Context-specificity in causal signaling networks revealed by phosphoprotein profiling

Steven M. Hill[1,12], **Nicole K. Nesser**[2,12], **Katie Johnson-Camacho**[2], **Mara Jeffress**[3], **Aimee Johnson**[4], **Chris Boniface**[2], **Simon E. F. Spencer**[5], **Yiling Lu**[6], **Laura M. Heiser**[7], **Yancey Lawrence**[2,13], **Nupur T. Pande**[8,9], **James E. Korkola**[7], **Joe W. Gray**[7,10,9], **Gordon B. Mills**[6], **Sach Mukherjee**[1,11,14,*], and **Paul T. Spellman**[2,15,*]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, CB2 0SR, UK

[2]Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR 97201, USA

[3]Kantar Health, Foster City, CA 94404, USA

[4]Bayer Healthcare North America, Berkeley, CA 94710, USA

[5]Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

[6]Department of Systems Biology, MD Anderson Cancer Center, Houston, TX 77030, USA

[7]Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR 97201, USA

[8]Department of Obstetrics and Gynecology, Women's Health Research Unit, Oregon Health and Science University, Portland, OR 97239, USA

[9]Knight Cancer Institute, Oregon Health and Science University, Portland, OR 97239, USA

[10]Center for Spatial Systems Biomedicine, Oregon Health and Science University, Portland, OR 97239, USA

[11]German Center for Neurodegenerative Diseases (DZNE), Bonn, 53127, Germany

Summary

Signaling networks downstream of receptor tyrosine kinases are among the most extensively studied biological networks, but new approaches are needed to elucidate causal relationships between network components and understand how such relationships are influenced by biological context and disease. Here, we investigate the context-specificity of signaling networks within a

---

[*]Correspondence: sach.mukherjee@dzne.de (S.M.), spellmap@ohsu.edu (P.T.S.).
[12]Co-first author
[13]Present address: New York Medical College, Valhalla, NY 10595, USA
[14]Present address: German Center for Neurodegenerative Diseases (DZNE), Bonn, 53127, Germany
[15]Lead contact

causal conceptual framework, using reverse-phase protein array time-course assays and network analysis approaches. We focus on a well-defined set of signaling proteins profiled under inhibition with five kinase inhibitors in 32 contexts—four breast cancer cell lines (MCF7, UACC812, BT20, and BT549) under eight stimulus conditions. The data, spanning multiple pathways and comprising ~70,000 phosphoprotein and ~260,000 protein measurements, provide a wealth of testable, context-specific hypotheses, several of which we experimentally validate. Furthermore, the data provide a unique resource for computational methods development, permitting empirical assessment of causal network learning in a complex, mammalian setting.

Introduction

The complexity of mammalian receptor tyrosine kinase (RTK) signaling continues to pose challenges for the understanding of physiological processes and aberrations that are relevant to disease. Networks, comprising nodes and linking directed edges, are widely used to summarize and reason about signaling. Obviously, signaling systems depend on the concentration and localization of their component molecules, so signaling events may be influenced by genetic and epigenetic context (Saez-Rodriguez et al., 2011; Good et al., 2009; Zalatan et al., 2012). In disease biology, and cancer in particular, an improved understanding of signaling in specific contexts may have implications for precision medicine by helping to explain variation in disease phenotypes or therapeutic response.

Genomic heterogeneity in disease has been well studied, notably in cancer, and heterogeneity is also manifested at the level of differential expression of components of signaling pathways downstream of RTKs (Akbani et al., 2014; Gerlinger and Swanton, 2010; Nickel et al., 2012; Szerlip et al., 2012). However, differences in average protein abundance (as captured in differential expression or gene set analyses) are conceptually distinct from differences in the edge structure of signaling networks, with the latter implying a change in the ability of nodes to causally influence each other. Causal relationships are also fundamentally distinct from statistical correlations: if there is a causal edge from node A to node B, then the abundance of B may be changed by inhibition of A, but A and B can be correlated with no causal edge linking them (see below for an illustrative example). For this reason, standard concepts from multivariate statistics (that in turn underpin many network analyses in bioinformatics) may not be sufficient for causal analyses (Pearl, 2009).

Canonical signaling pathways and networks (as described for example in textbooks and online resources) typically summarize evidence from multiple experiments, conducted in different cell types and growth conditions and therefore such networks are not specific to a particular context. Many well-known links in such networks most likely hold widely and so canonical networks remain a valuable source of insights. However, if causal signaling depends on context then using canonical networks alone will neglect context-specific changes, with implications for reasoning, modeling and prediction. A large literature has focused on the question of inferring molecular networks from data (see e.g. De Smet and Marchal, 2010 and Marbach et al., 2010 for reviews). The potential for molecular networks to depend on context has motivated efforts to tailor network models in a data-driven manner (Marbach et al., 2016; Petsalaki et al., 2015; Will and Helms, 2016). Our approach is in this

vein but with an emphasis on interventional data and a principled causal framework. Unbiased "interactome" approaches (e.g. Rolland et al., 2014) expand our view of the space of possible signaling interactions. However, due to the nature of genetic, epigenetic and environmental influences, such approaches cannot in general identify signaling events specific to biological context, e.g. specific to a certain cell type under defined conditions.

We study context-specific signaling using human cancer cell lines. The data span 32 contexts each defined by the combination of (epi)genetics (breast cancer cell lines MCF7, UACC812, BT20 and BT549) and stimuli. In each of the 32 *(cell line, stimulus)* contexts, we carried out time-course experiments using kinase inhibitors as interventions (note that as used here the inhibitors do not contribute to defining the context). Reverse-phase protein arrays (RPPA; Tibes et al. 2006) were then used to interrogate signaling downstream of RTKs. We used more than 150 high-quality antibodies targeting mainly total and phosphorylated proteins (see Table S1).

The inhibitors applied in each context allowed elucidation of context-specific causal influences between inhibited and downstream phosphoproteins. The extent of context-specificity seen can be summarized as follows: on average, across all kinase inhibitors and pairs of contexts in the study, approximately 1 in 5 phosphoproteins show changes in abundance under inhibition in one context that are not seen in the other. We also modeled the data using recently developed methods rooted in probabilistic graphical models to reconstruct context-specific networks intended to capture causal interplay between all measured phosphoproteins (and not just interplay related to inhibited nodes).

Thus, we show that causal signaling networks depend on context, with the pattern of changes under inhibition dependent on biological background. This is supported by independent validation experiments. Furthermore, we advance a conceptual view of signaling networks as causal networks (Pearl, 2009). In addition, this paper adds to available resources in two ways. First, the data we present acts as a valuable data resource, spanning all combinations of context, inhibitor and time, and allowing for a very wide range of analyses, including, but not limited to, analyses of the kind presented here. The data complement available patient datasets (see e.g. Akbani et al., 2014) by providing interventional readouts under defined conditions and provide a wealth of testable hypotheses regarding potentially novel and context-specific signaling links. Second, the data serve as a resource for computational biology benchmarking. Network reconstruction has long been a core topic in computational biology but performance with respect to learning of causal links has mainly been benchmarked using simulated data that may not adequately reflect the challenges of real data and relevant biology. A previous study established a small, five node synthetic network in yeast that was valuable to the computational biology community as it provided a gold-standard network in a biological model (Cantone et al., 2009). The design of our experiments allows for systematic testing of causal network learning in a complex mammalian setting and provides a unique resource for development of computational biology methods. The data presented here were used in the recent HPN-DREAM (Heritage Provider Network – Dialogue for Reverse Engineering Assessment and Methods) network inference challenge. The challenge focused on causal networks and the data were used to

score more than 2000 submitted networks (full details of the challenge are described in Hill et al., 2016).

Results

Causal molecular networks and context-specificity

We first define causal molecular networks at a conceptual level. Consider a specific cell line grown under defined conditions. We refer to the complete biological setting (including genetic/epigenetic background and growth/environmental conditions) as the context c. If, in this setting, we observe a change in molecule B under inhibition of molecule A, we can conclude that there exists a causal pathway (i.e. a sequence of mechanistic events, possibly involving additional molecular species) between A and B in context c. Conceptually, performing all possible inhibition experiments on a set of molecules (including in combinations) would allow construction of a directed network G_c, with nodes corresponding to the molecules, and edges encoding causal relationships between nodes. Specifically, an edge in G_c indicates that in context c, inhibition of the parent node can lead to a change in the child node that is not mediated via any other node in the network. We refer to G_c as the context-specific causal network and to edges therein as causal edges (Figure 1A).

Due to the large number of potentially relevant molecular species, it is likely that in any specific study there will be variables that are unmeasured but that nonetheless have a causal influence on one or more measured variables. Suppose there is no causal pathway between A and B, but the nodes are correlated due to co-regulation by an unobserved node C that is not represented in the graph (Figure 1B). Then, since inhibition of A would not be capable of changing B, an edge from A to B would not be contained in the ground truth network G_c as defined above, regardless of the strength of any correlation or statistical dependence between A and B (Figure 1C). A contrasting case is that of a missing variable that is intermediate in a causal pathway, e.g. if A influences B via an unmeasured molecule C. Then, using the definition above, we would consider the edge A->B to be a correct representation of the causal influence. However, if C were observed, the correct model would be A->C->B (Figure 1C). Thus, the definition we use is compatible with missing variables whilst correctly encoding the effect of interventions on observed nodes, but the edges are not intended to encode physically direct influences only. We note that there are many subtle and still open aspects of the epistemology of interventions and causation; for a wider discussion see Woodward (2013).

The definition of causal molecular networks above is rooted in changes under inhibition but is not restricted to any particular mechanism. We focus on kinase inhibitors, phosphoprotein nodes and relatively short-term changes (up to four hours after inhibition) and to that extent our focus is on signaling, but we note that changes seen in our data could be due to a number of mechanisms, including transcription, translation or protein stability. In considering causal influences, it is important to specify a relevant timeframe, because under the same intervention, different changes may occur over different time periods (see also Discussion). Note also that even if one assumes a very large sample size and neglects statistical issues entirely, a notion of magnitude (of change under inhibition) remains implicit in the network definition itself and influences the sparsity of the ground truth network.

Overview of approach

We sought to investigate causal signaling networks in specific biological contexts. We considered four breast cancer cell lines (MCF7, UACC812, BT20 and BT549) derived from distinct epigenetic states and harboring different genomic aberrations (these cell lines have been extensively characterized; see Barretina et al., 2012; Garnett et al., 2012; Heiser et al., 2012; Neve et al., 2006). Each cell line was serum starved for 24 hours and then at time t=0min stimulated with one of eight different stimuli (Figure 2A). For each *(cell line, stimulus)* context, we carried out RPPA time-course assays comprising a total of seven time points spanning four hours, and under five different kinase inhibitors plus DMSO as a control (Figure 2A and STAR Methods; the assays included additional, later time points that were not used in our analyses, but are available in Data S1). To ensure that targets of the kinase inhibitors were effectively blocked, cells were treated with inhibitors for two hours before stimulus. Low concentrations of each inhibitor were used to minimize off-target effects (see STAR Methods). Due to the functional significance of phosphorylation, the analyses presented below focus on the 35 phosphoproteins that were measured in all cell lines (see STAR Methods and Table S1; Data S1 contains measurements for all antibodies). Context-specific changes under intervention were summarized as *causal descendancy matrices* (Figure 2B; see below). Machine learning methods were used to integrate the interventional data with known biology to reconstruct context-specific signaling networks (Figure 2C).

Interventional time-course data specific to biological context

Comparing time-course data between inhibitor and control (DMSO) experiments allowed us to detect changes to phosphoprotein nodes caused by kinase inhibition (see STAR Methods for details). These changes are visualized in a global manner for cell line MCF7 in Figure 3B, with DMSO time courses shown in Figure 3A. In Figure 3B, the color coding indicates direction of effect (see examples in Figure 3C): green indicates a decrease under inhibition relative to control (consistent with positive regulation) and red an increase under inhibition (consistent with negative regulation). Corresponding visualizations for UACC812, BT20 and BT549 are shown in Figure S1.

Many effects, including many classical ones, are not stimulus-dependent. For example, phospho-p70S6K is reduced relative to control under mTOR inhibition (inhibitor AZD8055; Figure 3C), in line with the known causal role of mTOR in regulating phosphorylation of p70S6K. Since mTOR signaling is already active in serum starved cells, the reduction in phospho-p70S6K under mTOR inhibition is seen at all time points, including t=0min (recall that the inhibitor is applied prior to stimulus). However, some changes under intervention are specific to individual stimuli. Some of these effects can be readily explained, such as the reduction in abundance of several phosphoproteins in the AKT and MAPK pathways under FGFR inhibition (inhibitor PD173074) for cell line MCF7 stimulated with FGF1. Other stimulus-specific changes are less expected, including the decrease in abundance of phospho-AKT (phosphorylated at threonine 308) in cell line MCF7 under inhibition of mTOR/PI3K (inhibitor BEZ235) that is observed in only four of the stimuli.

Causal descendancy matrices summarize changes under intervention across multiple contexts

Changes seen under inhibition of mTOR (catalytic inhibitor AZD8055) are summarized in Figure 4A (with phosphoproteins in rows and the 32 contexts in columns). Here, a filled-in box for phosphoprotein p in context c indicates a salient change under mTOR inhibition (see STAR Methods), consistent with a causal influence of mTOR on phosphoprotein p in context c. This could occur via a causal pathway involving other (measured or unmeasured) nodes. In other words, an entry in location (p,c) in the matrix indicates that phosphoprotein p is a *descendant* of mTOR in the causal network G_c for context c; we therefore refer to this matrix as a *causal descendancy matrix* for mTOR. For comparison, an additional column shows proteins that are descendants of mTOR according to a canonical signaling network (Figure 4B; STAR Methods). Many classical signaling links are conserved across cell lines and stimuli. But there are also many examples of influences that are both non-canonical and context-specific; for example, phospho-p38 is elevated in UACC812 cells treated with the mTOR inhibitor AZD8055 under serum stimulation whereas there is no change in BT549 cells under the same conditions. Similarly, we obtained causal descendancy matrices for each of the other inhibitors in our study (Figure S3). On average across all kinase inhibitors and pairs of contexts, 8 out of 35 phosphoproteins show salient changes under inhibition in one context that are not seen in the other (mean number of differences = 8.14). Considering only pairs of cell lines under the same stimulus the mean number of differences is 8.58, while considering pairs of stimuli for the same cell line the corresponding value is 6.38. This suggests that the differences in (epi)genetic background between the cell lines have a relatively pronounced effect.

We sought to validate some of the observed causal effects by western blot analysis (STAR Methods). Observations were selected for validation across both inhibitors and antibodies, and included instances of increase and decrease under inhibition, as well as instances where no effect was observed (Table S2). A summary of the number of observations tested for each cell line and inhibitor regime, and of validation success rate in independent experiments (i.e. new lysates) is shown in Figure 4C. Overall we validated 78% of observations tested (104 out of 134 observations). There were 25 *(antibody, inhibitor)* combinations that for the same stimulus showed differing effects across cell lines in the RPPA data (and which were also tested by western blotting); 17 of these instances of heterogeneity across cell lines validated (68%). The corresponding validation rate for *(antibody, inhibitor)* combinations that for the same cell line showed differing effects across stimuli was only 3 out of 13 (23%). Failures to validate could represent biological variability, differential sensitivity between RPPA and western blotting, use of different antibodies or other technical issues.

Machine learning of signaling networks

We used dynamic Bayesian networks to learn context-specific causal networks over all measured phosphoprotein nodes (including those not intervened upon). To do so, we exploited several recent methodological advances that allow integration of interventional data and simultaneous network learning across multiple related problem instances (here, contexts; see STAR Methods and references therein for details). Known biology was incorporated using a prior network (Figure S4).

Figure 5 summarizes networks across all contexts by averaging across the eight stimulus-specific networks for each of the four cell lines. We see that while many edges, including several classical ones, are near universal, others are cell line-specific, mirroring, via a global analysis, the inhibition data reported above (Figure 4A). The networks contained edges included in the prior network as well as many edges that were not. Across the 32 contexts, networks contained an average of 49 edges (at a threshold of 0.2 applied to the edge probabilities that are the output of the learning procedure) and, on average, 40% of edges in each network were not in the prior network (Table S3). We discuss potentially novel edges that were not in the prior below. As discussed in Hill et al. (2016), the challenging nature of causal network learning means that empirical performance assessment is important. We used an extended variant of the train-and-test procedure described in Hill et al. (2016) to systematically assess causal network learning (see STAR Methods). We found that the models were able to achieve significant agreement with unseen test interventional data in most of the contexts (Figure S5). However, we note that empirical assessment is a frontier topic in causal inference and the assessment procedure used here is subject to a number of caveats (see Discussion).

Validation of context-specific signaling hypotheses

We identified 235 edges in the inferred networks that were not in the prior network. These potentially novel edges shared 35 parent proteins, 4 of which were inhibited in the original dataset. Five edges with parent nodes not among those inhibited in the original RPPA data were selected for validation by western blot. Edge selection was done on the basis of biological interest and availability of sufficiently specific inhibitors for the parent nodes (Figure 6). We note that our computational approach predicts presence/absence of each (directed) edge, but not sign (activating or inhibiting).

For each of the five edges, we tested contexts in which the edge was predicted as well as those in which the edge was not predicted. We inhibited the parent node and observed whether this altered abundance of the predicted child node. We found evidence supporting each of the five predicted causal edges, but with often complex context-dependence. These results – and their agreement and disagreement with context-specific predictions from network modeling – are summarized in Figure 6F,G.

An edge from Chk2_pT68 to p38_pT180/Y182 (for phosphoproteins we give the protein name before an underscore which is followed by the phosphorylation site(s)) was predicted only in cell line BT549 (Figure 5). We inhibited Chk2 with AZD7762 in BT549 cells, and saw decreases in phospho-p38 under serum (FBS) and NRG1, where the edge was predicted, as well as under insulin, where the edge was not predicted (Figure 6A). In contrast, there is no change in phospho-p38 in BT20 cells under AZD7762 treatment, consistent with the absence of the edge in the BT20 networks. Here we see that the edge validates in a cell line-specific but not stimulus-specific manner. However, it is important to note that AZD7762 inhibits Chk1 and Chk2 with equal potency and also demonstrates activity, albeit lower, against other kinases.

The networks predicted an edge from p38_pT180/Y182 to JNK_pT183/T185 in BT549 and BT20 cells under stimulus with FBS. We inhibited p38 with VX702 in BT549, BT20 and

UACC812 cells stimulated with FBS. In line with network predictions, we observed an increase in phospho-JNK in BT549 and BT20 cells (Figure 6B), but also observed a modest increase in phospho-JNK in UACC812 cells, where the edge was not predicted.

An edge from Src_p416 to NFκβ-p65_pS536 was predicted only in BT20 cells stimulated with EGF. Upon inhibition of Src with KX2391 both before and after stimulation with EGF, an increase in the abundance of phospho-NFκβ was observed in BT20 cells, consistent with the presence of a causal link (Figure 6C). The connection between phospho-Src and phospho-NFκβ was also observed in MCF7, where the edge was not predicted.

An edge from p70S6K_pT389 to p27_pT198 was predicted in all of the UACC812 and BT549 networks. The edge was also predicted in MCF7 networks for PBS, insulin, FGF, NRG1, and IGF1 and in the BT20 NRG1 network. When p70S6K was inhibited in UACC812 cells with PF4708671, a change in phospho-p27 was observed only at the zero time point before stimulus was added (Figure 6D). In MCF7 cells stimulated with HGF, phospho-p27 decreased in abundance under p70S6K inhibition; however, the edge was not predicted in this context. When PF4708671-treated MCF7 cells were stimulated with IGF, a context in which the edge was predicted with high probability, no change in phospho-p27 was observed. Similarly, there was no change in phospho-p27 in BT20 cells that had been treated with PF4708671 and stimulated with HGF.

In BT549 an edge was predicted from Chk2_pT68 to YAP_pS127 under HGF and insulin. BT549 cells treated with the Chk2 inhibitor AZD7762 exhibit an increase in phospho-YAP (Figure 6E). This edge was not predicted in any other cell line tested. However, in both UACC812 and MCF7 cells treated with AZD7762, a decrease in the abundance of phospho-YAP is observed. Active Chk2 appears to decrease phospho-YAP in BT549 cells (where the edge was predicted) and increase phospho-YAP in UACC812 and MCF7 cells (where the edge was not predicted). These results are consistent with the existence of a causal influence of phospho-Chk2 on phospho-YAP in all of these cell lines, and not just in BT549 as predicted.

Discussion

The data and analyses presented here support the notion that causal molecular networks can depend on context. We focused on signaling proteins and breast cancer cell lines. The cell lines represent contexts that are genetically perturbed but with a shared cancer type. The heterogeneity that we observed in causal networks suggests that substantial differences could exist between, for example, samples from different tissue types or divergent environmental conditions. This strongly argues for a need to refine existing regulatory models for specific contexts, not least in disease biology.

Given the range of potentially relevant contexts – spanning combinations of multiple factors, including genetic, epigenetic and environmental – we do not believe that characterization of causal networks across multiple contexts can feasibly be done using classical approaches in a protein-by-protein manner. Rather, it will require high-throughput data acquisition and computational analysis. Such a program of research requires an appropriate conceptual

framework, rich enough to capture regulatory relationships, but still tractable enough for large-scale investigation. Furthermore, for practical application, such an approach needs to be sufficiently robust to missing or unknown variables. Causal models of the kind we discussed may provide an appropriate framework because, unlike purely correlational models, they allow for reasoning about change under intervention and are, to a certain extent, robust to missing variables. In particular, causal descendancy matrices (Figures 4A and S3) are robust to missing variables in the sense that addition of a protein (row) to the matrix would not change the existing entries. We expect that a systematic program of investigation into context-specific causal networks will be important in many disease areas, and perhaps especially those that have to date not been well represented in the literature.

Our results extend the well-established notion of genomic intertumoral heterogeneity in cancer to the level of signaling phenotype. We found that cell line-specific findings were more reliable than stimulus-specific findings. This may be due to the magnitude of epigenetic and genetic differences between cell lines being more marked than differences between stimuli, all of which activate closely related cell surface receptors.

Our approach relied on inhibitor specificity, but we note that even at relatively low concentrations, off-target effects cannot be entirely ruled out. However, if the inhibitors were highly non-specific, the relatively good results seen in the train-and-test analysis would likely not be possible, since the analysis relies on assumed inhibitor targets. In the future it may be relevant to consider models that allow uncertainty in the inhibitor targets themselves.

We highlighted the need to specify a relevant timeframe in defining a causal graph. Indeed, an inhibitor may in the short-to-medium term induce changes to specific molecules but over the longer term the same inhibitor might induce adaptive changes to the cells themselves, e.g. via changes to epigenetic state (Duncan et al., 2012; Lee et al., 2012). We did not consider such "rewiring" in this paper, but note that the methods we discussed could be used to study rewiring, e.g. by comparing networks before and after adaptation.

In common with most protein profiling studies, including both low- and high-throughput techniques, our experiments were based on bulk assays and can therefore only elucidate signaling heterogeneity at the level of cell populations; we did not consider cell-to-cell heterogeneity, tumor stromal interactions, nor the spatial heterogeneity of tumors that plays an important role *in vivo* (Bedard et al., 2013; González-García et al., 2002). However, our data have implications for inter- and intra-tumoral heterogeneity because they suggest the possibility that *in vivo* causal signaling networks – and in turn the cell fates and disease progression events that they influence – may depend on the local micro-environment. Further work will be needed to elucidate such dependence and to draw out its implications.

In the future, causal molecular networks may start to play a role in precision medicine, for example by helping to inform rational assignment of targeted therapies. An implication of the context-specificity we report is that such analyses may require models that are learned, or at least modified, for individual samples (or subsets of samples). Although causal models are in some ways simpler than fully dynamical ones, causal inference remains fundamentally challenging and is very much an open area of research. For this reason, alongside advances

in relevant assays, a personalized, network-based approach will require suitable empirical diagnostics. Hill et al. (2016) used the data presented here to score, in an automated manner, over 2000 networks (~70 methods each applied to infer 32 context-specific networks) submitted to the HPN-DREAM network inference challenge and we used an extended version of this assessment procedure here. Such assessment procedures might allow for automated quality control, for example rejecting networks not sufficiently consistent with unseen interventional readouts (e.g. we did not obtain statistically significant performance under any test inhibitor for the *(BT549, EGF)* context; see Figure S5). However, as discussed in Hill et al. (2016) the assessment procedure remains limited in several ways and this argues for caution in interpreting the relatively good performance reported here. Of particular relevance to context-specificity, we note that the procedure focuses on global agreement with held-out interventional data and not specifically on identification of differences between contexts. Indeed, our validation experiments showed that although all novel edges that were tested validated in one or more contexts, network predictions were not accurate with respect to the precise context(s) in which changes were seen.

Recently, Carvunis and Ideker (2014) proposed a view of cellular function involving hierarchies of elements and processes and not just networks. Building detailed dynamical or biophysical models over hierarchies spanning multiple time and spatial scales may prove infeasible. A more tractable approach may be to extend coarser causal models of the kind used here in a hierarchical direction, for example by allowing causal links to cross scales and subsystems. Thus, the approach we pursued – of causal models based on context-specific interventional data – could in the future be used to populate models over biological hierarchies.

STAR Methods

Key Resources Table

Contact for Reagent and Resource Sharing

Further information and requests for reagents and resources may be directed to, and will be fulfilled by, the corresponding author Paul T. Spellman (spellmap@ohsu.edu).

Experimental Model Details

Breast epithelial cells in log-phase of growth were harvested, diluted in the appropriate media (DMEM (with phenol red) for UACC812, BT20 and MCF7; RPMI (with phenol red) for BT549) containing 10% fetal bovine serum, and then seeded into 6 well plates at an optimized cell density (to give 60-75% confluence at time of lysis). BT20 cells were plated at 230,000 cells/well; BT549 cells were plated at 175,000 cells/well; MCF7 cells were plated at 215,000 cells/well; and UACC812 cells were plated at 510,000 cells/well. After 24 hours of growth at 37°C and 5% CO2 in complete medium, cells were synchronized by incubating with serum-free medium for an additional 24 hours (serum starvation was also necessary to control the presence of stimuli in the medium). The medium was then exchanged with fresh serum-free medium containing either: 15nM AZD8055, 50nM GSK690693, 50nM BEZ235, 150nM PD173074, 10nM GSK1120212 in combination with

50nM GSK690693, or vehicle alone (0.05% DMSO) and incubated for two hours prior to stimulation. Cells were then either harvested (0 time point) or stimulated by addition of 200μL per well of 10X stimulus (either PBS, fetal bovine serum, 100 ng/mL EGF, 200ng/mL IGF1, 100nM insulin, 200ng/mL FGF1, 1μg/mL NRG1, or 500 ng/mL HGF) for 0, 5, 15, 30 or 60 minutes, or 2, 4, 12, 24, 48 or 72 hours prior to protein harvest.

All cell lines have been authenticated by performing STR analysis and matching to reference STR profiles at 15 different loci. STR analysis was performed by Genetica Cell Line Testing.

Method Details

Preparation of Cells for Reverse-Phase Protein Array (RPPA) Analysis—Cells were grown as described above, then washed twice with PBS and lysed by adding lysis buffer obtained from MD Anderson Functional Proteomics RPPA Core Facility (Houston, Texas; lysis buffer comprised 1% Triton X-100, 50mM HEPES, pH 7.4, 150mM NaCl, 1.5mM MgCl2, 1mM EGTA, 100mM NaF, 10mM Na pyrophosphate, 1mM Na3VO4, 10% glycerol; protease and phosphatase inhibitors were freshly added on the day of the experiment). Volume of lysis buffer used was optimized for each cell line (to ensure lysates were not too dense for the BCA assay; see below) and varied between 50μL and 100μL. Lysates were collected by scraping after 20 minutes incubation on ice. Lysates were spun at 4°C in a tabletop centrifuge at 15,000 RPM for 10 minutes and soluble proteins contained in the supernatant were collected. Protein concentration was determined by the Pierce BCA Protein Assay according to manufacturer's protocol. Protein was then diluted to 1 mg/mL and 30μL of the diluted lysate was mixed with 10μL 4X SDS sample buffer (obtained from MD Anderson Functional Proteomics RPPA Core Facility; comprised 40% Glycerol, 8% SDS, 0.25M Tris-HCL, pH 6.8; 10% v/v 2-mercaptoethanol was added fresh) and boiled for 5 minutes prior to freezing and shipment to MD Anderson Cancer Center Functional Proteomics RPPA Core Facility for RPPA analysis (Tibes et al., 2006).

RPPA methodology—RPPA methodology has been described previously (see e.g. Akbani et al., 2014); an outline is also provided below. Lysates were diluted in five two-fold serial dilutions with lysis buffer. An Aushon Biosystems 2470 arrayer (Burlington, MA) was used to print 1056 samples and control lysates on nitrocellulose-coated slides (Grace Bio-Labs). Each slide was probed with a primary antibody and a biotin-conjugated secondary antibody. Antibodies go through a validation process as previously described (Hennessy et al., 2010) to assess specificity, quantification and dynamic range. Each of the 183 primary antibodies was assigned a label based on this validation process (at the time the assay was performed): "validated", "use with caution" or "under evaluation" (see Table S1). Samples were split across three batches and some antibodies were used only in a subset of these batches (Table S1). A DakoCytomation-catalyzed system and DAB colorimetric reaction was used to capture signal. Following scanning of slides, spot intensities were analyzed and quantified using Microvigene software (VigeneTech Inc., Carlisle, MA). The EC50 values of the proteins in each dilution series were estimated using the SuperCurve software (Coombes et al., 2012), available at http://bioinformatics.mdanderson.org/Software/supercurve/. This uses the non-parametric, monotone increasing B-spline model (Hu et al., 2007) to fit a single

curve ("supercurve") using all samples on a slide, with signal intensity as the response variable and dilution step as the independent variable. The fitted curve is plotted with the signal intensities on the y-axis and the $\log_2$ protein concentrations on the x-axis for diagnostic purposes. A quality control (QC) metric, between zero and one was calculated for each slide (Coombes et al., 2012) and slides with values less than 0.8 were excluded. Within each batch, measurements were normalized for protein loading by median centering across antibodies (Hu et al., 2014; Li et al., 2013). This normalization process, performed on $\log_2$ concentrations, comprised the following steps:

1. For each antibody, calculate the median across samples and subtract from each value (i.e. median-center each antibody).

2. For each sample, calculate the median across antibodies to obtain a correction factor (CF).

3. For each sample, take the original $\log_2$ concentration values and subtract the corresponding CF (from step 2).

Normalized values, on a linear scale, are provided in Data S1.

Western Blot Analysis—Cells were grown as described above. For the novel edge validations in Figure 6, additional inhibitors were used to generate lysates following the protocol laid out above. The inhibitors, all used at 1μM, were AZD7762, KX2-391, PF4708671, and VX-702 (see Figure 6 for targets). Lysates were harvested 15 minutes after stimulation and protein concentrations quantified as described above. Denatured lysates were separated by PAGE on 4–12% Bis-Tris gradient gels (Invitrogen) along with Precision Plus Protein Standards (BioRad) using MOPS SDS NuPAGE Running Buffer (Invitrogen) and NuPAGE LDS Sample Buffer (Invitrogen) on ice at 200 V for 45 minutes. Gels were transferred to immobilin-FL PVDF membranes (Millipore) using NuPAGE Transfer Buffer (Invitrogen) on ice at 30 V for 1.5 hours before being washed 3x 5 min. with 5% Tween-TBS (TTBS, Amresco & Invitrogen) at room temperature (RT) with agitation and blocked with 5% BSA (Sigma) in TTBS for 1 hour at RT with agitation. Blots were again washed 3x 5 minutes in TTBS at RT with agitation before being incubated in primary antibody in 5% BSA in TTBS overnight at 4C with agitation. Blots were washed 3x for 5 minutes at RT with agitation and then transferred to HRP-conjugated secondary antibody in 5% BSA in TTBS and incubated at RT for 1.5 hrs. See Table S5 for primary antibodies and HRP-conjugated secondary antibodies used in western blot validations. Blots were washed again as previously described and visualized using SuperSignal West Pico Chemiluminescent HRP Substrate Kit (Thermo Scientific) and CL-X Posure Film (Thermo Scientific) and changes in protein abundance under inhibition were determined by visual inspection of exposed film.

Quality Control and Preprocessing of RPPA Data

Batch Normalization Procedure for Cell Line UACC812: The UACC812 data were split across two RPPA experiments with each batch containing different inhibitors (BEZ235, PD173074 and GSK690693&GSK1120212 in one batch; AZD8055 and GSK690693 in the other). DMSO control samples were common to both batches. The two batches were combined and normalized to obtain a single data set for UACC812.

The steps of the batch normalization procedure were as follows:

1. Any antibodies not included in both batches were removed.

   For each antibody, perform steps 2 and 3 below.

2. Using $\log_2$-transformed data (after normalization for protein loading; see above), the mean and standard deviation of the DMSO samples in each batch were calculated, giving values (μ_1, σ_1) and (μ_2, σ_2) for batch 1 and batch 2 respectively. Note that, for each batch, there are 16 replicates for DMSO, 0min samples (all other DMSO conditions consist of a single replicate). These 16 replicates were averaged prior to calculation of mean values and standard deviations.

3. All samples in batch 2 (for the given antibody) were then scaled and centered so that the mean and standard deviation of the batch 2 DMSO samples agreed with the corresponding batch 1 quantities (μ_1, σ_1). That is, a sample in batch 2 with value x became

$$\mu_1 + \frac{\sigma_1(x - \mu_2)}{\sigma_2}.$$

   This scaling and centering was applied to each individual replicate and not to replicate-averaged data.

4. The two batches were then combined to get a single data set for UACC812.

Data for the two batches and the final normalized data set are provided in Data S1 on a linear scale.

Samples Excluded from Analyses: Samples identified as outliers were excluded from our analyses. These samples were identified using the following criteria:

- Normalization for protein loading resulted in a correction factor (CF) for each sample (see above). Samples with $CF > 2.5$ or $CF < 0.25$ were regarded as outliers.

- Variance across all antibodies was calculated for each sample. Values greater than 40 were regarded as outliers.

- We used the replicates at time $t = 0$ to calculate the signal-to-noise ratio (SNR) for each cell line and phosphoprotein antibody under each inhibitor (mean of replicates divided by standard deviation of replicates). The mean across all calculated SNRs was 10.68 (s.d. = 5.8). SNR values less than 1 were investigated further to determine whether the poor SNR was caused by outlier replicates.

For cell line UACC812, these criteria were applied to the batch-normalized data set.

In addition to the above, data for the combination of inhibitors GSK690693 & GSK1120212 (AKTi & MEKi) for cell lines BT549 (all stimuli) and BT20 (PBS and NRG1 stimuli only) were excluded since none of the expected effects of MEKi were observed in these samples.

All samples excluded from analyses are shown in Table S6 and also indicated in the data files in Data S1.

Antibodies Included in Analyses: To facilitate comparisons between cell lines, the analyses presented here focused on the set of phosphoprotein antibodies common to all four lines. This set contained two highly correlated pairs of antibodies ($r > 0.9$ for all cell lines), consisting of phosphoforms of the same protein: GSK3αβ_pS21_pS9, GSK3_pS9 and S6_pS235_S236, S6_pS240_S244. Since highly correlated variables can lead to a reduction in the utility of network inference results, only one antibody out of each pair was included in analyses, resulting in a final set of 35 phosphoprotein antibodies. A full list of antibodies can be found in Table S1, where the 35 phosphoproteins included in the analyses are also indicated.

Final Preprocessing Steps: Data were $\log_2$ transformed and replicates (only present for $t = 0$ samples and some DMSO samples) were averaged. Prior to input into our network inference pipeline, imputation was performed for missing data by linear interpolation of adjacent time points.

Identification of Changes Under Kinase Inhibition—We used a procedure centered on paired *t*-tests to determine which phosphoproteins show a salient change in abundance under each kinase inhibitor. Details are described in Hill et al. (2016), but also outlined below for completeness.

For each phosphoprotein, inhibitor regime and *(cell line, stimulus)* context, a paired *t*-test was used to assess whether mean phosphoprotein abundance under DMSO control is significantly different to mean abundance under the inhibitor regime (mean values calculated over seven time points). As discussed above, some phosphoproteins show a clear response to the stimulus under DMSO control, with abundance increasing and then decreasing over time (a "peak" shape), while others show a less clear response due to signal already being present prior to stimulus. For phosphoproteins falling into the former category (according to a heuristic), paired *t*-tests were repeated, but this time restricted to intermediate time points within the peak. This focuses on the portion of the time course where an inhibition effect, if present, should be seen. The *p*-value from the repeated test was retained if smaller than the original *p*-value. For each *(cell line, stimulus)* context and inhibitor regime, the resulting set of *p*-values (one *p*-value for each phosphoprotein) were corrected for multiple testing using the adaptive linear step-up procedure for controlling the FDR (Benjamini et al., 2006).

For each *(cell line, stimulus)* context, a phosphoprotein was deemed to show a salient change under a given inhibitor regime if two conditions were satisfied. First, the corresponding FDR value had to be less than 5% and, second, the effect size ($\log_2$ ratio between DMSO control and inhibitor conditions) had to be sufficiently large relative to replicate variation (see Figure S2). The latter condition is an additional filter to remove small effects. Replicate variation was quantified by calculating the pooled replicate standard deviation at each time point of the DMSO and inhibitor time courses, and then averaging these values across time points. The phosphoproteins satisfying these criteria are depicted in

Figures 3B, 4A, S1, S2 and S3. We note that the overall procedure is heuristic and that the FDR values should not be interpreted formally.

A phosphoprotein p showing a salient change under an inhibitor is consistent with a node targeted by the inhibitor having a causal effect on the phosphoprotein. Since this effect can be direct or indirect, phosphoprotein p can be regarded as a descendant of the inhibitor target node in the underlying signaling network. That is, there exists a directed path starting from the node targeted by the inhibitor and ending at phosphoprotein p.

Network Learning—Networks were learned for each of the 32 *(cell line, stimulus)* contexts using dynamic Bayesian networks (DBNs), a type of probabilistic graphical model for time-course data (see e.g. Hill et al., 2012; Husmeier, 2003; Murphy, 2002). Specifically we used a recently proposed variant called *interventional DBNs* or iDBNs (Spencer et al., 2015), that uses ideas from causal inference (Pearl, 2009; Spirtes et al., 2000) to model interventions and thereby improve ability to infer causal relationships; model specification followed Spencer et al. (2015). Although interested in learning context-specific networks, we expect a good proportion of agreement between contexts. Therefore, rather than learn networks for each context separately, we used a recently developed joint learning approach to solve all the problem instances together (Oates et al., 2014). A prior network was used (Figure S4); this was curated manually with input from literature (Weinberg, 2013) and online resources. The extent to which context-specific networks are encouraged to agree with each other and with the prior network is controlled by two parameters, λ and η respectively, as described in detail in Oates et al. (2014). These parameters were set (to $\lambda = 3$ and $\eta = 15$) by considering a grid of possible values and selecting an option that provides a reasonable, but conservative amount of agreement, allowing for discovery of context-specific edges that are not in the canonical prior network. The network learning approach resulted in a score (edge probability) for each possible edge in each context-specific network. The network estimates were robust to moderate data deletion and precise specification of the biological prior network and its strength (Figure S6). Furthermore, the analyses were computationally efficient, requiring approximately 30 minutes to learn all 32 context-specific networks using serial computation on a standard personal computer (Intel i7-2640M 2.80GHz processor, 8GB RAM).

Assessing Performance of Causal Network Learning—The ability of our network learning approach to estimate context-specific causal networks was systematically assessed using a train and test scheme proposed by Hill et al. (2016) in the context of the HPN-DREAM network inference challenge associated with the RPPA data presented here. Due to factors specific to the challenge setting, Hill et al. (2016) used only a single iteration of train and test. In contrast, we were able to perform several iterations, as described below.

In each iteration, the data were divided into two sets: (i) a test data set, consisting of time courses for all 32 *(cell line, stimulus)* contexts under a *single* inhibitor regime, and (ii) a training data set, consisting of time courses (again for all 32 contexts) for a subset of the remaining five inhibitor regimes (Figure S5A). We refer to the single inhibitor regime in the test data as the *test inhibitor* (although note that one regime contains more than one kinase inhibitor: GSK690693 & GSK1120212). Thirty-two context-specific networks were learned

on the training data set and then each network was assessed as to how well it agreed with changes observed, for the same context, under the test inhibitor (in the test data set). For each test inhibitor, the set of phosphoproteins that show, for a given context, a salient change in abundance were determined as described above (and shown in Figures 4A & S3), resulting in context-specific "gold-standard" descendant sets. We then compared, for each context, predicted descendants of the test inhibitor target node(s) according to the network inferred from training data, against the corresponding "gold-standard" descendant set. This resulted in a number of correctly predicted descendants (true positives, TPs) and a number of incorrectly predicted descendants (false positives, FPs). Our network learning approach outputs edge probabilities, from which a network can be obtained using a threshold value. The TP and FP values were therefore a function of this threshold value, resulting in an ROC (receiver operating characteristic) curve. Our final assessment metric was then the area under this curve (AUROC), which we calculated for each context and test inhibitor (Figure S5B). The statistical significance of the AUROC scores was determined using an empirical null distribution, generated by calculating AUROC scores for sets of uniformly random edge probabilities.

The assessment procedure requires that nodes targeted by the test inhibitor are present in the network model so that their descendants can be determined. Also, it is important that the training data only contains inhibitor regimes that target nodes which are not also targeted by the test inhibitor. There were three train and test data splits that satisfied these criteria (while also maximizing the sample size of the training data set), and we assessed performance for all three (Figure S5B).

Quantification and Statistical Analysis

Replicates were averaged prior to carrying out statistical analysis and the time courses shown in Figures 3 and S1 were plotted using replicate-averaged data. The number of replicates were as follows: ~16 replicates for samples at $t = 0$, except for UACC812, BT20 and MCF7 DMSO $t = 0$ samples which were replicated ~32 times; 2 replicates for the majority of DMSO samples at $t > 0$, except for BT20 DMSO samples; all other samples had a single replicate.

Details of statistical procedures are provided in the methods section above or in figure legends. Analyses were performed using Matlab R2012a software.

Data and Software Availability

Software—Scripts for the computational and statistical analyses presented here are available at https://github.com/Steven-M-Hill/causal-signaling-networks-CellSystems2016. These scripts include identification of changes under kinase inhibition, network learning and assessment of network learning performance.

Data Resources—RPPA data, including additional time points and antibodies that were not used in the analyses presented here, are provided in Data S1. Time-course plots for all of the antibodies are provided in Data S2.

Additional Resources

HPN-DREAM network inference challenge associated with the RPPA data presented here: https://www.synapse.org/HPN_DREAM_Network_Challenge.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Akbani R, Ng PKS, Werner HMJ, Shahmoradgoli M, Zhang F, Ju Z, Liu W, Yang J-Y, Yoshihara K, Li J, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. Nat Commun. 2014; 5:3887. [PubMed: 24871328]

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012; 483:603–307. [PubMed: 22460905]

Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature. 2013; 501:355–364. [PubMed: 24048068]

Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. Biometrika. 2006; 93:491–507.

Cantone I, Marucci L, Iorio F, Ricci MA, Belcastro V, Bansal M, Santini S, di Bernardo M, di Bernardo D, Cosma MP. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. Cell. 2009; 137:172–181. [PubMed: 19327819]

Carvunis A-R, Ideker T. Siri of the cell: what biology could learn from the iPhone. Cell. 2014; 157:534–538. [PubMed: 24766803]

Coombes KR, Neeley S, Joy C, Hu J, Baggerly K, Roebuck P. SuperCurve: SuperCurve R Package. 2012

Duncan JS, Whittle MC, Nakamura K, Abell AN, Midland AA, Zawistowski JS, Johnson NL, Granger DA, Jordan NV, Darr DB, et al. Dynamic Reprogramming of the Kinome in Response to Targeted MEK Inhibition in Triple-Negative Breast Cancer. Cell. 2012; 149:307–321. [PubMed: 22500798]

Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012; 483:570–575. [PubMed: 22460902]

Gerlinger M, Swanton C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. Br J Cancer. 2010; 103:1139–1143. [PubMed: 20877357]

González-García I, Solé RV, Costa J. Metapopulation dynamics and spatial heterogeneity in cancer. Proc Natl Acad Sci U S A. 2002; 99:13085–13089. [PubMed: 12351679]

Good M, Tang G, Singleton J, Reményi A, Lim WA. The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation. Cell. 2009; 136:1085–1097. [PubMed: 19303851]

Heiser LM, Sadanandam A, Kuo W, Benz SC, Goldstein TC, Ng S, Gibb WJ, Wang NJ, Ziyad S, Tong F, et al. Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc Natl Acad Sci U S A. 2012; 109:2724–2729. [PubMed: 22003129]

Hennessy BT, Lu Y, Gonzalez-Angulo AM, Carey MS, Myhre S, Ju Z, Davies MA, Liu W, Coombes K, Meric-Bernstam F, et al. A Technical Assessment of the Utility of Reverse Phase Protein Arrays

for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. Clin Proteomics. 2010; 6:129–151. [PubMed: 21691416]

Hill SM, Lu Y, Molina J, Heiser LM, Spellman PT, Speed TP, Gray JW, Mills GB, Mukherjee S. Bayesian inference of signaling network topology in a cancer cell line. Bioinformatics. 2012; 28:2804–2810. [PubMed: 22923301]

Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, Zhang Y, Sokolov A, Paull EO, Wong CK, et al. Inferring causal molecular networks: empirical assessment through a community-based effort. Nat Methods. 2016; 13:310–318. [PubMed: 26901648]

Hu J, He X, Baggerly Ka, Coombes KR, Hennessy BTJ, Mills GB. Non-parametric quantification of protein lysate arrays. Bioinformatics. 2007; 23:1986–1994. [PubMed: 17599930]

Hu Y, Yan C, Hsu C, Chen Q, Niu K, Komatsoulis Ga. A Comprehensive Comparison of Normalization Methods for Loading Control and Variance Stabilization of Reverse-Phase Protein Array Data. Cancer Inform. 2014; 13:109–117.

Husmeier D. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. Bioinformatics. 2003; 19:2271–2282. [PubMed: 14630656]

Lee MJ, Ye AS, Gardino AK, Heijink AM, Sorger PK, MacBeath G, Yaffe MB. Sequential Application of Anticancer Drugs Enhances Cell Death by Rewiring Apoptotic Signaling Networks. Cell. 2012; 149:780–794. [PubMed: 22579283]

Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, Yang J-Y, Broom BM, Verhaak RGW, Kane DW, et al. TCPA: a resource for cancer functional proteomics data. Nat Methods. 2013; 10:1046–1047.

Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G. Revealing strengths and weaknesses of methods for gene network inference. Proc Natl Acad Sci U S A. 2010; 107:6286–6291. [PubMed: 20308593]

Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods. 2016; 13:366–370. [PubMed: 26950747]

Murphy, K. Dynamic Bayesian Networks: Representation, Inference and Learning. University of California; Berkeley, CA: 2002.

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe J-P, Tong F, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006; 10:515–527. [PubMed: 17157791]

Nickel GC, Barnholtz-Sloan J, Gould MP, McMahon S, Cohen A, Adams MD, Guda K, Cohen M, Sloan AE, LaFramboise T. Characterizing mutational heterogeneity in a glioblastoma patient with double recurrence. PLoS One. 2012; 7:e35262. [PubMed: 22536362]

Oates CJ, Korkola J, Gray JW, Mukherjee S. Joint estimation of multiple related biological networks. Ann Appl Stat. 2014; 8:1892–1919.

Pearl, J. Causality: Models, Reasoning, and Inference. Cambridge University Press; 2009.

Petsalaki E, Helbig AO, Gopal A, Pasculescu A, Roth FP, Pawson T. SELPHI: correlation-based identification of kinase-associated networks from global phospho-proteomics data sets. Nucleic Acids Res. 2015; 43:W276–W282. [PubMed: 25948583]

Rolland T, Ta An M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, et al. A Proteome-Scale Map of the Human Interactome Network. Cell. 2014; 159:1212–1226. [PubMed: 25416956]

Saez-Rodriguez J, Goldsipe A, Muhlich J, Alexopoulos LG, Millard B, Lauffenburger Da, Sorger PK. Flexible informatics for linking experimental data to mathematical models via DataRail. Bioinformatics. 2008; 24:840–847. [PubMed: 18218655]

Saez-Rodriguez J, Alexopoulos LG, Zhang M, Morris MK, Lauffenburger Da, Sorger PK. Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. Cancer Res. 2011; 71:5400–5411. [PubMed: 21742771]

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–2504. [PubMed: 14597658]

De Smet R, Marchal K. Advantages and limitations of current network inference methods. Nat Rev Microbiol. 2010; 8:717–729. [PubMed: 20805835]

Spencer SEF, Hill SM, Mukherjee S. Inferring network structure from interventional time-course experiments. Ann Appl Stat. 2015; 9:507–524.

Spirtes, P., Glymour, CN., Scheines, R. Causation, Prediction, and Search. MIT Press; 2000.

Szerlip NJ, Pedraza A, Chakravarty D, Azim M, McGuire J, Fang Y, Ozawa T, Holland EC, Huse JT, Jhanwar S, et al. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. Proc Natl Acad Sci U S A. 2012; 109:3041–3046. [PubMed: 22323597]

Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, Kornblau SM. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Mol Cancer Ther. 2006; 5:2512–2521. [PubMed: 17041095]

Weinberg, R. The Biology of Cancer. Garland Science; 2013.

Will T, Helms V. PPIXpress: construction of condition-specific protein interaction networks based on transcript expression. Bioinformatics. 2016; 32:571–578. [PubMed: 26508756]

Woodward, J. Causation and Manipulability. The Stanford Encyclopedia of Philosophy. Zalta, EN., editor. 2013.

Zalatan JG, Coyle SM, Rajan S, Sidhu SS, Lim WA. Conformational control of the Ste5 scaffold protein insulates against MAP kinase misactivation. Science (80-.). 2012; 337:1218–1222.
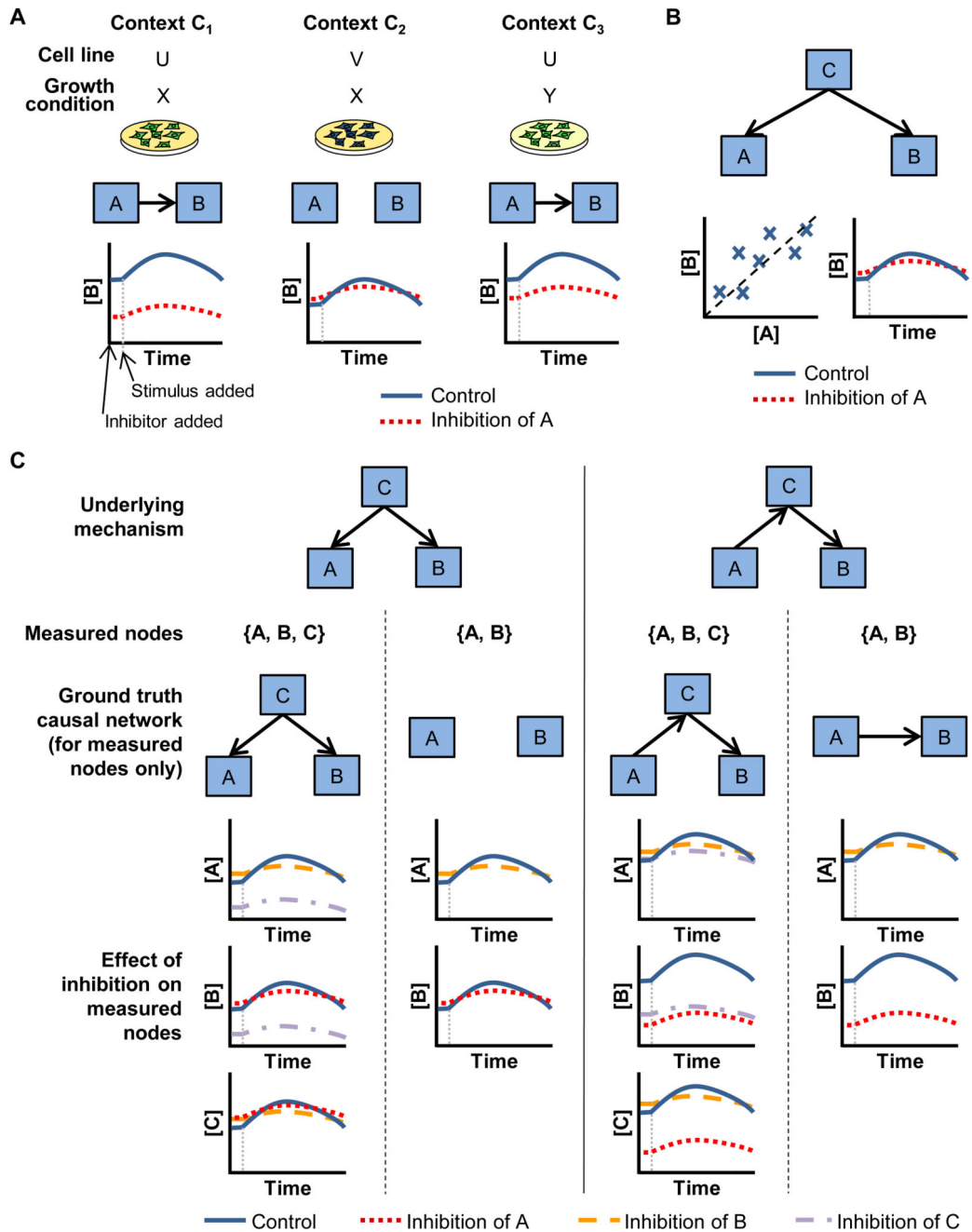
Figure 1. Context-specific Causal Networks.

(A) Context-specific causal influences. Node A has a causal influence on node B in contexts c_1 and c_3, but not c_2, encoded by the presence of a causal edge between A and B in c_1 and c_3 only. This reflects the outcome of experiments where A is inhibited. Here, each context is defined by the combination of cell line and growth condition. (B) Correlation and causation. The abundance of node A is correlated with that of node B due to regulation by the same node C. However, as there is no causal influence (direct or indirect) of A on B, inhibition of A does not result in a change in the abundance of B, no matter how strong the correlation or

statistical dependence. (C) Causal networks and missing nodes. In the first example, node *C* regulates both nodes *A* and *B* (as in panel (B)). In the formulation used here, if *C* is not observed and not included in the network, but *A* and *B* are, we would regard the network with no edge between *A* and *B* in either direction as the correct or ground truth causal network, in line with the results of experimental inhibition of these nodes, as shown. In the second example, the underlying mechanism is that *A* influences *C*, and *C* in turn influences *B*. In the formulation used here, if *C* is not measured and not included in the network, an edge from *A* to *B* would be regarded as correct, in line with the results of experimental inhibition of the nodes. However, if all three nodes were included, the correct network would match the underlying mechanism. Although abundance of *B* changes under inhibition of *A*, an edge from *A* to *B* would be regarded as incorrect here because the influence of *A* on *B* is fully mediated via another network node (i.e. *C*). See text for further details of the causal formulation and its interpretation.
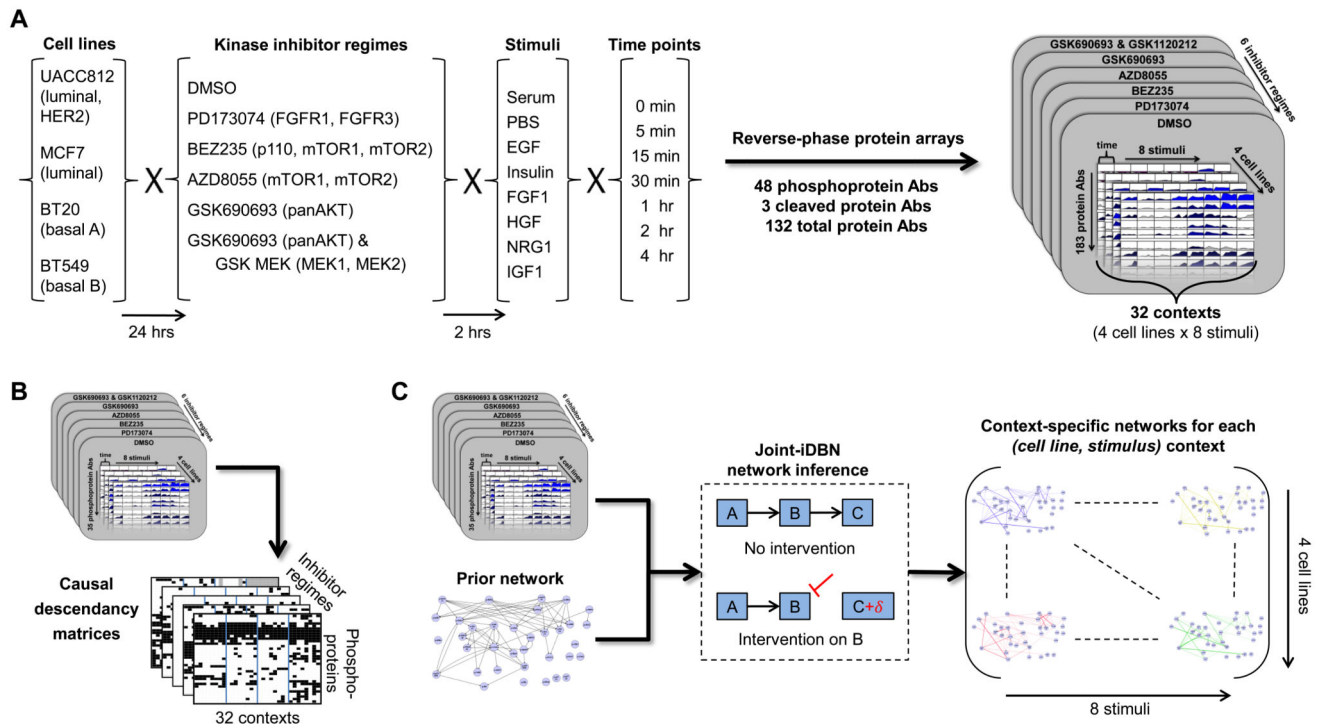
Figure 2. Data-Driven Reconstruction of Context-Specific Causal Signaling Networks.
(A) Overview of experimental approach. Reverse-phase protein arrays (RPPA) were used to investigate protein signaling in four human breast cancer cell lines under eight different stimuli. The combinations of cell line and stimulus defined 32 *(cell line, stimulus)* contexts. Prior to stimulus, cell lines were serum starved and treated with kinase inhibitors or DMSO control. RPPA assays were performed for each context at multiple time points post-stimulus, using more than 150 high-quality antibodies to target specific proteins, including approximately 40 phosphoproteins (the precise number of antibodies varies across cell lines; see STAR Methods and Table S1). (B) Causal descendancy matrices (CDMs). CDMs summarizing changes under intervention across all contexts were constructed for each intervention (see text for details). (C) Overview of causal network learning procedure. Interventional time-course data for each context were combined with existing biological knowledge in the form of a prior network to learn context-specific phosphoprotein signaling networks. Networks were learned using a variant of dynamic Bayesian networks designed for use with interventional data and that allowed joint learning over all 32 contexts at once (see STAR Methods).
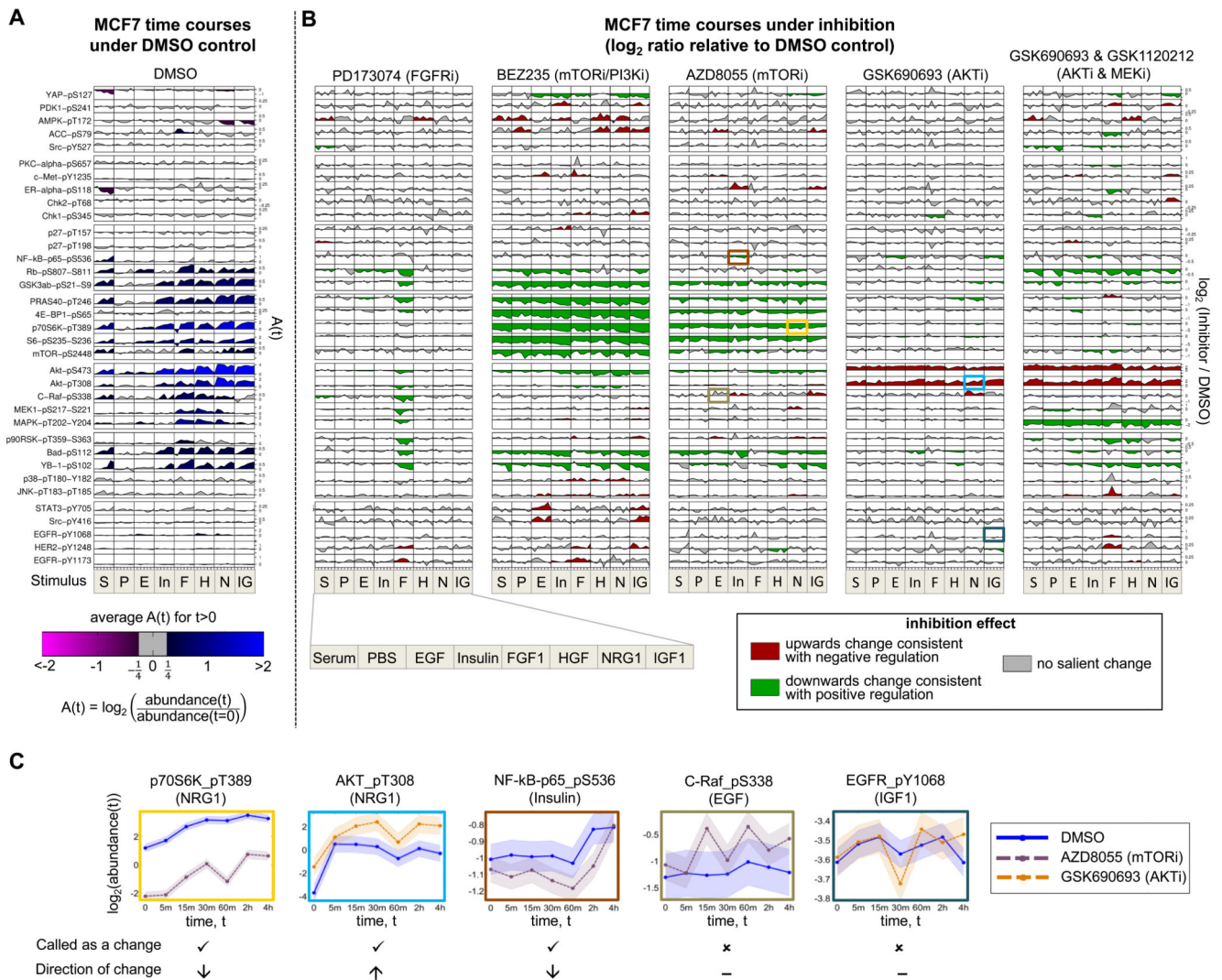
Figure 3. Phosphoprotein Time-Course Data and Context-Specific Changes Under Inhibition for Breast Cancer Cell Line MCF7.

(A) Phosphoprotein time courses under DMSO control. Rows correspond to 35 phosphoproteins (a subset of the full set of 48; see STAR Methods for details) and columns correspond to the eight stimuli. Each time course shows $\log_2$ ratios of phosphoprotein abundance relative to abundance at $t = 0$. Shading represents average $\log_2$ ratio for $t > 0$. (B) Phosphoprotein time courses under kinase inhibition. Each of the five vertical blocks corresponds to a different inhibition regime. Within each block, rows and columns are as in (A). Each time course shows $\log_2$ ratios of phosphoprotein abundance under inhibition relative to abundance under DMSO control. Shading represents direction of changes in abundance due to inhibitor: Green denotes a decrease in abundance, red denotes an increase and gray denotes no salient change (see examples in panel (C)). See STAR Methods for details of statistical analysis. For both (A) and (B), plots were generated using a modified version of the DataRail software (Saez-Rodriguez et al., 2008). Each phosphoprotein is plotted on its own scale and phosphoproteins are ordered by hierarchical clustering of all

data. See Figure S1 for corresponding plots for cell lines UACC812, BT20 and BT549. (C) Selected examples from (B) showing control (DMSO) and inhibitor time courses separately; box color identifies the source cell in (B). Examples are shown for (from left to right): a clear decrease in abundance ; a clear increase in abundance; a decrease in abundance that is borderline under the criteria we use; a borderline case called negative (i.e. called as no change) and a clear negative case. Shaded region indicates time-averaged replicate standard deviation. See also Figure S2.

Figure 4. Non-Canonical and Context-Specific Signaling.

(A) Causal descendancy matrix showing causal effects observed under mTOR inhibitor AZD8055 in each of the 32 *(cell line, stimulus)* contexts. Rows represent phosphoproteins and columns represent contexts (see Figure 3). Black boxes indicate phosphoproteins that show a salient change under mTOR inhibition in a given context (see STAR Methods) and can therefore be regarded as causal descendants of mTOR in the signaling network for that context. The final column on the right indicates phosphoproteins that are descendants of mTOR in the canonical mTOR signaling pathway shown in (B). Phosphoproteins are ordered first by canonical column and then by hierarchical clustering of all data. See Figure S3 for causal descendancy matrices for the other inhibitor regimes. (B) Canonical mTOR signaling pathway. Blue nodes are descendants of mTOR in the network and white nodes are non-descendants. The pathway shown is a subnetwork of the prior network used within the network inference procedure (Figure S4). Full node names, including phosphorylation sites,

are provided in Table S4. (C) Summary of western blot validations of causal effects observed in RPPA data. A number of observations from the causal descendancy matrices were chosen for validation via western blot analysis. The number of phosphoprotein validations attempted ("Tested") and the number of these that successfully validated ("Validated") are presented for various *(cell line, stimulus, inhibitor*) combinations. Summary totals are also presented for each cell line, each inhibitor and across all validation experiments. See also Table S2.
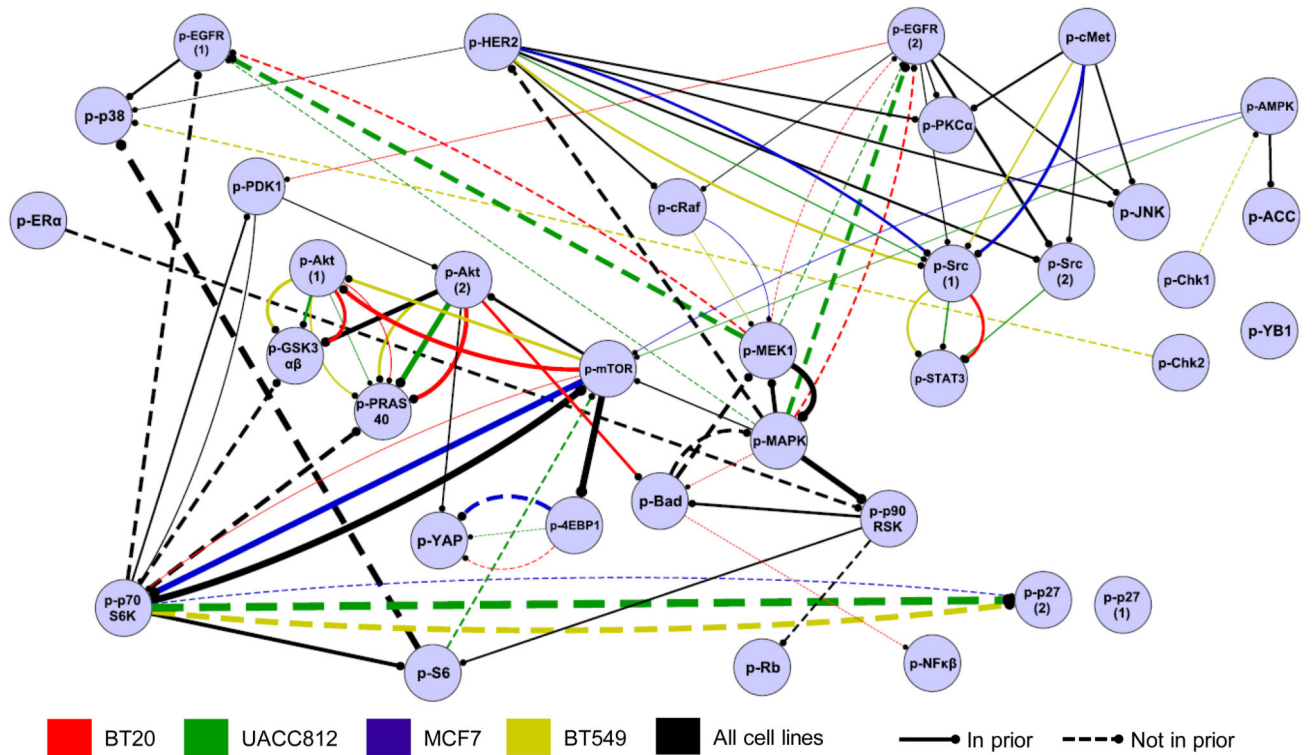
Figure 5. Context-Specific Signaling Networks Reconstructed Using a Machine Learning Approach.

Data for 35 phosphoproteins were analyzed using a machine learning approach based on interventional dynamic Bayesian networks, integrating also known biology in the form of a prior network (Figure S4). This gave a set of scores (edge probabilities) for each possible edge in each *(cell line, stimulus)* context (see STAR Methods). For each cell line, a summary network was obtained by averaging edge probability scores across the eight stimulus-specific networks for that cell line. Edge color denotes cell line. Only edges with average probabilities greater than 0.2 are shown. A black edge indicates an edge that appears (i.e. is above the 0.2 threshold) in all four cell lines. Edge thickness is proportional to the average edge probability (average taken across all 32 contexts for black edges). Solid/dashed edges were present/not present in the prior network respectively. Edges are directed with the child node indicated by a circle. Edge signs are not reported; the modeling approach does not distinguish between excitatory and inhibitory causal effects. Full nodes names, including phosphorylation sites, are provided in Table S4. Network visualized using Cytoscape (Shannon et al., 2003). See also Table S3.
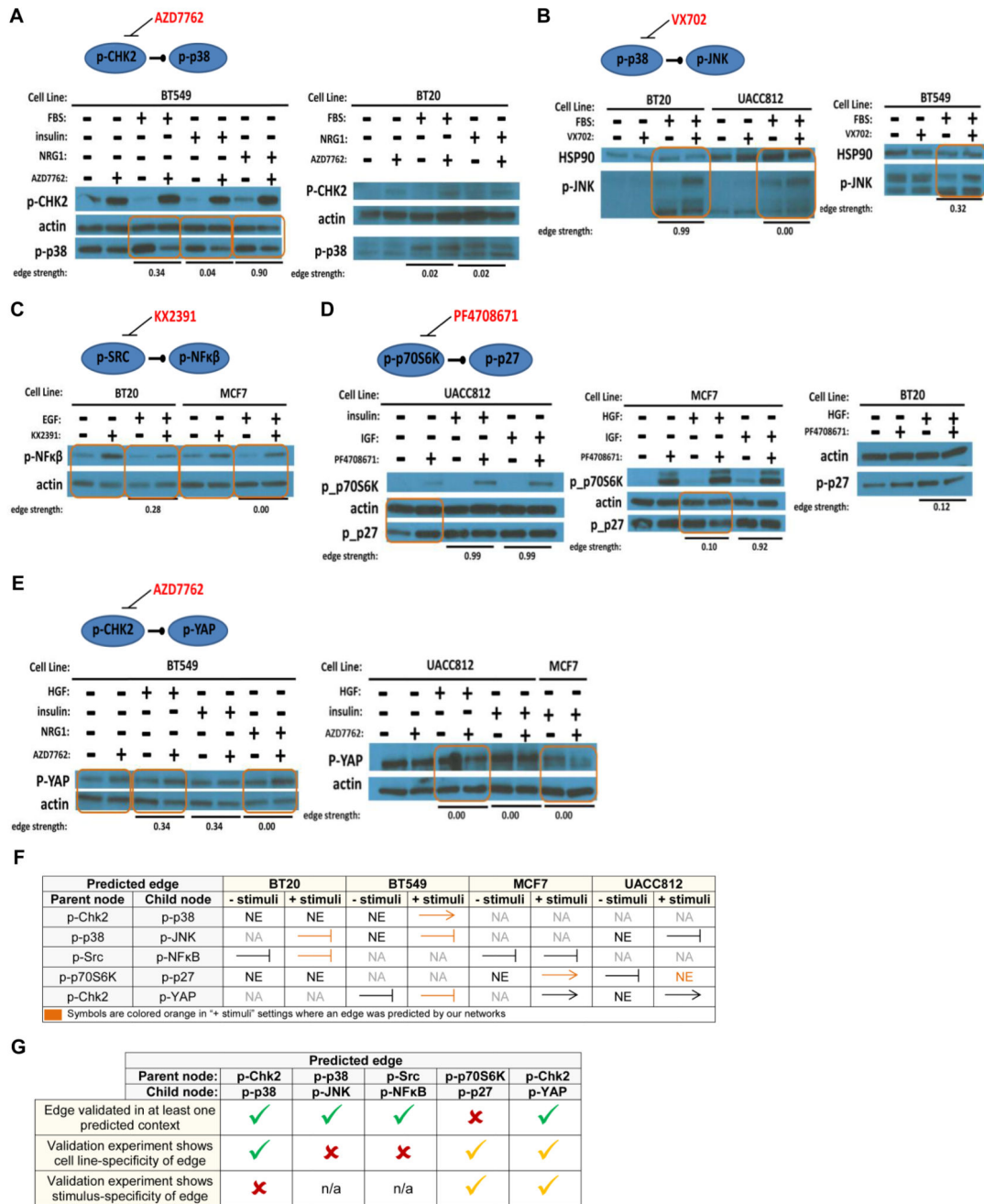
Figure 6. Validation of Network Edges.

Western blot analysis of selected context-specific network edges that were not in the prior network. Edges tested were: (A) phospho-Chk2 to phospho-p38 ; (B) phospho-p38 to phospho-JNK; (C) phospho-Src to phospho-NFκB; (D) phospho-p70S6K to phospho-p27; and (E) phospho-Chk2 to phospho-YAP. Orange boxed areas indicate observed changes in abundance of the predicted child node under inhibition of the parent node in a single *(cell line, stimulus)* context (changes in abundance are determined by visual inspection of the bands). Edge probabilities output by the network learning procedure are shown for each

context tested ("edge strength"). (F) A summary of the validation experiments. 'NA' denotes "not applicable" – the experiment was not run. 'NE' denotes "no effect" – there was no change in child node abundance upon inhibition of the parent node. An arrow indicates results consistent with an activating parent node. A stunted line represents results consistent with an inhibitory edge. Symbols are colored orange to indicate that an edge was predicted for the corresponding cell line under one of the stimuli tested. (G) Summary of agreement and disagreement between predicted edges and validation experiments. First row indicates whether validation experiments showed evidence for the edge in a *(cell line, stimulus)* context in which it was predicted. Second and third rows concern the cell line- and stimulus-specificity of each edge respectively: a green tick denotes specificity in (partial) agreement with predictions from inferred networks; an orange tick denotes specificity, but not in agreement with predictions in terms of the precise contexts in which effects were seen; a red cross indicates that specificity was not observed in the validation experiments, despite being predicted by the networks.