

Systems Analysis of the NCI-60 Cancer Cell Lines by Alignment of Protein Pathway Activation Modules with "-OMIC" Data Fields and Therapeutic Response Signatures

Giulia Federici^{1,3}, Xi Gao⁴, Janusz Slawek⁴, Tomasz Arodz⁴, Amanuel Shitaye³, Julia D. Wulfkuehle³, Ruggero De Maria², Lance A. Liotta³, and Emanuel F. Petricoin III³

Abstract

The NCI-60 cell line set is likely the most molecularly profiled set of human tumor cell lines in the world. However, a critical missing component of previous analyses has been the inability to place the massive amounts of "-omic" data in the context of functional protein signaling networks, which often contain many of the drug targets for new targeted therapeutics. We used reverse-phase protein array (RPPA) analysis to measure the activation/phosphorylation state of 135 proteins, with a total analysis of nearly 200 key protein isoforms involved in cell proliferation, survival, migration, adhesion, etc., in all 60 cell lines. We aggregated the signaling data into biochemical modules of interconnected kinase substrates for 6 key cancer signaling pathways: AKT, mTOR, EGF receptor (EGFR), insulin-like growth factor-1 receptor (IGF-1R), integrin, and apoptosis signaling. The net activation state of these protein network modules was correlated to available individual protein, phosphoprotein, mutational, metabolomic, miRNA, transcriptional, and drug sensitivity data. Pathway activation mapping identified reproducible and distinct signaling cohorts that transcended organ-type distinctions. Direct correlations with the protein network modules involved largely protein phosphorylation data but we also identified direct correlations of signaling networks with metabolites, miRNA, and DNA data. The integration of protein activation measurements into biochemically interconnected modules provided a novel means to align the functional protein architecture with multiple "-omic" data sets and therapeutic response correlations. This approach may provide a deeper understanding of how cellular biochemistry defines therapeutic response. Such "-omic" portraits could inform rational anticancer agent screenings and drive personalized therapeutic approaches. *Mol Cancer Res*; 11(6): 676–85. ©2013 AACR.

Introduction

The NCI-60 cell lines are a set of human tumor cell lines collected by the National Cancer Institute (NCI) over the last 20 years to accomplish drug screening tests of more than 100,000 chemical compounds and natural extracts (1). This panel represents the most common solid and soft tumors derived from 9 different tissues such as blood, lung, colon, kidney, breast, skin, prostate, ovary, and central nervous system. The NCI-60 cell line set is used in labo-

raries throughout the world as *in vitro* tumor models, thanks to their advantages of reproducibility, availability, and representation of tumor site lineages. Most recently, the NCI-60 cell lines have been characterized by a number of high-throughput molecular profiling efforts through DNA mutations (2), RNA (3), single-nucleotide polymorphisms (4), miRNAs (5), metabolomic, proteomic (6), and karyotyping (7) screens, which have led to a better understanding of the biology of these cell lines, and an increased understanding concerning the relationships between therapeutic resistance/sensitivity and the underpinning biology. Many scientists are now focusing on protein-based analysis for these types of studies because of the proximity of the proteome to the mechanism of action of most therapeutics and their primary relationship to cellular biochemistry. Past efforts used techniques such as 2-dimensional PAGE, mass spectrometry-based profiling, and both antibody and protein microarrays (8).

Reverse-phase protein microarray (RPPA) technology (9), in particular, has been used to gain better insights into the expression profiles of the NCI-60 set due to its ability to quantitatively measure a large number of protein analytes at once with extremely high sensitivity, and was used to

Authors' Affiliations: ¹Department of Hematology, Oncology and Molecular Medicine, Istituto Superiore di Sanità; ²Regina Elena National Cancer Institute, Rome, Italy; ³Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas; and ⁴Department of Computer Science, School of Engineering, Virginia Commonwealth University, Richmond, Virginia

Note: Supplementary data for this article are available at Molecular Cancer Research Online (<http://mcr.aacrjournals.org/>).

Corresponding Author: Emanuel F. Petricoin, George Mason University, 10900 University Blvd, Bull Run Hall, Room 325, Manassas, VA 20110. Phone: 703-993-8646; Fax: 703-993-8606; E-mail: epetrico@gmu.edu

doi: 10.1158/1541-7786.MCR-12-0690

©2013 American Association for Cancer Research.

measure a subscribed set of proteins and phosphoproteins from the NCI-60 cell lines (6, 10). However, these past efforts have failed to systematically interrogate the signaling architecture of this important set using broad-scale, functional phosphoproteomic mapping within the context of a focused network-oriented approach. In our work described herein, we used RPPA to measure the activation/phosphorylation state of 135 key signaling proteins, and a total of nearly 200 protein endpoints (i.e., cleaved, phosphorylated, and/or total protein isoforms) involved in many aspects of tumorigenesis and metastatic progression and representing drug targets for a number of current and experimental therapeutics across the entire 60 lines that comprise the NCI-60 set. We aggregated these individual protein measurements in a system-based, pathway-oriented manner that was then used to interrogate these pathway activity profiles more deeply within the context of the myriad of "-omic" measurements and coordinate enormous existing drug sensitivity data sets. This analysis uncovered new biochemical linkages between drug sensitivities, RNA, DNA, protein, phosphoprotein, and metabolomic profiles, and could provide a beginning step to fully link drug sensitivity and "-omic"-based interconnections to a network-focused view of tumor biology.

Materials and Methods

Protein lysate preparation

Three independent sets of the NCI-60 cell lines were obtained as frozen, nonviable cell pellets from the Developmental Therapeutics Program (DTP), NCI (Frederick, MD; $N = 180$). The cell pellets were lysed in buffer containing T-PER reagent (Thermo-Fisher Scientific), 300 mmol/L NaCl, 1 mmol/L orthovanadate, 2 mmol/L Pefabloc (Roche), 5 $\mu\text{g}/\text{mL}$ aprotinin (Sigma), 5 $\mu\text{g}/\text{mL}$ pepstatin A (Sigma), and 5 $\mu\text{g}/\text{mL}$ leupeptin (Sigma), and incubated on ice for 20 minutes. Samples were centrifuged at $9,300 \times g$ for 5 minutes, and the supernatant transferred to fresh tubes. Protein concentrations were measured using Coomassie protein assay reagent (Thermo Fisher Scientific). The lysates were then diluted for printing in extraction buffer containing 50% T-PER (Thermo Fisher Scientific), 47.5% $2\times$ SDS (Invitrogen), and 2.5% β -mercaptoethanol (Thermo Fisher Scientific) to concentrations of 0.5 $\mu\text{g}/\mu\text{L}$ and 0.125 $\mu\text{g}/\mu\text{L}$.

Reverse-phase protein array analysis

RPPAs were constructed and analyzed as previously described (9). Briefly, samples from the replicate sets of the 60 cell lines were printed in triplicate spots on nitrocellulose-coated glass slides (GRACE Bio-Labs) using an Aushon 2470 arrayer equipped with 185 μm pins (Aushon Biosystems), according to the manufacturer's instructions. Reference standard lysates, composed of HeLa + Pervanadate (BD), Jurkat + Etoposide (Cell Signaling), and Jurkat + Calyculin A (Cell Signaling) cell lysates, were printed in 10-point dilution curves as procedural controls and as positive controls for antibody staining. Each reference standard curve

was printed in triplicate at concentrations of 0.5 $\mu\text{g}/\mu\text{L}$ and 0.125 $\mu\text{g}/\mu\text{L}$. A selected subset of the printed array slides were stained with Sypro Ruby Protein Blot Stain (Invitrogen) to estimate sample total protein concentration, and the remaining slides were stored desiccated at -20°C . Just before antibody staining, printed slides were treated with $1\times$ ReBlot Mild Solution (Chemicon) for 15 minutes, washed 2 times for 5 minutes with PBS (Invitrogen), and incubated for 1 hour in blocking solution [2% I-Block (Applied Biosystems), 0.1% Tween-20 in PBS]. Immunostaining was completed on an automated slide stainer using a catalyzed signal amplification kit (DAKO). The arrays were probed with a library of almost 200 antibodies against total, cleaved, and phosphoprotein endpoints (Supplementary Table S1). Primary antibody binding was detected using a biotinylated goat anti-rabbit immunoglobulin G (IgG) H+L (1:7500; Vector Laboratories) or rabbit anti-mouse IgG (1:10; DAKO) followed by streptavidin-conjugated IRDye680 fluorophore (LI-COR Biosciences). Before use, primary antibodies were extensively validated for single-band specificity by Western immunoblotting with complex cellular lysates. Negative control slides were incubated with secondary antibody only. All Sypro and immunostained slides were scanned using a Revolution 4550 scanner (Vidar Corp.), and acquired images were analyzed with MicroVigene v4.0.0.0 (VigeneTech), which conducted spot detection, local background subtraction, negative control subtraction, replicate averaging, and total protein normalization, producing a single value for each sample. Unsupervised hierarchical clustering was conducted with JMP v5.1 (SAS Institute). Endpoint relative intensity correlation plots were conducted with GraphPad Prism v5 (GraphPad Software Inc.).

Protein pathway activation score determination

Protein signaling pathway activation modules for AKT, mTOR, EGF receptor (EGFR), insulin-like growth factor-1 receptor (IGF-1R), integrin, and apoptosis signaling were defined on the basis of known biochemical linkages between the individual phosphoproteins quantitatively measured by RPPA. Pathway activation module scores were calculated by first scaling the relative intensity values within each endpoint to the sample with the highest value, resulting in values ranging from 1 to 0 that were designated as the "single endpoint score". Second, final pathway activation module scores for each sample were generated by summing the single individual phosphoprotein score for each endpoint component in a given module. These scores, referred to as an "overall module score" or "pathway activation score," represent the whole activation status of each of the 6 pathways under consideration for each cell line.

"-Omic" network analysis

We downloaded the following normalized datasets for the NCI-60 cell lines from CellMiner (11): mRNA expression measured using Affymetrix HG-U133, miRNA expression measured using Agilent Human miRNA microarray and OSU V3 chip, DNA copy number measured with OncoBAC DNA microarrays, and DNA mutation data. From the

same source, we also downloaded drug responses measured as $-\log(GI_{50})$ for datasets A118. From NCI DTP (12), we downloaded the Metabolon metabolomic dataset, Sequenom DNA methylation dataset, and drug responses for 97 U.S. Food and Drug Administration-approved anticancer drugs.

From the mRNA expression data, we inferred coexpression networks using ARACNe (13), CLR (14), MIRNET (15), C3NET (16), and GENIE3 (17). In the first 3 methods, we constructed 2 networks for each method, using Spearman rank coefficient and mutual information as a relationship measure. From the 8 resulting gene coexpression networks, we constructed a single consensus network, by taking edges that were present in all 8 networks, and also edges that were present in at least 6 networks if the 2 genes for the edge were significantly positively or negatively correlated at 2-tailed $P \leq 0.05$ after Bonferroni correction. We repeated the same procedure with phosphoprotein data, arriving with a consensus protein coexpression network.

From the inferred cell-wide protein and gene consensus coexpression networks, we narrowed our focus to networks centered around individual pathway activation scores. Each such network consists of proteins, genes, and other entities that were directly correlated with the score. We calculated Pearson correlation coefficients, with Bonferroni-adjusted 2-tailed $P = 0.05$ cutoff, to detect genes (mRNA expression), DNA mutation, methylation and copy number, metabolites, and drug sensitivities that are directly correlated with each pathway activation score. For DNA mutation, a dichotomous variable, we measured the point-biserial correlation coefficient. We also detected correlations between the scores and all individual total, phospho, and cleaved proteins, using a more strict Bonferroni-adjusted $P = 0.01$ cutoff. On top of proteins and genes directly correlated with the score, we also included their immediate neighbors in the inferred consensus coexpression networks, and any edges between them that existed in those networks.

Results

Pathway activation profiling of the NCI-60 tumor cell lines

To understand the basal signaling network architecture of the NCI-60 panel, we conducted an initial broad-scale protein pathway activation mapping analysis of 194 proteins and phosphoproteins, the largest number measured for the set to date. The full RPPA dataset can be accessed at: http://dtpsearch.ncifcrf.gov/WEBDATA_PETRICOIN_PROTEIN.zip. The results of this mapping effort are shown in Fig. 1 and revealed that clustering, composed of 3 major groups, was largely independent of the cell line origin classification (see also Supplementary Fig. S1) but was defined by the underpinning pathway activation. Other subgroups were readily apparent, composed of multiple tumor origin types and characterized by activation of several members of the EGFR superfamily as well as the downstream mTOR signaling (see also Supplementary Fig. S1).

Network module construction and analysis

On the basis of the pathway-centric nature of the signaling architecture revealed, and to more fully interrogate the relationships between phosphorylation-driven signaling at the pathway level and other "-omic" and drug sensitivity relationships, we took advantage of the fact that we measured the activation/phosphorylation of a number of key signaling proteins that spanned specific signaling pathways and developed biochemically linked signaling "modules" for systems-level analysis. We focused our "pathway module" analyses on 6 important signal transduction pathways that are central both to tumorigenic mechanisms as well as to therapeutics being evaluated in many ongoing cancer clinical trials as well as predictive marker analysis: EGFR, integrin, IGF-1R, mTOR, AKT, and apoptosis signaling (Table 1; Fig. 2 and Supplementary Fig. S2), and for which we were able to measure a large number of biochemically

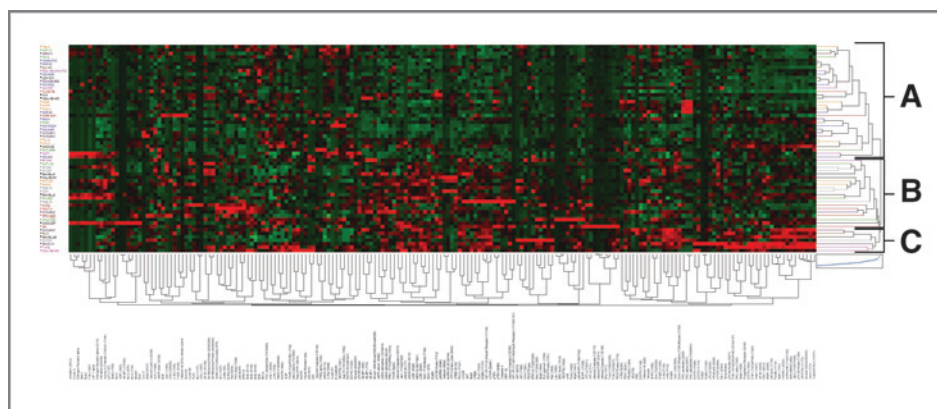


Figure 1. Two-way, unsupervised hierarchical clustering representing the 60 tumor cell lines (vertical axis) and the analyzed total, phospho- and cleaved proteins (horizontal axis). The clustering reveals the formation of 3 major groups (A, B, and C) that are largely independent upon the cell line origin classification but are defined by the underpinning pathway activation. Within the heatmap, red color represents higher levels of relative activity/expression; black represents intermediate levels, and green represents lower levels of relative activity/expression. The 60 tumor cell lines have been highlighted by colors indicating the tissue origin: orange – kidney, blue – lung, pink – breast, red – blood, green – colon, brown – prostate, purple – ovary, black – skin, and gray – central nervous system.

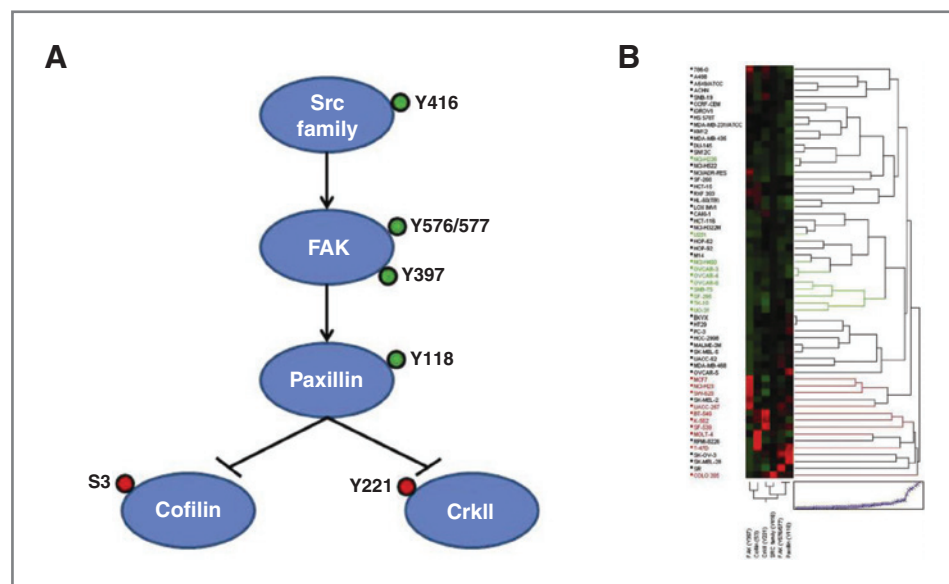
Table 1. Signaling proteins included in signaling module scores

Integrin	IGF-1R	Apoptosis	mTOR	AKT	EGFR
Src family (Y416)	IGF-1R (Y1135/1136)/Insulin Receptor (Y1150/1151)	BAD	mTOR (S2448)	PDK1 (S241)	EGFR (Y1148)
FAK (Y576/577)	IGF-1R (Y1131)/Insulin Receptor (Y1146)	BAX	mTOR (S2481)	AKT (S473)	EGFR (Y845)
FAK (Y397)	IRS-1 (S612)	Smac/diablo	4E-BP1 (S65)	AKT (T308)	EGFR (Y1173)
Paxillin (Y118)	SHC (Y317)	XIAP	4E-BP1 (T70)	GSK3 α/β (S21/9)	Ras-GRF1 (S916)
Cofilin (S3)	GAB1 (Y627)	Cleaved Caspase-9 (D315)	eIF4E (S209)	PRAS40 (T246)	AKT (S473)
CrkII (Y221)	SHP2 (Y580)	Cleaved Caspase-9 (D330)	eIF4G (S1108)	Tuberin/TSC2 (Y1571)	AKT (T308)
		Cleaved Caspase-3 (D175)	p70S6K (T389)		SHC (Y317)
			p70S6K (T412)		C-Raf (S338)
			p70S6K (S371)		C-Raf (S259)
			S6 Ribosomal Protein (S240/244)		B-Raf (S445)
					A-Raf (S299)
					MEK1/2 (S217/221)
					ERK1/2 (T202/Y204)
					p90RSK (S380)

linked kinase substrate signaling proteins. Pathway activation scores were determined for each module such that every cell line had a continuous variable integer determinant for each pathway module. The individual proteins/phosphoproteins that were used to underpin each pathway module are shown in Table 1. In each case, we considered only the most directly involved first-order kinase–substrate relationships to minimize interactions with other signaling or pathway feedback loops. The images in Fig. 2, with the integrin pathway shown as an example, show the correlation

between this ranking and the hierarchical clustering and clearly indicate which cell lines had high overall pathway activation scores and those that did not, and that these scores were driven by systemic activation of most of the biochemically linked individual signaling proteins. This pattern was also revealed in the other modules constructed (Supplementary Fig. S2). Interestingly, when comparing the signaling architecture derived from the pathway activation modules (Supplementary Fig. S3), a number of the lung-derived cell lines clustered together, characterized by relatively low-level

Figure 2. Identification and characterization of activated integrin pathway module by signal transduction representation (A) and unsupervised hierarchical clustering (B). For the pathway representation, the inhibitory phosphorylations considered in the analysis are shown in red and the activating phosphorylations in green. For the unsupervised hierarchical clustering, the complete panel of NCI-60 is shown (vertical axis). Specific endpoint relative intensity values have been used to create the heatmap. After overall score calculation, we highlighted the cell lines with the top 10 pathway scores (red), and the cell lines with the lowest 10 pathway activation scores (green).



signaling activation of the analyzed phosphoproteins that comprised the modules (Table 1).

Systems level "-omic" analysis

As we defined the module-driven pathway activation states for the NCI-60 cell lines, we sought to understand all the possible relationships between our pathway module activation results and the very large and full complement of publicly available molecular data for the NCI-60 set as well as the individual RPPA-driven phosphoproteomic data we derived. In particular, the DTP serves as a vital resource for preclinical information and research materials, including web-accessible data and tools and the molecular analysis and drug sensitivity screening of the NCI-60 tumor cell line, among other information. Thus, we were able to correlate the individual RPPA phosphoprotein and protein data we generated, as well as the pathway activation scores for our 6 signal transduction networks, to DNA copy number and mutation data, methylation status, mRNA and miRNA expression, along with drug and metabolite responses. From the cell-wide protein and gene consensus coexpression networks, we selected as network nodes all of the genes and proteins that were identified as statistically correlated with the pathway activation score and linked them to the pathway score node (Fig. 3A and B and Supplementary Fig. S4). In the figures, we also included these nodes' direct neighbors in the inferred consensus networks, that is, for each node directly correlated with pathway score, we included all genes or proteins linked to that node in the inferred protein and gene coexpression networks, even if they were not significantly correlated with the pathway score. We included all edges between the selected nodes in the inferred consensus networks. For all the gene nodes in the resulting score-centered network, we searched for genes that had significantly correlated DNA mutation, methylation, and copy number, as well as for miRNAs with correlated expression, metabolites with correlated levels, and drugs with correlated growth inhibition profiles, and included them in the network. In Fig. 3 and Supplementary Fig. S4, we used different colors to denote different types of analytes (e.g., protein, gene, metabolite, etc.), and used solid or dashed lines to indicate positive or negative correlations, respectively. For the phosphoprotein nodes, we repeated the same search for correlations using the expression of genes encoding these proteins. Also, we looked for direct correlations between phosphoprotein levels and DNA mutation, metabolite levels, and drug responses. We did not observe any RPPA endpoints that correlated negatively with pathway score, although some other types of variables (e.g., microRNA expression) with negative correlations were observed. However, the possibility of negative correlations between pathway scores and RPPA endpoints outside the pathway modules technically exists. To account for this possibility, we considered both positive and negative correlations in our analyses, and used 2-tailed statistical tests.

We also constructed a unifying joint network (Fig. 3C), attempting to link all 6 of the pathway modules together, but in this instance, we only included phosphoproteins that were

significantly correlated with at least 2 different pathway scores to focus on the strongest linkages. Then, we looked for significant correlations of those phosphoproteins, or of genes encoding them, to DNA mutation, methylation, and copy number data, as well as to metabolite and drug data. For all pathway activation modules analyzed, we found that 83% (109/131) of the direct linkages to the central pathway activation score were established by individually RPPA-generated phosphoprotein measurements (brown edges), even though we found correlations with all the other types of data. Moreover, specific statistically significant linkages were identified between RPPA-generated phosphoproteins and proteins and the other "-omic" data measured (Fig. 3 and Supplementary Fig. S4).

To assess drug sensitivity–network correlations, we considered 2 publicly available drug sensitivity databases: the A118 database and a database composed of a panel of 97 FDA-approved targeted drugs. The A118 database is composed of 118 of the most common chemotherapeutic drugs in use, and along with the 97 FDA-approved targeted drugs have been tested by DTP on the NCI-60 to evaluate their sensitivity and resistance as defined by LC_{50} and GI_{50} determinants. Considering all the networks together, we were able to define both positive (solid line) and negative (dash line) correlations between drug $-\log(GI_{50})$ s and several proteins/genes. Focusing on targeted therapies, we found significant correlations with 5 drugs specifically: imatinib, vemurafenib, sunitinib, nilotinib, and crizotinib. Interestingly, when the correlation between a drug sensitivity and a phosphoprotein was positive, the same phosphoprotein was directly and positively correlating as well with the mutation of the gene (and only with that one gene) against which the drug is specific. For example, each and every time imatinib correlated with phospho-PYK2 (or phospho-eNOS or phospho-STAT5), these endpoints correlated with the mutation of platelet-derived growth factor receptor α (*PDGFR α*), the known target of imatinib. This seemed to be true for every positive correlation between phosphoproteins and targeted drugs. For vemurafenib, an inhibitor of the V600E-mutated form of B-RAF, which is a component of our apoptosis, IGF-1R, mTOR, and EGFR pathway modules, we found positive correlations with phospho-p90RSK and/or phospho-PKC δ (vemurafenib and phospho-PKC δ correlate only in the EGFR network). These 2 phosphoproteins in turn correlated with the mutated B-RAF in the same 4 networks (10/11 NCI-60 cell lines carry the V600E B-RAF mutation; Fig. 3A and B and Supplementary Fig. S4). These results indicate that phosphoprotein levels provide a significant degree of linkage to both therapeutic sensitivity and gene mutation attributes. More extensive experimentation will be needed, using both chemical and genomic knockin/knockout, to determine the ordering and linkage of the molecular data with and between each other.

Discussion

Because the mechanism of action of nearly all oncology-based therapeutics, especially the new classes of targeted

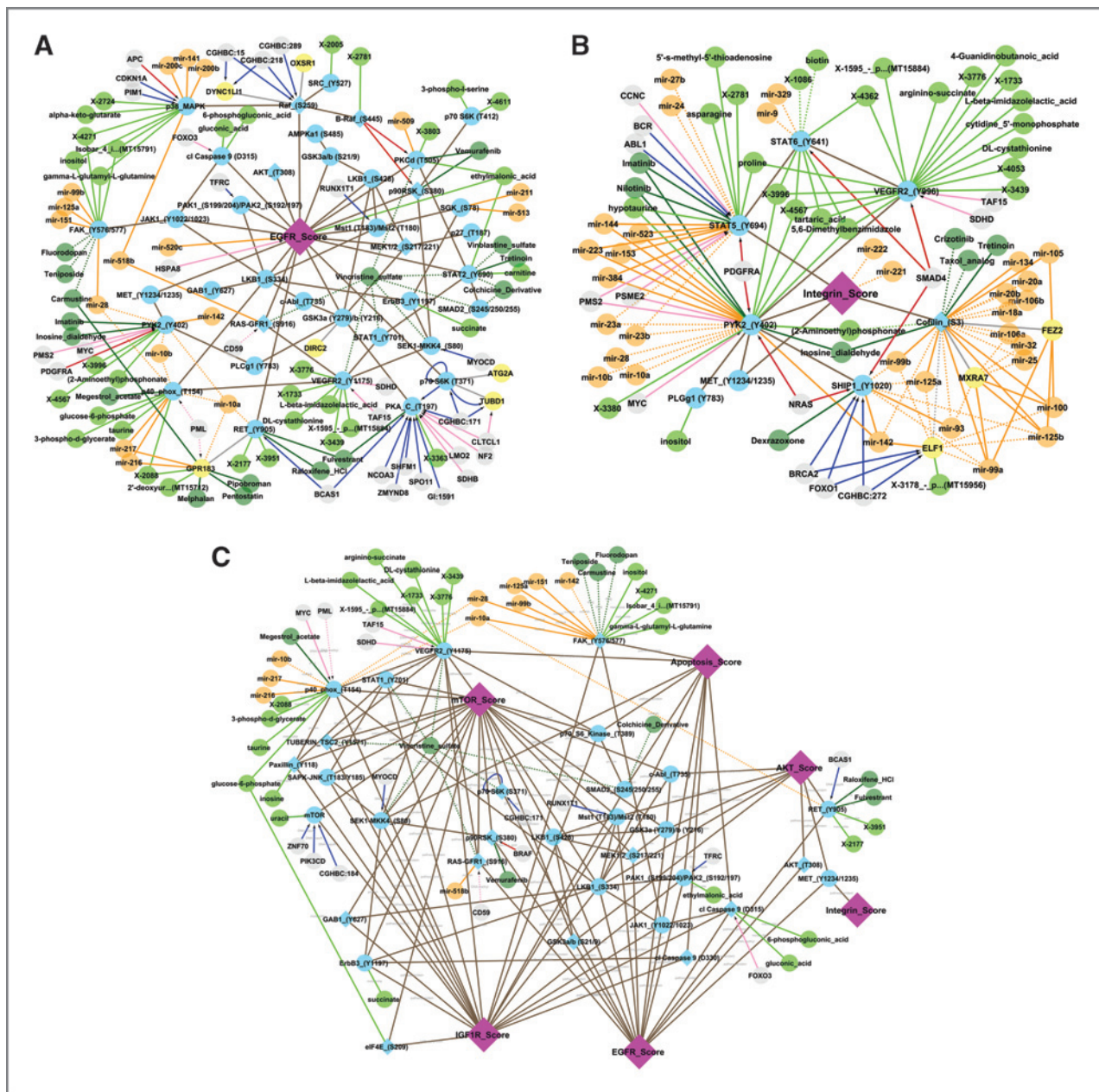


Figure 3. Systems networks of EGFR (A) and integrin (B) pathway activation, together with a unified pathway activation network (C). Node color: Violet, node representing the pathway score; Blue, phosphoproteins (linked to pathway score or to other proteins/genes at the protein level); Yellow, genes (linked to pathway score or to other proteins/genes at the mRNA level); Orange, miRNA; Dark green, drug; Light green, metabolite; Gray, other. Node shape: Diamond, pathway score node, or phosphoprotein that is used in calculating the pathway score; Circle, other nodes. Edge color and label: Brown, relationship inferred on the basis of phosphoprotein level (either with level other phosphoprotein or with pathway score); Gray, relationship inferred on the basis of gene mRNA expression (either with mRNA of other gene or with pathway score); Dark green, phosphoprotein level or gene expression (mRNA) is significantly correlated with drug response measured as $-\log(GI_{50})$; Light green, phosphoprotein level or gene expression (mRNA) is significantly correlated with metabolite concentration; Red, phosphoprotein level or gene expression (mRNA) is significantly correlated with mutation of other gene. Arrow points from mutation gene to the mRNA gene; Pink, gene expression (mRNA) is significantly correlated with methylation of other gene. Arrow points from methylation gene to the mRNA gene; Dark blue, gene expression (mRNA) is significantly correlated with copy number of other gene. Arrow points from copy number gene to the mRNA gene; Orange, gene expression (mRNA) is significantly correlated with expression of microRNA. Edge line style: Solid, positive correlation; Dashed, negative correlation.

therapeutics, is based largely on the modulation of protein activity/protein expression as the principal drug target, and it is known that cancer is largely driven by dysregulated and

hyperactivated signaling networks (9, 18), as proteomic technologies have matured, their use in these molecular characterization efforts has become increasingly important.

To date, a central missing component of the NCI-60 profiling efforts has been the lack of broad-scale protein activation/phosphorylation analysis. Past integrative biology efforts that incorporated proteomic analysis of the NCI-60 set have either used a more limited number of individual proteins ($N = 74$; ref. 10), compared with our analysis ($N = 135$) for defined drug sensitivity analysis, or evaluated total protein levels in an attempt to link gene and protein expression levels together (and found little concordance between these datasets; ref. 6). Moreover, our own analysis of the total levels and corresponding phosphorylation levels of a number of important cytoplasmic signaling proteins such as extracellular signal-regulated kinase (ERK), AKT, and receptor proteins such as ERBB3 revealed complete lack of concordance (Supplementary Fig. S5), which indicated that direct measurements of the phosphorylation levels of a given analyte are necessary and cannot be inferred from measuring the total levels alone. Most importantly, these past efforts used a one-analyte at-a-time approach for correlation and integrated analysis, and because we now know that cancer is a pathway-oriented disease (9, 18), we postulated that a more detailed molecular analysis that used a network-oriented approach would provide new insights into how different molecular compartments may link together and better characterize drug sensitivity markers. Here, we report the integration of individual protein activation/phosphorylation measurements into biochemically connected pathway modules that provide a new means to align functional protein architecture with multiple "-omic" data sets and therapeutic response correlations.

In this study, we have developed a novel systems biology-based approach to "multiomic" network analysis through the implementation of a phosphorylation-driven pathway activation score to integrate functional signaling data from RPPA analysis with all other "multiomic" data. Orientation of systems biology analysis toward pathway/network-focused function despite individual genes, proteins, or phosphoproteins will be a critical next step in translational research. This network-based approach is useful not only as a new means to integrate genomic, metabolomic, and proteomic data together, but also to delve into the molecular basis through which a drug acts and discover new potential inhibitory mechanisms for drugs with unknown mechanisms of action.

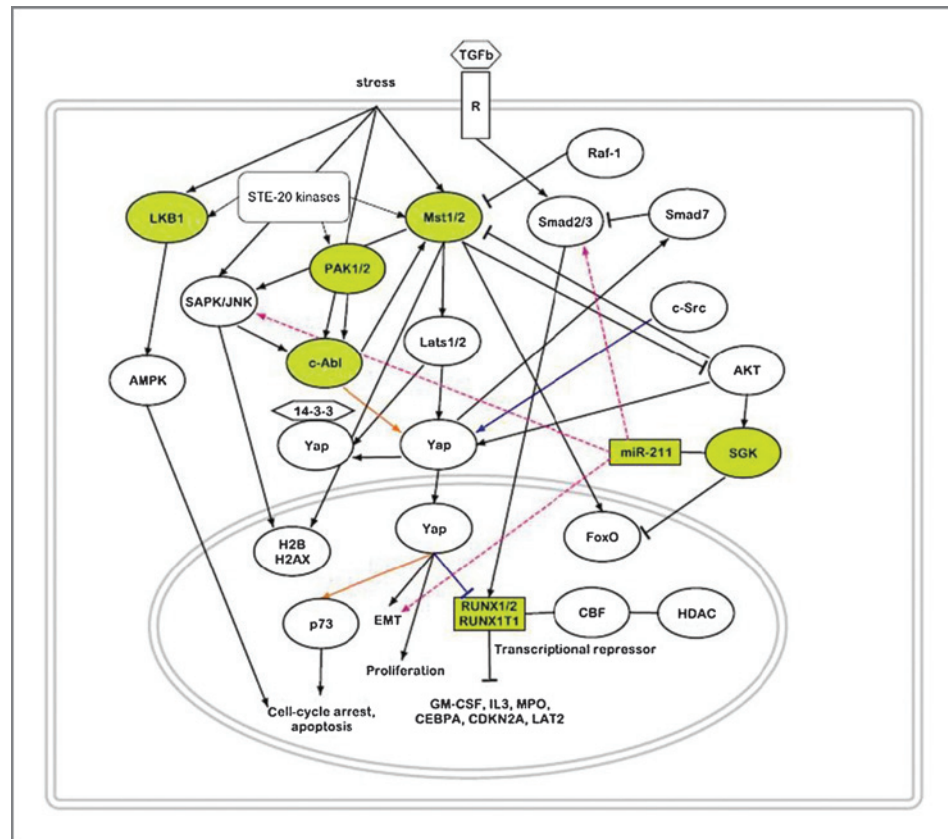
Despite the complex relationships revealed by our analysis (Fig. 3, Supplementary Fig. S4), our approach provided an opportunity to interrogate and more fully characterize some of the specific linkages observed and provided evidence for biochemical rationale of the correlations uncovered. In several pathway activation networks (IGF-1R, EGFR, AKT, mTOR, and apoptosis) we studied, we found a consistent linkage of MST1/MST2, PAK1/PAK2, LKB1, SGK1, and ABL phosphoproteins, *RUNX1T1* gene methylation status, and miRNA-211. The MST protein family is known to be involved in the Hippo pathway, a signaling cascade that controls organ size through the regulation of cell proliferation and apoptosis (19, 20). As many cancers are marked by unchecked cell division, Hippo family signaling has become

increasingly significant in the study of human cancers; however, how this family integrates within the context of other biochemical events still remains unclear. It has already been shown that mir-204/211 are involved in the repression of TGF β R and SNAIL signaling, and also with other genes implicated in cancer progression (21).

The transcription factor RUNX1T1 (together with RUNX1) is known to be involved in several cancers, above all hematopoietic malignancies, due to its repressor activity that involves histone deacetylases and the ability to undergo chromosomal translocation, where the *t8;21* is the most well known, and results in expression of a leukemia-specific chimeric transcription factor (22). On the basis of our analysis, we propose a model of interactions (Fig. 4) based on both previous data (18–29) and the interactions we found. This example provides evidence that our approach has identified "-omic" linkages that, although new, have basis in biochemical rationale and are supported, in part, by past research (18–29).

Another aspect of our analysis we can consider is the prevalence of miRNA correlations with gene and protein expression patterns. In particular, it was notable that many miRNAs correlated within the integrin pathway module. It is well known that miRNAs have a relevant role in several cellular mechanisms regulating the expression of genes and proteins, as well as in tumorigenesis. Within the integrin signaling network, we found that phosphorylation of cofilin has direct connections with several miRNAs reported to have roles in oncologic malignancies. For example, the ectopic expression of miR-17, -20, -93, and -106 increases proliferation, colony outgrowth, and replating capacity of myeloid progenitors and results in enhanced phospho-ERK levels (30). Meenhius and colleagues found that these miRNAs are endogenously and abundantly expressed in myeloid progenitors and downregulated in mature neutrophils and they identified sequestosome 1 (SQSTM1), a ubiquitin-binding protein and regulator of autophagy-mediated protein degradation, as a major target for these miRNAs in myeloid progenitors, showing that these miRNAs promote cell expansion, replating capacity, and signaling in hematopoietic cells by interference with SQSTM1-regulated pathways. Nishida and colleagues showed that the 2 miRNA clusters, miR-17-92a and miR-106b-25, composed of miR-17, -18a, -19a, -20a, -19b-1, -91a-1, and miR-106b, -93, -25, respectively, were upregulated in colorectal cancer stromal tissues compared with normal stroma. Gene expression profiles of the same stromal tissue samples revealed that putative targets of these miRNA clusters, such as *TGF β R2*, *SMAD2*, and *BMP* family genes, were significantly downregulated in cancer stromal tissue. Other downregulated putative targets were found to be involved in cytokine interaction and cellular adhesion, strengthening the evidence for involvement of these miRNAs in adhesion, including the integrin pathway (31). miR-17-92 has been shown to inhibit collagen synthesis by targeting *TGF β R2*, in mice (32), as well as it is known that miR-93 promotes angiogenesis and tumor growth by targeting *integrin β 8* (33), and together with miR-106a, seems to directly target the *Cofilin2* gene (34).

Figure 4. Hippo-based model. Arrows point out the direction of signaling: in orange, the YAP signaling effects on p73 driven by c-Abl, in blue the YAP signaling effects on the transcription factor RUNX1 driven by c-Src. In pink, the involvement of miR-211 in stress-activated protein kinase (SAPK)/c-jun-NH2-kinase (JNK) and SMAD signaling, as well as in epithelial-mesenchymal transition (EMT) is shown. Signaling pathway model made with PathVisio v2.0 (37).



The correlations we found, concerning miRNAs within both the integrin and other modules, underpin the growing relevance of these regulatory nucleic acids in governing many cellular processes. The ability to identify new miRNA targets could be enhanced by generating a systems-level analysis that could predict new, possible biologic correlations using "multi-omic" datasets and bioinformatic tools. Because many investigators are excited by the potential regulatory framework of miRNA and protein expression, orienting connectivity analysis with functional protein network activation as the proximal center could provide a firm basis of such exploration.

We were also able to use our network analysis to identify correlations concerning sensitivity to specific chemotherapeutics and targeted drugs. Of particular interest was the interconnection between phosphoproteins directly connected to the central score with targeted drugs and the mutation of the gene encoding the drug target. As stated previously, we found both positive and negative correlations with 5 targeted therapeutics: imatinib, vemurafenib, sunitinib, nilotinib, and crizotinib. When these drugs had positive correlations with the gene or the phosphoprotein encoded by the gene, the same phosphoprotein was also positively correlated with the mutated gene targeted by that drug. For example, phospho-PYK2 or phospho-STAT5 correlated with both imatinib and mutated PDGFR α . Also, phospho-p90RSK correlated with vemurafenib and the mutation of B-RAF (Fig. 3A and B and Supplementary Fig. S4). Of course, *in vitro* experiments are needed to study the

cellular responses to these drugs, and to further confirm the existence of biologic effects involving our predicted phosphoprotein-driven networks. This approach may provide a deeper understanding of how cellular biochemistry defines therapeutic response. Such "-omic" portraits could inform rational anticancer agent screenings and drive personalized therapeutic approaches.

A significant finding of the overall analysis described herein is the paucity of first-order connections to the pathway activation core with DNA genomic or RNA transcriptomic data. Most of the highly correlated analytes are either the individual phosphoproteins that comprise the activation score itself, which would be expected, or other phosphoproteins that do not underpin the score. These results indicate that gene expression and DNA mutation profiles may often not accurately predict functional protein signaling states. This has important implications in the context of personalized medicine where DNA mutation analyses are often currently used for patient stratification to drugs that target protein activity. Recent reports (35) have shown a lack of concordance of tumor mutations such as PTEN with response to inhibitors that target PI3K/mTOR signaling. Although these effects can also be underpinned by coordinate activation of compensatory pathways such as mitogen-activated protein kinase, our data show that the DNA mutation status of *PTEN*, while correlated with AKT activation (Supplementary Fig. S4A), has no direct correlation with mTOR pathway activation module (Supplementary Fig. S4D) in the NCI-60 cell line set.

These results from our analysis are in keeping with our recent published data concerning pathway activation mapping of primary human T-cell acute lymphoblastic leukemia tumor cells, which showed no statistical linkage of *PTEN* DNA mutations with downstream mTOR protein pathway activation state (36), indicating that a direct measurement of signaling pathway activation is required for accurate determination of the state of the drug target.

There are certainly limitations to our current study. Because of the complexity of cell signaling and the large volume of data we integrated, we cannot simply assume that the connections we found are all clinically relevant and many likely do not occur in every patient/tumor. A critical problem for systems-level analysis of large datasets is the overfitting of the data. By using a network/signaling module-based approach, which we feel more accurately reflects the signaling architecture, we hope to minimize inaccurate conclusions because our analysis uses this as the central hub of the network and works outward. Any interconnections, however, would need to be extensively validated and verified. We also used statistical approaches that minimize overfitting, but these efforts can only dampen chances of random associations being found and not eliminate them. We could verify the correctness of these models, such as the consistent connections of the Hippo family, with *in vitro* chemical and genomic knockout analysis, as well as expanding our phosphoproteomic analysis to include other proteins that were not considered in the first experimental design but could be tied to these specific networks. Our approach was based on a "hub" model philosophy whereby we chose broad-scale analyses centered on interrogating signaling proteins known to be integration points within the signaling architecture, under the postulate that even with incomplete coverage, we would be measuring the aggregate signaling inputs and outputs of a larger series of proteins. By aggregating these individual protein nodes with upstream and downstream direct first-order kinase-substrate relationships, we generate more complete data about the true nature of the signaling activation state rather than just an assortment of phosphoproteins randomly distributed across the kinase space. Moreover, our analysis is based solely on data from a limited number of cell lines, which most certainly neither recapitulate the heterogeneous state of human cancer *in vivo*, nor are representative of all of the cancer cell line sets in the public domain. Finally, despite the broad-scale pathway activation mapping conducted, the totality of the kinome and signaling architecture was not queried and the networks chosen for exploration, although important to cancer biology and therapy, were not comprehensive to include every known pathway and drug target. Thus, we present these data as both a guide to future "-omic" analyses as well as first steps

into a protein network-centered journey into future analyses rather than as a definitive dataset.

In conclusion, we used a novel, functional network-oriented approach that incorporated the most comprehensive protein pathway activation mapping analysis conducted to date on the NCI-60 cell line set to interrogate connectivity with other "-omic" data previously generated. Inclusion of the phosphoprotein data in the public domain as a growing information archive and knowledge base will provide an expanding opportunity for a systems biology view of tumorigenesis and related drug sensitivity correlations. Despite the known limitations of cell lines to recapitulate *in vivo* biology, cell line sets such as the NCI-60 provide a unique opportunity to investigate molecular correlations tied to causality within the defined context of the system used. Coupling extensive molecular architecture surrounding the genome, proteome, phosphoproteome, metabolome, and other molecular archives with drug sensitivity screens provide a unique "sandbox" to exploit the aggregate efforts of hundreds of scientists to more fully understand cancer biology. In the future, we hope that our pathway-oriented approach can generate new data concerning optimal therapeutic combinations, identify new predictive markers for therapeutic efficacy, and uncover new insights about how intra- and intercellular communications occur.

Disclosure of Potential Conflicts of Interest

J.D. Wulfschle has ownership interest (including patents) in Theranostics Health, LLC. L.A. Liotta has ownership interest (including patents) and is a consultant/advisory board member of Theranostics Health. E.F. Petricoin has ownership interest (including patents) and is a consultant/advisory board member of Theranostics Health, Inc.

Authors' Contributions

Conception and design: J.D. Wulfschle, L.A. Liotta, E.F. Petricoin
Development of methodology: T. Arodz, J.D. Wulfschle, L.A. Liotta, E.F. Petricoin
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): G. Federici, A.B. Shitaye, J.D. Wulfschle
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): G. Federici, X. Gao, J. Slawek, T. Arodz, A.B. Shitaye, L.A. Liotta, E.F. Petricoin
Writing, review, and/or revision of the manuscript: G. Federici, T. Arodz, J.D. Wulfschle, R. De Maria, E.F. Petricoin
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): G. Federici, L.A. Liotta, E.F. Petricoin
Study supervision: G. Federici, J.D. Wulfschle, R. De Maria, E.F. Petricoin

Acknowledgments

The authors thank Dr. Vikas Chandhoke and the College of Life Sciences for their generous support at George Mason University.

Grant Support

This work was partly supported by the Italian Istituto Superiore di Sanità within the framework Italy/USA cooperation agreement between the U.S. Department of Health and Human Services, George Mason University, and the Italian Ministry of Public Health.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 14, 2012; revised February 26, 2013; accepted March 28, 2013; published OnlineFirst May 1, 2013.

References

- Holbeck SL, Collins JM, Doroshow JH. Analysis of Food and Drug Administration-approved anticancer agents in the NCI60 panel of human tumor cell lines. *Mol Cancer Ther* 2010;9:1451-60.
- Ikediodi ON, Davies H, Bignell G, Edkins S, Stevens C, O'Meara S, et al. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther* 2006;5:2606-12.

3. Sausville EA, Holbeck SL. Transcription profiling of gene expression in drug discovery and development: the NCI experience. *Eur J Cancer* 2004;40:2544–9.
4. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, et al. Integrative genomic analyses identify MTF1 as a lineage survival oncogene amplified in malignant melanoma. *Nature* 2005;436:117–22.
5. Blower PE, Verducci JS, Lin S, Zhou J, Chung JH, Dai Z, et al. MicroRNA expression profiles for the NCI-60 cancer cell panel. *Mol Cancer Ther* 2007;6:1483–91.
6. Nishizuka S, Charboneau L, Young L, Major S, Reinhold WC, Waltham M, et al. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci U S A* 2003;100:14229–34.
7. Roschke AV, Tonon G, Gehlhaus KS, McTyre N, Bussey KJ, Lababidi S, et al. Karyotypic complexity of the NCI-60 drug-screening panel. *Cancer Res* 2003;63:8634–47.
8. Myers TG, Anderson NL, Waltham M, Li G, Buolamwini JK, Scudiero DA, et al. A protein expression database for the molecular pharmacology of cancer. *Electrophoresis* 1997;18:647–53.
9. Improta G, Zupa A, Fillmore H, Deng J, Aieta M, Musto P, et al. Protein pathway activation mapping of brain metastasis from lung and breast cancers reveals organ type specific drug target activation. *J Proteome Res* 2011;10:3089–97.
10. Park ES, Rabinovsky R, Carey M, Hennessy BT, Agarwal R, Liu W, et al. Integrative analysis of proteomic signatures, mutations, and drug responsiveness in the NCI 60 cancer cell line set. *Mol Cancer Ther* 2010;9:257–67.
11. Available from: <http://discover.nci.nih.gov/cellminer/>. Accessed 4/26/2013.
12. Available from: <http://dtp.nci.nih.gov>. Accessed 4/26/2013.
13. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006;7 Suppl 1:S7.
14. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, et al. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007;5:e8.
15. Meyer PE, Kontos K, Lafitte F, Bontempi G. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J Bioinform Syst Biol* 2007;79879.
16. Altay G, Emmert-Streib F. Inferring the conservative causal core of gene regulatory networks. *BMC Syst Biol* 2010;4:132.
17. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* 2010;5.
18. Mammano E, Galdi F, Pierobon M, Tessari E, Deng J, Pucciarelli S, et al. Multiplexed protein signal pathway mapping identifies patients with rectal cancer that responds to neoadjuvant treatment. *Clin Colorectal Cancer* 2012;11:268–74.
19. Bao Y, Hata Y, Ikeda M, Withanage K. Mammalian Hippo pathway: from development to cancer and beyond. *J Biochem* 2011;149:361–79.
20. Zeng Q, Hong W. The emerging role of the hippo pathway in cell contact inhibition, organ size control, and cancer development in mammals. *Cancer Cell* 2008;13:188–92.
21. Wang FE, Zhang C, Maminishkis A, Dong L, Zhi C, Li R, et al. MicroRNA-204/211 alters epithelial physiology. *FASEB J* 2010;24:1552–71.
22. Ptasinska A, Assi SA, Mannari D, James SR, Williamson D, Dunne J, et al. Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* 2012;26:1829–41.
23. Cinar B, Fang PK, Lutchman M, Di Vizio D, Adam RM, Pavlova N, et al. The pro-apoptotic kinase Mst1 and its caspase cleavage products are direct inhibitors of Akt1. *EMBO J* 2007;26:4523–34.
24. Bertini E, Oka T, Sudol M, Strano S, Blandino G. YAP: at the crossroad between transformation and tumor suppression. *Cell Cycle* 2009;8:49–57.
25. Duque-Afonso J, Yalcin A, Berg T, Abdelkarim M, Heidenreich O, Lubbert M. The HDAC class I-specific inhibitor entinostat (MS-275) effectively relieves epigenetic silencing of the LAT2 gene mediated by AML1/ETO. *Oncogene* 2011;30:3062–72.
26. Jang SW, Yang SJ, Srinivasan S, Ye K. Akt phosphorylates Mst1 and prevents its proteolytic activation, blocking FOXO3 phosphorylation and nuclear translocation. *J Biol Chem* 2007;282:30836–44.
27. Lee KK, Yonehara S. Identification of mechanism that couples multisite phosphorylation of Yes-associated protein (YAP) with transcriptional coactivation and regulation of apoptosis. *J Biol Chem* 2007;282:9568–78.
28. Ling P, Lu TJ, Yuan CJ, Lai MD. Biosignaling of mammalian Ste20-related kinases. *Cell Signal* 2008;20:1237–47.
29. Xiao L, Chen D, Hu P, Wu J, Liu W, Zhao Y, et al. The c-Abl-MST1 signaling pathway mediates oxidative stress-induced neuronal cell death. *J Neurosci* 2011;31:9611–9.
30. Meenhuis A, van Veelen PA, de Looper H, van Bostel N, van den Berge IJ, Sun SM, et al. MiR-17/20/93/106 promote hematopoietic cell expansion by targeting sequestosome 1-regulated pathways in mice. *Blood* 2011;118:916–25.
31. Nishida N, Nagahara M, Sato T, Mimori K, Sudo T, Tanaka F, et al. Microarray analysis of colorectal cancer stromal tissue reveals upregulation of two oncogenic miRNA clusters. *Clin Cancer Res* 2011;18:3054–70.
32. Li L, Shi JY, Zhu GQ, Shi B. MiR-17-92 cluster regulates cell proliferation and collagen synthesis by targeting TGF β pathway in mouse palatal mesenchymal cells. *J Cell Biochem* 2012;113:1235–44.
33. Fang L, Deng Z, Shatseva T, Yang J, Peng C, Du WW, et al. MicroRNA miR-93 promotes tumor growth and angiogenesis by targeting integrin-beta8. *Oncogene* 2011;30:806–21.
34. Available from: <http://mirdb.org/cgi-bin/search.cgi?searchType=miRNA&full=mirbase&searchBox=MIMAT0000093>. Accessed 4/26/2013.
35. Janku F, Broadus R, Bakkar R, Hong DS, Stepanek VMT, Naing A, et al. PTEN assessment and PI3K/mTOR inhibitors: importance of simultaneous assessment of MAPK pathway aberrations. *J Clin Oncol* 2012 (suppl); abstr 10510.
36. Zuurbier L, Petricoin EF III, Vuerhard MJ, Calvert V, Kooi C, Buijs-Gladdines JG, et al. The significance of PTEN and AKT aberrations in pediatric T-cell acute lymphoblastic leukemia. *Haematologica* 2012;97:1405–13.
37. van Iersel MP, Kelder T, Pico AR, Hanspers K, Coort S, Conklin BR, et al. Presenting and exploring biological pathways with PathVisio. *BMC Bioinformatics* 2008;9:399.

Molecular Cancer Research

Systems Analysis of the NCI-60 Cancer Cell Lines by Alignment of Protein Pathway Activation Modules with "-OMIC" Data Fields and Therapeutic Response Signatures

Giulia Federici, Xi Gao, Janusz Slawek, et al.

Mol Cancer Res 2013;11:676-685. Published OnlineFirst May 1, 2013.

Updated version Access the most recent version of this article at:
doi:[10.1158/1541-7786.MCR-12-0690](https://doi.org/10.1158/1541-7786.MCR-12-0690)

Supplementary Material Access the most recent supplemental material at:
<http://mcr.aacrjournals.org/content/suppl/2013/04/19/1541-7786.MCR-12-0690.DC1>

Cited articles This article cites 29 articles, 11 of which you can access for free at:
<http://mcr.aacrjournals.org/content/11/6/676.full#ref-list-1>

Citing articles This article has been cited by 2 HighWire-hosted articles. Access the articles at:
<http://mcr.aacrjournals.org/content/11/6/676.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://mcr.aacrjournals.org/content/11/6/676>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.