Global Proteome Analysis of the NCI-60 Cell Line Panel

Amin Moghaddas Gholami,1,4,* Hannes Hahne,1,4 Zhixiang Wu,1,4 Florian Johann Auer,3 Chen Meng,1 Mathias Wilhelm,1 and Bernhard Kuster1,2,3,*

1Proteomics and Bioanalytics, Technische Universität München, Emil-Erlenmeyer-Forum 5, 85354 Freising, Germany
2Center for Integrated Protein Science Munich, Department of Chemistry and Biochemistry, Butenandtstr. 5–13, 81377 Munich, Germany
3German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
4These authors contributed equally to this work
*Correspondence: amin@tum.de (A.M.G.), kuster@tum.de (B.K.)
http://dx.doi.org/10.1016/j.celrep.2013.07.018
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

SUMMARY

The NCI-60 cell line collection is a very widely used panel for the study of cellular mechanisms of cancer in general and in vitro drug action in particular. It is a model system for the tissue types and genetic diversity of human cancers and has been extensively molecularly characterized. Here, we present a quantitative proteome and kinome profile of the NCI-60 panel covering, in total, 10,350 proteins (including 375 protein kinases) and including a core cancer proteome of 5,578 proteins that were consistently quantified across all tissue types. Bioinformatic analysis revealed strong cell line clusters according to tissue type and disclosed hundreds of differentially regulated proteins representing potential biomarkers for numerous tumor properties. Integration with public transcriptome data showed considerable similarity between mRNA and protein expression. Modeling of proteome and drug-response profiles for 108 FDA-approved drugs identified known and potential protein markers for drug sensitivity and resistance. To enable community access to this unique resource, we incorporated it into a public database for comparative and integrative analysis (http://wwz.tum.de/proteomics/nci60).

INTRODUCTION

Cell lines derived from human tumors are very widely used model systems for the study of cancer biology and drug discovery. Particularly in combination with system-level explorative profiling technologies, the comparative analysis of cancer cell lines can reveal distinct similarities and differences in biological processes between cancer cells that can be exploited in many different ways. For instance, the 59 cancer cell lines (NCI-60 panel) of the National Cancer Institute’s (NCI’s) Developmental Therapeutics Program (DTP; http://dtp.nci.nih.gov) are an established tool for in vitro drug screening. The collection represents, at least to some extent, the tissue type and genetic diversity of human cancers (Shoemaker, 2006). Since its inception, the NCI-60 panel has led to many important discoveries, including a general advance in the understanding of cancer mechanisms (Boyd and Paull, 1995; Weinstein, 2006), the identification of mechanisms of action of drugs, and the approval of new chemotherapeutic agents (e.g., bortezomib). Hundreds of thousands of potential anticancer agents have by now been screened using the NCI-60 panel (Holbeck et al., 2010; Shoemaker, 2006), and multiple technology platforms have been used to characterize the cells on the molecular level including, but not limited to, array comparative genomic hybridization (Bussey et al., 2006), karyotype analysis (Rochette et al., 2003), DNA mutational analysis (Abaan et al., 2013; Ikediobi et al., 2006), DNA fingerprinting (Lorenzi et al., 2009), microarrays for transcript expression (Scherf et al., 2000; Shankavaram et al., 2007), microarrays for microRNA expression (Blower et al., 2008; Liu et al., 2010), single-nucleotide polymorphism arrays to identify DNA copy number alterations (Garraway et al., 2005), and DNA methylation (Ehrlich et al., 2008). Although proteins carry out virtually all cellular processes and represent the vast majority of anticancer drug targets, very few studies have focused on the analysis of protein expression across the NCI-60 panel (Nishizuka et al., 2003; Park et al., 2010; Shankavaram et al., 2007). In particular, reverse-phase protein microarrays from cellular lysates have been employed in this context, and although these studies focused on a rather confined number of proteins, their results highlight the potential of systematic protein expression analyses for cancer research in general and drug discovery in particular. Mass spectrometry (MS)-based proteomics has undergone rapid progress in past years (Aebersold and Mann, 2003; Mallick and Kuster, 2010), and systematic analyses can now be carried out to identify and quantify the majority of proteins expressed in a human cell line (Beck et al., 2011; Burkard et al., 2011; Geier et al., 2012; Lundberg et al., 2010; Nagaraj et al., 2011). In addition, important oncogene classes such as kinases, which are often of low cellular abundance, can now be systematically queried using MS-based proteomics (Bantscheff et al., 2007; Wu et al., 2011). Protein kinases are key players of intracellular signal transduction, and dysregulation of protein kinases can be cause or consequence of cancer and therefore are among the most important anticancer drug targets today (Cohen,
2002; Knapp et al., 2013). More generally speaking, proteomic analyses are now valuable tools for molecular and clinical cancer research (Hanash and Taguchi, 2010) and drug discovery (Schirle et al., 2012) and will be more and more used in concert with DNA/RNA-level investigations in the future.

In the present study, we employed a comprehensive analysis using MS-based proteomics to obtain quantitative proteome profiles of all 59 cell lines of the NCI-60 panel. Collectively, 10,350 proteins, including 375 protein kinases, were identified and 6,003 proteins were consistently quantified in at least 5 out of 59 cell lines. Unsupervised bioinformatic analysis revealed strong cell line clusters according to tissue type, and further analyses disclosed 522 differentially regulated proteins and several pathways predominant in certain tissue types. Integration with public transcriptome data showed a significant degree of correlation between messenger RNA (mRNA) and protein expression, and the integration with profiles of 108 drugs approved by the US Food and Drug Administration (FDA) revealed known and potentially novel protein markers involved in mediating drug resistance and sensitivity. To increase tissue-specific proteome coverage, one representative cell line from each of the nine tissue groups was analyzed in more depth and resulted in the identification of 8,443 proteins (deep proteomes; Figures S1A and S1E). Kinase profiles were obtained for the complete NCI-60 panel using immobilized nonselective kinase inhibitors (kinobeads) followed by MS-based protein identification (Bantscheff et al., 2007; Figures S1B, S1D, and S2A). The kinobead approach resulted in the identification of 220 protein kinases of which 106 were not identified in the other two approaches, thus accessing a part of the proteome of too low abundance for conventional shotgun proteomics. In addition, 155 protein kinases, which have low or no affinity for kinobeads (Bantscheff et al., 2007; Lemeer et al., 2013), were exclusively identified in the proteome profiling experiments. Collectively, the three data sets comprise 10,350 distinct proteins corresponding to 8,739 unique genes and representing 46% of the protein-coding human genome (Figure 1B). Protein identifications are spread evenly across autosomes with an average coverage of 44%. Interestingly, while a similar coverage was also obtained for the X chromosome, we found only five proteins encoded by Y-chromosomal genes (12% coverage) in cell lines of male origin, which is consistent with previous findings (Geiger et al., 2012).

The large number of cell lines investigated allowed us to analyze expression profiles of proteins systematically across cell lines. To identify proteins ubiquitously expressed in cancer cells of different origins, we reconstructed the common proteome for all nine tissue groups. A protein was considered as part of the core proteome if it was identified in at least one cell line of every single tissue group. The resulting 5,578 proteins can thus be regarded as the core cancer proteome. The remaining ~5,000 proteins show a more distinct expression pattern between tissues and each tissue contributing ~2,000 proteins not contained in the core proteome (Figure 1C).

RESULTS AND DISCUSSION

Proteomic Analysis of the NCI-60 Panel Identifies the Core Cancer Proteome
The NCI-60 panel comprises 59 individual cancer cell lines derived from nine different tissues (brain, blood and bone marrow, breast, colon, kidney, lung, ovary, prostate, and skin), which we analyzed using three proteomic approaches (Figures 1A and S1A–S1C). The proteome profiles of all individual cell lines followed a conventional one-dimensional PAGE followed by in-gel digestion and liquid chromatography-tandem mass spectrometry (GeLC-MS/MS) approach (Schirle et al., 2003) and yielded 8,113 proteins (Figure S1D). To increase tissue-specific proteome coverage, one representative cell line from each of the nine tissue groups was analyzed in more depth and resulted in the identification of 8,443 proteins (deep proteomes; Figures S1A and S1E). Kinase profiles were obtained for the complete NCI-60 panel using immobilized nonselective kinase inhibitors (kinobeads) followed by MS-based protein identification (Bantscheff et al., 2007; Figures S1B, S1D, and S2A). The kinobead approach resulted in the identification of 220 protein kinases of which 106 were not identified in the other two approaches, thus accessing a part of the proteome of too low abundance for conventional shotgun proteomics. In addition, 155 protein kinases, which have low or no affinity for kinobeads (Bantscheff et al., 2007; Lemeer et al., 2013), were exclusively identified in the proteome profiling experiments. Collectively, the three data sets comprise 10,350 distinct proteins corresponding to 8,739 unique genes and representing 46% of the protein-coding human genome (Figure 1B). Protein identifications are spread evenly across autosomes with an average coverage of 44%. Interestingly, while a similar coverage was also obtained for the X chromosome, we found only five proteins encoded by Y-chromosomal genes (12% coverage) in cell lines of male origin, which is consistent with previous findings (Geiger et al., 2012).

The large number of cell lines investigated allowed us to analyze expression profiles of proteins systematically across cell lines. To identify proteins ubiquitously expressed in cancer cells of different origins, we reconstructed the common proteome for all nine tissue groups. A protein was considered as part of the core proteome if it was identified in at least one cell line of every single tissue group. The resulting 5,578 proteins can thus be regarded as the core cancer proteome. The remaining ~5,000 proteins show a more distinct expression pattern between tissues and each tissue contributing ~2,000 proteins not contained in the core proteome (Figure 1C).
Proteomics Detects Proteins between 100 and 10 Million Copies per Cell

We used intensity-based label-free quantification for the relative quantification of proteins across cell lines in which the summed protein intensity from proteome profiling experiments (not deep proteomes or kinomes) served as a proxy for protein abundance within a cell line (Luber et al., 2010). Protein abundance distributions of all 59 cell lines are generally very similar and span at least five orders of magnitude (Figures 2A and S2B), which is consistent with recent estimates of protein copy numbers in mammalian cell lines (Beck et al., 2011; Geiger et al., 2012; Nagaraj et al., 2011; Schwanhäusser et al., 2011).

To identify proteins that appear uniformly among the most and least abundant proteins, we calculated median abundance values across all cell lines (Figures 2B and S2B) and estimated protein copy numbers. Assuming a median number of ~10,000 copies per protein (Beck et al., 2011) and a linear correlation between measured protein abundance and copy number (Beck et al., 2011; Malmström et al., 2009), we estimate that the identified proteins have copy numbers between 100 and 10,000,000 copies per cell. The 10% most highly expressed proteins contain mostly structural proteins and proteins involved in basic cellular machineries that are known to be much more abundant than regulatory proteins (Figures 2C and S3A–S3D). For instance, proteins involved in transport processes, as classified by Gene Ontology (GO) annotation (Ashburner et al., 2000), formed a tight cluster at the top end of the distribution of protein expression levels. Similarly, proteins with roles in protein folding, molecular transport, and translation are significantly enriched among the most abundant proteins. Conversely, among the 10% least abundant proteins are mainly regulatory proteins, membrane proteins, and a large proportion of as-yet-uncharacterized proteins. The correlation of protein abundance of all NCI-60 cell lines was relatively high (average Pearson correlation of R = 0.81; Figure S3E). Despite this strong correlation, there are clearly also proteins that differ by orders of magnitude in expression between cell lines (Figure S3F). Taken together, these observations are consistent with previous comparative studies of multiple cell lines (Geiger et al., 2012) as well as with studies of single cell lines (Beck et al., 2011; Nagaraj et al., 2011).
We next examined the similarity of individual cell line proteomes using unsupervised hierarchical clustering using 6,003 proteins that were quantified in at least 5 out of the 59 cell lines (Figure 3A). In general, cell lines originating from the same tissue converged into the same or closely related clusters. Hierarchical clustering revealed subclusters of colon (seven out of seven), leukemia (five out of six), CNS (five out of six), and melanoma (six out of eight) cell lines in which the samples segregated largely according to their tissue of origin. Interestingly, one melanoma line that did not cluster (LOX IMVI) has been reported to lack melanin production (Stinson et al., 1992), which is a strong determinant for the melanoma cluster. Cell lines from breast and ovarian cancers showed a more pronounced distribution across multiple clusters, indicating that their protein expression patterns are quite heterogeneous. We observed, for instance, that the estrogen receptor (ER)-negative breast cancer cell line Hs578T clustered with the stromal/mesenchymal cluster of glioblastoma and renal tumor cell lines. In contrast, the ER-positive breast cancer lines MCF-7 and T47D clustered to colon cancer lines displaying an epithelial phenotype, which is, for instance, characterized by the expression of proteins involved in cell-junction signaling.

We separately performed hierarchical clustering on kinase profiles of all NCI-60 cell lines (220 protein kinases; Figure 3B), which revealed five distinct clusters at an average intercluster correlation coefficient of $\geq 0.6$. While CNS (six out of six), leukemia (six out of six), prostate (two out of two), and melanoma (six out of eight) cell lines clustered on single branches, ovarian, and non-small-cell lung cancer lines diverged into multiple clusters, likely reflecting a more heterogeneous molecular phenotype than the aforementioned tumor entities. Interestingly, the clustering analysis shows that the similarity of cell lines according to their tissue of origin is less pronounced on the kinase level compared to the proteome level. On the one hand, this suggests that cancer cell lines preserve, at least to a significant extent, the basic molecular makeup and biological functions of their tissue of origin and thus may be valuable model systems for studying
tissue-specific functions and processes. On the other hand, the striking kinase expression heterogeneity observed for some cell lines may reflect molecularly diverged signal-transduction mechanisms among and between cancer cell types.

Given that genomic instability is a hallmark of cancer cells, it is of note that particularly highly abundant proteins, such as those involved in cellular maintenance (e.g., energy metabolism and protein synthesis), evolve more slowly than proteins of lower abundance, which include most regulatory proteins (Beck et al., 2011; Pál et al., 2001; Subramanian and Kumar, 2004). Highly expressed proteins are apparently under strong selective pressure, likely because of energy constraints (Lane and Martin, 2010) and/or because of a requirement for translational robustness (i.e., minimizing the risk for protein aggregation and toxicity; Drummond et al., 2005). Consistent with the common notion that aberrant signal transduction is an important driver of cancer development and progression, the lower selective pressure of low-abundance proteins, in turn, might represent a general mechanism of how regulatory functions of cancer cells evolve and diverge.

To identify proteins differentially expressed between different tissue categories, we grouped cell lines accordingly. Overall, the correlation of protein expression between tissue groups is strong (R > 0.8; Figure S4A). Proteins significantly changing between groups were detected by fitting a linear model to the normalized data and calculating empirical Bayes and moderated F and t statistics. In total, this test determined 522 such proteins (q < 0.05) in at least one of the groups across the whole NCI-60 panel. The majority of these proteins are of high abundance (Figures S4B and S4C), which is in accordance with the observation that abundant proteins can be consistently quantified across multiple cell lines.

Principal component analysis (PCA) of these proteins also revealed convergence of cell lines according to their tissue of origin (Figure 3C). Again, the separation of cell lines was less pronounced in the kinase profiles than in the proteome profiles, which is consistent with the hierarchical clustering results. Two cell lines are particularly notable in this context. The MDA-MB-435 cell line, derived from the pleural effusion of a patient with breast cancer, co-clustered with melanoma cell lines. This cell line was originally reported as a breast carcinoma cell line, but more recent evidence indicates that it is a derivative of the M14 melanoma cell line (Rae et al., 2007). This is supported by the proteomic data, which suggests that MDA-MB-435 is indeed a melanoma line. We also observed high similarity between MDA-MB-435 and the ovarian line IGROV1, raising the possibility that the latter may also originate from an (occur) melanoma.

Hierarchical clustering of differentially expressed proteins and cell lines disclosed that significantly enriched biological functions and biochemical pathways of selected protein clusters are consistent with actual biological functions of the respective tissue type (Figure 3D). For instance, cluster 3 comprises proteins highly expressed in melanoma cell lines, and the only biological function associated with proteins in this cluster is melanin biosynthesis and pigmentation. In contrast, proteins in cluster 5 are highly expressed in leukemic cell lines and participate, for instance, in intracellular signaling pathways of immune cells (e.g., the tyrosine-protein kinase BTK or the proto-oncogene Vav). We also note that the leukemia lines all show markedly reduced levels of proteins involved in processes more relevant for solid tumors (e.g., cell junction, cell adhesion, blood vessel development, and others).

Protein kinases are the major constituents of cellular signal transduction. Among the 220 kinases identified from kinobead purifications, 49 are differentially expressed between tissue groups. Kinases were identified across all branches of the phylogenetic tree and, in particular, tyrosine kinases (TKs) display lineage-specific differential expression (Figure 4). The TK family comprises multiple targets of anticancer drugs, and many of them show significant differences between tissue types, such as EGFR, EPHA2 and EPHB2, SRC, or MET.

Moreover, the profiles of differentially expressed kinases underscore known drug/target combinations and suggest yet-unexplored potential targets or applications of known drugs. For instance, BTK represents an attractive drug target as it has been found solely in leukemia cell lines. This is consistent with recent results that indicate that the BTK inhibitor ibrutinib has significant activity and is well tolerated in patients with relapsed or refractory B cell malignancies (Advani et al., 2013). Our results also highlight platelet-derived growth factor receptor alpha (PDGFRα), which has been exclusively identified in CNS cell lines, indicating that the inhibition of angiogenesis via PDGFRα might represent a potential target in malignant glioblastomas. A yet-unexplored drug target, for instance, might be the ribosomal protein S6 kinase alpha-4 (RPS6KA4), which has been found to be significantly overexpressed in prostate cancer cell lines. It is downstream of important signaling pathways and involved in the phosphorylation and regulation of various transcription factors including activation of the proto-oncogenes c-Fos and c-Jun (FOS and JUN; Pierrat et al., 1998; Soloaga et al., 2003).

Comparative Analysis of Proteome and Transcriptome Profiles

Transcriptome data are frequently used for global gene expression analysis of cancer cells, and the approach has also been applied to the NCI-60 panel (Pfister et al., 2009). Given that proteins are translated from mRNA templates, it is logical to compare mRNA and proteome profiling data as molecular descriptors of gene expression. The normalized transcriptome data (Figures S5A–S5C) mapped to 13,741 genes above the detection limit, and of all the expressed transcripts, 8,065 (59%) were detected as proteins with a clear bias toward higher-abundance transcripts (Figures 5A and 5B). While 81% of the most abundant mRNAs (first quartile) could be identified on the protein level, this was only the case for 21% of the least expressed transcripts (last quartile). Proteins and transcripts were equally well identified across subcellular localizations (Figure 5C). This is of note as it is a common (albeit probably inaccurate) notion that membrane proteins are underrepresented in proteomic data.

To identify trends and relationship between mRNA and protein abundances across the NCI-60 cell lines, we performed multivariate analyses. Coinertia analysis (CIA) was used to visualize relationships between transcriptome and proteome profiles (Figure 5D). In CIA plots, each arrow represents one of the NCI-60
Figure 4. Kinome Tree of Identified and Differentially Expressed Kinases

Mass spectrometric analysis of kinobead purifications from 59 NCI-60 cell lines identified 220 protein kinases across all branches of the phylogenetic tree. Identified and differentially expressed (q < 0.05, ANOVA) protein kinases are indicated in blue and red, respectively. Red dots are sized according to the negative logarithmic q value.

Illustration reproduced courtesy of Cell Signaling Technology (http://www.cellsignal.com).
cell lines. The base of each arrow represents the position of that cell line in transcriptome space, and the tip of each arrow gives the position in proteome space. The length of an arrow represents the divergence between transcriptome and proteome of a given cell line (the shorter the arrow, the higher the level of concordance between the mRNAs and proteins for a particular cell line). Generally, the previously described cell line clusters are also observed in the CIA plot (for transcriptome and proteome hierarchical clustering clusters, see Figures S5D and 3A, respectively) and represent the same or similar biological functions. This indicates that the high-level biological information content of transcriptomes and proteomes is similar and equally well suited to cluster cell lines into the correct tissue space. We also note that the more heterogeneous cell line groups (e.g., breast cancer cell lines) do not exhibit a considerably higher degree of divergence between their transcriptome and proteome profiles, indicating that the observed spread reflects the heterogeneity on cellular level.

The position of cell lines along the axes of the CIA plot in Figure 5D can be explained by distinct cellular properties. The horizontal axis separates cell lines according to their proliferation rate (Spearman rank correlation of $p = 4.7 \times 10^{-5}$ and $p = 6.4 \times 10^{-6}$ for transcriptome and proteome, respectively; Figure S6A) and the vertical axis separates according to the tissue type. While the majority of cell lines isolated from carcinomas and sarcomas do not exhibit strong divergence, leukemia and melanoma cell lines form distinct clusters toward one end of the vertical axis. We note that many cell lines are projected.
further in protein space than in gene space, indicating that the protein expression profiles of these cell lines contain more information than the corresponding mRNA profiles and thus contribute more to the trends on the axes. To disclose mRNAs and proteins responsible for the separation of tissue clusters, we selected highly correlated mRNAs and proteins of the same PCA direction (Figure 5E). While the majority of genes exhibits considerable variation between transcriptome and proteome expression, 629 genes are highly correlated (R > 0.7; Figure 5E, black dots) and are strongly expressed on both the transcriptome (Figure 5E, left panel) and proteome levels (Figure 5E, right panel) of colon, leukemia, melanoma, CNS, or renal cell lines. For instance, the integrated analysis of mRNA and protein expression disclosed several markers for leukemia cancer cell lines, including protein tyrosine phosphatase PTPRC (CD45), the hematopoietic lineage cell-specific protein HCLS1, and the RacGTPase-activating protein ArhGAP15, all of which fulfill well-studied functions in immune cells (Costa et al., 2011; Rhee and Veillette, 2012; Yamanashi et al., 1993).

Taken together, despite the moderate depth of proteome coverage per single cell line (again, 6,003 proteins were quantified in at least 5 out of 59 cell lines), proteome and transcriptome data both appear to be powerful molecular descriptors of cancer cells, and their integrative analysis enables a more comprehensive view on the multiple layers of cellular regulation than any technique alone.

Identification of Protein Signatures for Drug Sensitivity and Resistance

In many cases, sensitivity or resistance of cell lines to anticancer therapeutics cannot be simply attributed to single genes or proteins (Garnett et al., 2012). To uncover protein signatures significantly associated with drug sensitivity and resistance, we employed elastic net regression analysis (Zou and Hastie, 2005), which disclosed cooperative interactions between protein expression, their mutational status (Reinhold et al., 2012) and the response signature of anticancer therapeutics (DTP; Rubinstein et al., 1990) across the NCI-60 panel, thereby revealing complex molecular signatures, which might be used as “panel biomarkers” (or “multifeature biomarkers”) to predict drug sensitivity or resistance. Overall, elastic net modeling identified 20,743 protein-drug associations from which 1,801 associations significantly associated with proteins involved in RNA processes and, hence, can be considered as a potential tumor suppressor (Georgitsi et al., 2007; Lorenzo and Susin, 2007).

Proteins often correlated with drug resistance exhibit a variety of cellular functions. The most frequently observed protein, the 14–3–3 protein zeta/delta (YWHAZ; 17/19), has recently been implicated in drug resistance of breast cancer patients (Li et al., 2010) and likely affects drug sensitivity through inhibition of apoptosis via sequestration of the proapoptotic Bcl-2-associated death promoter protein (BAD; Hermeking, 2003; Neal et al., 2009; Niemannsverdriet et al., 2008). Thus, 14–3–3 zeta/delta represents an attractive target to overcome resistance to a broad range of anticancer agents. A notable group of proteins recurrently associated with drug resistance is involved in membrane trafficking and regulation and includes four members of the Rab family (Rab5B, 11/11; Rab1B, 10/13; Rab11, 10/12; and Rab14, 7/10) as well as Sec22B (9/11), indicating that these proteins mediate resistance by facilitating degradation and extrusion of drug molecules, thus keeping intracellular levels of the active drug low. We note that protein kinases were only rarely associated with sensitivity to more than one drug, underscoring the notion that aberrantly expressed and regulated protein kinases represent selective drug targets for selected cancer subtypes. Drug resistance, however, might be more frequently mediated by protein kinases, such as PAK-4 (B/9), which may, again, modulate drug sensitivity through the inhibition of apoptosis (Gnesutta and Minden, 2003; Gnesutta et al., 2001).

Elastic net analysis enabled the identification of drug sensitivity and resistance features for both targeted drugs as well as chemotherapeutics. For instance, sensitivity to paclitaxel, a mitotic drug targeting tubulin, is strongly associated with high expression of aryl hydrocarbon receptor interacting protein (AIP; Figures 6B and 6F), for which several lines of evidence suggest a role as a tumor suppressor (Georgiti et al., 2007; Nord et al., 2010) and proapoptotic protein (Kang and Altieri, 2006). Despite numerous associated mutations, the strongest correlates for paclitaxel resistance were actually the expression of glutamate dehydrogenase 1 (GluD1) and Rab5B (Figure 6B). Dasatinib sensitivity, in contrast, was notably associated with expression of Src kinase, one of the primary targets of dasatinib, and accompanied by expression of numerous Src substrates, such as STAT1 or the small subunit 1 of the calcium-dependent protease calpain (CAPNS1) as well as integrin beta-1 (ITGB1), a Src activator (Figure 6C). Consistent with this, although STAT1 activation is frequently increased in tumors and cell lines, its functional role has previously been associated with growth suppression and, hence, can be considered as a potential tumor suppressor (Bromberg et al., 1996). Resistance to dasatinib is significantly associated with proteins involved in RNA processing as well as cellular compromise, such as the apoptosis-inducing factor (AIF; Figure 6C). AIF plays a central role in caspase-independent cell death; it has not yet been ascribed a potential role in resistance to targeted cancer drugs, but it might represent valuable therapeutic potential in resistant tumor cells (Lorenzo and Susin, 2007).
The NCI-60 Proteome Database Enables Access to This Comprehensive Resource

This current proteome resource of 59 commonly employed cancer cell lines enables a wide range of analyses, which are beyond the scope of this study. For instance, the data can be used to obtain reference expression profiles for proteins of interest both within and across a range of cell line proteomes. It may facilitate selecting the appropriate cell line for the study of a particular biological phenomenon of interest without the need for additional experiments such as RNA deep sequencing (Danielsson et al., 2013), or might alternatively be used as reference for targeted proteome analyses (Picotti and Aebersold, 2012).

To enable efficient use of these data by the scientific community, the data were incorporated into a database, which is accessible via a user-friendly web interface. The database contains protein expression profiles along with details about protein and peptide identification information, including the matched tandem mass spectra. The data are cross-referenced to Uniprot, Ensemble, and several other major resources to facilitate additional information browsing.

Conclusions

In the present study, we provide a comprehensive resource of NCI-60-wide protein expression profiles for more than 10,000 proteins, including more than 350 protein kinases. Our results indicate that, despite the moderate depth of proteome coverage per single cell line, the broad biological information contained in transcriptomics and proteomics data is similar, but each technique provides complementary information. While transcriptomics enables genome-wide investigations of mRNAs and has proven very useful, it can by nature not be utilized to study post-transcriptional processes of cellular regulation, such as protein expression, posttranslational modifications, protein degradation, protein interaction, or protein activities, which are areas largely confined to proteome analysis (some of which were highlighted here). Although the proteomic data acquired in this study...
EXPERIMENTAL PROCEDURES

Protein Preparation and Kinobead Affinity Purification
Cell pellets obtained from DTP of the NCI were lysed and kinobead pull-downs were performed and prepared for in-gel digestion as previously described (Wu et al., 2011, 2012). For full proteomes, 50 µg from each kinobead flow-through were reduced, alkylated, and separated via an LDS-PAGE gel. It is of note that the kinobead procedure does not result in significant kinase depletion from the flow-through (Bantscheff et al., 2007). In-gel trypsin digestion was performed according to standard procedures (Shevchenko et al., 1996).

Protein Identification and Quantification
Nanoflow LC-MS/MS analyses of tryptic peptides were conducted on an Eksigent nanoLC-Ultra 1D+ (Eksigent) coupled to an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific). The mass spectrometers were operated in data-dependent mode and raw MS spectra were processed using Maxquant (version 1.3.0.3; Cox and Mann, 2008). Tandem mass spectra were searched with Andromeda (Cox et al., 2011) against the IPI human database (version 3.68; 87,061 sequences) and a maximum false discovery rate (FDR) of 1% for peptides and proteins was required. Protein abundance was estimated based on summed peptide intensities from proteome profiling experiments (but not deep proteomes and kinomes), and label-free quantification was used for comparisons between samples (Luber et al., 2010).

Statistical and GO Enrichment Analyses
Statistical analysis of quantified proteins was performed using R (version 2.12.1; Team, 2012). Differential expression was assessed via ANOVA and p values were corrected for multiple hypothesis testing to control the FDR at 5% (Benjamini and Hochberg, 1995). Cluster analyses including hierarchical clustering and PCA were performed using a variety of algorithms and metrics. Classification and functional enrichment as well as pathway membership were analyzed using BiNGO (Maere et al., 2005) and Ingenuity Pathway Analysis (Ingenuity Systems). The kinase dendrogram was an adapted and reproduced courtesy of Cell Signaling Technology using modified version of the of Human Kinome application of the Tripod project (http://tripod.nih.gov/).

Comparison of Proteomics and Transcriptomics
Normalized gene expression data for NCI-60 cell lines were obtained from the Gene Expression Omnibus (accession number GSE32474; Barrett et al., 2009; Pfister et al., 2009). Significant differences were identified applying a Bayesian approach using the limma package (Bioconductor 2.7; Smyth, 2004). A threshold of an adjusted p value <0.05 was used to identify significant changes. CIA (Culhane et al., 2003; Doledec and Chessel, 1994) was used to analyze statistical relationships between protein and gene expression patterns.

Protein-Drug Associations
Elastic net regression (Zou and Hastie, 2005) was used to identify associations between proteins and drug response across NCI-60 cell lines. Protein-drug associations have been assessed for 108 FDA-approved drugs, and drug between proteins and drug response across NCI-60 cell lines. Protein-drug associations have been assessed for 108 FDA-approved drugs, and drug between proteins and drug response across NCI-60 cell lines. Elastic net regression (Zou and Hastie, 2005) was used to identify associations

ACCESION NUMBERS
The MS raw files associated with this manuscript are available for download (https://www.proteomicsdb.org/proteomicsdb/#projects/35). Processed MS files (MaxQuant output files) are available for download (http://wwz.tum.de/proteomics/nci60).

SUPPLEMENTAL INFORMATION
Supplemental Information includes Extended Experimental Procedures and six figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.07.018.

ACKNOWLEDGMENTS
We gratefully acknowledge Fiona Pachi for mass spectrometric analyses, Guillaume Médard and Thomas Kuehne for programming, and Michaela Kroetz-Fahning, Andrea Hubauer, and Andreas Klaus for excellent laboratory assistance.

Received: February 14, 2013
Revised: May 23, 2013
Accepted: July 18, 2013
Published: August 8, 2013

REFERENCES


