signaling by Insulin-like growth factor receptors, which are also present in these cells) from a metabolic function in normal liver (31) to a pro-survival function in tumors that involves elevated PI3K/AKT and GSK3 phosphorylation. Increased AKT activity is also expected in tumors due to downregulation of proteins such as the p85 subunit of PI3K, a common feature of HCC (32).

Overall we conclude that direct comparison of Boolean models was successful in identifying activity-dependent differences among normal and diseased cells that are hard to capture in PINs and PKNs assembled from physical interaction or steady-state proteomic and expression data. At the same time, we note that the Boolean models in this paper capture signaling at a single time point only, and only describe ligand-driven changes in phosphorylation. The absence of an IR→PI3K link in the inferred map for hepatocyte signaling (despite the known function of IR in the liver (32)) might arise because basal levels of AKT phosphorylation are high in these cells, making it difficult to detect ligand-dependent super-activation, or because we assayed cells at the wrong point in time (methods for incorporating time-series data into calibrated Boolean models are in development). However, Boolean modeling correctly inferred an EFGR→PI3K link in both transformed and primary cells and follow-up experiments suggest that there is indeed a greater propensity of tumor cells to respond to insulin by activating AKT.

Identification and testing of model-inferred novel pathway interactions

The model training described above focused on eliminating interactions present in the PKN but not supported by data. However, it is likely that PKNs also lack interactions that are supported by data. Indeed, we observed the single largest error in all models with respect to data involved an observed inhibition of Stat3-Y705 phosphorylation by the IKK inhibitor TPCA-1 under conditions of IL6 stimulation (Fig. 3b). TPCA-1 is reported to be a potent and selective inhibitor of human Iκbkinase-2 (IC_{50} = 18 nM for IKK-2 and 400nm for IKK-1) and was originally identified by GlaxoSmithKline in a drug-discovery effort focused on rheumatoid arthritis and airway inflammation (13, 14). The inhibition of Stat3-Y705 phosphorylation by TPCA-1 can be explained in a Boolean model by adding an interaction between IKK and Stat3 (see Fig. 4); this reduced the MSE error of all model families by ~5%. While an IKK → Stat3 edge might represent a physical interaction, an alternative explanation is that prior knowledge about TPCA-1 is incorrect and the drug actually inhibits Jak2, the known kinase for Stat3. To distinguish among these possibilities we performed in vitro activity assays of purified Jak2 and IKK-2 in the presence of TPCA-1 or BMS345541, an IKK-2 inhibitor having a distinct chemical structure (IC_{50}~300 nM). BMS345541 does not compete with ATP (it binds to homologous allosteric sites on IKK-1 and IKK-2) and presumably has different off-target effects. We found TPCA-1 to be nearly as potent an inhibitor of Jak2 *in vitro* (K_i ~9 nM) as of IKK-2 (K*i*~1.6 nM) its known target but BMS345541 was IKK-selective. Moreover, in IL6 stimulated cells, BMS345541 reduced phosphorylation of the IKK substrate Iκbα on Ser32/Ser36 but had no detectable effect on the level of phosphorylated Stat3- Y705 (Fig. 7). We conclude that Jak2 is a target of TPCA-1 (consistent with a recent kinome profile (33), and that Boolean network inference therefore identified a new target for the drug rather than a new protein-protein interaction.

DISCUSSION

Despite the relative crudeness of two-state logical approximations of biochemical reactions, this paper demonstrates that is possible to use Boolean modeling in combination with high-throughput cell-response data to automate discovery of biochemical differences in signal transduction among tumor and normal cell types. Apply the approach to primary human

hepatocytes and four HCC cell lines revealed consistent differences in the apparent logic and activities of growth factor receptor and intracellular kinase cascade in response to different ligands. Among the inferred differences between normal and transformed cells are several involving the strength or logic of signaling among IR, PI3K, AKT and NFκB, all molecules that have been implicated in the development of HCC. An unexpected pharmacological insight was the identification of Jak-Stat signaling as a target for TPCA-1, an IκB kinase inhibitor developed to treat arthritis and airway inflammation. Detecting this polypharmacology required comparison of a computable network model against data across a landscape of treatment conditions, thereby allowing multi-variate effects to be linked to specific causes. Intriguingly, TPCA-1 is significantly more potent than other IKK inhibitors in assays for airway inflammation. Both Jak2/Stat3 and IKK/NFκB play a role in inflammation (13, 14) and TPCA-1 would therefore appear to a “dirty” drug that is superior to a drug that binds specifically to the nominal target (34). More generally, the approach to modeling described in this paper may constitute a general means to study polypharmacology that is complementary to methods for investigating drug mechanism based on transcriptional data and protein interaction networks (35).

Our method focuses on eliminating interactions in the PKN that do not fit data. Because the number of potential edges in an ~80 node network exceeds ~10^{40} it is currently impossible to perform a comprehensive search for new edges that improve the fit to data. However, in the current work simple inspection sufficed to identify a potential AND-gated edge connecting IKK → Stat3 that was absent from the PKN. Implementing a rigorous approach to finding new edges will require efficient means to search models locally or to make more intelligent use of prior knowledge. A variety of network-inference methods other than Boolean logic are likely to be effective for analyzing biochemical data, including differential equations (36–38) Bayesian-networks (39) and fuzzy logic (40). Continued development of *CellNOpt* on our part as well as the creation of alternative modeling approaches fostered by efforts such as DREAM (41) is likely to improve the efficiency and accuracy of network inference from biochemical data. Moreover, improvements in affinity capture assays, protein arrays (42) and mass spectrometry (43, 44) should make it possible to collect much more extensive training data than we report here.

Deep sequencing and other approaches to genome analysis support the idea that it is networks and pathways rather than individual genes that are the functional objects of oncogenic mutation and selection in human cancer (45). However, approaches for constructing large-scale protein interaction graphs (4–6) are rarely cell-type specific, and traditional ‘bottom-up’ experiments that focus on cellular responses rather than network topologies are effective at uncovering such differences but they cannot easily incorporate network information in a formal way. Comparative network inference is needed to close the gap, but current efforts focus on gene regulatory networks, in large part because available data sets involve expression signatures, gene sequences and transcription factor binding sites (9, 46). Our results with Boolean logic and biochemical data constitute an encouraging proof-of-principle that the biochemistry of signaling networks, including the states and activities of proteins important in modern drug discovery, can be also be inferred and studied systematically. Because the product of our analysis is a computable model, it is amenable to continuous improvement and extension (with new data and interactions, for example) in much the same way as networks inferred from genome data. In contrast, it is difficult to account for new data using conventional, informal descriptions.

However, comparison of contemporary approaches to studying gene regulatory network in cancer (e.g. Carro et al. (9)) with our work serves to illustrate a fundamental difference between genome-scale data and “high throughput” biochemistry. Genomic data sets tend to contain many data points and identifying regulatory interactions and biologically meaningful

co-variation is a major challenge. In contrast, even “systematic” sets of biochemical data are much smaller. However, biochemical data are also pre-selected to contain relevant signaling information and the primary challenge is creating a framework that is effective at modeling relatively sparse data from multiple cell types. The rather narrow purview of biochemical models also makes it likely that many important responses are missed because they are not measured. The future clearly lies in creating hybrid models that fuse biochemical data on immediate early signaling with data on sequence variation, gene expression and transcription factor binding. Such models would provide unique insight into mechanisms of oncogenic transformation and would have practical uses in drug discovery and tumor classification. For example, it has been demonstrated that generic PINs and PKNs represent valuable prior knowledge for tumor classification from transcriptional profiles (47) and it seems likely that more accurate tumor-specific hybrid “interactomes” would prove even more useful in this context.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grants GM68762 and CA112967. We thank W. Chen, M. Niepel, J. Muhlich, S. Milton, S. Mirschel and members of Pfizer RTC for support and useful discussions.

References

1. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, et al. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. Mol Syst Biol. 2009; 5:331. [PubMed: 19953085]
2. Goldberg, D. Genetic Algorithms in Search, Optimization, and Machine Learning. Boston, MA, USA: Addison-Wesley; 1989.
3. Kim HD, Shay T, O'Shea EK, Regev A. Transcriptional regulatory circuits: predicting numbers from alphabets. Science. 2009; 325:429–432. [PubMed: 19628860]
4. Pieroni E, de la Fuente van Bentem S, Mancosu G, Capobianco E, Hirt H, de la Fuente A. Protein networking: insights into global functional organization of proteomes. Proteomics. 2008; 8:799–816. [PubMed: 18297653]
5. Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. Mol Syst Biol. 2009; 5:290. [PubMed: 19638971]
6. Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis A-R, Simonis N, et al. Literature-curated protein interaction datasets. Nat Meth. 2009; 6:39–46.
7. Przytycka TM, Singh M, Slonim DK. Toward the dynamic interactome: it's about time. Brief Bioinformatics. 2010; 11:15–29. [PubMed: 20061351]
8. Cusick ME, Klitgord N, Vidal M, Hill DE. Interactome: gateway into systems biology. Hum Mol Genet. 2005; 14:R171–R181. Spec No. 2. [PubMed: 16162640]
9. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, et al. The transcriptional network for mesenchymal transformation of brain tumours. Nature. 2009; 463:318–325. [PubMed: 20032975]
10. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. Nat Cell Biol. 2006; 8:1195–1203. [PubMed: 17060902]
11. Parkin DM, Bray F, Ferlay J, Pisani P. Global Cancer Statistics, 2002. CA Cancer J Clin. 2005; 55:74–108. [PubMed: 15761078]
12. Llovet JM, Bruix J. Molecular targeted therapies in hepatocellular carcinoma. Hepatology. 2008; 48:1312–1327. [PubMed: 18821591]
13. Podolin PL, Callahan JF, Bolognese BJ, Li YH, Carlson K, Davis TG, et al. Attenuation of Murine Collagen-Induced Arthritis by a Novel, Potent, Selective Small Molecule Inhibitor of IkB Kinase

2, TPCA-1 (2-[(Aminocarbonyl)amino]-5-(4-fluorophenyl)-3-thiophenecarboxamide), Occurs via Reduction of Proinflammatory Cytokines and Antigen-Induced T Cell Proliferation. Journal of Pharmacology and Experimental Therapeutics. 2005; 312:373–381. [PubMed: 15316093]

14. Birrell MA, Hardaker E, Wong S, McCluskie K, Catley M, De Alba J, et al. Ik-B Kinase-2 Inhibitor Blocks Inflammation in Human Airway Smooth Muscle and a Rat Model of Asthma. Am J Respir Crit Care Med. 2005; 172:962–971. [PubMed: 16002568]
15. Alexopoulos LG, Saez-Rodriguez J, Cosgrove B, Lauffenburger DA, Sorger PK. Networks inferred from biochemical data uncover differences in TLR signaling between normal and transformed hepatocytes. mol cell proteomics. 2010 inpress.
16. Lee HC, Tian B, Sedivy JM, Wands JR, Kim M. Loss of Raf kinase inhibitor protein promotes cell proliferation and migration of human hepatoma cells. Gastroenterology. 2006; 131:1208–1217. [PubMed: 17030190]
17. ATCC. http://www.atcc.org/Portals/1/Pdf/CellBiologyStandards.pdf.
18. Saez-Rodriguez J, Goldsipe A, Muhlich J, Alexopoulos LG, Millard B, Lauffenburger DA, et al. Flexible Informatics for Linking Experimental Data to Mathemati-cal Models via DataRail. Bioinformatics. 2008; 24:840–847. [PubMed: 18218655]
19. DataRail. Available from: http://code.google.com/p/sbpipeline.
20. Saez-Rodriguez J, Mirschel S, Hemenway R, Klamt S, Gilles ED, Ginkel M. Visual setup of logical models of signaling and regulatory networks with ProMoT. BMC Bioinformatics. 2006; 7:506. [PubMed: 17109765]
21. Ingenuity. Available from: http://www.ingenuity.com.
22. CellNOpt. Available from: http://www.ebi.ac.uk/saezrodriguez/software.html.
23. Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK. A compendium of signals and responses triggered by prodeath and prosurvival cytokines. Mol Cell Proteomics. 2005; 4:1569–1590. Epub 2005 Jul 18. [PubMed: 16030008]
24. Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov. 2002; 1:727–730. [PubMed: 12209152]
25. Fuchs BC, Fujii T, Dorfman JD, Goodwin JM, Zhu AX, Lanuti M, et al. Epithelial-to-mesenchymal transition and integrin-linked kinase mediate sensitivity to epidermal growth factor receptor inhibition in human hepatoma cells. Cancer Res. 2008; 68:2391–2399. [PubMed: 18381447]
26. Lee J-S. Thorgeirsson SS. Functional and genomic implications of global gene expression profiles in cell lines from human hepatocellular cancer. Hepatology. 2002; 35:1134–1143. [PubMed: 11981763]
27. Yuzugullu H, Benhaj K, Ozturk N, Senturk S, Celik E, Toylu A, et al. Canonical Wnt signaling is antagonized by noncanonical Wnt5a in hepatocellular carcinoma cells. Mol Cancer. 2009; 8:90. [PubMed: 19849855]
28. Yasuda E, Kumada T, Takai S, Ishisaki A, Noda T, Matsushima-Nishiwaki R, et al. Attenuated phosphorylation of heat shock protein 27 correlates with tumor progression in patients with hepatocellular carcinoma. Biochem Biophys Res Commun. 2005; 337:337–342. [PubMed: 16199221]
29. Frame S, Cohen P. GSK3 takes centre stage more than 20 years after its discovery. Biochemical Journal. [Review]. 2001; 359:1–16.
30. Khamzina L, Gruppuso PA, Wands JR. Insulin signaling through insulin receptor substrate 1 and 2 in normal liver development. Gastroenterology. 2003; 125:572–585. [PubMed: 12891559]
31. Saad MJA, Araki E, Miralpeix M, Rothenberg PL, White MF, Kahn CR. Regulation of Insulin-Receptor Substrate-1 in liver and muscle of animal-models of insulin resistance. Journal of Clinical Investigation. [Article]. 1992; 90:1839–1849.
32. The phosphoinositide 3-kinase regulatory subunit p85alpha can exert tumor suppressor properties through negative regulation of growth factor signaling. 2010
33. Bamborough P, Drewry D, Harper G, Smith GK, Schneider K. Assessment of Chemical Coverage of Kinome Space and Its Implications for Kinase Drug Discovery. J Med Chem. 2008; 51:7898–7914. [PubMed: 19035792]

34. Petrelli A, Giordano S. From single- to multi-target drugs in cancer therapy: when aspecificity becomes an advantage. Curr Med Chem. 2008; 15:422–432. [PubMed: 18288997]
35. Berger S, Iyengar R. Network Analyses in Systems Pharmacology. Bioinformatics (Oxford, England). 2009
36. Nelander S, Wang W, Nilsson B, She Q-B, Pratilas C, Rosen N, et al. Models from experiments: combinatorial drug perturbations of cancer cells. Mol Syst Biol. 2008; 4:11.
37. Mendoza L, Xenarios I. A method for the generation of standardized qualitative dynamical systems of regulatory networks. Theoretical biology & medical modelling. 2006; 3:13. [PubMed: 16542429]
38. Wittmann DM, Krumsiek J, Saez-Rodriguez J, Lauffenburger DA, Klamt S, Theis FJ. Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. BMC systems biology. 2009; 3:98. [PubMed: 19785753]
39. Sachs K, Gifford D, Jaakkola T, Sorger P, Lauffenburger DA. Bayesian Network Approach to Cell Signaling Pathway Modeling. Sci STKE. 2002; 2002 pe38-
40. Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger DA. Fuzzy Logic Analysis of Kinase Pathway Crosstalk in TNF/EGF/Insulin-Induced Signaling. PLoS Comput Biol. 2009; 5 e1000340.
41. Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, et al. Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges. PLoS ONE. 2010; 5:e9202. [PubMed: 20186320]
42. Ciaccio MF, Wagner JP, Chuu C-P, Lauffenburger DA, Jones RB. Systems analysis of EGF receptor signaling dynamics with microwestern arrays. Nat Methods. 2010; 7:148–155. [PubMed: 20101245]
43. Gstaiger M, Aebersold R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. Nat Rev Genet. 2009; 10:617–627. [PubMed: 19687803]
44. Jørgensen C, Sherman A, Chen GI, Pasculescu A, Poliakov A, Hsiung M, et al. Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. Science. 2009; 326:1502–1509. [PubMed: 20007894]
45. Jones S, Zhang X, Parsons DW, Lin JC-H, Leary RJ, Angenendt P, et al. Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. Science. 2008; 321:1801–1806. [PubMed: 18772397]
46. Prediction of a gene regulatory network linked to prostate cancer from gene expression, microRNA and clinical data. 2010
47. Chuang H-Y, Lee E, Liu Y-T, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol. 2007; 3:10.
48. GraphViz. Available from: www.graphviz.org.
49. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–2504. [PubMed: 14597658]

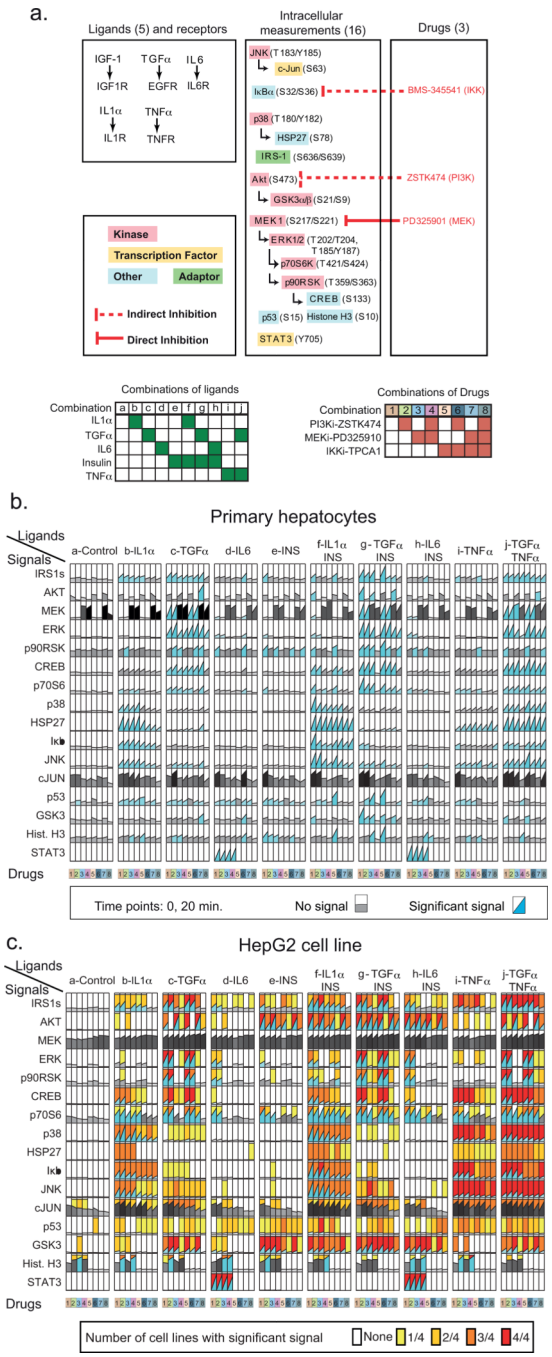

Figure 1. Experimental design and primary data

(**a**) Outline of the experimental approach showing five ligand cues and their receptors, 16 intracellular “signals” assayed using phospho-specific antibodies and xMAP technology (the relevant phosphorylated residues are indicated, as are the likely upstream kinases) and three small molecule kinase inhibitors **(b)** Dataset from human hepatocytes. Rows represent intracellular signals assayed immediately prior to ligand addition and 25 minutes thereafter, and columns represent different ligand combinations (see legend). For each combination of ligands, one of three kinase inhibitors was applied, as well as all possible combinations of them. Data are color-coded to highlight induction (relative to basal activity). **(c)** Data from HepG2 cells following the same design as in (a). The intensity of background color for each

box (yellow to red) indicates how the proportion of tumor cell lines that also responded to a particular drug-ligand combination; complete data can be found in Fig. S1. Data was processed using *DataRail* software (18).

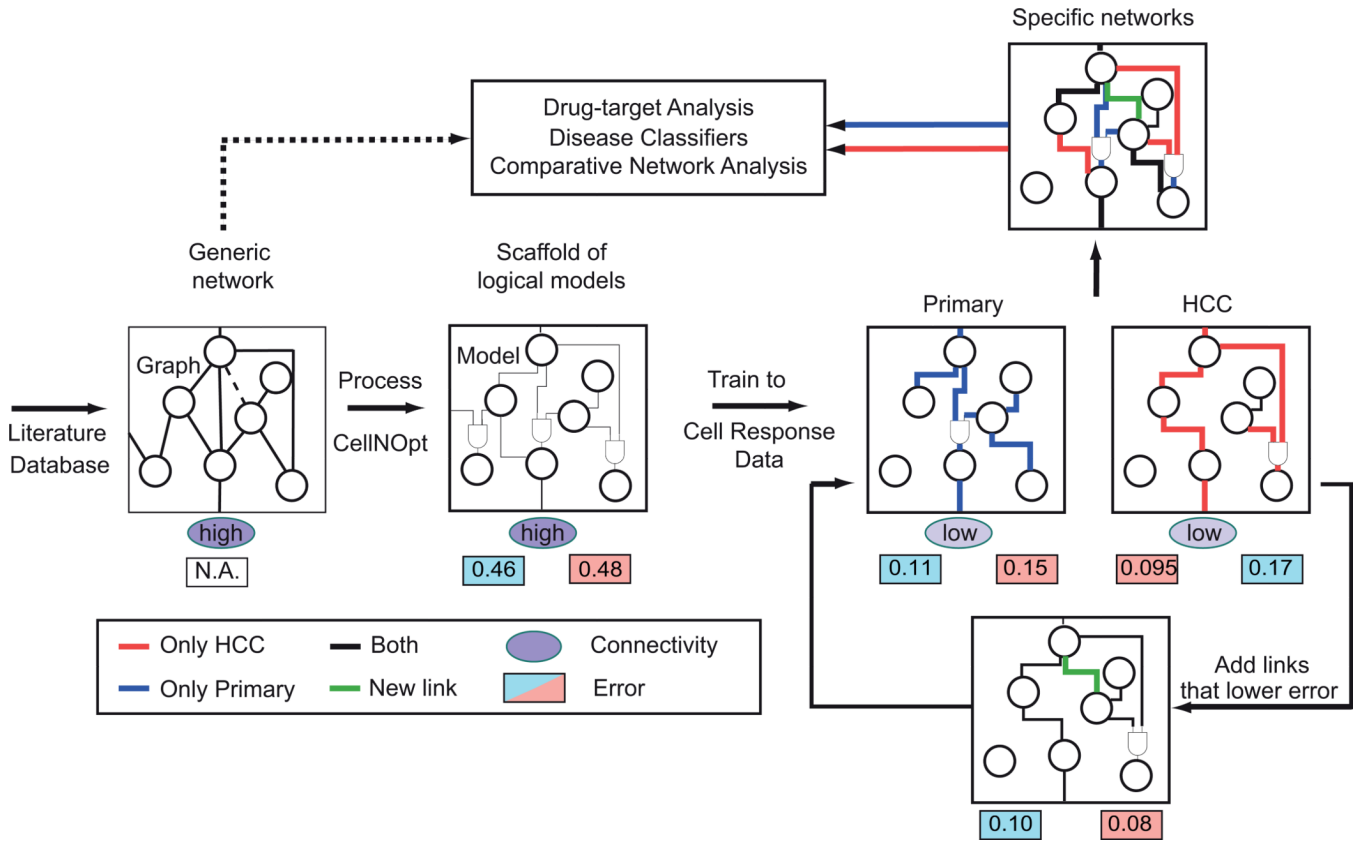

Figure 2. General workflow and goodness of fit of the cell-specific models

Schematic of workflow of model assembly, training, validation and extension. A PKN derived from interaction data is imported into *CellNOpt* and converted into an assembly of all possible logical interactions. The assembly is then trained against experimental data, generating cell-specific models having lower MSE error. In general, this involves removing interactions present in the PKN that are not supported by the data. Interactions absent from the PKN can then be added to see if they reduce residual error (green line). Once a satisfactory model is found, is is analyzed to identify differences among cell lines.

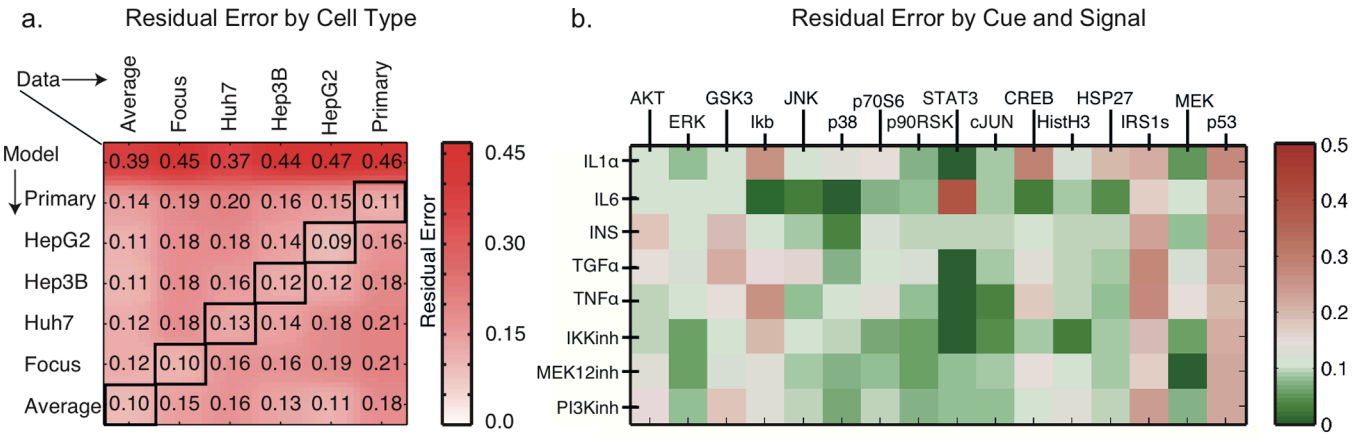

Figure 3. Residual error (MSE deviation, see methods) of the models

(**a**) MSE for the ensemble of all possible models compatible with the PKN and pairwise combinations of trained, cell-specific models with various data sets. The magnitude of the error is mapped to the intensity of the background (with red highest). Trained models fit corresponding data best, but all trained models exhibit lower error with all data sets than the starting ensemble. **(b)** Average MSE error of models across all cell types for specific experimental conditions and biochemical assays. For a single experiment, the error varies between 0 and 1; the maximal error was observed for phospho-STAT3 in cells treated with IL6 (MSE ~0.4).

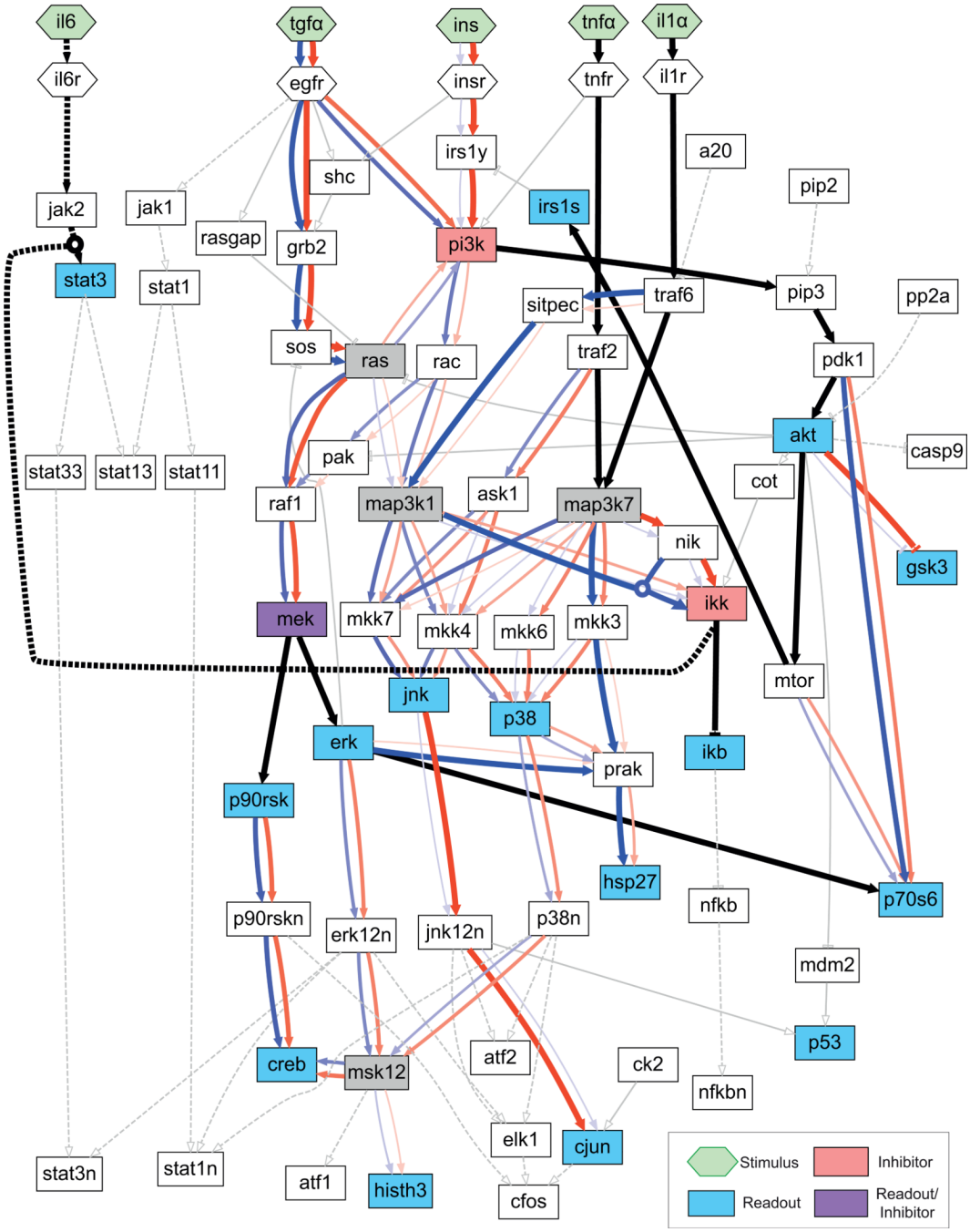

Figure 4. Inferred models of immediate-early signaling downstream in primary and transformed hepatocytes

A graph representing nodes and edges present in a set of trained Boolean models deviating by 1% in MSE and considered to be indistinguishable. The graph was created by exporting a GraphML file from from ProMoT (20) and modifying it with GraphViz (48). Line thickness denotes the frequency OR-gated edges found in trained models of primary hepatocytes (blue) or an average of the four HCC models (red); solid black edges were found in both primary and HCC cell lines. Thin grey edges were absent from all models, and dashed grey edges removed in the network preprocessing step (1). Green hexagons denote stimuli, red rectangles targets of kinase inhibitors, and blue rectangles readouts. Rectangles with

magenta fill were both measured and subjected to inhibition using small molecule drugs. Rectangles with white fill were compressed during graph processing. An AND gate is present upstream of IKK only in models of primary hepatocytes (see text), no other AND gate was consistently identified. The dashed black line denotes an interaction missing from the starting PKN but whose inclusion reduced the MSE of both hepatocyte and HCC models. This interaction implies inhibition of Stat3 phosphorylation by IKK in IL6-stimulated cells (see text).

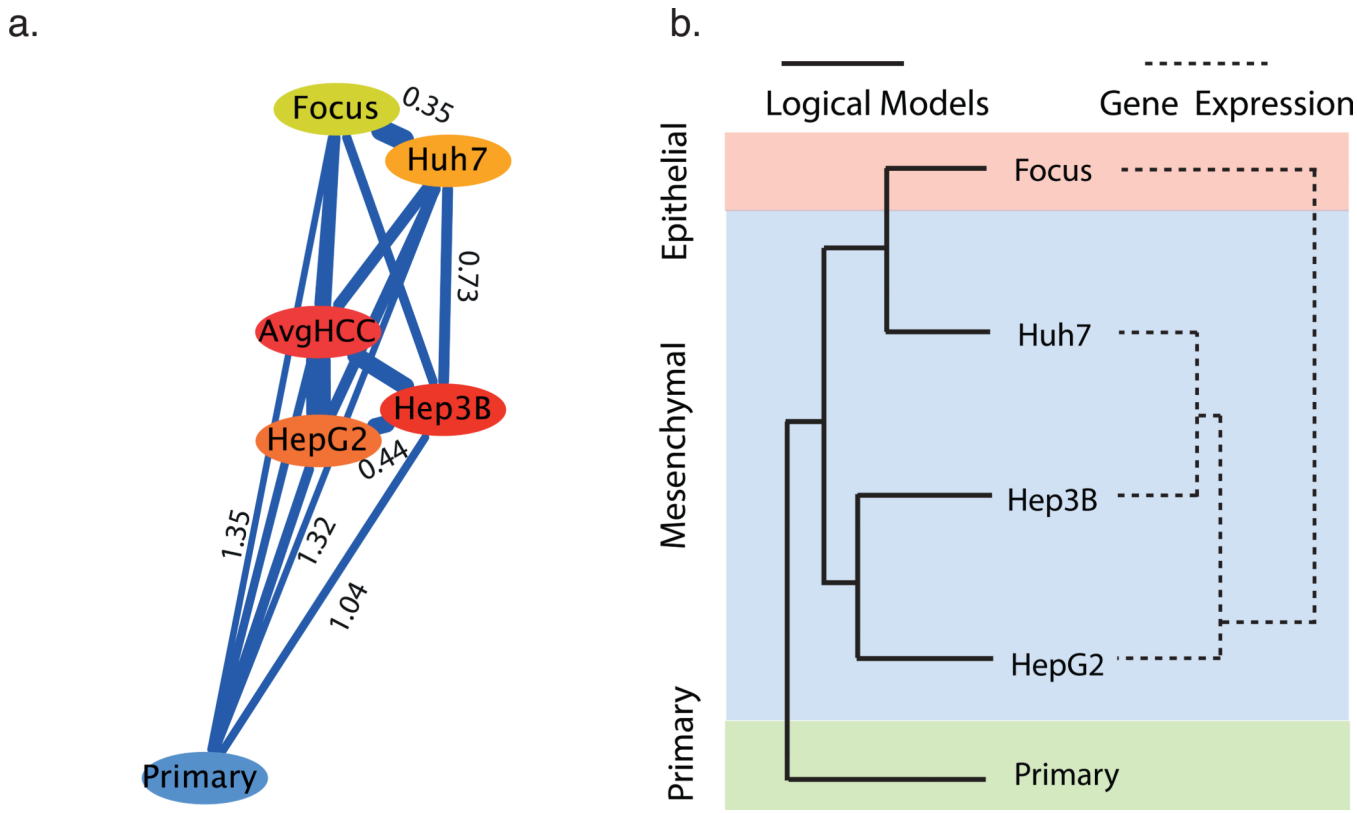

Figure 5. Topological clustering of logic-based models of primary and transformed hepatocytes

(a) Distances between primary hepatocytes and the four HCC cell lines (Huh7, Focus, Hep3B, and HepG2) and a 'average HCC cell line' (AvgHCC) based on the mean difference in the topologies of Boolean models as defined by the distance d^2 (see methods) and plotted using the edge-weighted spring embedded algorithm in Cytoscape (49). The full range of distances is listed in the supplementary materials. **(b)** Hierarchical clustering of primary and transformed hepatocytes based on the logical models, and a comparison to previous published clusters based on gene expression (the dendogram was manually redrawn from reference (26) and distances are approximate). A classification of cell lines as epithelial or mesenchymal (25) is also shown (see Fig. S8).

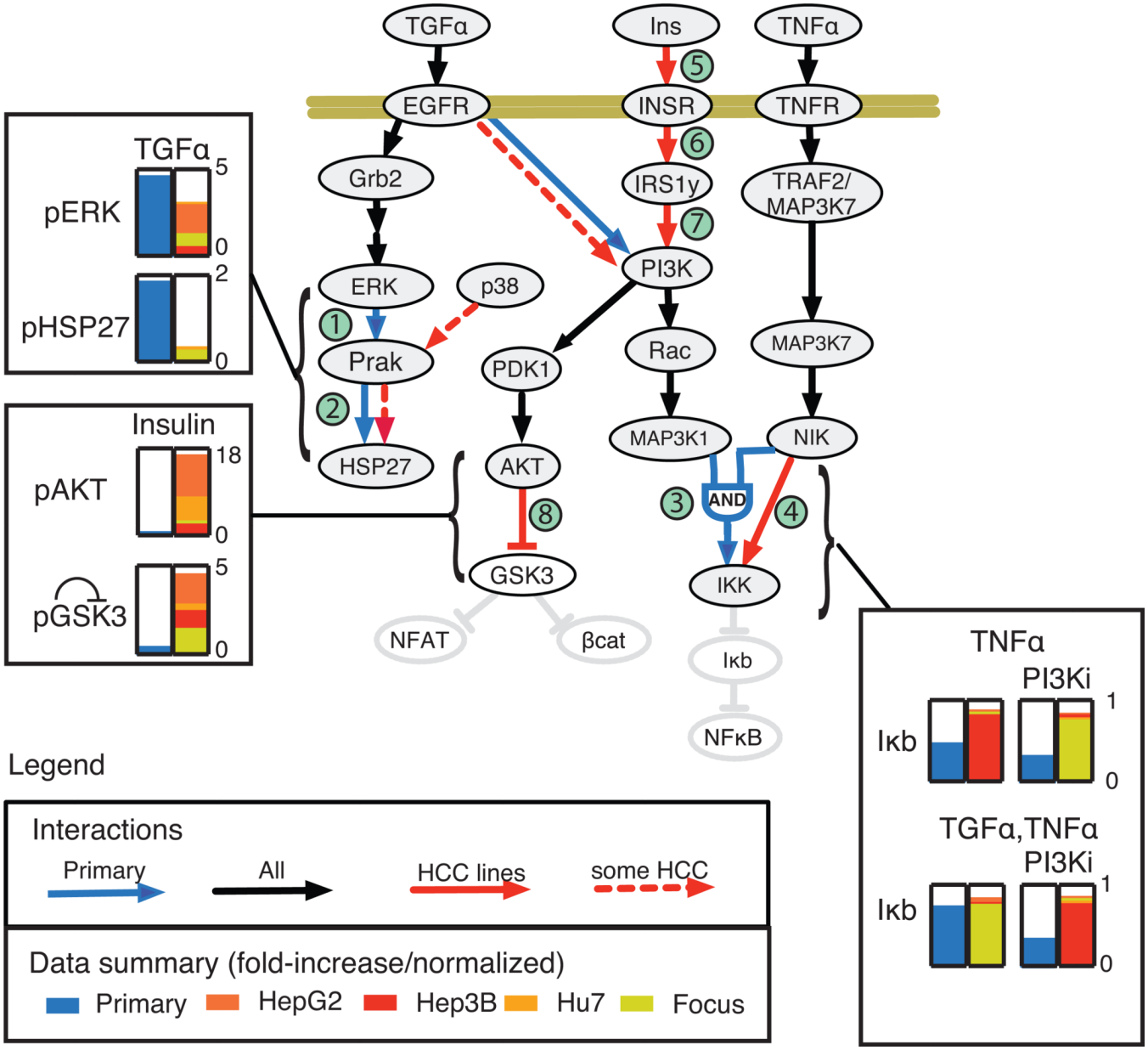

Figure 6. Key differences uncovered by logic-based modeling

Schematic of inferred differences between HCC cell lines and hepatocytes. The callouts show data associated with each topological difference with data from primary hepatocytes as blue bars and HCC as overlaid red bars. Fold-increase in actual values are shown, except in the case of IκB, where normalized data are compared to predictions from Boolean models (for clarity). Numbered green circles label differences between primary and HCC, cells.

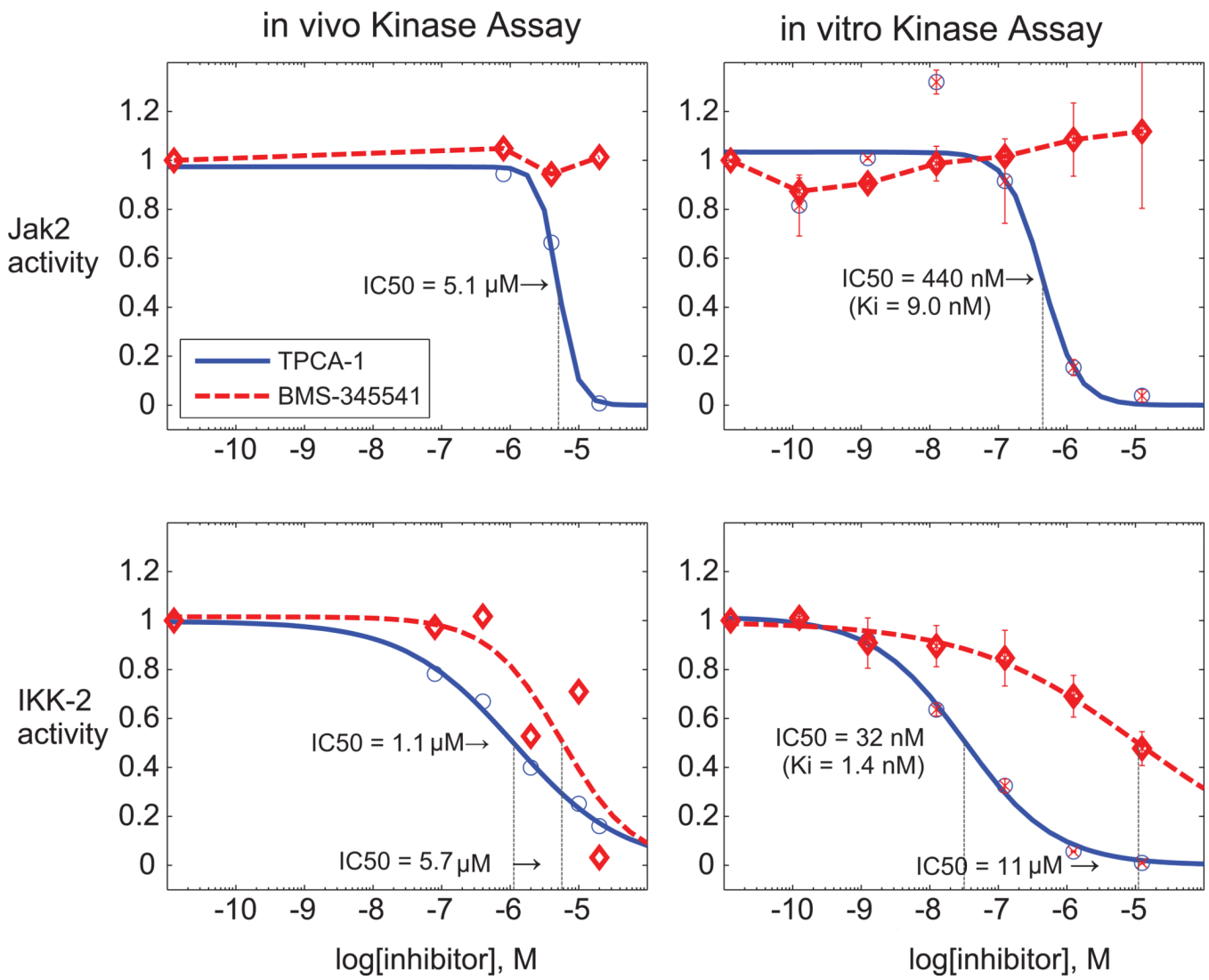

Figure 7. Probing the biochemistry of an inferred IKK-Stat3 interaction

For in vivo experiments, cells were pretreated with 2 to 20 μM TPCA-1or BMS345541. The concentration range of TPCA-1 was 10 fold lower than BMS345541 to match published IC50 values for inhibition of Iκb phosphorylation. Cells were pretreated with drug for one hour and then stimulated with IL6; the phosphorylation of Stat3Y705 was measured 25 minutes thereafter. In vitro experiments were performed with recombinant kinases and phosphorylation levels measured using a fluorescence assay.