Integrated analysis of breast cancer cell lines reveals unique signaling pathways

Laura M. Heiser[1], Nicholas J. Wang[1], Carolyn L. Talcott[2], Keith R. Laderoute[2], Merrill Knapp[2], Yinghui Guan[1], Zhi Hu[1], Safiyyah Ziyad[1], Barbara L. Weber[4], Sylvie Laquerre[4], Jeffrey R. Jackson[4], Richard F. Wooster[4], Wen-Lin Kuo[1], Joe W. Gray[1, 3] and Paul T. Spellman[1]*

1. Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

2. SRI International, Inc., 333 Ravenswood Ave., Menlo Park, CA 94025, USA

3. Comprehensive Cancer Center, 2340 Sutter Street, University of California, San Francisco, San Francisco, CA 94143, USA

4. Oncology CEDD, GlaxoSmithKline, 709 Swedeland Rd, King of Prussia, PA 19406, USA

* To whom correspondence should be addressed:
Paul T. Spellman, Ph.D.
1 Cyclotron Rd. MS-977R225A
Berkeley, CA 94720
Phone: 510.486.5213
Fax: 510.486.5714
Email: ptspellman@lbl.gov

Abstract

Background: Cancer is a heterogeneous disease resulting from the accumulation of genetic defects that negatively impact control of cell division, motility, adhesion and apoptosis. Deregulation in signaling along the EGFR-MAPK pathway is common in breast cancer, though the manner in which deregulation occurs varies between both individuals and cancer subtypes.

Results: We were interested in identifying subnetworks within the EGFR-MAPK pathway that are similarly deregulated across subsets of breast cancers. To that end, we mapped genomic, transcriptional and proteomic profiles for 30 breast cancer cell lines onto a curated Pathway Logic symbolic systems model of EGFR-MEK signaling. This model was comprised of 539 molecular states and 396 rules governing signaling between active states. We analyzed these models and identified several subtype specific subnetworks, including one that suggested PAK1 is particularly important in regulating the MAPK cascade when it is over-expressed. We hypothesized that PAK1 over-expressing cell lines would have increased sensitivity to MEK inhibitors. We tested this experimentally by measuring quantitative responses of 20 breast cancer cell lines to three MEK inhibitors. We found that PAK1 over-expressing luminal breast cancer cell lines are significantly more sensitive to MEK inhibition as compared to those that express PAK1 at low levels. This indicates that PAK1 over-expression may be a useful clinical marker to identify patient populations that may be sensitive to MEK inhibitors.

Conclusions: All together, our results support the utility of symbolic system biology models for identification of therapeutic approaches that will be effective against breast cancer subsets.

Background

Cancer is a heterogeneous disease that results from the accumulation of multiple genetic and epigenetic defects [1-4]. These defects lead to deregulation in cell signaling and ultimately impacts control of cell division, motility, adhesion and apoptosis [5]. The MAPK (mitogen-activated protein kinase/ERK) pathway plays a central role in cell comunication: it orchestrates signaling from external receptors to internal transcriptional machinery, which leads to changes in phenotype [6, 7]. This pathway has been implicated in the origin of multiple carcinomas, including those of the breast [8-10]. Activation of MAPK is initiated by one of the four ERBB receptors (ERBB1/EGFR, ERBB2-4), which leads to signaling through RAF, MEK and ERK. In addition, the ERBB receptors integrate a diverse array of signals, both at the cell surface level and through cross-talk with other pathways, such as the phosphoinositide 3-kinase (PI3K) pathway [11]. Both EGFR and ERBB2 are overexpressed in a substantial fraction of breast cancers and are recognized targets for breast cancer therapy [12-16]. In addition, MEK has long been studied as a therapeutic target, and many drugs that inhibit it are currently under development [17-20].

Among breast cancers, unique subsets can be defined at the genomic, transcriptional and proteomic levels. For many years, breast cancers were classified by whether or not they express various receptors, namely estrogen (ER/ESR1), progesterone (PR/PGR) and ERBB2 [21-25]. This key insight has been used to tailor therapies to individual patients [22, 26]. Of particular interest is the finding that ER-negative tumors frequently show elevated signaling along the MAPK pathway as compared to ER-positive cancers [27]. DNA amplification at various loci can also be used to stratify

patients, and, importantly, has prognostic value as well [28, 29]. For example, amplification at 8p12 and 17q12 are both associated with poor outcome [28, 30]. The emergence of expression profiling technology led to the seminal observation that breast cancers can be systematically classified at the transcriptional level [23-25]. More recently, interest has turned toward the analysis of somatic mutations [31]. Different cancer types show common patterns of mutation, implying that a few key mutations play a pivotal role in tumorigenesis. All together, these studies indicate the value of identifying unique subsets of cancers, both for understanding the origin of the disease as well as identification of appropriate therapeutics.

A critical question remaining is how to identify meaningful subsets of cancers that differ in their cell signaling pathways. One approach to this problem is to identify gene expression signatures that reflect the activation status of oncogenic pathways [32, 33]. While it is possible to stratify cancers into unique populations based on their expression patterns of these signatures, a key challenge lies in interpreting the meaning of the various genes within these signatures [34]. Here, we used an alternate approach in which we explored subtype-dependent behavior in genes that make up known signaling pathways.

Our goal was to identify signaling pathway modules that are deregulated in particular cancer subtypes. To that end, we populated a well-curated cell signaling model with molecular information from a panel of breast cancer cell lines. We used a combination of transcriptional, proteomic and mutational data to create a unique signaling network for each cell line. Specifically, we discretized transcript and protein data and used it to populate the network models; genes or proteins that are differentially

expressed across the cell lines were evaluated as present in some cell lines and absent from others. The resultant network models can be viewed as a statistical formalism of the pathways activated in each of the cell lines.

We created our network models with Pathway Logic [35-38], a system designed to build discrete, logical (rule-based) models of signal transduction pathways [39]. Logical models are directly related to the canonical schematic diagrams ("cartoons") commonly used to show functional relationships among proteins, and as such, are easily interpretable in the context of biological systems (see Figure 1B) [40]. The two critical elements of a Pathway Logic model are a rule set and an initial state. The rules represent biochemical reactions, and the initial state is a representation of all proteins present in a particular cell line. Our model contains a rich rule set: the interactions between proteins have all been individually curated from primary literature sources and therefore represent well-characterized signaling biology. In short, we used our collection of molecular data to identify active states in each cell line, and the rules to define signaling between these active states. The resultant networks are static coarse graphical representations of signaling that can be used to generate hypotheses about key signaling events in subsets of the cell lines.

We focused our modeling on the ERBB/MAPK pathway because deregulation along this pathway is both frequent in breast cancers and heterogeneous across them [12, 41]. Further, it is involved in a complex web of signaling that results from cross-talk with other pathways [42]. Our model system includes rules that describe: (i) interactions between the ERBB receptors and their ligands; (ii) direct association of intracellular signaling proteins with the phosphorylated ERBB receptors; (iii) signaling along the

canonical RAF-MEK-ERK pathway; (iv) cross-talk with PI3K and JAK/STAT pathways; (v) activation of immediate-early transcription factors (e.g., JUN, FOS); and (vi) signaling from other receptors that influence MAPK signaling, including EPHA2 and Integrins.

Our panel of cell lines captures many features of biological variation found in primary breast tumors [43]. Both the cell lines and tumors cluster into basal (ESR1-negative, CAV1-positive) and luminal (ESR1-positive, ERBB3-positive) expression subsets. These two subtypes – basal and luminal – also show distinct biological characteristics, including differences in morphology and invasive potential [23, 25]. In addition, the cell lines show a broad response to pathway-targeted drugs (Gray, et al, in preparation). Overall, the genomic heterogeneity in the cell lines mirrors that observed in a large population of primary tumors, and en ensemble constitute a useful model of the molecular diversity of primary tumors [43].

We generated signaling network models for our panel of cell lines with the goal of identifying subnetworks that are active in particular subsets of cell lines. We found that the discretized data used to populate the initial states of the networks showed only a small amount of variation. Specifically, only 13% of the components in the initial state of the networks varied across the cell lines. Even with this small amount of variation, the discretized data used in the initial states could be clustered into basal and luminal cell line groups. Surprisingly, over half of the protein interactions predicted to occur varied across the cell line network models. In order to identify active subnetworks, we clustered the network features of our models. This clustering resulted in three main groups of cell lines: basal, luminal and a third mixed group composed of both basal and luminal cell

lines. In addition, we identified several network modules active in specific subsets of the cell lines. One module in particular implicated PAK1 (p21 protein (Cdc42/Rac)-activated kinase 1) as a key regulator of the RAF-MEK-ERK pathway in the subset of PAK1 over-expressing cell lines. We found that among luminal cell lines, the over-expression of PAK1 was significantly associated with sensitivity to MEK inhibition. Taken together, these results indicate that our modeling approach can be used to identify signaling subnetworks that are particularly important in subsets of breast cancer cell lines.

Results

Data clustering and model initialization

Our goal was to create a unique signaling network model for each cell line in our panel. In generating these models, we must accommodate two fundamental biological principles. First, the ERBB network results from the integration of many diverse signals, and second, most cell signaling occurs through protein-protein interactions. Ideally, then, we would create large networks populated with protein data. Because the acquisition of comprehensive protein abundance data for multiple cell lines is not technically feasible, we used transcript data to infer protein levels when protein data were unavailable. An example of one of these large computed networks is shown in **Figure 1A**.

A key feature of Pathway Logic models is that they are discrete, so components are considered either present or absent. In order to populate our network models, we first discretized the transcript and protein data (see Methods, **Figure 1C, D, E**). Following discretization, we determined which components (proteins) are present in the initial state

of each cell line. We considered genes and proteins that are differentially expressed across the cell lines to be present in some cell lines and absent from others. Genes and proteins that showed little variation in expression were considered present in all cell lines. Although this approach is coarse, we can use it to assess which pathways may be most critical in each of the cell lines. That is, we can identify the pathways that may be highly up- or down-regulated in particular cell lines. This discretization algorithm captured many well-documented differences in expression across the cell lines. For example, the transcript data for ESR1 yields three clusters, which parallels the observation that primary breast tumors show varied expression of this protein (**Figure 1E;** [44, 45]).

The initial states were constructed from a population of 286 signaling components. We had expression data alone for 191 of these components, both protein and expression data for 25, and no available data for the 70 remaining components. Following discretization, 13/25 (52%) proteins and 19/191 (10%) transcripts form both present and absent groups. For the remaining protein and transcript data, a single group best describes the distribution of expression values. To explore the transcript and protein data further, we compared the clustering results for the 25 components that had both protein and transcript data available. Approximately two-thirds of these components show a high level of concordance between the two discretized datasets: 9 yield a single present group for both datasets; 8 yield a present and absent group for both datasets (mean Pearson's r=0.603). The remaining 8 components form a single group in one dataset and two groups in the other. For 6 of these, the transcript data yield a single group while the protein data form two groups (**Table 1**).

We used the Sanger COSMIC database to identify mutations to KRAS, PTEN and PIK3CA in our cell lines, and included these data in the initial states [46]. We focused on mutations in these three proteins for two reasons: first, they influence MAPK signaling, and second, the mutations have a known functional impact, so it is possible to computationally model them. Specifically, a G13D point mutation in KRAS causes it to become constitutively active [47, 48]. A frameshift mutation in PTEN leads to premature termination and an inactive protein [49]. Three common point mutations in PIK3CA (E542K, E545K and H1047R) lead to increased lipid kinase activity [50, 51]. PIK3CA is the most frequently mutated gene in our cell line panel (6/30, 20%), a finding that parallels other reports [52].

Initial states reflect the known biology

We found that 39/286 (13%) of the components vary across the initial states of the cell lines (**Figure 2**). This includes both the effect of data discretization, as well as differences in mutational status for KRAS, PTEN and PIK3CA. The components that vary are located throughout the network and include receptors, GTPases and transcription factors. We used unsupervised hierarchical clustering to analyze the variable components in the initial states [53]. In accordance with our previous studies, we found that the site of origin, basal or luminal epithelium, largely defines the two major clusters [43]. We achieved a similar result when we clustered the data with a partitioning around medoids (PAM) algorithm that searched for two groups in the discretized data. Specifically, most of the cell lines (26/30) correctly segregated into basal or luminal groups. This finding demonstrates that our modeling system has some of the genes that influence this

phenotypic difference. Further, it indicates that the discretized data used to populate the network models recapitulates some of the known cell biology associated with the origins of the breast cancer cell lines.

The network models are highly variable

A principal interest in modeling these pathways was to determine how the network topology differs across the set of cell lines. To address this question, we determined which components and rules were present in each of the networks. The network models contained an average of 334 (8.29 s.e.m.) rules and 218 (4.55 s.e.m.) unique state changes. Over 55% of the rules and state changes differ across the 30 models, indicating that the networks are highly variable (**Table 2**). This result was surprising at first, considering that the initial states have 87% of the components in common.

To explore this finding further, we examined the connectivity of individual components by determining the number of rules in which each component is involved. The majority of the components participate in only one or two rules, whereas a few components participate in many rules (**Figure 3A**). EGFR, the most highly connected component, is involved in 22 rules. When we plotted these data on a log-log plot, a robust linear relationship was revealed, indicating that the connectivity follows a power-law (**Figure 3B**). Interestingly, some of the most highly connected components vary across the initial states of the cell lines, namely EGFR, SRC, PI3K, and KRAS **(Table 3)**. These proteins have a particularly large role in shaping network topology. If they are

omitted from the initial state, many rules will fail to fire and many pathways in the resultant network will be truncated.

We were interested in whether the cell line models could be grouped by their network properties. We addressed this by performing an unsupervised hierarchical clustering of the network features (i.e., the components in the initial state, rules, and components that underwent state changes) that differed across the cell lines. This clustering resulted in three major groups for the cell line models: basal, luminal and a third group comprised of both basal and luminal cell lines (**Figure 4**). The observation that there is a mixed group of basal and luminal networks indicates that the cell lines may be segmented by their signaling pathways, rather than by site of origin alone.

Unique signaling modules are active in particular subsets of the network models

We next asked how the network structure varies across the cell lines. To answer this question, we used partitioning around medoids (PAM) clustering to partition the network features into 30 clusters. Each cluster represents a unique "signaling module" that is present in some cell line models and absent from others. A summary of these signaling modules provides an overview of the variable network features **Table 4**. Each signaling module is driven by the presence of particular components in the initial state. For example, the ERBB4 module is present in 10 cell lines, 9 of which are luminal and 1 that is basal, reflecting the fact that ERBB4 is present in the initial state of these 10 cell lines. The signaling modules average 8 rules each, though they vary in size from a single rule up to 76 rules for the SRC/RAC1 module.

The RHOB (ras homolog gene family, member B) module is largely responsible for the segmentation of the basal and luminal cell line models, and is present in all the luminals and absent from all the basals. RHOB interacts with NGEF to activate many downstream targets that go on to regulate a diverse array of cellular functions including cell motility, cell adhesion and cell cycle progression [54, 55]. In accordance with this, we have found that the basal cell lines are far more invasive than the luminal cell lines [43]. Furthermore, RHOB levels have been shown to decrease as cancer progresses [56-58].

Clustering of the "mixed" group of cell lines is strongly driven by the three SRC modules (**Figure 4**). SRC is one of the most highly connected components in the network (18 rules), and serves to integrate a variety of signals. This module, which results from the omission of SRC from the initial state, is present in all cell lines except two, namely basaloid MDAMB435 and luminal MDAMB453. The other two SRC modules are dependent on the presence of either EGFR or RAC1. The SRC/EGFR module includes SRC-dependent activation of EGFR; if either component is missing from the initial state, signaling along this cascade is compromised. The SRC/EGFR module is absent only from the mixed group of networks: four are missing EGFR, one is missing SRC, and the other is missing both EGFR and SRC.

One small signaling module is related to the presence of CAV1 in the initial state. One of the rules in this module describes activation of SHC that is dependent on FYN, CAV1 and Integrin (ITGB1) (**Figure 5A**). Both the transcript and protein data indicate that the presence of CAV1 is bimodal, and is clearly present at either very low or very high levels (**Figure 5B, C**). This module is only present in basal cell lines, and further,

most of the cell lines that contain it are of the most aggressive basal B subtype [43]. This signaling module provides a direct feed into the RAF-MEK-ERK pathway, suggesting that these cell lines have an alternate route available for ERK activation (**Figure 5A**). This interaction may help to explain why these basal cell lines are particularly aggressive.

PAK1 plays a pivotal role in the network models

In our model, PAK1 is required for the activation of MEK and ERK (**Figure 6A**). Specifically, PAK1 phosphorylates MEK, which in turn facilitates signaling along the RAF-MEK-ERK cascade [59]. It follows, then, that network models with PAK1 omitted from the initial state fail to activate ERK. Across the cell lines, the distribution of PAK1 transcript levels is highly skewed, so our discretization algorithm yields two clusters, a large group centered at -0.26, and a small group centered at 2.16 (**Figure 6B**). PAK1 was present in the initial state of the cell lines with high expression and absent from the others. The four cell lines with high PAK1 transcript levels, MDAMB134, 600MPE, SUM52PE and SUM44PE, are all of luminal origin.

Based on the observations that PAK1 directly regulates MAPK signaling, and that its expression pattern shows substantial variation, we hypothesized that PAK1 differentially regulates MAPK signaling across our panel of cell lines. We tested this hypothesis experimentally. The first issue we addressed was whether PAK1 protein levels vary across the cell lines. We found highly variable expression of total PAK1 protein. Specifically, three of the four cell lines with elevated PAK1 transcript levels have concordantly high PAK1 protein levels. In addition, a handful of other cell lines also show over-expression of PAK1 protein. PAK1 transcript and protein levels are

significantly correlated (Pearson's r = 0.78, $p < 0.0001$, **Figure 6C**). While this relationship is largely dependent on the cell lines that highly express PAK1, it nonetheless supports the idea that elevated transcript levels affect protein expression levels. Focal changes in copy number are thought to convey a selective advantage for tumor growth, so we next asked whether PAK1 is amplified in any of our cell lines. The four cell lines that over-express PAK1 show high-level amplification (> 8.7 copies, see Methods) of the PAK1 amplicon (11q13.5-q14, [60], **Figure 6D**); none of the other cell lines show this amplification. In addition to PAK1 amplification, three of these cell lines also show amplification at CCND1, though in all cases there are distinct peaks at each locus.

If PAK1 indeed regulates MAPK signaling, we would expect to find a correlation between PAK1 and phospho-MEK levels. To address this, we quantified isoform-specific phospho-MEK levels in our cell lines (see Methods). We found a small but significant correlation between total PAK1 and percent MEK1-S298 (Pearson's r = 0.32, $p<0.05$, **Figure 6E**). Although the correlation is somewhat weak, it is clear that high PAK1 levels are always associated with elevated phospho-MEK1. In accordance with the observation that the interaction between PAK1 and MEK is specific to MEK1 [61], we found no correlation between PAK1 and percent phospho-MEK2 ($p>>0.05$).

The above findings suggest that elevated PAK1 levels provide a foothold into regulation of the MAPK cascade, and led us to hypothesize that PAK1 over-expressing luminal cell lines would be particularly sensitive to MEK inhibition. To test this, we measured the response of 20 luminal cell lines to three MEK inhibitors: CI-1040, UO126 and GSK1120212. We compared growth inhibition (GI_{50}, the drug concentration

required to inhibit growth by 50%) following drug exposure between cell lines that over-express PAK1 (n=3) and those that do not (n=17). The two groups of cell lines had significantly different mean expression of both the PAK1 transcript and protein (t-test, $p < 0.01$). The three PAK1 over-expressing cell lines (MDAMB134, SUM52PE and 600MPE) were significantly more sensitive to MEK inhibition as compared to the non-PAK1 over-expressing cell lines (GSK1120212, $p < 0.005$; CI-1040, $p < 0.05$; UO126, $p < 0.05$; (*t*-test), **Figure 7**). This result indicates that PAK1 over-expression may be a useful clinical marker to determine whether a particular tumor will be responsive to MEK inhibition.

Discussion

Cancer arises from deregulation in any of a multitude of genes, but exactly how this deregulation impacts cell signaling is not well understood. Here, we leveraged a rich dataset of transcriptional and protein profiles with a computational modeling system in order to gain a greater understanding of the critical signaling pathways associated with breast cancer. By creating a unique network model for individual cell lines, we were able to identify signaling pathways that are particularly important in subsets of the cell lines. Our modeling led to new insight about the importance of PAK1 as a modulator of the MAPK cascade.

Approaches to computational modeling

There are many approaches to computationally modeling biological systems, ranging from high-level statistical models to low-level kinetic models [62]. We used a

simplified mid-level scheme to construct network models from transcript and protein profiles for two reasons. First, we were able to create a unique model for each cell line, rather than a single network that represents "breast cancer." We used this approach to examine how a collection of genomic and proteomic changes in individual cell lines affects its network architecture. In contrast, other approaches, such as Bayesian reconstruction, are designed to describe ensemble behavior, rather than behavior of individual cell lines [63, 64]. A key attribute of our modeling system is that it can be used to identify specific biological instances of cell signaling that can be used to generate hypotheses. Our observations about PAK1 are a key example of this feature. The second reason for using this mid-level modeling scheme is that the computational algorithm is relatively simple; logical operators define relationships between signaling components. It is therefore possible to create networks that are quite large, which provides the opportunity to examine multiple inputs that impinge upon the central signaling pathway of interest. In comparison, kinetic models that offer more detail about signaling components are quite computationally demanding, so it is only feasible to examine a limited number of components [65, 66]. As a "hypothesis generator," our modeling system could be used to guide the development of dynamic modeling systems by identifying key signaling components to include in them.

One limitation of our modeling system is that it operates in a totally discrete manner: components are either present or absent, and rules fire with absolute certainty or not at all. This is a simplification of true biological systems in which the levels of signaling components show a wide dynamic range, and the probability that a reaction will occur changes as a function of the concentration of individual proteins. We captured the

variation in the concentration of signaling components by individually discretizing the data for each component in the initial state and then assigning each cell line to a "present" or "absent" group. With this approach, we examined how signaling is affected by extreme changes in protein levels, therefore homing in on key signaling events. We found that even with this simplified approach, we were able to make insights into key signaling events in subsets of our cell lines. Hybrid modeling approaches, which combine continuous dynamical systems with discrete transition systems, have been developed to overcome this limitation [67, 68]. Modification of the current model system to a hybrid system would allow for a more detailed examination of cell signaling over smaller changes in protein concentrations.

Modeling results

We found that the network connectivity follows a power law relationship in which most components have low connectivity while a few components are highly connected (Figure 3). The relationship we observed reflects not only intrinsic connectivity, but also curation bias, as literature relevant to EGFR/MAPK signaling was preferentially surveyed during creation of the rule set. Nonetheless, this "scale free" relationship has been described in more thorough surveys of protein-protein interactions [69, 70]. The observation that our network models also have this scale free property supports the idea that they are biologically relevant representations. Further, this pattern of connectivity implies that the few highly connected components may be most critical for regulating cell signaling along these pathways -- these components serve as promising candidates for more detailed study at both the computational and experimental levels. Those that also

show substantial variation across the cell lines (e.g. EGFR, SRC, PI3K, and KRAS) may be particularly relevant in the context of breast cancer.

Traditionally, the site of origin has been one of the primary features with which to classify breast cancers [23-25]. The full transcriptional profiles of our cell line panel show this characteristic split between basal and luminal subtypes [43], which we could largely recapitulate in our construction of the initial states (Figure 2). Here, we have shown that ERBB/MAPK signaling systematically varies across our panel of cell lines. Specifically, we found that the cell line networks could be classified into three groups (Figure 4). The basal and luminal network groups reflect the split we observed in the components of the initial state, while the third mixed group is largely defined by signaling related to SRC. SRC acts as a well-connected signaling hub, so it is particularly important in shaping network architecture. It also interacts with several key proteins in the MAPK cascade, including EGFR and its targets, ERK, and CDC42 [71, 72]. SRC has been studied as a therapeutic target in a wide range of cancers, including cancers of the breast, lung and pancreas [73, 74].

The basal and luminal networks could be well-differentiated by the RHOB signaling module, which is present in the luminal cell lines and absent from the more aggressive basal cell lines (Figure 4). A number of reports have indicated that loss of RHOB expression is frequently associated with cancer progression [58]. Furthermore, suppression of RHOB is a critical step leading to transformation in a variety of cancers, including those of the lung and cervix [75]. These observations bolster the idea that modulation of the RHOB pathway may serve as a useful therapy in the basal cell lines. Among the basal cell line networks, the CAV1/ITGB1 signaling module was primarily

found in the most aggressive basal B cell lines. In accordance with this, CAV1 has been shown to have a role in carcinogenesis, though its mechanism may vary with cancer type [76, 77].

PAK1 impacts signaling along the MAPK cascade

Through an analysis of our breast cancer network models, we identified PAK1 as a putative differential regulator of the MAPK cascade in our cell lines. PAK1, a serine/threonine kinase, has long been studied as a regulator of cytoskeletal remodeling and cell motility [78, 79], but more recently has been shown to regulate both proliferation [80] and apoptosis [81]. The PAK family of proteins has been implicated in a variety of cancers, including those of the breast [80, 82, 83]. In particular, PAK1 hyperactivation has been shown to cause mammary-gland tumors in mice [84].

Across our panel of cell lines, PAK1 is differentially expressed at the copy number, transcript and protein levels (Figure 6). The finding of elevated PAK1 expression in some of our cell lines mirrors the observation that PAK1 is sometimes upregulated in breast tumors [80]. The correlation between PAK1 and phospho-MEK1 levels (Figure 6C) suggests that across the cell lines, PAK1 differentially modulates activation of the MAPK cascade. Although statistically significant, this correlation was not perfect: high PAK1 levels are always associated with high phospho-MEK1 levels, while a more variable relationship emerges when PAK1 is low. This observation implies that when PAK1 levels are high, it dominates the regulation of phospho-MEK1, whereas at low PAK1 levels, alternate proteins must serve as the principle regulator of phospho-MEK1. For example, KSR1 and SPRY are both involved in regulation of the MAPK

cascade, and may be particularly important in the cell lines that express PAK1 at low levels [85, 86]. Based on this finding, we hypothesized that the luminal cell lines that over-express PAK1 would be particularly sensitive to MEK inhibition. Indeed, the PAK1 over-expressing cell lines were significantly more sensitive to three MEK inhibitors than the non-PAK1 over-expressing cell lines (Figure 7). The observation that all three drugs showed the same pattern indicates that the inhibition is quite robust and not due to off-target effects. These results indicate that PAK1 over-expression may be a useful clinical marker to determine which patient populations may be sensitive to MEK inhibitors.

Conclusions

Breast cancer is a remarkably heterogeneous disease that results from the accumulation of various genetic defects. We were interested in identifying signaling subnetworks that may be particularly important in generating oncogenic phenotypes. To address this, we generated a discrete, static network model for a panel of 30 breast cancer cell lines. The resultant network models were highly variable: of the protein interactions predicted to occur, over half of them varied across the cell lines. We searched for active subnetworks by clustering the network features of our models. This clustering yielded three main groups of cell lines, a basal group, a luminal group, and a third mixed group composed of both basal and luminal cell lines. In addition, we identified several network modules active in specific subsets of the cell lines. One signaling module implicated PAK1 (p21 protein (Cdc42/Rac)-activated kinase 1) as a key regulator of the RAF-MEK-ERK pathway in the cell lines that over-express it. Based on this observation, we hypothesized that luminal cell lines that over-express PAK1 would be particularly

responsive to MEK inhibition. In support of this idea, we found that among luminal cell lines, the over-expression of PAK1 was indeed significantly associated with sensitivity to three MEK inhibitors. All together, these results indicate the utility of symbolic systems modeling for the identification of key cell signaling events in the context of cancer.

Materials and Methods

Cell lines: The complete panel contains 51 breast cancer cell lines that have been previously described [43]. We assembled our panel of breast cancer cell lines from the ATCC and the laboratories of Drs. Steve Ethier and Adi Gazdar. All cell lines have been carefully maintained in culture, and we have stored stocks of the earliest-passage cells. We assure quality control by careful analysis of morphology, growth rates, gene expression and protein levels over time. All extracts were made from subconfluent cells in the exponential phase of growth in full media. Information about biological characteristics and culture conditions is available elsewhere [87]. We generated network models for the 30 well-characterized cell lines with the complete datasets described below.

Protein abundance data: We measured the abundance of 25 proteins associated with ERBB/MAPK signaling in our network model. These abundances were assayed and quantitated as previously described [43]. Briefly, proteins were measured by Western blots of cells lysed in 1% Nonidet-P40, 50 mM HEPES (pH 7.5), 150 mM NaCl, 25 mM b-glycerophosphate, 25 mM NaF, 5 mM EGTA, 1 mM EDTA, 15 mM pyrophosphate, 2

mM sodium orthovanadate, 10 mM sodium molybdate, leupeptin (10 mg/ml), aprotinin (1 0mg/ml), and 1 mM PMSF.

We quantified protein levels by measuring the emitted chemiluminescence or infrared radiation recorded from labeled antibodies using Scion Image [88] or Odyssey software [89]. For each protein, the blots were made for 4 sets of 11 cell lines, where each set included the same pair (SKBR3 and MCF12A) to permit intensity normalization across sets. We performed a basic multiplicative normalization by fitting a linear mixed-effects model to log intensity values, and adjusted within each set to equalize the log intensities of the pair of reference cell lines across the sets.

Transcriptional profiles: Total RNA was prepared from samples using Trizol reagent (GIBCO BRL Life Technologies), and quality was assessed on the Agilent Bioanalyser 2100. Preparation of in vitro transcription (IVT) products, oligonucleotide array hybridization, and scanning were performed according to Affymetrix (Santa Clara, California) protocols. In brief, 5 μg of total RNA from each breast cancer cell line and T7-linked oligo-dT primers were used for first-strand cDNA synthesis. IVT reactions were performed to generate biotinylated cRNA targets, which were chemically fragmented at 95°C for 35 min. Fragmented biotinylated cRNA (10 μg) was hybridized at 45°C for 16 hr to Affymetrix high-density oligonucleotide array human HG-U133A chip. The arrays were washed and stained with streptavidin-phycoerythrin (SAPE; final concentration 10 μg/ml). Signal amplification was performed using a biotinylated anti-streptavidin antibody. The array was scanned according to the manufacturer's instructions (2001 Affymetrix Genechip Technical Manual). Scanned images were inspected for the

presence of obvious defects (artifacts or scratches) on the array. Defective chips were excluded, and the sample was reanalyzed.

We generated probe set based gene expression measurements from quantitated Affymetrix image files with the RMA algorithm [90] from the BioConductor tools suite [91] and annotated with Unigene annotations from the July 2003 mapping of the human genome [92]. All 51 CEL files were analyzed simultaneously, yielding a data matrix of probe sets by cell lines in which each value is the calculated log abundance of each gene probe set for each cell line. Gene expression values were centered by subtracting the mean value of each probe set across the cell line set from each measured value.

Mutation data: We searched the Sanger Catologue Of Somatic Mutations In Cancer (COSMIC) website for reported mutations in our cell lines [46]. We incorporated mutations to KRAS, PTEN and PIK3CA into our models through the construction of rules that reflect the functional impact of each mutation.

Copy number profiles: We measured copy number profiles with molecular inversion probes (MIPs). The MIP assay was performed as previously described [93]. Briefly, test DNA samples were diluted to 16 ng/ml. All DNA quantification was done using PicoGreen dsDNA Assay Kit (Molecular Probes/Invitrogen, Carlsbad, CA, USA, P7589). We used 96- or 384-well plates whenever possible to reduce variation. For day 1 overnight annealing, 4.7 µl of DNA samples (75 ng total), 0.75 µl of Buffer A, 1.1 µl of the 53 K probe pool (200 amol/µl/probe) and 0.045 µl of Enzyme A were mixed well in a 384-well plate on ice. The reaction was incubated at 20°C for 4 minutes, 95°C for 5 minutes, then 58°C overnight. On day 2, 13 µl of Buffer A was added to each well with 1.25 µl of Gapfill Enzyme mix, then 9 µl of this was put in each of two wells in a 96-well

plate. MIP probes were circularized with 4 µl of dinucleotide (dATP with dTTP, dCTP with dGTP) and mixed at 58°C for 10 minutes. The uncircularized probes and genomic DNA were eliminated by addition of 4 µl of Exonuclease Mix and incubation at 37°C for 15 minutes, followed by heat-killing of enzymes. The circularized probes were linearized by the addition of Cleavage Enzyme Mix at 37°C for 15 minutes, then subjected to universal primer amplification for 18 cycles at 95°C for 20 s, 64°C for 40 s, and 72°C for 10 s. For the labeling reaction, the product was further amplified with the label primers for 10 cycles, and then subjected to cleavage by Digest Enzyme Mix at 37°C for 2 h. To hybridize, the cleaved MIP products were mixed with hybridization cocktail, denatured and hybridized to 70 K Universal Taq arrays at 39°C for 16 h (two arrays per sample). The overnight hybridized arrays were washed on GeneChip® Fluidics Station FS450 and stained by SAPE at 5 ng/ml (Invitrogen). Copy number estimation was obtained from the hybridization signals as previously described [93].

We filtered the dataset to eliminate MIP probes missing from more than 5% of the samples. We used the previously described amplicon boundaries to compute average copy number across all the probes in the PAK1 and CCND1 amplicons [60]. We defined high-level amplification as M+3*IQR, where M represents the median copy number and IQR represents the interquartile range, each computed across all amplicons and cell lines.

Quantitative analysis of MEK: We used high-resolution capillary isoelectric focusing technology to quantify the abundance of individual phosphoforms and isoforms of MEK. We used MEK1 (Upstate) and MEK2 (Cell Signaling) antibodies for this assay, which has been described in detail elsewhere [94].

Cell growth inhibition assay and data analysis: Cells were plated at proper density in 96-well plates such that they would remain in log growth at the end of assay time. The cells were allowed to attach overnight before being exposed to MEK inhibitor CI-1040, UO126 or GSK1120212 for 72 h. Drugs were dissolved in DMSO as 10 mM stock, and a set of 9 doses in 1:5 serial dilution was added in triplicate wells. The final DMSO concentration in the treated well was 0.3% or less. The cell growth was determined using Cell Titer Glo (CTG) assay (CellTiter-Glo Luminescent Cell Viability Assay, Promega, Madison, WI), with slight modification from the manufacturer's protocol at day 0 (time when drug was added) and day 3 of drug exposure. Briefly, CTG reagent was diluted with PBS (1:1, volume:volume) and the culture media was removed from the 96-well plate prior to adding 50 ul per well of the diluted CTG reagent. Luminescence from the assay was recorded using BIO-TEK FLx800.

Data calculations were made according to the method described by the NCI/NIH DTP Human Tumor Cell Line Screen Process [95] and as previously described [96]. The percent growth curve is calculated as $[(T-T_0)/(C- T_0)]$ x100, where T_0 is the cell count at day 0, C is the vehicle control (e.g. 0.3% DMSO without drug) cell count at day 3, T is the cell count at the test concentration. The GI_{50} and TGI value are determined as the drug concentration that results in a 50% and 0% growth at 72 hr drug exposure.

Pathway Logic modeling system

Pathway Logic [97] is a system for building discrete, logical models of biological systems [35, 36]. The construction of a Pathway Logic model requires two key elements: a set of *rules* and an *initial state*. Each rule represents a statement of a precisely defined

biological transformation or biochemical reaction. For example, the rule below describes the activation of the ERBB2 receptor by activated EGFR:

```
rl[793.ERBB2.on]:
  {CLm | clm [EGFR - act] ERBB2          }
  =>
  {CLm | clm [EGFR - act][ERBB2 - act] } .
```

The first term on each line represents a cellular location. In this case, CLm indicates that EGFR and ERBB2 are located in the cell membrane. A reaction will occur ("fire") only if the components are located in the specified cellular compartment. Most rules in our database describe changes to the state of a protein, such as activation, exchange of GDP for GTP, or translocation to a different cellular compartment. In total, the relevant rule database contains 396 rules, all of which have been individually curated from primary literature sources.

The initial state specifies the model components present in a cell, as well as their locations. We created the initial states for each network model from a set of 286 components. Models are generated by "rewrites." In a simple rewrite, the initial state is presented to the rules. Whenever the state meets the conditions required by a rule input, the state is adjusted in accordance with the rule. The new state is then presented to the rules and more adjustments are made. This iterative process continues until either no further alterations can be made, or a user-defined condition is reached. We visualize the result of these rewrites as a Petri net, a directed bipartite graph that contains places, transitions, and directed arcs that connect the places and transitions [98]. In Petri net models of cell signaling, places represent proteins and transitions represent chemical reactions. Petri nets are a useful representation because they closely resemble hand-drawn cartoon models of cellular signaling pathways.

Data discretization

We discretized the protein and transcript data in order to determine which components were present in (or absent from) the initial state of each cell line network model. Conceptually, the idea was to analyze the expression data for each protein in the initial state in order to decide if it showed differential expression across the panel of cell lines. Proteins that showed a highly variable expression pattern across the panel of cell lines were considered present in some cell lines and absent from others. Our approach to discretization and creation of the initial states was quite conservative. That is, we did not omit a component from the initial state unless there was strong evidence that it is absent from a particular cell line. We chose a conservative approach because in discrete networks such as these, errant omission of a component from the initial state can lead to significant effects on the structure of the network, in the form of truncated signaling pathways (see **Figure 1B**).

We developed the following discretization method and applied it to both the protein and transcript data. First, for each gene or protein, we used partitioning around medoids (PAM) clustering and a mean split silhouette (MSS) statistic to determine whether the log-transformed expression values are best represented as 1, 2 or 3 groups of cell lines [99]. We searched for 1, 2 or 3 groups because the distributions of expression values appear unimodal (i.e., 1 group, see **Figure 1C**), bimodal (i.e., 2 groups, see **Figure 1D**), or tri-modal (see **Figure 1E**). We used the MSS statistic for three reasons: first, it can be used to classify the expression values as a single group, whereas most algorithms (e.g., k-means) require a minimum of two groups; second, it accurately classified both

one-tailed and two-tailed distributions; and finally, because it could identify small clusters in the data.

Next, for genes that clustered into 2 or 3 groups, we compared the mean expression levels of the groups. If the expression levels between the highest and lowest group differed by less than a 4-fold change, we collapsed the groups together. This ensured that expression differences between the groups were great enough to be meaningful. We assigned proteins to the initial states in the following way. If a single group best described the distribution of expression values, the protein was considered present in all the cell lines. For distributions that yielded more than one group, the protein was considered absent from the initial state of the cell lines with the lowest mean expression; the protein was present in the initial state of cell lines in the highest group(s). We considered the protein present in the two clusters with highest mean expression in order to avoid erroneous omissions from the initial state of cell lines in the middle expression group. Finally, if we had no data available from which to estimate the initial state, we considered the protein present in all cell lines.

For model components that had both transcript and protein data available, we used the clustered protein (rather than transcript) data to populate the model. To ensure that we made the most robust initial state assignments possible, we used data from as many of the 51 cell lines for the discretization step, even if we ultimately did not create a network model for the cell line. We performed the analyses above in R with the hopach package, available as part of the BioConductor tools suite [100].

Analysis of network topology

We used the following method to compare the networks. First, we decomposed each network into a list of all the components and rules contained within it. This list describes all the state changes (e.g., phosphorylation) and reactions in each network. We clustered the network features with PAM and an MSS statistic that searched for the optimal number of clusters, up to a maximum of 40. Each cluster can be considered a unique "signaling module" that represents a small portion of the total network. We compared the presence or absence of these signaling modules across the panel of cell lines.

Hierarchical clustering and data visualization

The discretized data used to populate the initial states was hierarchically clustered using an average linkage algorithm and a Pearson correlation for the distance measure [53]. We also used this algorithm to cluster the cell line network models. We used Java TreeView to visualize the clustered data in Figures 2 and 4 [101].

Abbreviations

CDC42: Cell division control protein 42 homolog; CAV1: Caveolin-1; EPHA2: Ephrin type-A receptor 2 precursor; EGFR: epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian); ER: estrogen/ESR1; ERBB2: v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian); ERBB3: v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian); ERBB4: v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian); FOS: FBJ murine osteosarcoma viral (v-fos) oncogene homolog; FYN: Proto-

oncogene tyrosine-protein kinase FYN; JAK: Tyrosine-protein kinase JAK1 (Janus kinase 1); JUN: Transcription factor AP-1/ V-jun avian sarcoma virus 17 oncogene homolog, p39l; ITGB1: Integrin beta-1 precursor; KRAS: Transforming protein p21 K-Ras 2/Ki-Ras/c-K-ras; KSR1: Kinase suppressor of ras-1; MAPK: mitogen-activated protein kinase/ERK; MEK: Mitogen-activated protein kinase kinase 1/2; NGEF: Ephexin, EPH receptor interacting exchange protein; PAK1: p21 protein (Cdc42/Rac)-activated kinase 1; PI3K: phospho-inositide-3-kinase; PIK3CA: PI3-kinase p110 subunit alpha; PTEN: Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase; PR: progesterone/PGR; RAC1: Ras-related C3 botulinum toxin substrate 1 p21-Rac1; RAF: RAF proto-oncogene serine/threonine-protein kinase; RHOB: ras homolog gene family, member B; SPRY1: sprouty homolog1, antagonist of FGF signaling; SRC: v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog; STAT: Signal transducer and activator of transcription 1-alpha/beta; MSS: mean split silhouette statistic; PAM: partitioning around medoids clustering.

Authors' contributions

LMH participated in the design of the study, constructed the computational models and drafted the manuscript. NJW conducted the PAK1 molecular studies. CLT created the Pathway Logic software and provided guidance of the computational modeling. MK curated the rule library. MK and KRL provided biological guidance of the modeling work. YG, ZH, SZ and WLK conducted the growth inhibition assays. BLW, SL, JRJ and RFW provided the MEK inhibitor GSK1120212. JWG and PTS conceived of the study and participated in its design and coordination.

Acknowledgements

This work was supported by the Director, Office of Science, Office of Biological & Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the National Institutes of Health, National Cancer Institute grants P50 CA 58207 Breast SPORE, the U54 CA 112970 (ICBP), and by the SmithKline Beecham Corporation grant to JWG.

References

1. Albertson DG, Collins C, McCormick F, Gray JW: **Chromosome aberrations in solid tumors**. *Nat Genet* 2003, **34**:369-376.
2. Baylin SB, Herman JG: **DNA hypermethylation in tumorigenesis: epigenetics joins genetics**. *Trends Genet* 2000, **16**:168-174.
3. Jones PA: **Overview of cancer epigenetics**. *Semin Hematol* 2005, **42**:S3-8.
4. Knuutila S, Autio K, Aalto Y: **Online access to CGH data of DNA sequence copy number changes**. *Am J Pathol* 2000, **157**:689.
5. Hanahan D, Weinberg RA: **The hallmarks of cancer**. *Cell* 2000, **100**:57-70.
6. Roberts PJ, Der CJ: **Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer**. *Oncogene* 2007, **26**:3291-3310.
7. Murphy LO, MacKeigan JP, Blenis J: **A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration**. *Mol Cell Biol* 2004, **24**:144-153.
8. Coutts AS, Murphy LC: **Elevated mitogen-activated protein kinase activity in estrogen-nonresponsive human breast cancer cells**. *Cancer Res* 1998, **58**:4071-4074.
9. El-Ashry D, Miller DL, Kharbanda S, Lippman ME, Kern FG: **Constitutive Raf-1 kinase activity in breast cancer cells induces both estrogen-independent growth and apoptosis**. *Oncogene* 1997, **15**:423-435.
10. Sivaraman VS, Wang H, Nuovo GJ, Malbon CC: **Hyperexpression of mitogen-activated protein kinase in human breast cancer**. *J Clin Invest* 1997, **99**:1478-1483.
11. Citri A, Yarden Y: **EGF-ERBB signalling: towards the systems level**. *Nat Rev Mol Cell Biol* 2006, **7**:505-516.
12. Mendelsohn J, Baselga J: **The EGF receptor family as targets for cancer therapy**. *Oncogene* 2000, **19**:6550-6565.
13. Baselga J, Tripathy D, Mendelsohn J, Baughman S, Benz CC, Dantis L, Sklarin NT, Seidman AD, Hudis CA, Moore J, Rosen PP, Twaddell T, Henderson IC, Norton L: **Phase II study of weekly intravenous recombinant humanized anti-p185HER2 monoclonal antibody in patients with HER2/neu-overexpressing metastatic breast cancer**. *J Clin Oncol* 1996, **14**:737-744.
14. Cobleigh MA, Vogel CL, Tripathy D, Robert NJ, Scholl S, Fehrenbacher L, Wolter JM, Paton V, Shak S, Lieberman G, Slamon DJ: **Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease**. *J Clin Oncol* 1999, **17**:2639-2648.
15. Kris MG, Ranson M, Ferry D, Hammond L, Averbuch S, Ochs J, Rowinsky E: **Phase I study of oral ZD1839, a novel inhibitor of epidermal growth factor receptor tyrosine kinase (EGFR-TK): evidence of good tolerability and activity**. *Clinical Cancer Research* 1999, **5**.
16. Ferry D, Hammond L, Ranson M, Kris MG, Miller V, Murray A, Tullo A, Feyereislova A, Averbuch S, Rowinsky E: **Intermittent oral ZD1839 (Iressa), a novel epidermal growth factor receptor tyrosine kinase inhibitor**

(EGFRTKI), shows evidence of good tolerability and activity: final results from a Phase I study. *Proceedings of the American Society for Clinical Oncology* 2000, **19**.

17. Dudley DT, Pang L, Decker SJ, Bridges AJ, Saltiel AR: **A synthetic inhibitor of the mitogen-activated protein kinase cascade**. *Proc Natl Acad Sci U S A* 1995, **92**:7686-7689.
18. Solit DB, Garraway LA, Pratilas CA, Sawai A, Getz G, Basso A, Ye Q, Lobo JM, She Y, Osman I, Golub TR, Sebolt-Leopold J, Sellers WR, Rosen N: **BRAF mutation predicts sensitivity to MEK inhibition**. *Nature* 2006, **439**:358-362.
19. Thompson N, Lyons J: **Recent progress in targeting the Raf/MEK/ERK pathway with inhibitors in cancer drug discovery**. *Curr Opin Pharmacol* 2005, **5**:350-356.
20. Yeh TC, Marsh V, Bernat BA, Ballard J, Colwell H, Evans RJ, Parry J, Smith D, Brandhuber BJ, Gross S, Marlow A, Hurley B, Lyssikatos J, Lee PA, Winkler JD, Koch K, Wallace E: **Biological characterization of ARRY-142886 (AZD6244), a potent, highly selective mitogen-activated protein kinase kinase 1/2 inhibitor**. *Clin Cancer Res* 2007, **13**:1576-1583.
21. Lapidus RG, Nass SJ, Davidson NE: **The loss of estrogen and progesterone receptor gene expression in human breast cancer**. *J Mammary Gland Biol Neoplasia* 1998, **3**:85-94.
22. Oh DS, Troester MA, Usary J, Hu Z, He X, Fan C, Wu J, Carey LA, Perou CM: **Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers**. *J Clin Oncol* 2006, **24**:1656-1664.
23. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, Lashkari D, Shalon D, Brown PO, Botstein D: **Distinctive gene expression patterns in human mammary epithelial cells and breast cancers**. *Proc Natl Acad Sci U S A* 1999, **96**:9212-9217.
24. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**:747-752.
25. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc Natl Acad Sci U S A* 2001, **98**:10869-10874.
26. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM: **The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes**. *Clin Cancer Res* 2007, **13**:2329-2334.
27. Creighton CJ, Hilger AM, Murthy S, Rae JM, Chinnaiyan AM, El-Ashry D: **Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in vivo molecular phenotype of estrogen receptor alpha-negative human breast tumors**. *Cancer Res* 2006, **66**:3903-3911.

28. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW: **Genomic and transcriptional aberrations linked to breast cancer pathophysiologies**. *Cancer Cell* 2006, **10**:529-541.
29. Fridlyand J, Snijders AM, Ylstra B, Li H, Olshen A, Segraves R, Dairkee S, Tokuyasu T, Ljung BM, Jain AN, McLennan J, Ziegler J, Chin K, Devries S, Feiler H, Gray JW, Waldman F, Pinkel D, Albertson DG: **Breast tumor copy number aberration phenotypes and genomic instability**. *BMC Cancer* 2006, **6**:96.
30. Letessier A, Sircoulomb F, Ginestier C, Cervera N, Monville F, Gelsi-Boyer V, Esterni B, Geneix J, Finetti P, Zemmour C, Viens P, Charafe-Jauffret E, Jacquemier J, Birnbaum D, Chaffanet M: **Frequency, prognostic impact, and subtype association of 8p12, 8q24, 11q13, 12p13, 17q12, and 20q13 amplifications in breast cancers**. *BMC Cancer* 2006, **6**:245.
31. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV *et al*: **The genomic landscapes of human breast and colorectal cancers**. *Science* 2007, **318**:1108-1113.
32. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA, Jr., Marks JR, Dressman HK, West M, Nevins JR: **Oncogenic pathway signatures in human cancers as a guide to targeted therapies**. *Nature* 2006, **439**:353-357.
33. Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, Cragun J, Cottrill H, Kelley MJ, Petersen R, Harpole D, Marks J, Berchuck A, Ginsburg GS, Febbo P, Lancaster J, Nevins JR: **Genomic signatures to guide the use of chemotherapeutics**. *Nat Med* 2006, **12**:1294-1300.
34. Nevins JR, Potti A: **Mining gene expression profiles: expression signatures as cancer phenotypes**. *Nat Rev Genet* 2007, **8**:601-609.
35. Eker S, Knapp M, Laderoute K, Lincoln P, Meseguer J, Sonmez K: **Pathway logic: symbolic analysis of biological signaling**. *Pac Symp Biocomput* 2002:400-412.
36. Talcott C, Eker S, Knapp M, Lincoln P, Laderoute K: **Pathway logic modeling of protein functional domains in signal transduction**. *Pac Symp Biocomput* 2004:568-580.
37. Talcott C: **Formal Executable Models of Cell Signaling Primatives**. In: *2nd International Symposium On Leveraging Applications of Formal Methods, Verification and Validation {ISOLA06}.* Edited by Tiziana Margaria APaBS; 2006: 303-307.
38. Talcott C: **Symbolic Modeling of Signal Transduction in Pathway Logic**. In: *2006 Winter Simulation Conference.* Edited by L. F. Perrone FPW, J. Liu, B. G. Lawson, D. M. Nicol and R. M. Fujimoto; 2006: 1656-1665.
39. Hlavacek WS, Faeder JR, Blinov ML, Posner RG, Hucka M, Fontana W: **Rules for modeling signal-transduction systems**. *Sci STKE* 2006, **2006**:re6.

40. Kohn KW, Aladjem MI, Weinstein JN, Pommier Y: **Molecular interaction maps of bioregulatory networks: a general rubric for systems biology**. *Mol Biol Cell* 2006, **17**:1-13.
41. Baselga J, Arteaga CL: **Critical update and emerging trends in epidermal growth factor receptor targeting in cancer**. *J Clin Oncol* 2005, **23**:2445-2459.
42. Yarden Y, Sliwkowski MX: **Untangling the ErbB signalling network**. *Nat Rev Mol Cell Biol* 2001, **2**:127-137.
43. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A, Gray JW: **A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes**. *Cancer Cell* 2006, **10**:515-527.
44. Holst F, Stahl PR, Ruiz C, Hellwinkel O, Jehan Z, Wendland M, Lebeau A, Terracciano L, Al-Kuraya K, Janicke F, Sauter G, Simon R: **Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer**. *Nat Genet* 2007, **39**:655-660.
45. Stierer M, Rosen H, Weber R, Hanak H, Spona J, Tuchler H: **Immunohistochemical and biochemical measurement of estrogen and progesterone receptors in primary breast cancer. Correlation of histopathology and prognostic factors**. *Ann Surg* 1993, **218**:13-21.
46. **Sanger COSMIC Website** [http://www.sanger.ac.uk/genetics/CGP/cosmic/]
47. Kozma SC, Bogaard ME, Buser K, Saurer SM, Bos JL, Groner B, Hynes NE: **The human c-Kirsten ras gene is activated by a novel mutation in codon 13 in the breast carcinoma cell line MDA-MB231**. *Nucleic Acids Res* 1987, **15**:5963-5971.
48. Rochlitz CF, Scott GK, Dodson JM, Liu E, Dollbaum C, Smith HS, Benz CC: **Incidence of activating ras oncogene mutations associated with primary and metastatic human breast cancer**. *Cancer Res* 1989, **49**:357-360.
49. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliaresis C, Rodgers L, McCombie R, Bigner SH, Giovanella BC, Ittmann M, Tycko B, Hibshoosh H, Wigler MH, Parsons R: **PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer**. *Science* 1997, **275**:1943-1947.
50. Gymnopoulos M, Elsliger MA, Vogt PK: **Rare cancer-specific mutations in PIK3CA show gain of function**. *Proc Natl Acad Sci U S A* 2007, **104**:5569-5574.
51. Ikenoue T, Kanai F, Hikiba Y, Obata T, Tanaka Y, Imamura J, Ohta M, Jazag A, Guleng B, Tateishi K, Asaoka Y, Matsumura M, Kawabe T, Omata M: **Functional analysis of PIK3CA gene mutations in human colorectal cancer**. *Cancer Res* 2005, **65**:4562-4567.
52. Campbell IG, Russell SE, Choong DY, Montgomery KG, Ciavarella ML, Hooi CS, Cristiano BE, Pearson RB, Phillips WA: **Mutation of the PIK3CA gene in ovarian and breast cancer**. *Cancer Res* 2004, **64**:7678-7681.
53. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns**. *Proc Natl Acad Sci U S A* 1998, **95**:14863-14868.

54. Hall A: **Rho GTPases and the control of cell behaviour**. *Biochem Soc Trans* 2005, **33**:891-895.
55. Shamah SM, Lin MZ, Goldberg JL, Estrach S, Sahin M, Hu L, Bazalakova M, Neve RL, Corfas G, Debant A, Greenberg ME: **EphA receptors regulate growth cone dynamics through the novel guanine nucleotide exchange factor ephexin**. *Cell* 2001, **105**:233-244.
56. Huang M, Prendergast GC: **RhoB in cancer suppression**. *Histol Histopathol* 2006, **21**:213-218.
57. Wheeler AP, Ridley AJ: **Why three Rho proteins? RhoA, RhoB, RhoC, and cell motility**. *Exp Cell Res* 2004, **301**:43-49.
58. Mazieres J, Antonia T, Daste G, Muro-Cacho C, Berchery D, Tillement V, Pradines A, Sebti S, Favre G: **Loss of RhoB expression in human lung cancer progression**. *Clin Cancer Res* 2004, **10**:2742-2750.
59. Coles LC, Shaw PE: **PAK1 primes MEK1 for phosphorylation by Raf-1 kinase during cross-cascade activation of the ERK pathway**. *Oncogene* 2002, **21**:2236-2244.
60. Snijders AM, Schmidt BL, Fridlyand J, Dekker N, Pinkel D, Jordan RC, Albertson DG: **Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma**. *Oncogene* 2005, **24**:4232-4242.
61. Park ER, Eblen ST, Catling AD: **MEK1 activation by PAK: a novel mechanism**. *Cell Signal* 2007, **19**:1488-1496.
62. Ideker T, Lauffenburger D: **Building with a scaffold: emerging strategies for high- to low-level cellular modeling**. *Trends Biotechnol* 2003, **21**:255-262.
63. Friedman N, Linial M, Nachman I, Pe'er D: **Using Bayesian networks to analyze expression data**. *J Comput Biol* 2000, **7**:601-620.
64. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA: **Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks**. *Pac Symp Biocomput* 2001:422-433.
65. Kholodenko BN, Demin OV, Moehren G, Hoek JB: **Quantification of short term signaling by the epidermal growth factor receptor**. *J Biol Chem* 1999, **274**:30169-30181.
66. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors**. *Nat Biotechnol* 2002, **20**:370-375.
67. Koh G, Teong HF, Clement MV, Hsu D, Thiagarajan PS: **A decompositional approach to parameter estimation in pathway modeling: a case study of the Akt and MAPK pathways and their crosstalk**. *Bioinformatics* 2006, **22**:e271-280.
68. Lincoln P, Tiwari A: **Symbolic systems biology: Hybrid modeling and analysis of biological networks**. In: *Hybrid Systems: Computation and Control, 7th International Workshop, LNCS 2993.* 2004.
69. Dokholyan NV, Shakhnovich B, Shakhnovich EI: **Expanding protein universe and its origin from the biological Big Bang**. *Proc Natl Acad Sci U S A* 2002, **99**:14132-14136.

70. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL: **The large-scale organization of metabolic networks**. *Nature* 2000, **407**:651-654.
71. Ishizawar R, Parsons SJ: **c-Src and cooperating partners in human cancer**. *Cancer Cell* 2004, **6**:209-214.
72. Russello SV, Shore SK: **Src in human carcinogenesis**. *Front Biosci* 2003, **8**:s1068-1073.
73. Benati D, Baldari CT: **SRC family kinases as potential therapeutic targets for malignancies and immunological disorders**. *Curr Med Chem* 2008, **15**:1154-1165.
74. Johnson FM, Gallick GE: **SRC family nonreceptor tyrosine kinases as molecular targets for cancer therapy**. *Anticancer Agents Med Chem* 2007, **7**:651-659.
75. Jiang K, Delarue FL, Sebti SM: **EGFR, ErbB2 and Ras but not Src suppress RhoB expression while ectopic expression of RhoB antagonizes oncogene-mediated transformation**. *Oncogene* 2004, **23**:1136-1145.
76. Williams TM, Hassan GS, Li J, Cohen AW, Medina F, Frank PG, Pestell RG, Di Vizio D, Loda M, Lisanti MP: **Caveolin-1 promotes tumor progression in an autochthonous mouse model of prostate cancer: genetic ablation of Cav-1 delays advanced prostate tumor development in tramp mice**. *J Biol Chem* 2005, **280**:25134-25145.
77. Williams TM, Medina F, Badano I, Hazan RB, Hutchinson J, Muller WJ, Chopra NG, Scherer PE, Pestell RG, Lisanti MP: **Caveolin-1 gene disruption promotes mammary tumorigenesis and dramatically enhances lung metastasis in vivo. Role of Cav-1 in cell invasiveness and matrix metalloproteinase (MMP-2/9) secretion**. *J Biol Chem* 2004, **279**:51630-51646.
78. Sells MA, Boyd JT, Chernoff J: **p21-activated kinase 1 (Pak1) regulates cell motility in mammalian fibroblasts**. *J Cell Biol* 1999, **145**:837-849.
79. Sells MA, Knaus UG, Bagrodia S, Ambrose DM, Bokoch GM, Chernoff J: **Human p21-activated kinase (Pak1) regulates actin organization in mammalian cells**. *Curr Biol* 1997, **7**:202-210.
80. Balasenthil S, Sahin AA, Barnes CJ, Wang RA, Pestell RG, Vadlamudi RK, Kumar R: **p21-activated kinase-1 signaling mediates cyclin D1 expression in mammary epithelial and cancer cells**. *J Biol Chem* 2004, **279**:1422-1428.
81. Schurmann A, Mooney AF, Sanders LC, Sells MA, Wang HG, Reed JC, Bokoch GM: **p21-activated kinase 1 phosphorylates the death agonist bad and protects cells from apoptosis**. *Mol Cell Biol* 2000, **20**:453-461.
82. Carter JH, Douglass LE, Deddens JA, Colligan BM, Bhatt TR, Pemberton JO, Konicek S, Hom J, Marshall M, Graff JR: **Pak-1 expression increases with progression of colorectal carcinomas to metastasis**. *Clin Cancer Res* 2004, **10**:3448-3456.
83. Schraml P, Schwerdtfeger G, Burkhalter F, Raggi A, Schmidt D, Ruffalo T, King W, Wilber K, Mihatsch MJ, Moch H: **Combined array comparative genomic hybridization and tissue microarray analysis suggest PAK1 at 11q13.5-q14 as a critical oncogene target in ovarian carcinoma**. *Am J Pathol* 2003, **163**:985-992.

84. Wang RA, Zhang H, Balasenthil S, Medina D, Kumar R: **PAK1 hyperactivation is sufficient for mammary gland tumor formation**. *Oncogene* 2006, **25**:2931-2936.
85. Hanafusa H, Torii S, Yasunaga T, Nishida E: **Sprouty1 and Sprouty2 provide a control mechanism for the Ras/MAPK signalling pathway**. *Nat Cell Biol* 2002, **4**:850-858.
86. Roy F, Laberge G, Douziech M, Ferland-McCollough D, Therrien M: **KSR is a scaffold required for activation of the ERK/MAPK module**. *Genes Dev* 2002, **16**:427-438.
87. **LBNL Breast Cancer Website** [http://cancer.lbl.gov/breastcancer/data.php]
88. **Scion Image** [http://www.scioncorp.com/]
89. **Odyssey software website** [http://www.licor.com/]
90. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data**. *Biostatistics* 2003, **4**:249-264.
91. **BioConductor Website** [http://www.bioconductor.org]
92. **UCSC Genome Browser** [http://genome.ucsc.edu/]
93. Wang Y, Moorhead M, Karlin-Neumann G, Wang NJ, Ireland J, Lin S, Chen C, Heiser LM, Chin K, Esserman L, Gray JW, Spellman PT, Faham M: **Analysis of molecular inversion probe performance for allele copy number determination**. *Genome Biol* 2007, **8**:R246.
94. O'Neill RA, Bhamidipati A, Bi X, Deb-Basu D, Cahill L, Ferrante J, Gentalen E, Glazer M, Gossett J, Hacker K, Kirby C, Knittle J, Loder R, Mastroieni C, Maclaren M, Mills T, Nguyen U, Parker N, Rice A, Roach D, Suich D, Voehringer D, Voss K, Yang J, Yang T, Vander Horn PB: **Isoelectric focusing technology quantifies protein signaling in 25 cells**. *Proc Natl Acad Sci U S A* 2006, **103**:16153-16158.
95. **NCI/NIH DTP Human Tumor Cell Line Screen Process Website** [http://dtp.nci.nih.gov/branches/btb/ivclsp.html]
96. Monks A, Scudiero D, Skehan P, Shoemaker R, Paull K, Vistica D, Hose C, Langley J, Cronise P, Vaigro-Wolff A, Gray-Goodrich M, Campbell H, Mayo J, Boyd M: **Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines**. *J Natl Cancer Inst* 1991, **83**:757-766.
97. **Pathway Logic Website** [http://pl.csl.sri.com]
98. Peterson RJL: **Petri Nets: Properties, Analysis and Applications**. New Jersey: Prentice-Hall; 1981.
99. Pollard KS, van der Laan MJ: **A method to identify significant clusters in gene expression data**. *UC Berkeley Division of Biostatistics Working Paper Series* 2002:Working Paper 107.
100. **BioConductor Hopach Package Website** [http://www.bioconductor.org/packages/2.3/bioc/html/hopach.html]
101. Saldanha AJ: **Java Treeview--extensible visualization of microarray data**. *Bioinformatics* 2004, **20**:3246-3248.

Figure Legends

Figure 1. The signaling networks include several hundred components, all connected in a discrete manner. A. Example network. Each circle represents a component in the network; lines represent connections between them (i.e., rules). Key signaling components are noted. B. A small subnetwork. C, D, E. Examples of data used to populate the model. Each histogram shows the distribution of expression values across the complete panel of cell lines. Data for each component in the model were clustered individually to determine whether or not the component should be included in the initial state. Components that clustered into two groups were present in the initial states of some cell lines and absent from others. A. RAF1 Transcript data yielded a single group. B. ERBB4 protein data yielded two groups. C. ESR1 yielded three groups.

Figure 2. Initial states recapitulate the known biology. Heatmap shows the components in the initial states that varied across the cell lines. Each column represents the initial state from a single cell line network; each row represents data for one component. Red indicates the component was present in the cell line; green indicates it was absent. Data are hierarchically clustered along both dimensions. Basal and luminal cell lines cluster into distinct groups.

Figure 3. Network connectivity follows a power-law relationship. A. Distribution of the number of rule connections for each component in the model. Most components have only a few rule connections. B. Log-log plot. Each dot represents the number of components in the model that have a particular number of rule connections. The line represents the least-squares fit to the data.

Figure 4. The network models cluster into basal, luminal and mixed groups of cell lines. Heatmap shows the network features that varied across the cell line network models. Each column represents data from one network model; each row represents data for one network feature (component in the initial state, rule or component that underwent a state-change). Red indicates the component was present in the cell line; green indicates it was absent. Hierarchical clustering along the vertical dimension reveals that the networks form basal, luminal and mixed clusters. Hierarchical clustering along the horizontal dimension yields 30 signaling modules, each of which represents a small subnetwork. Signaling modules of particular interest, along with the key components in the initial state, are noted along the right side.

Figure 5. CAV1/ITGB1 signaling module is present in basal cell lines. A. Signaling module. CAV1, ITGB1 and FYN interact to activate SHC, which leads to activation of the MAPK cascade. B, C. Distribution of CAV1 transcript (B) and protein (C) levels across the cell lines. Both datasets show a bimodal distribution of CAV1.

Figure 6. PAK1 is a critical component of the MAPK cascade in our network models. A. Subnet shows that PAK1 leads directly to activation of RAF, MEK and ERK. B. Distribution of PAK1 transcript levels used in construction of the initial states. PAK1 yielded two clusters: a lower 'absent' cluster centered at -0.26 and an upper 'present' cluster centered at 2.16. C. PAK1 protein and transcript levels are correlated. Protein

abundance is plotted on the x-axis; transcript data (log2 scale) is plotted along the y-axis. Line represents the least-squares fit to the data. D. Copy number profiles for the region around the PAK1 amplicon on chromosome 11. The vertical lines represent the locations of CCND1 (69Mb) and PAK1 (76Mb). E. PAK1 protein levels are correlated with percent phospho-MEK1. Each dot represents data from one cell line. The line represents the least-squares fit to the data.

Figure 7. PAK1 over-expression predicts responsiveness to MEK inhibitors. Each pair of boxplots represents the average GI50 for luminal cell lines that over-express PAK1 (PAK1+, blue) and those that express it at normal levels (PAK1-, red). Within each box, the line represents the median; upper and lower boundaries represent the first and third quartiles, respectively. The vertical lines extend to +/- 1.5IQR. For all three drugs, PAK1+ cell lines are significantly more sensitive than PAK1- cell lines.

Table 1. Comparison of discretized protein and transcript data

	Protein clusters	**Transcript clusters**	**Pearson's correlation**
IRS1	2	2	0.0354
EGFR	2	2	0.491
ERBB3	2	2	0.491
CAV1	2	2	0.523
CD44	2	2	0.6
CAV2	2	2	0.882
ESR1	2	2	0.883
CDH1	2	2	0.923
AKT1	1	1	-
GRB2	1	1	-
HRAS	1	1	-
IGF1R	1	1	-
JAK1	1	1	-
KRAS	1	1	-
MAPK1	1	1	-
MAPK3	1	1	-
PTK2	1	1	-
ERBB2	1	2	-
GRB7	1	2	-
CTNNB1	2	1	-
EFNA1	2	1	-
ERBB4	2	1	-
RELA	2	1	-
SRC	2	1	-
JUN	2	1	-

Table 2. Summary of network features for the cell line models

	Total	**Number variant**	**Percent variant**
Rules	396	248	60
State changes	253	141	55
Initial state	286	39	13

Table 3. The most highly connected components in the network model

Component	Number of rule connections	Variable across initial state
EgfR	22	yes
Pi3k	20	yes
Src	18	yes
Kras	17	yes
Rhob	17	yes
Rhoa	17	no
Cbl	16	no
Cdc42	16	no
Rac1	16	no
Erk1	15	no
Erk2	15	no
Hras	14	no
Grb2	13	no
PIP2	13	no
Raf1	13	no
Smad2	12	no
Acta1	11	no
Epha2	11	no
Pkca	11	no

Table 4. Summary of signaling modules

	Number of rules	Key component(s)	Summary of key events
1	1	Pi3k, Erbb4	ErbB4 activation of Pi3k
2	1	Snca	Pyk2 activation of Snca
3	1	Caml, Rsk	Rsk activation of Caml
4	1	Stat3	Stat3 activation by EgfR
5	1	Irs, Pi3k	Irs activation of Pi3k
6	1	Rela	Formation of Ikba, Nfkb1, Rela complex
7	2	Pik3ca-mut	Akt signaling through Pi3k mutant
8	2	Mef2c	Camk activation of Mef2c
9	2	IL11R, Jak	IL11R activation of Jak
10	2	Elmo, Rac1	Elmo activation of Rac1
11	2	Abi1, Pir121	Wave1 activation dependent on Abi1 and Pir121
12	3	Mylk	Mylk activation of Mlc
13	3	Rhob	Rhob activation
14	3	EsR1, Bcat	EsR1 activation by Rsk; Bcat activation
15	3	Fos	Fos activation by Erk
16	3	Bcat	Activation and degredation of Bcat
17	4	Cav1, UpaR	Integrin/Cav1 activation of Shc; UpaR activation
18	5	Pten, Kras, Pik3ca	Mutation rules
19	5	ErbB4	ErbB4 activation of ErbB2, Shc; Grb2 relocation
20	6	PrlR	PrlR signaling
21	7	Irs1	Irs1 activation; Grb2 translocation
22	8	Pi3k	8 ways to activate Pi3k
23	12	Rhob	Rhob activation of first-order effectors
24	12	Cbl	Cbl-related signaling, including Rap1a, Crk, Dock
25	14	Src	Src-related signaling, including Fak, Pax, Cas
26	15	EgfR	First-order EgfR interactions including ErbB2, Grb2, Cbl
27	16	Efna1	Epha2/Efna1 signaling; Integrin deactivation by Epha2
28	27	ErbB3	ErbB3 activation by Nrg1 and ErbB2; ErbB3 activation of Shc
29	32	EgfR, Src	Src-dependent activation of EgfR; Cdc42 signaling; activation of Src effectors
30	76	Src, Rac1-GTP	Rac1 signaling; MAPK activation

Fig. 1

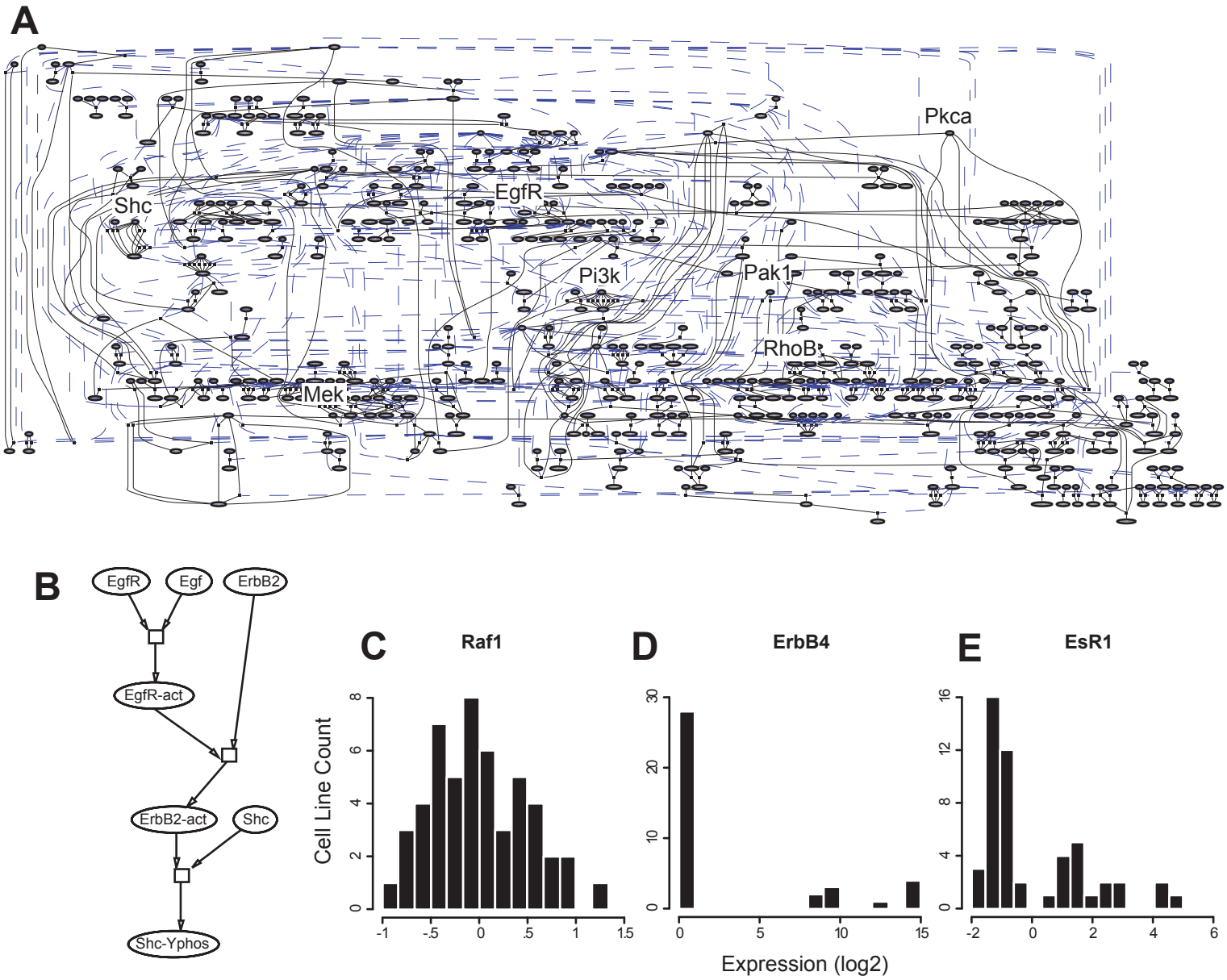

A
Pkca
Shc
EgfR
Pi3k
Pak1
RhoB
Mek
B
EgfR
Egf
ErbB2
EgfR-act
ErbB2-act
Shc
Shc-Yphos
C
Raf1
D
ErbB4
E
EsR1
Cell Line Count
Expression (log2)

Fig. 2

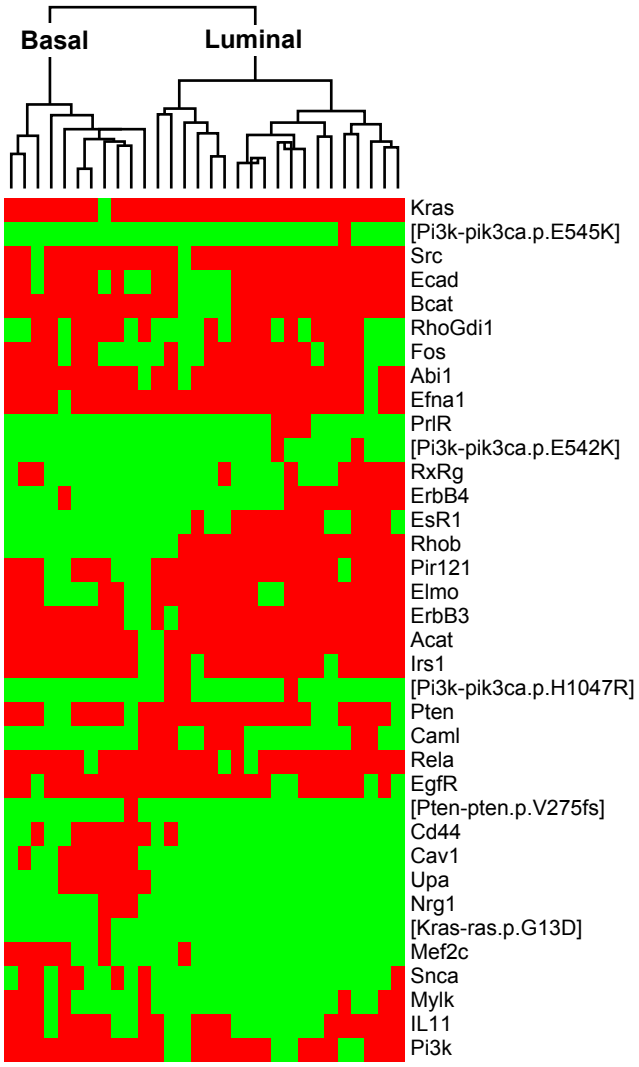

Basal
Luminal
Kras
[Pi3k-pik3ca.p.E545K]
Src
Ecad
Bcat
RhoGdi1
Fos
Abi1
Efna1
PrlR
[Pi3k-pik3ca.p.E542K]
RxRg
ErbB4
EsR1
Rhob
Pir121
Elmo
ErbB3
Acat
Irs1
[Pi3k-pik3ca.p.H1047R]
Pten
Caml
Rela
EgfR
[Pten-pten.p.V275fs]
Cd44
Cav1
Upa
Nrg1
[Kras-ras.p.G13D]
Mef2c
Snca
Mylk
IL11
Pi3k

Fig. 3

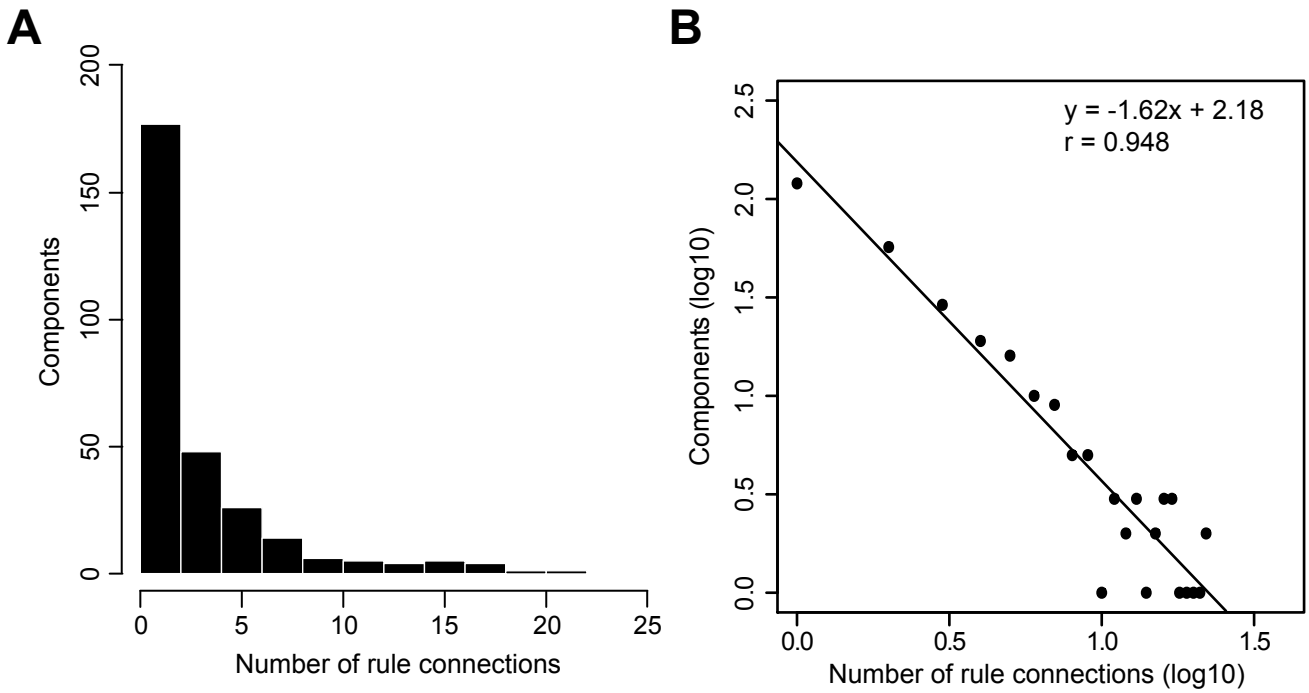

A
Components
0
50
100
150
200
0
5
10
15
20
25
Number of rule connections
B
y = -1.62x + 2.18
r = 0.948
Components (log10)
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.5
1.0
1.5
Number of rule connections (log10)

Fig. 4

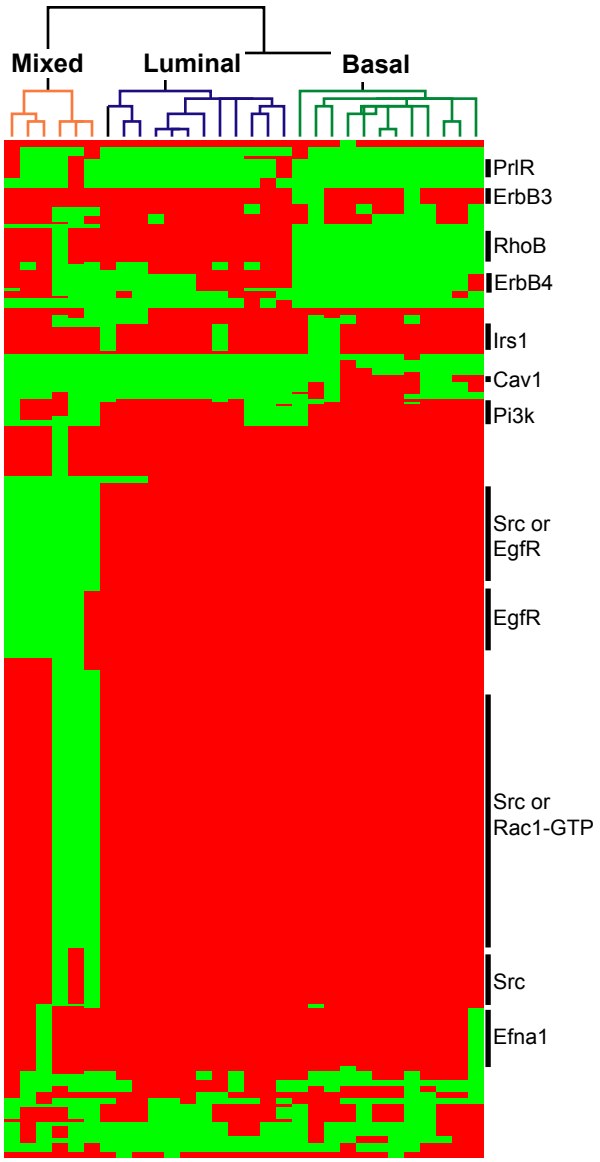

Mixed
Luminal
Basal
PrlR
ErbB3
RhoB
ErbB4
Irs1
Cav1
Pi3k
Src or
EgfR
EgfR
Src or
Rac1-GTP
Src
Efna1

Fig. 5

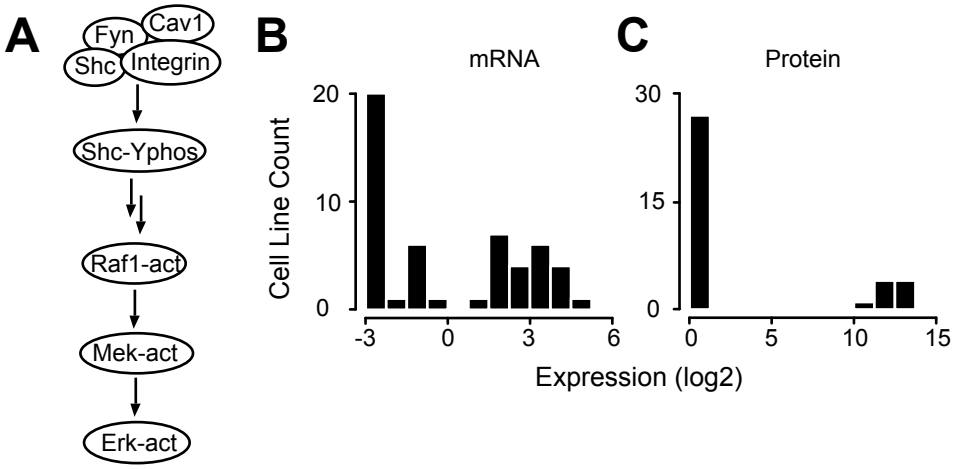

A
Fyn
Cav1
Shc
Integrin
Shc-Yphos
Raf1-act
Mek-act
Erk-act
B
mRNA
Cell Line Count
20
10
0
-3
0
3
6
C
Protein
30
15
0
0
5
10
15
Expression (log2)

Fig. 6

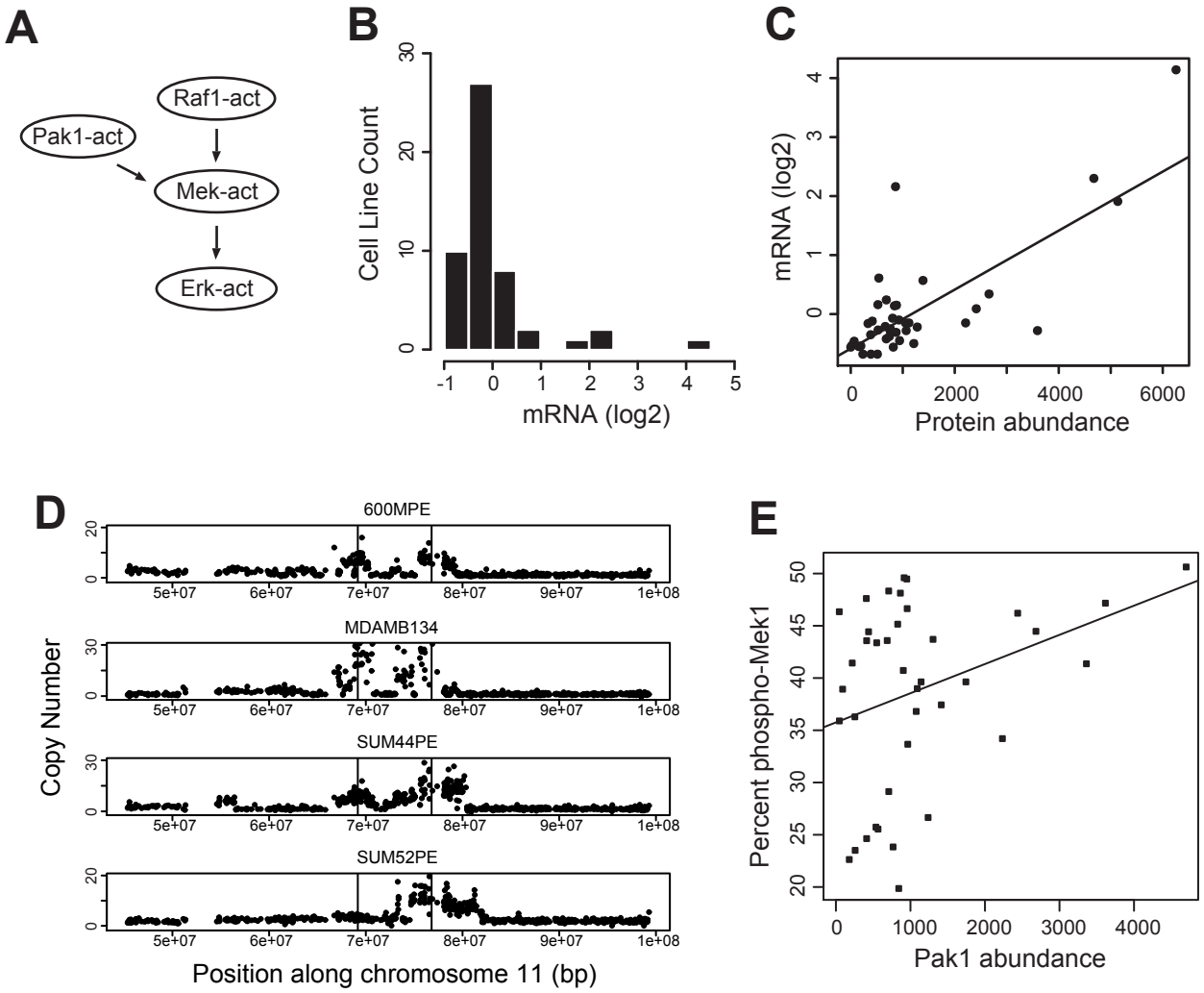
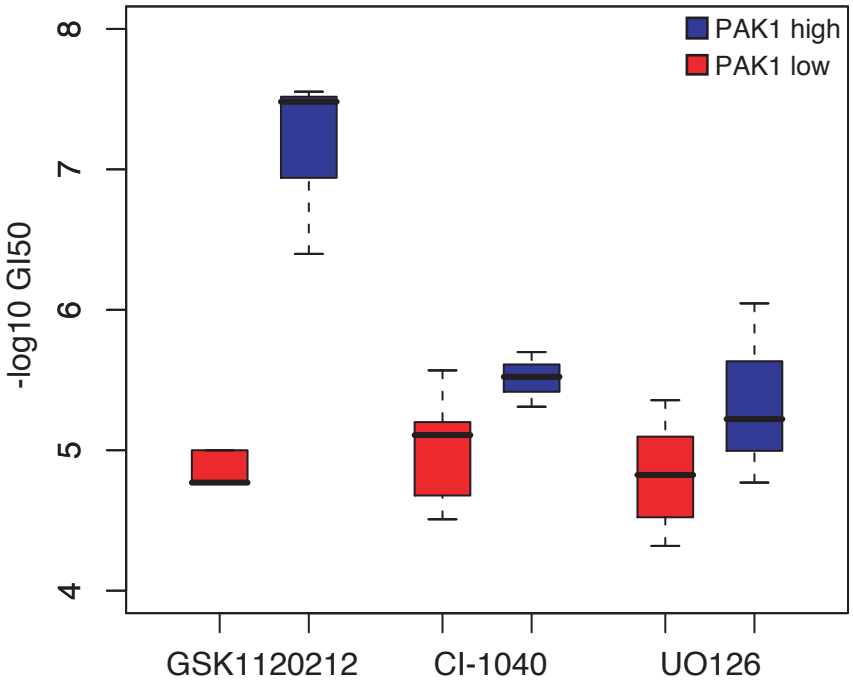

A
Raf1-act
Pak1-act
Mek-act
Erk-act
B
Cell Line Count
mRNA (log2)
C
mRNA (log2)
Protein abundance
D
600MPE
MDAMB134
SUM44PE
SUM52PE
Copy Number
Position along chromosome 11 (bp)
E
Percent phospho-Mek1
Pak1 abundance

Fig. 7


PAK1 high
PAK1 low
-log10 GI50
4
5
6
7
8
GSK1120212
CI-1040
UO126