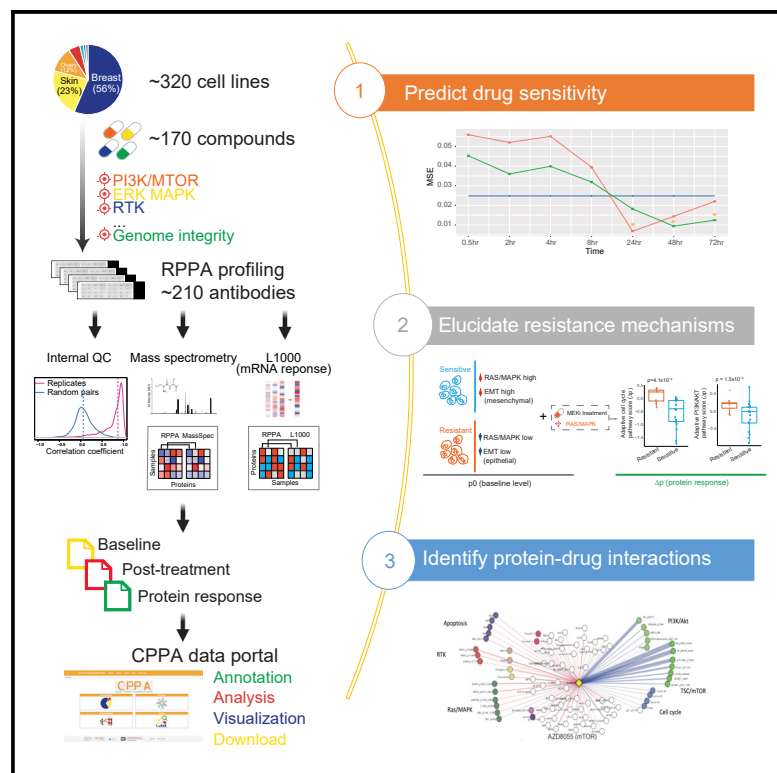# Large-Scale Characterization of Drug Responses of Clinically Relevant Proteins in Cancer Cell Lines

Graphical Abstract

Authors

Wei Zhao, Jun Li, Mei-Ju M. Chen, ...,
Yiling Lu, Gordon B. Mills, Han Liang

Correspondence

gmills@ohsu.org (G.B.M.),
hliang1@mdanderson.org (H.L.)

In Brief

Zhao et al. profile the protein responses of a large collection of cancer cell lines to drug perturbations using the RPPA platform and build a systematic protein-drug connectivity map. The integration of perturbed protein responses provides better prediction of drug sensitivity and insights into drug-resistance mechanisms and combination therapies.

Highlights

- A large collection of cancer cell line protein responses to drug perturbations

- Perturbed protein responses greatly increase predictive power for drug sensitivity

- A systematic map of protein-drug connectivity was built based on response profiles

- A user-friendly, interactive data portal was developed for community use

CellPress

Cancer Cell

CellPress

Article

Large-Scale Characterization of Drug Responses of Clinically Relevant Proteins in Cancer Cell Lines

Wei Zhao,[1,2,12] Jun Li,[1,12] Mei-Ju M. Chen,[1] Yikai Luo,[1,3] Zhenlin Ju,[1] Nicole K. Nesser,[4] Katie Johnson-Camacho,[4] Christopher T. Boniface,[4] Yancey Lawrence,[4] Nupur T. Pande,[4] Michael A. Davies,[5] Meenhard Herlyn,[6] Taru Muranen,[7] Ioannis K. Zervantonakis,[7,8] Erika von Euw,[9] Andre Schultz,[1] Shwetha V. Kumar,[1] Anil Korkut,[1] Paul T. Spellman,[4] Rehan Akbani,[1] Dennis J. Slamon,[9] Joe W. Gray,[10] Joan S. Brugge,[7] Yiling Lu,[2] Gordon B. Mills,[11,*] and Han Liang[1,2,3,13,*]

[1]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[2]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[3]Graduate Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX 77030, USA
[4]Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR 97201, USA
[5]Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[6]Molecular and Cellular Oncogenesis Program, Wistar Institute, Philadelphia, PA 19104, USA
[7]Department of Cell Biology, Ludwig Center at Harvard, Harvard Medical School, Boston, MA 02115, USA
[8]Department of Bioengineering, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA 15219, USA
[9]Division of Hematology-Oncology, Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA
[10]Center for Spatial Systems Biomedicine, Department of Biomedical Engineering, Oregon Health & Science University, Portland, OR 97201, USA
[11]Knight Cancer Institute and Cell, Developmental and Cancer Biology, Oregon Health & Science University, Portland, OR 97201, USA
[12]These authors contributed equally
[13]Lead Contact
*Correspondence: gmills@ohsu.org (G.B.M.), hliang1@mdanderson.org (H.L.)
https://doi.org/10.1016/j.ccell.2020.10.008

SUMMARY

Perturbation biology is a powerful approach to modeling quantitative cellular behaviors and understanding detailed disease mechanisms. However, large-scale protein response resources of cancer cell lines to perturbations are not available, resulting in a critical knowledge gap. Here we generated and compiled perturbed expression profiles of ~210 clinically relevant proteins in >12,000 cancer cell line samples in response to ~170 drug compounds using reverse-phase protein arrays. We show that integrating perturbed protein response signals provides mechanistic insights into drug resistance, increases the predictive power for drug sensitivity, and helps identify effective drug combinations. We build a systematic map of "protein-drug" connectivity and develop a user-friendly data portal for community use. Our study provides a rich resource to investigate the behaviors of cancer cells and the dependencies of treatment responses, thereby enabling a broad range of biomedical applications.

INTRODUCTION

Cancer is a highly heterogeneous disease encompassing many tissue types and diverse oncogenic drivers, with treatment responses that are often variable in distinct tumor contexts. Over the past decade, extensive efforts have been made to characterize the tremendous heterogeneity of human cancers at the molecular level (Hutter and Zenklusen, 2018; Jiang et al., 2019). A real challenge in cancer research, however, is to obtain a systematic understanding of causality and mechanisms underlying the behaviors of cancer cells with the eventual goal of improving patient outcomes (Wise and Solit, 2019). To address this challenge, perturbation experiments provide a powerful approach in which cells are modulated by perturbagens and downstream consequences are monitored (Korkut et al., 2015;

Molinelli et al., 2013; Ng et al., 2018). The longitudinal data thus obtained provide considerably greater information on both the basal biological network wiring and its associated changes under stress, thereby leading to a deeper understanding of mechanisms underlying cell survival under stress. Recently, large-scale compendia of the phenotypic and cellular effects of perturbed cancer cell lines have been established. For example, large-scale pharmacologic perturbation studies, cell viability measurements upon different drug treatments across many cell lines, have been published (Barretina et al., 2012; Basu et al., 2013; Garnett et al., 2012; Iorio et al., 2016); several studies have built genome-wide "cancer dependency maps" across a large number of cell lines using loss-of-function small interfering RNA, short hairpin RNA, or CRISPR-Cas9 screens (McDonald et al., 2017; Tsherniak et al., 2017); and a "connectivity map"

Cancer Cell
Article

CellPress
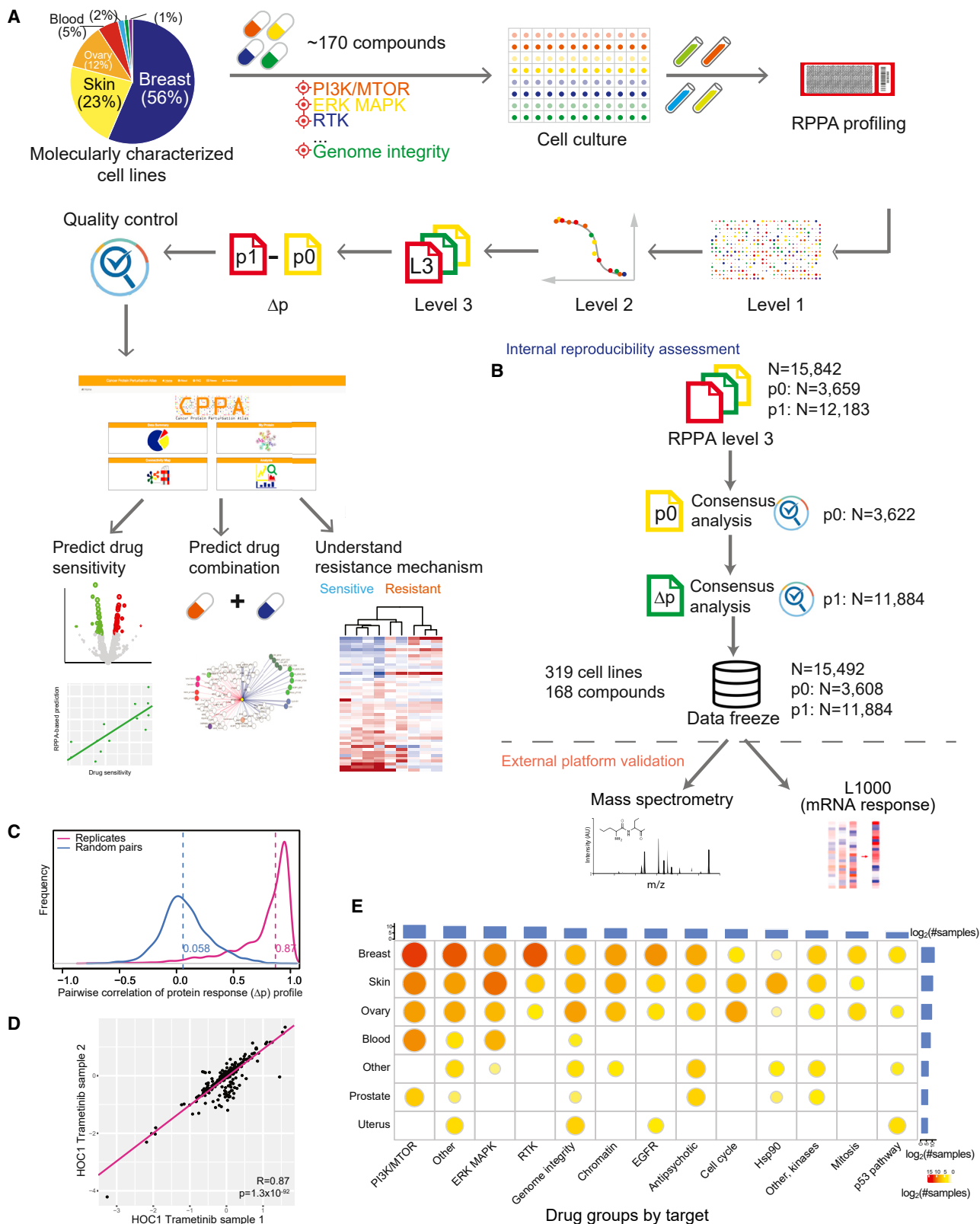
of profiled mRNA responses of cancer cell lines to diverse perturbations using an efficient, robust RNA measurement platform, L1000, has been developed (Subramanian et al., 2017). These studies provide valuable resources for gaining a systems-level understanding of cancer mechanisms and phenotypes. However, similar large-scale resources for analysis and integration of protein responses of perturbed cancer cell lines have yet to be established. This knowledge gap is even more striking, considering that proteins comprise the basic functional units in biological processes and represent the major targets for cancer therapy.

To fill this gap, we generated and compiled a large compendium of perturbed protein expression profiles of cancer cell lines in response to a diverse array of clinically relevant drugs using reverse-phase protein arrays (RPPAs). RPPA is a quantitative antibody-based approach to assess protein markers in a large number of samples in a high-throughput, cost-effective, sensitive manner (Hennessy et al., 2010; Nishizuka et al., 2003; Tibes et al., 2006). This platform depends on antibodies for the detection of proteins, and currently, there is a limited but rapidly growing number of proteins for which high-quality antibodies exist that give an analyzable signal. We have applied this technology to quantify protein expression levels of large patient cohorts (e.g., The Cancer Genome Atlas) (Akbani et al., 2014) and cancer cell lines (e.g., MD Anderson Cell Lines Project and Cancer Cell Line Encyclopedia) (Ghandi et al., 2019; Li et al., 2017). The current antibody repertoire covers key oncogenic pathways such as PI3K/AKT, RAS/MAPK, Src/FAK, TGF-β/SMAD, JAK/STAT, DNA damage repair, Hippo, cell cycle, apoptosis, histone modification, and immune oncology. Compared with proteome-wide mass spectrometry approaches, our RPPA-based approach has several advantages. First, although the number of protein markers in the RPPA readout is much smaller (200–300), this highly select protein set is enriched in therapeutic targets and biomarkers, thereby greatly increasing the ability to generate clinically relevant hypotheses and make translational impacts. Statistically, this more focused assessment also substantially reduces multiple testing, a major challenge in identifying significant hits from unbiased proteomic searches (if no pre-filters are applied). Second, one RPPA slide can measure up to 1,000 samples simultaneously. Thus, the high throughput and cost effectiveness make RPPA a practical platform for assessing a large number of samples (e.g., >10,000), which is simply not feasible for alternative proteomic approaches. Third, protein-level responses, particularly changes in post-translational modifications, more likely reflect how cancer cells rewire their signaling pathways to adapt and survive a specific drug treatment, as most targeted therapies act by modulating protein phosphorylation and activity. The superior ability of RPPA to

quantify some key post-translationally modified proteins has the potential to capture such adaptive responses and can provide stronger predictors of therapy response or resistance mechanisms (Mertins et al., 2014). Indeed, our recent studies have demonstrated the value of RPPA-based adaptive responses in the rational design of combination therapies (Fang et al., 2019; Iavarone et al., 2019; Korkut et al., 2015; Krepler et al., 2016, 2017; Kwong et al., 2015; Molinelli et al., 2013; Muranen et al., 2012; Sun et al., 2017, 2018), with several of these translated to the clinic with patient benefit.

RESULTS

A Large, High-Quality Collection of Perturbed RPPA Profiles of Cancer Cell Lines

To generate a high-quality resource of perturbed protein responses, we measured RPPA-based protein expression profiles of cancer cell lines in response to ~170 preclinical and clinical therapeutics (often across multiple time points), generated normalized RPPA data (including baseline level p0 and post-treatment level p1) and protein response to perturbation ($\Delta p = p1 - p0$) profiles using a standardized data processing pipeline, and made the data public through a user-friendly data portal (Figure 1A). We developed a multiple-stage, multiple-platform quality control (QC) pipeline in which we first evaluated the reproducibility of both baseline and protein response profiles of our samples within the RPPA platform, and then validated the replicability of our data using independent platforms (i.e., protein by mass spectrometry and mRNA by L1000) (STAR Methods, Figure 1B). By comparing protein response (Δp) correlations of replicate samples to those of random sample pairs, we demonstrated that replicate samples showed higher correlations across protein markers (mean R = 0.87) than random pairs (mean R = 0.058) (Figures 1C and 1D), indicating high reproducibility of our RPPA data. After excluding ~2.2% low-confidence samples, the final compendium contained QC-passed RPPA profiles (~210 total and phosphorylated protein markers) of 15,492 samples (11,884 drug-treated samples and 3,608 control samples related to perturbation of 168 compounds in 319 cell lines) in total.

The cancer cell lines in our dataset come from several lineages, including breast, ovarian, uterus, skin, blood, and prostate; and the drug compounds target a broad range of cancer-related processes, including PI3K/mTOR signaling, ERK/MAPK signaling, RTK signaling, EGFR signaling, TP53 pathway, genome integrity, cell cycle, antipsychotic drugs, and chromatin remodeling (Figure 1E and Table S1). Due to time and cost constraints and the clinical relevance of different drugs, instead of profiling all possible perturbations across all cell line and drug

Figure 1. Summary of the Perturbed RPPA Profiling Data in This Study
(A) Overview of the RPPA profiling experiments and data processing of cell line perturbations. The pie chart shows the lineage distribution of cancer cell lines profiled (n = 319).
(B) The RPPA quality control pipeline, which contains within-platform assessment and external validation using independent platforms.
(C) Reproducibility of perturbed RPPA data based on protein response (Δp) profiles of technical replicates (n = 2,753 pairs).
(D) A representative scatterplot showing the correlation of Δp between two replicate samples across protein markers.
(E) The distribution of drug-treated samples by cell lineage and drug group. The bar plots show the number of samples profiled for each lineage or drug group, and the size of the circle is proportional to the number of samples profiled for each lineage-drug combination.
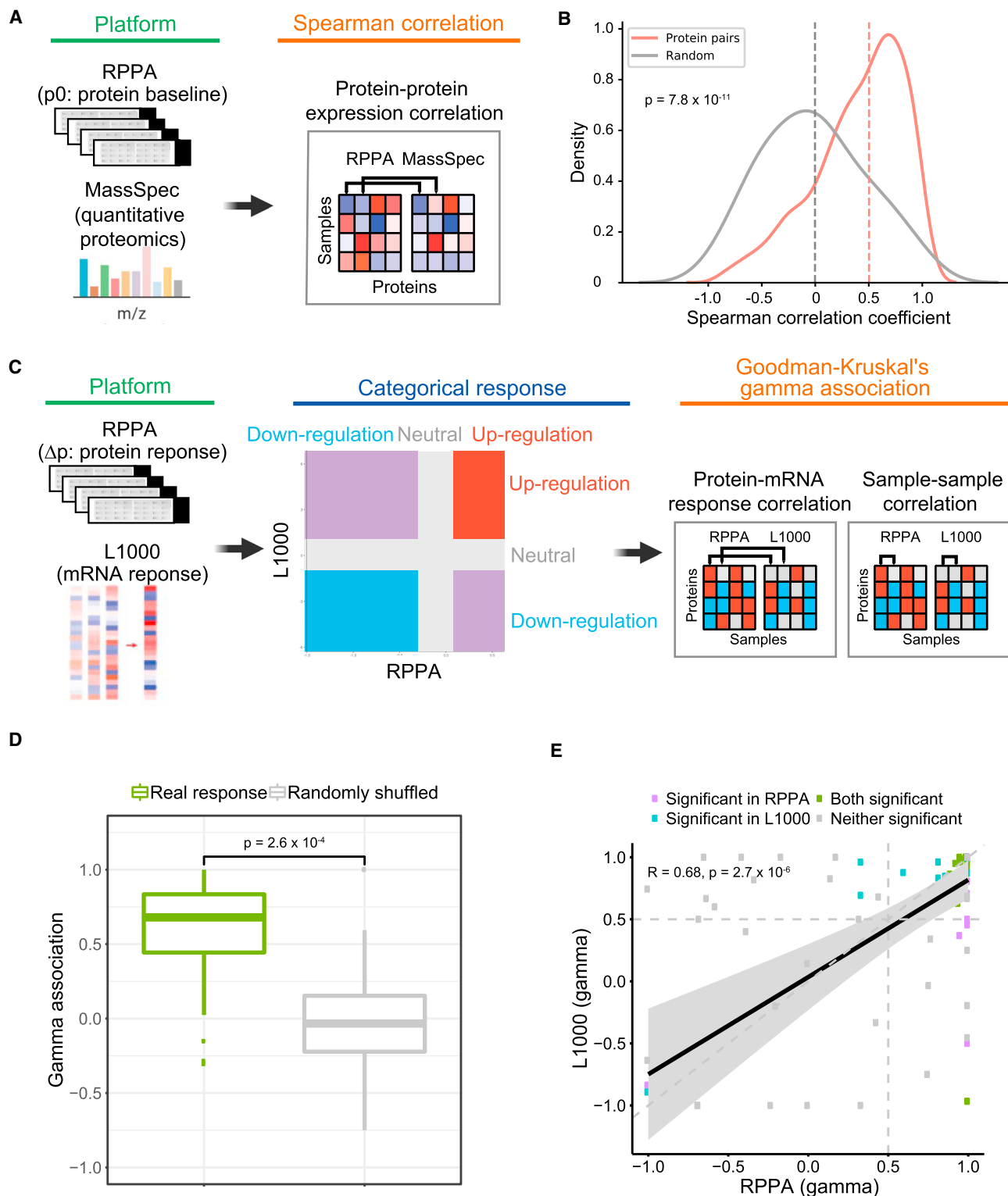See also Figures S1 and S6 and Table S1.

Figure 2. Quality Assessment of RPPA-Based Protein Expression Data Using Independent Platforms

(A) Overview of the comparison between RPPA-based and mass spectrometry-based protein expression data.

(B) The distribution of correlation coefficients of matched and random protein pairs between RPPA and mass spectrometry. The median values are marked by dash lines.

(C) Overview of the comparison between RPPA-based protein response and L1000-based mRNA response (see STAR Methods for details).

Cancer Cell
Article

CellPress

combinations, we took a more pragmatic approach in which some cell lineages and drug groups were more frequently profiled but still represent an extensive survey of drug perturbations (Figure 1E). Our sample set is highly enriched in responses from a subset of common, well-characterized cancer cell lines that have rich molecular profiling and drug response data in public resources (Figure S1A). For example, >1,500 drug-treated samples were from MCF7, and >250 drug-treated samples were from BT20, SKBR3, MDA-MB-468, BT549, UACC812, BT474, SKOV3, and HCC1954 (Figure S1B). For drug treatment, ~86% of the samples were treated with monotherapy, and ~1,700 samples were treated with double- or triple-drug combinations (Figure S1C). Among the drug compounds used, 23 compounds had >150 treated samples, with lapatinib (480 samples, HER2 inhibitor) and GSK690693 (454 samples, AKT inhibitor) being the top two drug treatments (Figure S1D). Importantly, for many of the therapeutic targets, we profiled multiple targeting agents, including those that target different members of the same pathway, to increase our ability to identify on-target activity.

Since high reproducibility in the same platform does not necessarily imply validity, to further confirm the quality of the RPPA data output, we sought to validate our protein response data using independent platforms (Figure 1B). First, we compared the baseline protein expression in our RPPA platform with baseline mass spectrometry data in a set of shared cell lines (Figure 2A). We found that the corresponding protein pairs between the two proteomic platforms showed substantially higher correlations across cell lines than random protein pairs (median correlation coefficient: 0.50 versus 0.0, paired Student's t test, $p = 7.8 \times 10^{-11}$, Figure 2B). Indeed, the majority of protein pairs showed high correlations between the RPPA and the mass spectrometry platforms. Second, since extensive data on protein changes in response to drug treatment are not available, we employed mRNA response data from the connectivity map (Subramanian et al., 2017). As this analysis is for different molecules (protein versus RNA), and across different platforms (RPPA versus L1000), we employed the Goodman-Kruskal's γ correlation to conduct a robust assessment. Based on the same cell lines perturbed by the same compounds (n = 46 unique cell line-drug perturbations), we converted the original continuous response scores into categorical response groups (i.e., upregulated, neutral, and downregulated) and compared the mRNA-protein response concordance by calculating mRNA-protein response association and sample-sample association (Figure 2C). We observed that the matched mRNA-protein responses from the same condition were highly associated with each other, with median $\gamma = 0.68$, which is significantly higher than that from the randomly shuffled background distribution (paired Student's t test, $p = 2.6 \times 10^{-4}$, Figure 2D). Then, we tested whether the sample-sample associations inferred from the RPPA-based protein responses were preserved in the

L1000-based mRNA responses. Among the significant sample-sample associations identified by either platform (false discovery rate [FDR] <0.01), the RPPA-based γ scores showed a strong, positive correlation with the L1000-based γ scores (Pearson's correlation, $R = 0.68$, $p = 2.7 \times 10^{-6}$, Figure 2E). Further, categorized RPPA-based associations are highly consistent with L1000-based associations (Fisher's exact test, $p = 3 \times 10^{-3}$). These external validations using cross-molecule, cross-platform, and cross-study comparisons strongly support the overall high quality of our protein response data.

Mechanistic Insights into Drug Sensitivity by Protein Responses

To assess whether the protein response data can provide meaningful insights into phenotypic consequences of drug treatments, we focused on MCF7, a breast cancer cell line with the highest number of treated samples in our dataset, for an in-depth analysis. We first extracted RPPA data from >1,500 MCF7 samples treated by a variety of compounds. In total, these samples were treated with 19 compounds or combinations, 9 stimuli, and DMSO, at multiple time points, which covered major MCF7 drug targets, including the estrogen receptor (ER), PI3K/mTOR, AKT, MEK, and EGFR. To elucidate the signaling pathways underlying different drug responses, we further summarized the protein response (Δp) data in two dimensions by (1) grouping proteins into major cancer functional pathways (e.g., PI3K/AKT, RAS/MAPK, and TSC/mTOR) and (2) categorizing compounds based on their target pathways (e.g., ER, PI3K/mTOR, and AKT). We next ranked the drug groups based on their median values of drug sensitivities in MCF7 using Cancer Therapeutics Response Portal v.2 (CTRPv2) (Basu et al., 2013) and found that MCF7 was significantly more sensitive to two of the eight drug groups (Figure 3A). The same pattern was confirmed using Genomics of Drug Sensitivity in Cancer 2 (GDSC2) (Garnett et al., 2012), another large-scale drug-sensitivity data resource (Figure S2A). In particular, MCF7, an ER-positive breast cancer cell line, showed the highest sensitivity to ER inhibitors, with PI3K/mTOR inhibitors being the second most effective group. We used the median of Δ pathway scores (Akbani et al., 2014) to represent the average pathway response to each drug group and found that, indeed, the two drug groups showed the most dramatic pathway responses (Figure 3B). Specifically, using pathway analysis as defined previously, ER inhibitors decreased TSC/mTOR and hormone_a, but increased EMT, core reactive, DNA damage, and hormone_b pathways; PI3K/mTOR compounds inhibited TSC/mTOR, PI3K/AKT, and cell cycle, but activated apoptosis and RTK signaling; other drug groups also specifically inhibited their target pathways, such as Abl/Src/c-Kit and MEK. We further compared the sensitive drug groups (ER and PI3K/mTOR) with others for each pathway and revealed several differentially altered response pathways (Figure 3C, 3D). Specifically, three pathways were inhibited in the sensitive

(D) Boxplots of protein-mRNA response associations between the RPPA and the L1000 platforms using the same perturbations (i.e., the same cell line and the same compound, n = 46). The γ associations from the real responses (green box) were compared with those from the randomly shuffled background distribution (gray box). The p value is based on a paired Student t test. The middle line in the box is the median, the bottom and top of the box are the first and third quartiles, and the whiskers extend to 1.5 interquartile range (IQR) of the lower and upper quartiles, respectively.

(E) Scatterplot showing the correlation of sample-sample γ associations from the RPPA (x axis) and L1000 (y axis) platforms. Only significant data points (γ associations) with FDR <0.01 in either platform are shown. Pearson's correlation coefficient and p value are shown.
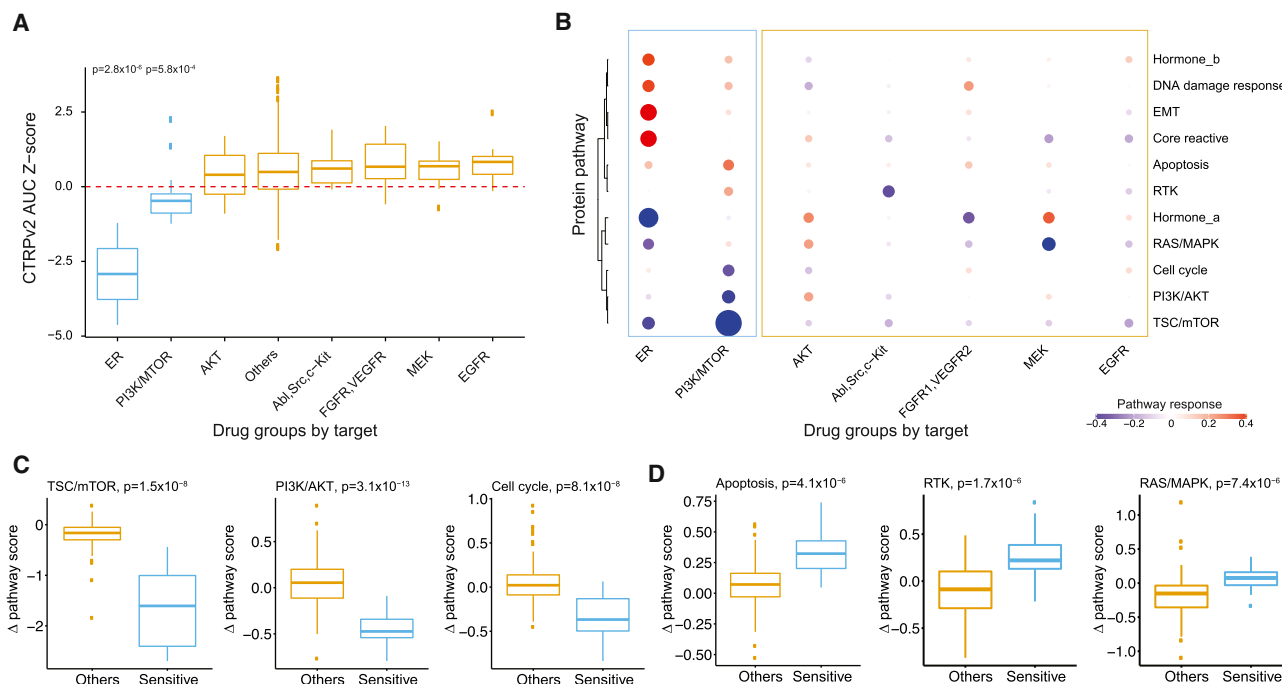
Figure 3. Protein Responses to Various Drug Treatments in MCF7 Cells
(A) A comparison of MCF7 drug sensitivity data between different drug groups using CTRPv2 data.
(B) Heatmap showing the pathway responses among drug groups. The size of the circle is proportional to the effect size of the protein changes.
(C and D) Boxplots showing significantly down- (C) and upregulated pathways (D) in sensitive drug groups.
The p values were calculated by Student's t test. See also Figure S2.

groups, i.e., TSC/mTOR (t test, $p = 1.5 \times 10^{-8}$), PI3K/AKT ($p = 3.1 \times 10^{-13}$), cell cycle ($p = 8.1 \times 10^{-8}$), and three pathways in the sensitive groups had significantly higher Δ pathway scores than other drugs, including apoptosis ($p = 4.1 \times 10^{-6}$), RTK ($p = 1.7 \times 10^{-6}$), and RAS/MAPK ($p = 7.4 \times 10^{-6}$). These observations not only indicate that our RPPA-based protein response data successfully captured important phenotypic effects of drug treatments, but also demonstrate the ability to uncover molecular mechanisms underlying drug sensitivity.

To further demonstrate how protein response could help elucidate drug-response mechanisms, we focused on the analysis of MEK inhibitors (MEKis) across different cell lines, using cobimetinib as an illustration example and considering both baseline (p0) and protein response levels (Δp) (Figure 4). Cell lines were divided into MEKi resistant (OVCAR432, RAS pathway wild type [WT]; OVCAR3, RAS pathway WT; and OAW28, MAP2K4 mutant) and MEKi sensitive (OV90, BRAF mutant; CAOV3, RAS pathway WT; ES2, BRAF and MEK mutant; OVCAR5, KRAS mutant; JHOM1, RAS pathway WT; and OVCAR8, KRAS mutant) based on response to multiple MEKis in our and publicly available data (CTRPv2 and GDSC2). As expected, cell lines with aberrations in the RAS/MAPK pathway had a higher propensity to RAS/MAPK baseline pathway activity and sensitivity to MEKi, as indicated by low BIM and high EGFR, DUSP4, transglutaminase, pYB1, p90RSK, pMAPK, pMEK, and pJun (Pohl et al., 2005) (Figure 4A). There was also a suggestion that cell state and, particularly, decreased epithelial characteristics, or epithelial-mesenchymal transition (EMT) (low E-cadherin, β-catenin,

RAB25, ERα, and GATA3 and high EPPK1, N-cadherin, AXL, PAI-1, and fibronectin), were associated with sensitivity to MEKi. The EMT characteristics were likely mediated, at least in part, by the effects of RAS/MAPK pathway activation noted above (Shao et al., 2014).

Sensitivity to cobimetinib was associated with evidence for a greater cobimetinib-induced decrease in RAS/MAPK pathway activity (decreased DUSP4, transglutaminase, FOXM1, p90RSK, pMAPK, pYB1, pS6, and pJun, and increased BIM), and decreased cell-cycle progression (decreased pRB, cyclinB1, CDK1, PLK1, cdc25c, and Chk1, and increased p16, p21, and p27), likely as a consequence of RAS/MAPK signaling inhibition (Figure 4B). Further, there was a marked shift to an epithelial phenotype, as indicated by increased EMA, EPPK1, Claudin1, and β-catenin (Figure 4B). Many of the associations with sensitivity to cobimetinib were identifiable in the pre-treatment samples, with the associations markedly accentuated and extended in cobimetinib-treated samples. The marked increase in BIM in response to MEKi has been identified previously and provides a biomarker for response to combined inhibition of MEKi and BCL2 family members (Cragg et al., 2008; Iavarone et al., 2019). We also performed a similar analysis using trametinib (Figure S3) and observed a marked overlap of potential biomarkers despite the analysis of different cell lines and different MEKis. We next calculated the pathway-level responses by aggregating the protein changes in the pathways (Akbani et al., 2014) and found that the adaptive pathway score changes associated with drug sensitivity include cell-cycle inhibition in
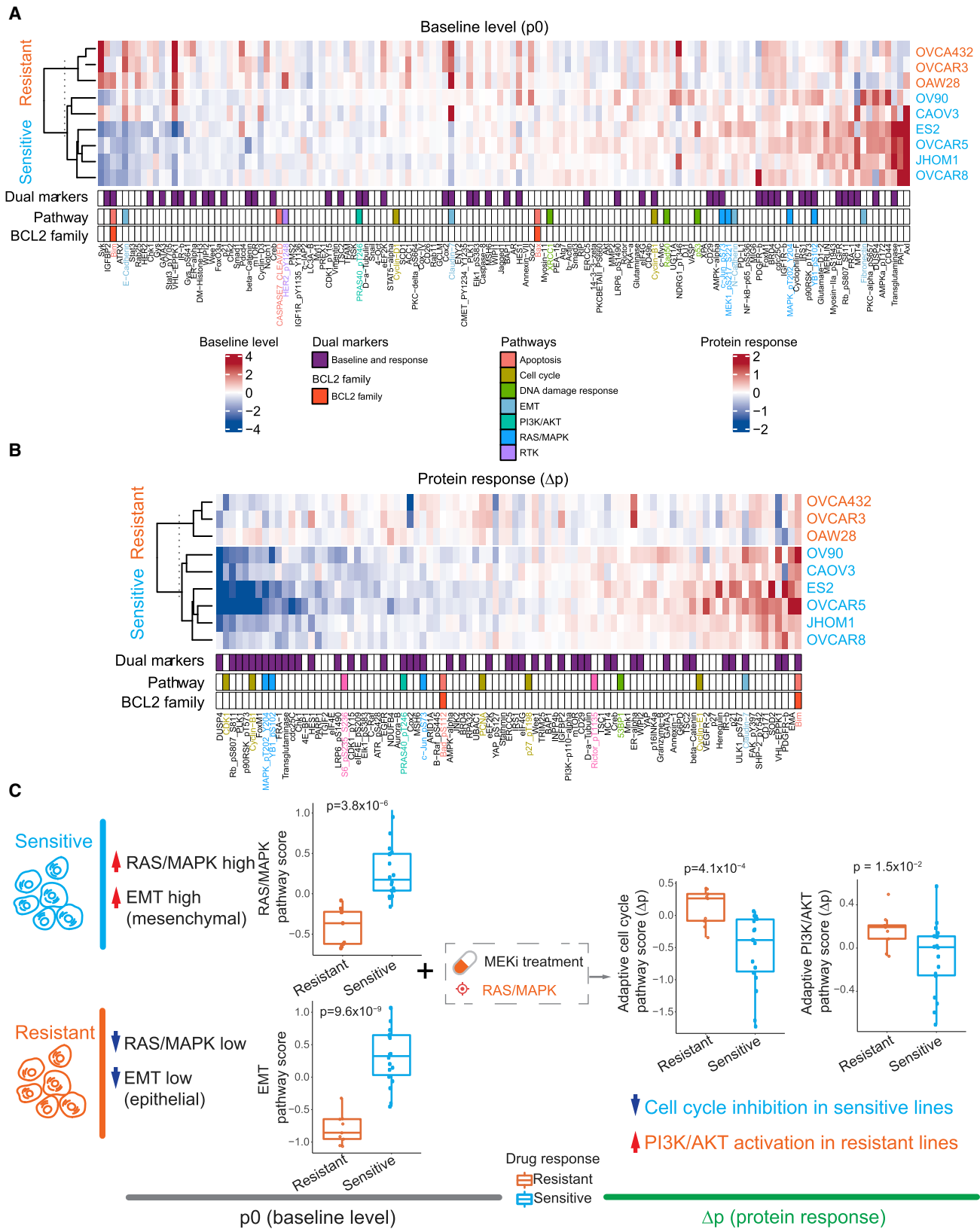
(legend on next page)

sensitive cell lines (t test, $p = 4.1 \times 10^{-4}$, Figure 4C) and PI3K/ AKT signaling activation in resistant cell lines (t test, $p = 0.015$, Figure 4C). Together, the results suggest that (1) sensitivity to RAS/MAPK pathway inhibition is associated with baseline pathway activity and cell state and (2) acquired resistance to MEKi may come from the adaptive activation of PI3K/AKT pathways in resistant cells (Mirzoeva et al., 2009; Westin et al., 2019).

Increased Predictive Power for Drug Sensitivity by Protein Response

Our previous study demonstrated that RPPA-based baseline protein levels showed considerable predictive power for drug sensitivity in cancer cell lines (Li et al., 2017). Here we performed two complementary analyses to further assess the predictive power of protein responses for drug sensitivity. First, we evaluated the associations of perturbed RPPA data with drug sensitivity based on individual proteins. We integrated our RPPA data and drug sensitivity data available in GDSC2 (Iorio et al., 2016) and identified seven drugs whose sensitivity and protein expression data were available for at least five different cell lines. Then, for each drug, we defined three types of protein markers that may be informative about drug sensitivity: (1) p0: the baseline level of a protein shows a significant correlation with the sensitivity to the drug across cell lines (Pearson's correlation, $p < 0.05$); (2) Δp only: the protein response shows a significant correlation with drug sensitivity (Pearson's correlation, $p < 0.05$); and (3) $\Delta p | p0$: given p0, the protein response shows additional information content in predicting drug sensitivity. Across all the drugs, the numbers of Δp-informative (Δp only + $\Delta p | p0$) protein markers were significantly higher than those of p0-based markers (paired t test, $p = 1.38 \times 10^{-3}$, $n = 7$ drugs, Figure 5A).

Second, we focused on two drugs, lapatinib and GSK690693, which had RPPA protein and drug sensitivity data available in at least 10 cell lines (Garnett et al., 2012; Daemen et al., 2013), in order to assess the overall performance of all RPPA protein markers for drug sensitivity prediction using a rigorous machine learning approach. For each drug, we had RPPA protein data available at seven post-treatment time points (Figures 5B–5E). For comparison, we developed predictive models using elastic net based on three sets of protein markers: (1) p0, baseline level; (2) p1, post-treatment level (averaged across different time points); and (3) joint p0 and p1 profiles. Based on leave-one-out cross-validation, models based on joint p0 and p1 profiles had superior performance compared with models based on p0 or p1 only, and the predictions were significantly correlated with drug sensitivity (Figures 5B and 5D, Pearson's correlation: lapatinib, $p = 9.7 \times 10^{-5}$; GSK690693, $p = 0.021$). Further, the p1-based models of both drugs showed better predictions than the p0-based models, and their associations with drug sensitivity were also significant (Pearson's correlation: lapatinib, $p = 2.7 \times 10^{-3}$; GSK690693, $p = 0.05$). The p0-based lapatinib model was significant, but to a lesser extent ($p = 0.016$), and

the GSK690693 model was not significant ($p = 0.4$). To obtain more insights into the model performance, we examined the predictive models based on the three sets of markers at each of the seven time points (Figures 5C and 5E). Overall, the p1 and p0 + p1 joint models showed much better performance at later time points (e.g., >8 h) than at earlier time points (e.g., ≤ 4 h). For both drugs, the prediction models based on the p0 + p1 joint profiles at 48 and 72 h had high predictive powers. The varied predictive performance at different time points suggests that during the initial stage after drug treatment, the protein responses reflect target inhibition and related collateral chaos, and it takes some time for cells to rewire signaling pathways to adapt to the treatment stress, and therefore, induced protein changes are more informative in predictive modeling. Collectively, the results in this section not only further support the high quality of the RPPA data, but also suggest that changes in protein levels on therapeutic challenge provide substantial additional information content beyond that provided by baseline protein levels for predicting treatment responses.

### A Systematic "Protein-Drug" Connectivity Map

To systematically evaluate the utility of our protein response data, we built a protein-drug connectivity map based on the RPPA data. In this map, each node represents a protein or a drug, protein-drug connections are based on whether the drug treatment caused a significant change in the protein, and drug-drug connections are based on whether the two drugs caused similar protein responses (Figure 6A). As expected, drugs for the same target are clustered together: for example, several MEKis and mTOR/PI3K inhibitors are highly connected, highlighting their similar downstream protein responses. This map also identifies intriguing connections: a PARP inhibitor showed both similar and opposite relationships with some drugs, suggesting potential additive or agonistic effects that could direct the development of rational drug combinations. Indeed, based on the assessment of functional proteomics changes as assessed by RPPA, we have validated the synergistic activity of PARP inhibitors and inhibition of PI3K pathway, MEK, ATR, and WEE1 inhibitors in preclinical and clinical studies (Fang et al., 2019; Shen et al., 2015; Sun et al., 2017, 2018).

We next studied protein-protein relationships on the map. For any given drug treatment, we classified RPPA proteins into perturbed proteins and other proteins. We found that perturbed proteins were more likely to interact than other proteins, based on the STRING database (Szklarczyk et al., 2019) (t test, $p = 3.2 \times 10^{-6}$), suggesting that proteins co-perturbed by a drug tend to be involved in the same biological processes and to interact as part of a signaling cascade (Figure 6B). This global assessment using prior protein interaction knowledge supports the utility of the approach to drive biological discoveries.

Using drug-centered protein neighborhoods, we initially focused on signaling through tyrosine kinases and their

---

Figure 4. Differentially Expressed Protein Markers between Cobimetinib-Sensitive and -Resistant Cell Lines
(A and B) Heatmaps showing baseline (A) and perturbed protein expression (B) with significant differences between sensitive and resistant cell lines (FDR <0.1). Each protein marker is annotated by whether it is a dual marker (i.e., significant in both p0 and Δp), it is a BCL-2 family member, or it belongs to a specific pathway. (C) Cartoon summary of baseline protein levels and adaptive protein responses to MEK inhibitors between the two cell groups. The difference in pathway scores between the two groups was assessed based on the Student t test.
See also Figure S3.
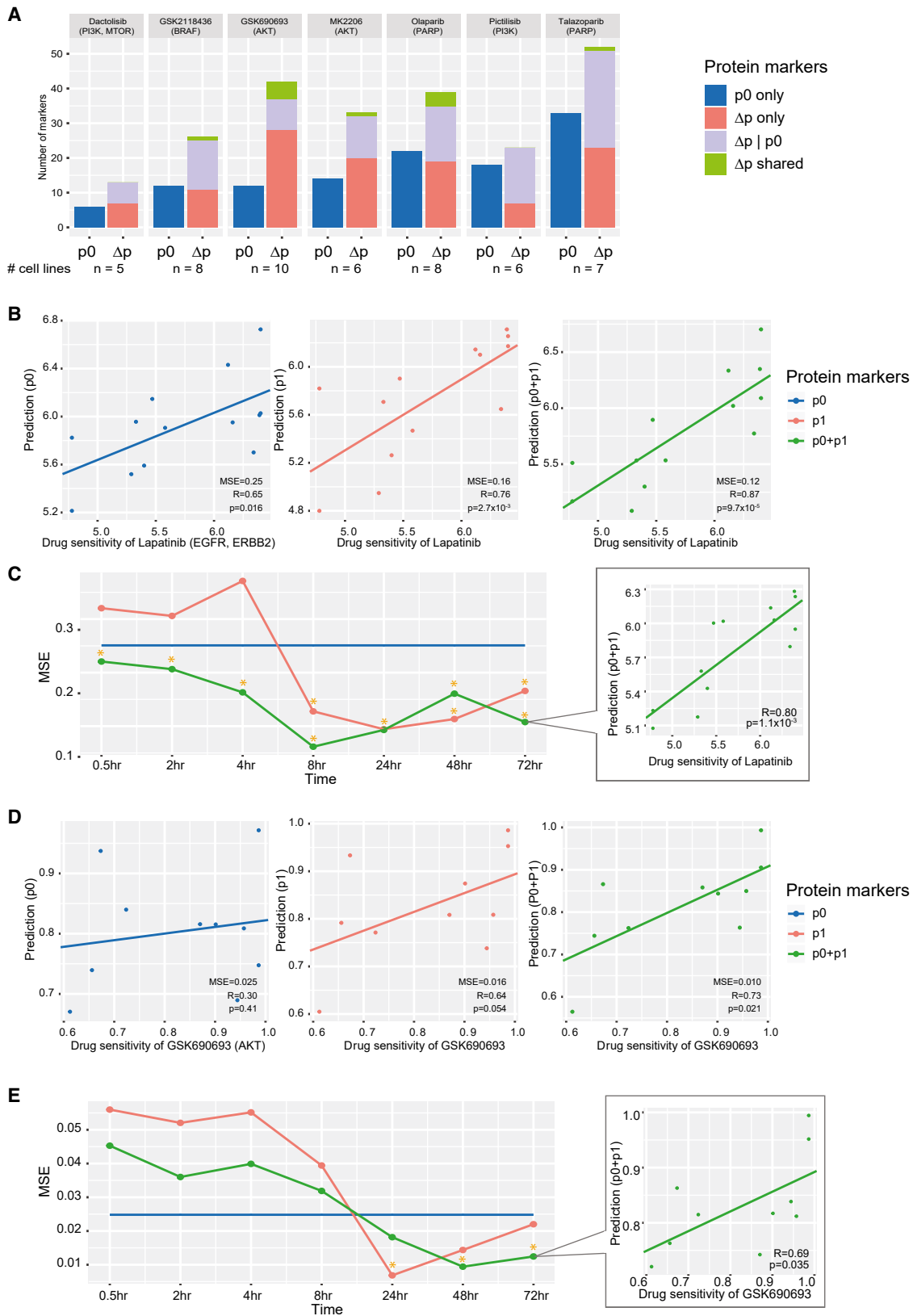
(legend on next page)

downstream networks, selumetinib (target: MEK) (Figure S4A), AZD8055 (target: mTOR) (Figure S4B), GSK1838705A (target: IGF1R/ALK) (Figure S4C), and sapitinib (target: EGFR/ERBB2) (Figure S4D), and demonstrated a marked overlap in protein networks in inhibitor-perturbed cells. Interestingly, the Hsp90 inhibitor (gamitrinib) protein neighborhood (Figure S5A) demonstrated similarities to the tyrosine kinase pathway inhibitors, potentially due to the role of Hsp90 in stabilizing multiple members of the tyrosine kinase signaling pathway. Indeed, the similarities in the protein networks argue that the major effects of Hsp90 are likely attributable to its effects on tyrosine kinase signaling pathways (Lee et al., 2017). In contrast, rabusertib (target: Chk1) (Figure S5B) and chlorpromazine (target: autophagy) (Figure S5C) demonstrated distinct protein neighborhoods consistent with markedly different mechanisms of action.

As described above, the MEKi protein neighborhood is strongly associated with signaling through the MAPK and mTOR pathways, cell-cycle progression, and cell state. Based on extensive validation of the relationships between these pathways and RAS/MAPK signaling, the associations with multiple other proteins in the neighborhood map (Figure S4A) are likely valid. Given that the MAPK pathway is a key regulator of the TSC1/2 complex that is upstream of mTORC1 signaling, the protein neighborhood of the mTOR inhibitor AZD8055 (Figure S4B) was indeed highly related to the selumetinib protein neighborhood. The most marked differences between the MEK and the mTOR inhibitor protein neighborhoods were represented in the upper components of the PI3K and MAPK pathway that appeared relatively independent of each other. Interestingly, the IGF1R/ALK inhibitor GSK1838705A protein neighborhood encompassed components of both the MEK and the mTOR protein neighborhoods, consistent with the IGF1R having input into both pathways. While the strong link to the PI3K pathway was expected, a link between the IGF1R and the MAPK pathway has been suggested but less studied (Molina-Arcas et al., 2013). The pan-EGFR family inhibitor sapitinib neighborhood reflects EGFR family receptors being the key regulators of the PI3K and MAPK pathways in epithelial cells (Akbani et al., 2014). The EGFR family has a stronger link than either mTOR or MEKis to the DNA damage repair pathway (i.e., 53BP1, Rad50, XRCC1, pChk1/2, and BRCA2), consistent with recent studies (Russo et al., 2019; Wang et al., 2013).

Prediction of Drug Combinations Based on Protein Response

To further demonstrate the utility of protein response data, we developed an integrated analysis to predict drug combinations based on pathway-level protein responses (Figure 7A). Briefly,

(1) we used a drug-centered connectivity map to infer upregulated and downregulated pathways; (2) we performed a correlation analysis to confirm whether the pathways identified in step (1) correlated with drug resistance; and (3) based on the pathway-drug matrix, we identified drugs that were affected by the resistance pathways identified in (2). Through this strategy, we identified 150 drug combinations for 9 specific drugs, including mTOR/PI3K, EGFR, AKT, PARP, and FGFR inhibitors (Figure 7B and Table S2). We then evaluated our predictions by determining whether the proposed drug combinations have been reported in the literature or have been employed in clinical trials, since each publication or clinical trial can be viewed as substantial evidence to support a specific prediction. On average, >50% of the predicted combinations had supporting evidence, and the validation rate ranged from 90% to 10% for each drug.

Next, we focused on the combination of selumetinib + MK-2206 for a detailed analysis and validation using the CTRPv2 sensitivity data of the two drugs and their combination. We compared the drug sensitivities of MK2206 (AKTi) and selumetinib (MEKi) with those of their combination and found that the drug sensitivity (area under the curve) of the individual drugs was significantly higher than that of their combination, indicating their synergistic effects (pairwise Wilcox sum rank test, $p < 2.2 \times 10^{-16}$, Figure 7C). To better understand the mechanisms underlying this synergistic effect, we further analyzed the protein response data for the individual drugs and their combination (Figure 7D). We observed three general patterns from the differential analysis: (1) both drugs had similar effects on the pathway, which were accentuated by the combination (TSC/mTOR, DNA damage response; RTK, hormone signaling); (2) one drug altered the pathway, and this was accentuated by the combination (apoptosis, cell cycle, and PI3K/AKT); and (3) each drug had the opposite effect (RAS/MAPK). The most significantly perturbed pathways were PI3K/AKT and TSC/mTOR, which were both inhibited by the combination treatment, especially the TSC/mTOR pathway. This result suggests that the synergistic effect could be due to effective inhibition of the PI3K/AKT and TSC/mTOR pathways. Notably, the cell-cycle pathway was also inhibited by the combination, which appeared to be mainly due to the effects of MK2206 (AKTi).

A User-Friendly Data Portal for Community Use of Protein Responses

To facilitate the utilization of our protein response data, we provided unrestricted access to the data through a user-friendly portal, called "Cancer Perturbed Proteomics Atlas," for fluent data exploration and analysis, which can be accessed at

---

Figure 5. Comparison of the Predictive Power of Protein Markers for Drug Sensitivity

(A) A summary of predictive markers based on baseline level (p0) and protein response (Δp) using drug response data from GDSC2. Given a specific drug, three types of predictive markers were identified: (1) proteins whose p0 level is significantly correlated with drug sensitivity; (2) proteins whose Δp level is significantly correlated with drug sensitivity; and (3) proteins whose Δp level is significantly correlated with drug sensitivity, given the p0 contribution. Protein markers identified based on both Δp only and Δp|p0 are called Δp shared. The number of cell lines for each compound is shown at the bottom.

(B and D) The scatterplots showing the correlations between predicted and measured drug sensitivity values of lapatinib (B) and GSK690693 (D) based on the multivariate models using three sets of protein markers (left, p0; middle, p1; and right, p0 + p1). Measured sensitivity data of lapatinib (n = 13 cell lines) and GSK690693 (n = 10 cell lines) were from Daemen et al. (2013) and GDSC2, respectively.

(C and E) The MSE curves of the three predictive models at different time points for lapatinib (C) and GSK690693 (E). The time points with significant correlations between predicted and measured values are indicated with *. The scatterplot at the last time point is shown.
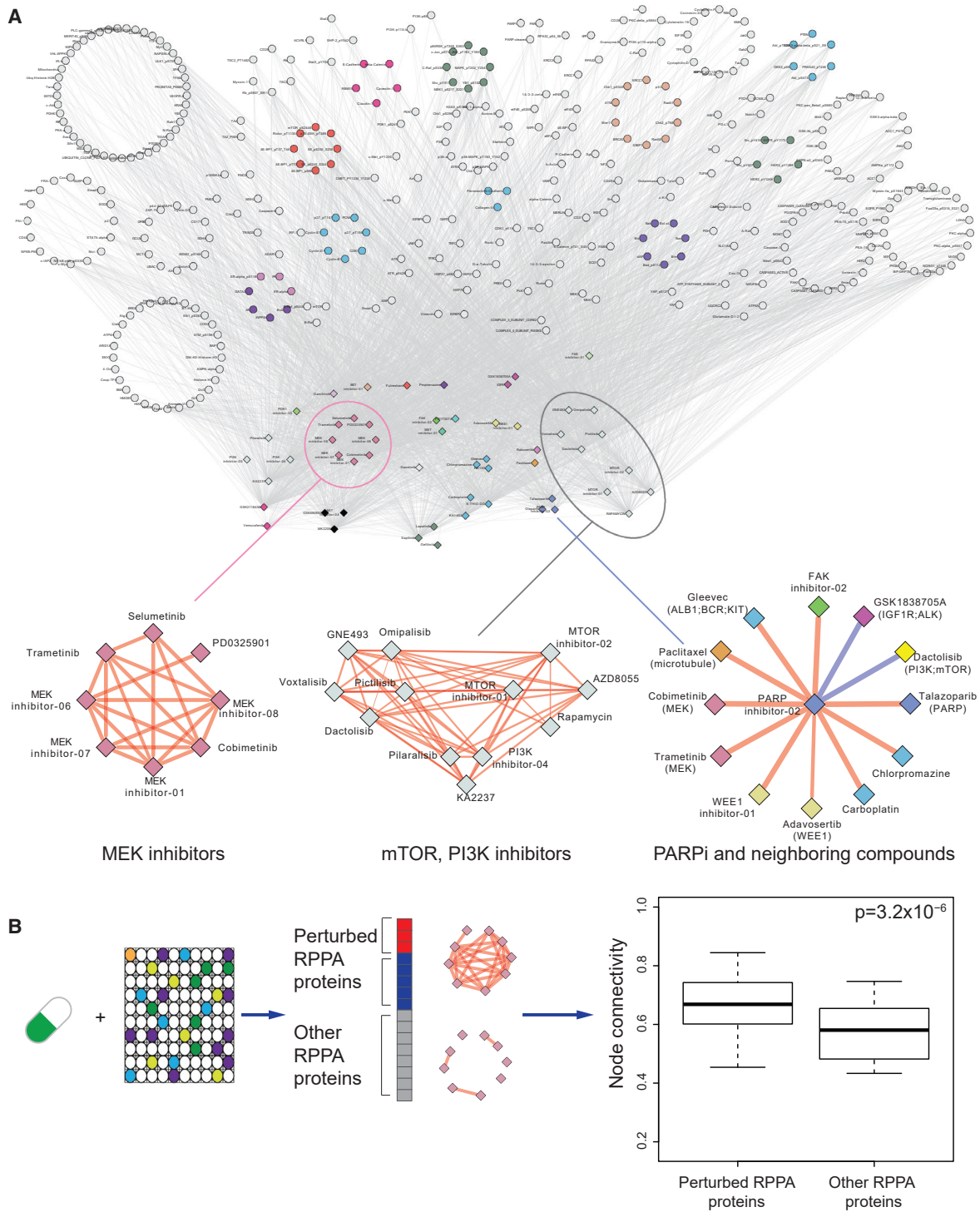
Cancer Cell
Article

CellPress

Figure 6. A "Drug-Protein" Connectivity Map Based on Protein Response Signals

(A) A global view of the drug-protein connectivity map with highlighted examples of drug-drug correlation networks (i.e., MEK inhibitors, mTOR/PI3K inhibitors, and neighboring drugs of a PARP inhibitor). Red/blue edges represent positive/negative drug-drug correlations, respectively. Proteins were grouped and colored by their related functional pathways. Drugs were grouped and colored by their targeted genes or pathways.

(B) Comparison of node connectivity between perturbed and neutral proteins in the protein interaction network. The p value was computed based on a paired Wilcoxon test. The middle line in the box is the median, the bottom and top of the box are the first and third quartiles, and the whiskers extend to 1.5 IQR of the lower and upper quartiles, respectively.
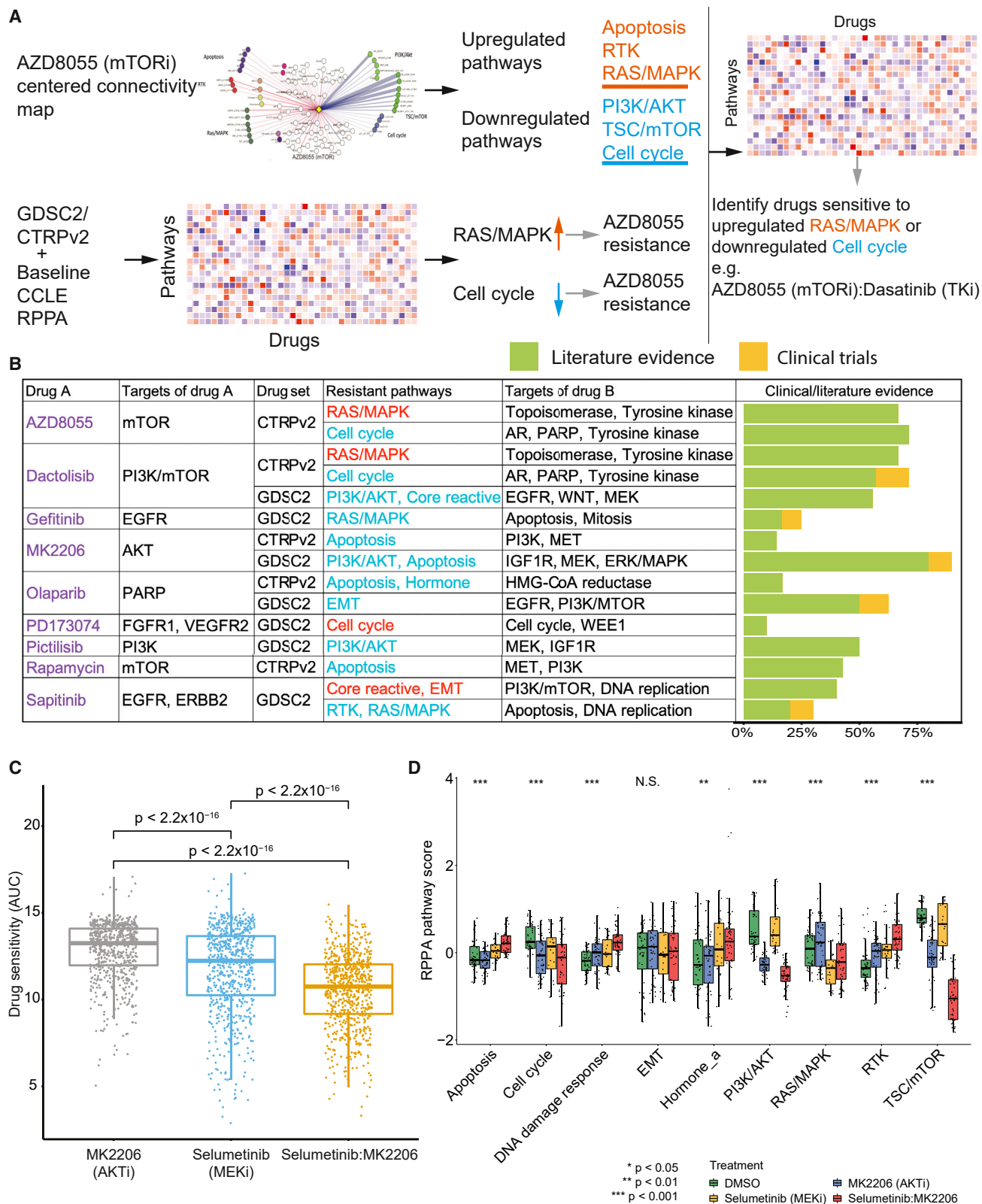
See also Figure S4 and S5.

Figure 7. Prediction of Drug Combinations Based on Connectivity Maps

(A) The workflow of drug combination prediction.

(B) Summary of predicted drug combinations and the corresponding literature/clinical evidence.

Cancer Cell
Article

CellPress

https://bioinformatics.mdanderson.org/public-software/cppa. The data portal provides four interactive modules: "Data Summary," "My Protein," "Connectivity Map," and "Analysis" (Figure S6). The Data Summary module provides detailed information about each sample (including cell line, compound, dose, time, and culture conditions). The datasets can be easily downloaded through a tree-view interface. The My Protein module provides annotation of RPPA protein markers, including the corresponding genes and antibody information. The Connectivity Map provides an interactive approach to exploring the map, through which protein-drug and drug-drug connectivity can be examined through different visual and layout styles. The Analysis module provides three common analyses through which users can explore protein responses associated with a drug/compound, including protein response (Δp) rank, volcano plots for the correlations between protein responses and drug sensitivity, and boxplots for differential protein responses between sensitive and resistant cell lines. Collectively, this data portal enables researchers to explore, analyze, and visualize RPPA-based protein response data intuitively and efficiently.

DISCUSSION

Here we present a large collection of protein responses (including total and post-translationally modified proteins) to drug treatments (>12,000 treated samples) using RPPA, which is several magnitudes larger than previously published studies. We validated the quality of our datasets in several ways. First, we demonstrated the high reproducibility of replicate samples using the same platform. Second, we established a high consistency between RPPA measurements and independent platforms such as mass spectrometry and L1000. Third, the quality of our dataset is also supported by the meaningful patterns observed in an MCF7-focused analysis and a systematic protein-drug connectivity map, such as the clustering of similar drugs and higher node connectivity of perturbed proteins annotated in the STRING protein interaction database. Our study represents a unique, high-quality compendium of protein responses of cancer cell lines to a diversity of compound perturbations available for use by a broad community.

The utility of our protein response dataset is severalfold. First, our dataset provides a basis for understanding cause-effect relationships that is complementary to correlation analyses and associations that can be obtained from patient cohorts. Based on these data, it will be possible to develop quantitative predictive models of how signaling networks function in intact cellular systems. Second, we show that while there is information content in biomarkers at baseline, the information content is markedly increased when baseline and response signals are combined. This is predicted by systems biology and engineering precepts, wherein perturbed systems contain more information than static analysis. Biomarkers designed to select treatment using baseline data frequently have a limited power to predict benefit, and our results suggest that adaptive protein responses after initial treatment could be highly informative in terms of treatment response and clinical benefit. Further clinical investigations are warranted to assess the potential benefit gains using such a strategy. Third, since protein responses reflect how cancer cells critically rewire their signaling pathways to survive and adapt to the stress of specific drug treatments, these protein signals provide a strong basis for the rational design of combination therapies, as we have demonstrated previously (Iavarone et al., 2019; Krepler et al., 2016, 2017; Kwong et al., 2015; Molinelli et al., 2013; Muranen et al., 2012; Sun et al., 2017, 2018).

We recognize some limitations of this study. First, compared with mass spectrometry-based protein-level or mRNA-level readout assays, the number of protein markers that can be effectively monitored by the RPPA technology is much smaller. However, the increased sensitivity (particularly for some key proteins and phosphoproteins), and cost considerations, makes RPPA a practical platform for generating such a large resource. In capturing protein responses, RPPA and mass spectrometry are complementary because of their different scopes and focuses. Second, although many perturbed protein response profiles were generated, some cell lines and drug treatments (including different dosages) are still sparsely sampled. As a result, we could assess the ability of perturbed RPPA data to predict drug sensitivity based on only a small number of drugs. Further efforts are required to obtain more comprehensive sets. However, machine learning approaches may have the potential to fill some of these gaps. Finally, as with other high-throughput technologies, there can be technical measurement errors for individual samples, and interesting observations from our study should be followed by further in-depth investigations.

We have provided an interactive, user-friendly data portal through which biomedical researchers can explore, visualize, and intuitively analyze these data. With this bioinformatics tool, we expect an effective translation of the large-scale perturbed protein data into biological knowledge and clinical utility. Together with recent efforts that have systematically characterized phenotypic and molecular responses to drug treatment, our study provides a rich resource for the research community to investigate the behaviors of cancer cells and the dependencies of treatment responses.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  ○ Lead Contact
  ○ Material Availability

(C) Boxplots showing CTRPv2 drug sensitivities of cell lines treated with MK2206 (AKT inhibitor), selumetinib (MEK inhibitor), and their combination (p values were calculated by Wilcoxon tests; n = 706 cell lines for each treatment).

(D) Protein pathway scores for samples treated with DMSO (n = 48 samples) or MK2206 (n = 48 samples), selumetinib (n = 24 samples), or their combination (p values were calculated by ANOVA; n = 48 samples). The middle line in the box is the median, the bottom and top of the box are the first and third quartiles, and the whiskers extend to 1.5 IQR of the lower and upper quartiles, respectively.

See also Table S2.

AUTHOR CONTRIBUTIONS

G.B.M. and H.L. conceived of the project. W.Z., J.L., M.M.C., Y. Luo, J.Z., A.S., S.V.K., A.K., R.A., G.B.M., and H.L. contributed to the data analysis. N.K.N., K.J.C., C.T.B., Y. Lawrence, N.T.P., M.A.D., M.H., T.M., I.K.Z., E.V.E., P.T.S., D.J.S., J.S.B., J.W.G., Y. Lu, and G.B.M. contributed to the experiments. M.M.C. and J.L. implemented the web portal. W.Z., J.L., M.M.C., Y. Luo, G.B.M., and H.L. wrote the manuscript, with input from other authors. H.L. supervised the whole project.

DECLARATION OF INTERESTS

REFERENCES

Akbani, R., Ng, P.K., Werner, H.M., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J.Y., Yoshihara, K., Li, J., et al. (2014). A pan-cancer proteomic perspective on the Cancer Genome Atlas. Nat. Commun. *5*, 3887.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *492*, 290.

Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E.V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell *154*, 1151–1161.

Cragg, M.S., Jansen, E.S., Cook, M., Harris, C., Strasser, A., and Scott, C.L. (2008). Treatment of B-RAF mutant human tumor cells with a MEK inhibitor requires Bim and is enhanced by a BH3 mimetic. J. Clin. Invest. *118*, 3651–3659.

Daemen, A., Griffith, O.L., Heiser, L.M., Wang, N.J., Enache, O.M., Sanborn, Z., Pepin, F., Durinck, S., Korkola, J.E., Griffith, M., et al. (2013). Modeling precision treatment of breast cancer. Genome Biol. *14*, R110.

Fang, Y., McGrail, D.J., Sun, C., Labrie, M., Chen, X., Zhang, D., Ju, Z., Vellano, C.P., Lu, Y., Li, Y., et al. (2019). Sequential therapy with PARP and WEE1 inhibitors minimizes toxicity while maintaining efficacy. Cancer Cell *35*, 851–867.e7.

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570–575.

Ghandi, M., Huang, F.W., Jane-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the cancer cell line Encyclopedia. Nature *569*, 503.

Hennessy, B.T., Lu, Y., Gonzalez-Angulo, A.M., Carey, M.S., Myhre, S., Ju, Z., Davies, M.A., Liu, W., Coombes, K., Meric-Bernstam, F., et al. (2010). A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. Clin. Proteomics *6*, 129–151.

Hutter, C., and Zenklusen, J.C. (2018). The cancer genome Atlas: creating lasting value beyond its data. Cell *173*, 283–285.

Ju, Z., Liu, W., Roebuck, P., Siwak, D., Zhang, N., Lu, Y., Davies, M., Akbani, R., Weinstein, J., Mills, G., et al. (2015). Development of a robust classifier for quality control of reverse-phase protein arrays. Bioinformatics *31*, 912–918.

Iavarone, C., Zervantonakis, I.K., Selfors, L.M., Palakurthi, S., Liu, J.F., Drapkin, R., Matulonis, U.A., Hallberg, D., Velculescu, V.E., Leverson, J.D., et al. (2019). Combined MEK and BCL-2/XL inhibition is effective in high-grade serous ovarian cancer patient-derived xenograft models and BIM levels are predictive of responsiveness. Mol. Cancer Ther. *18*, 642–655.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Goncalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A landscape of pharmacogenomic interactions in cancer. Cell *166*, 740–754.

Jiang, Y.Z., Ma, D., Suo, C., Shi, J., Xue, M., Hu, X., Xiao, Y., Yu, K.D., Liu, Y.R., Yu, Y., et al. (2019). Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. Cancer Cell *35*, 428–440.e5.

Korkut, A., Wang, W., Demir, E., Aksoy, B.A., Jing, X., Molinelli, E.J., Babur, O., Bemis, D.L., Onur Sumer, S., Solit, D.B., et al. (2015). Perturbation biology nominates upstream-downstream drug combinations in RAF inhibitor resistant melanoma cells. eLife *4*, e04640.

Krepler, C., Sproesser, K., Brafford, P., Beqiri, M., Garman, B., Xiao, M., Shannan, B., Watters, A., Perego, M., Zhang, G., et al. (2017). A comprehensive patient-derived xenograft collection representing the heterogeneity of melanoma. Cell Rep. *21*, 1953–1967.

Krepler, C., Xiao, M., Sproesser, K., Brafford, P.A., Shannan, B., Beqiri, M., Liu, Q., Xu, W., Garman, B., Nathanson, K.L., et al. (2016). Personalized preclinical trials in BRAF inhibitor-resistant patient-derived xenograft models identify second-line combination therapies. Clin. Cancer Res. *22*, 1592–1602.

Cancer Cell
Article

CellPress

Kwong, L.N., Boland, G.M., Frederick, D.T., Helms, T.L., Akid, A.T., Miller, J.P., Jiang, S., Cooper, Z.A., Song, X., Seth, S., et al. (2015). Co-clinical assessment identifies patterns of BRAF inhibitor resistance in melanoma. J. Clin. Invest. *125*, 1459–1470.

Lee, H., Saini, N., Parris, A.B., Zhao, M., and Yang, X. (2017). Ganetespib induces G2/M cell cycle arrest and apoptosis in gastric cancer cells through targeting of receptor tyrosine kinase signaling. Int. J. Oncol. *51*, 967–974.

Li, J., Zhao, W., Akbani, R., Liu, W.B., Ju, Z.L., Ling, S.Y., Vellano, C.P., Roebuck, P., Yu, Q.H., Eterovic, A.K., et al. (2017). Characterization of human cancer cell lines by reverse-phase protein arrays. Cancer Cell *31*, 225–239.

McDonald, E.R., 3rd, de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project drive: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. Cell *170*, 577–592 e510.

Mertins, P., Yang, F., Liu, T., Mani, D.R., Petyuk, V.A., Gillette, M.A., Clauser, K.R., Qiao, J.W., Gritsenko, M.A., Moore, R.J., et al. (2014). Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. Mol. Cell Proteomics *13*, 1690–1704.

Mirzoeva, O.K., Das, D., Heiser, L.M., Bhattacharya, S., Siwak, D., Gendelman, R., Bayani, N., Wang, N.J., Neve, R.M., Guan, Y., et al. (2009). Basal subtype and MAPK/ERK kinase (MEK)-phosphoinositide 3-kinase feedback signaling determine susceptibility of breast cancer cells to MEK inhibition. Cancer Res. *69*, 565–572.

Molina-Arcas, M., Hancock, D.C., Sheridan, C., Kumar, M.S., and Downward, J. (2013). Coordinate direct input of both KRAS and IGF1 receptor to activation of PI3 kinase in KRAS-mutant lung cancer. Cancer Discov. *3*, 548–563.

Molinelli, E.J., Korkut, A., Wang, W., Miller, M.L., Gauthier, N.P., Jing, X., Kaushik, P., He, Q., Mills, G., Solit, D.B., et al. (2013). Perturbation biology: inferring signaling networks in cellular systems. PLoS Comput. Biol. *9*, e1003290.

Muranen, T., Selfors, L.M., Worster, D.T., Iwanicki, M.P., Song, L., Morales, F.C., Gao, S., Mills, G.B., and Brugge, J.S. (2012). Inhibition of PI3K/mTOR leads to adaptive resistance in matrix-attached cancer cells. Cancer Cell *21*, 227–239.

Ng, P.K., Li, J., Jeong, K.J., Shao, S., Chen, H., Tsang, Y.H., Sengupta, S., Wang, Z., Bhavana, V.H., Tran, R., et al. (2018). Systematic functional annotation of somatic mutations in cancer. Cancer Cell *33*, 450–462.e10.

Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W.C., Waltham, M., Kouros-Mehr, H., Bussey, K.J., Lee, J.K., Espina, V., et al. (2003). Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. Proc. Natl. Acad. Sci. U S A *100*, 14229–14234.

Nusinow, D.P., Szpyt, J., Ghandi, M., Rose, C.M., McDonald, E.R., 3rd, Kalocsay, M., Jane-Valbuena, J., Gelfand, E., Schweppe, D.K., Jedrychowski, M., et al. (2020). Quantitative proteomics of the cancer cell line Encyclopedia. Cell *180*, 387–402 e316.

Otasek, D., Morris, J.H., Boucas, J., Pico, A.R., and Demchak, B. (2019). Cytoscape Automation: empowering workflow-based network analysis. Genome Biol. *20*, 185.

Pohl, G., Ho, C.L., Kurman, R.J., Bristow, R., Wang, T.L., and Shih Ie, M. (2005). Inactivation of the mitogen-activated protein kinase pathway as a potential target-based therapy in ovarian serous tumors with KRAS or BRAF mutations. Cancer Res. *65*, 1994–2000.

Russo, M., Crisafulli, G., Sogari, A., Reilly, N.M., Arena, S., Lamba, S., Bartolini, A., Amodio, V., Magri, A., Novara, L., et al. (2019). Adaptive mutability of colorectal cancers in response to targeted therapies. Science *366*, 1473–1480.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shao, D.D., Xue, W., Krall, E.B., Bhutkar, A., Piccioni, F., Wang, X., Schinzel, A.C., Sood, S., Rosenbluh, J., Kim, J.W., et al. (2014). KRAS and YAP1 converge to regulate EMT and tumor survival. Cell *158*, 171–184.

Shen, J., Peng, Y., Wei, L., Zhang, W., Yang, L., Lan, L., Kapoor, P., Ju, Z., Mo, Q., Shih Ie, M., et al. (2015). ARID1A deficiency impairs the DNA damage checkpoint and sensitizes cells to PARP inhibitors. Cancer Discov. *5*, 752–767.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell *171*, 1437–1452.e17.

Sun, C., Fang, Y., Yin, J., Chen, J., Ju, Z., Zhang, D., Chen, X., Vellano, C.P., Jeong, K.J., Ng, P.K., et al. (2017). Rational combination therapy with PARP and MEK inhibitors capitalizes on therapeutic liabilities in RAS mutant cancers. Sci. Transl. Med. *9*, eaal5148.

Sun, C., Yin, J., Fang, Y., Chen, J., Jeong, K.J., Chen, X., Vellano, C.P., Ju, Z., Zhao, W., Zhang, D., et al. (2018). BRD4 inhibition is synthetic lethal with PARP inhibitors through the induction of homologous recombination deficiency. Cancer Cell *33*, 401–416.e8.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Tibes, R., Qiu, Y., Lu, Y., Hennessy, B., Andreeff, M., Mills, G.B., and Kornblau, S.M. (2006). Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Mol. Cancer Ther. *5*, 2512–2521.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency map. Cell *170*, 564.

Wang, Y., Yuan, J.L., Zhang, Y.T., Ma, J.J., Xu, P., Shi, C.H., Zhang, W., Li, Y.M., Fu, Q., Zhu, G.F., et al. (2013). Inhibition of both EGFR and IGF1R sensitized prostate cancer cells to radiation by synergistic suppression of DNA homologous recombination repair. PLoS One *8*, e68784.

Westin, S.N., Sill, M.W., Coleman, R.L., Waggoner, S., Moore, K.N., Mathews, C.A., Martin, L.P., Modesitt, S.C., Lee, S., Ju, Z., et al. (2019). Safety lead-in of the MEK inhibitor trametinib in combination with GSK2141795, an AKT inhibitor, in patients with recurrent endometrial cancer: an NRG Oncology/GOG study. Gynecol. Oncol. *155*, 420–428.

Wise, H.C., and Solit, D.B. (2019). Precision oncology: three small steps forward. Cancer Cell *35*, 825–826.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Perturbed RPPA data	CPPA data portal	https://bioinformatics.mdanderson.org/public-software/cppa
Quantitative mass spectrometry data of CCLE	Nusinow et al., 2020	Table S2
Connectivity map L1000 phase 1 gene expression assay data	Gene Expression Omnibus; Subramanian et al., 2017	http://www.ncbi.nlm.nih.gov/geo/; GEO:GSE92742
CTRPv2 drug sensitivity data	Cancer Therapeutics Response Portal v2	https://portals.broadinstitute.org/ctrp/
GDSC2 drug sensitivity data	Genomics of Drug Sensitivity in Cancer	https://www.cancerrxgene.org
Daemen et al. drug sensitivity data	Daemen et al., 2013	Table S1
Protein-protein interaction network data	STRING; Szklarczyk et al., 2019	https://string-db.org
Experimental Models		
Cell lines	This study	Table S1
Software and Algorithms		
SuperCurve	Ju et al., 2015	https://bioinformatics.mdanderson.org/public-software/supercurve/
Cytoscape	Otasek et al., 2019; Shannon et al., 2003	https://cytoscape.org
DataTables		https://datatables.net
HighCharts		https://www.highcharts.com

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Han Liang (hliang1@mdanderson.org).

Material Availability
This study did not generate new unique reagents.

Data and Code Availability
The RPPA data generated in this study can be found at the CPPA data portal: https://bioinformatics.mdanderson.org/public-software/cppa. The quantitative mass spectrometry data of CCLE were downloaded from Table S2 of Nusinow et al. (2020). The L1000 gene expression data were downloaded from Gene Expression Omnibus series GSE92742. The CTRPv2 drug sensitivity data were downloaded from CTRPv2 (https://portals.broadinstitute.org/ctrp/). The GDSC2 drug sensitivity data were downloaded from GDSC Release 6.1 (https://www.cancerrxgene.org). The drug sensitivity data of lapatinib were downloaded from Table S1 of Daemen et al. (2013). The protein-protein interaction network data were downloaded from STRING (https://string-db.org).

All software supporting the analysis in this study can be found in public repositories. SuperCurve is available at https://bioinformatics.mdanderson.org/public-software/supercurve/; Cytoscape is available at https://cytoscape.org; DataTables is available at https://datatables.net; and HighCharts is available at https://www.highcharts.com.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines
We collected cancer cell lines through the MD Anderson Cancer Center (MDACC) CCSG-supported Cell Line Characterization Core Facility (Houston, TX, USA) and from several outside collaborations (see Table S1 for details). All cell lines prepared at MDACC were confirmed by short tandem repeat (STR) analysis in the core per institutional policy, and the outside collaborators also routinely confirmed cell lines by STR analysis.

Cancer Cell
Article

CellPress

METHOD DETAILS

RPPA Experiments

RPPA experiments were performed at the RPPA core facility at MDACC. Cell line samples were prepared, and antibodies were validated by comparison with immunoblotting, as previously described (Hennessy et al., 2010; Li et al., 2017). Briefly, lysates were manually serial-diluted in 5 two-fold dilutions with lysis buffer and printed on nitrocellulose-coated slides using an Aushon Biosystems 2470 arrayer. Slides were probed with validated primary antibodies, followed by detection with appropriate biotinylated secondary antibodies (Biotin conjugated-Goat anti-Rabbit IgG, Goat anti-Mouse IgG, or Rabbit anti-Goat IgG, from Vector Lab). The signal obtained was amplified using a Dako Cytomation–catalyzed system of avidin-biotin-peroxidase (from Agilent) binding to the secondary antibody and catalyzing a tyramide-biotin conjugation to form insoluble biotinylated phenols. Stained RPPA slides were first quantified using ArrayPro (Media Cybernetics) to generate signal intensities (level 1). Then, the SuperCurve software processed spots from all horizontal samples on the slide to determine the relative protein level for each sample (level 2). Standard parameters were used in this step, including nonparametric curve fitting via monotone increasing B-spline, and use of spatial adjustment for regional correction. Finally, protein measurements were corrected for loading using median polish (level 3). RPPA slide quality was assessed by a quality control classifier in the R package "SuperCurve" (Ju et al., 2015), which was trained in curated RPPA data sets using a generalized linear model and logistic function. Only slides with a quality score above 0.8 (range: 0-1) were retained for further analysis. In total, we generated RPPA data from 15,842 samples, including 12,183 treated cell line samples and 3,659 baseline samples (e.g., treated with DSMO).

Quality Assessment within the RPPA Platform

Internal quality assessment of RPPA data was first performed in baseline samples. We examined the consensus of the RPPA signals in baseline samples of the same cell line and excluded the samples for which average correlation coefficients with other baseline samples were < 0.5. Next, we performed a quality assessment of post-treatment samples. For each pair of post-treatment replicate samples, we generated protein response (Δp) profiles using the corresponding baseline samples that passed the consensus test and compared the Δp profile of the replicate samples (n = 2,753 pairs). Samples pairs with low Pearson correlation were removed from the analysis. In total, 11,884 post-treatment samples and 3,608 baseline samples passed the internal quality assessment. For each treated sample, we calculated Δp for each protein by deducting its protein level in the corresponding control sample. When replicates (technical or biological) were available, we used the average level across replicates. By combining replicates, we generated 1,916 unique baseline protein profiles and 7,941 unique protein response profiles of post-treatment samples. To generate the time-independent Δp profile for a specific treatment in a cell line, we merged the protein responses across different time points by taking the average.

Quality Assessment Using Independent Platforms

We obtained normalized CCLE protein expression data, generated through quantitative mass spectrometry, from a recent study (Nusinow et al., 2020). For each of the baseline RPPA sets that had an overlap with the proteomics dataset of $\geq$ 8 cell lines, we calculated the Spearman correlation coefficient between the normalized expressions generated by these two assays across the common cell lines, and in total, 169 unique proteins were included in the analysis. A background distribution of correlations was built by computing the Spearman correlation coefficient between each pair of different proteins across the same cell lines in the same sets.

We downloaded the level-5 data of L1000 phase 1 from the GEO database (GSE92742). For a fair comparison, we collected data from the same cell lines perturbed by the same compound. In total, 46 "perturbation-cell-line" IDs (60 samples) and 347 genes/proteins (total proteins but phosphoprotein if a total protein was not available) commonly shared by the two platforms were used in the subsequent analyses. For a perturbation-cell-line ID with multiple concentration and/or time points, we adopted the median value across all conditions as the representative response score. For each platform, we first converted the continuous response to a categorical response: upregulated, downregulated, or neutral. Random events were defined by the global median ± 35% quantile, calculated from the full matrix. Next, we excluded the random events and computed Goodman-Kruskal's gamma (γ) to estimate sample associations across genes. We evaluated the concordance between RPPA and L1000 platforms through two analyses. (i) Protein-mRNA response associations: for each sample, a γ association between the two platforms was computed across genes/proteins when at least 12 genes showed up/downregulation. To generate the background distribution, we randomly shuffled protein labels and computed the response associations between the shuffled proteins and mRNAs (the seed used for randomization is "1234"). Then, a paired Student's t-test was used to evaluate the statistical significance of the group difference between the real and matched randomly shuffled responses. (ii) Sample-sample associations: in our RPPA dataset, a perturbation-cell-line ID might have replicate samples. Here, we only retained the one with the best protein-mRNA response association from the previous analysis. Next, for samples that showed up/downregulation of > 3 genes, γ associations for every two such samples were computed within each platform (within the same batch). Then, Pearson's correlation between the significant γ associations (FDR < 0.01 for each platform) was used to evaluate the consistency between the protein and mRNA responses.

Analysis of Predictive Protein Markers of Drug Sensitivity

We collected drug sensitivity data from public resources including GDSC2, CTRPv2, and Daemen et al. (2013). For MCF7 analysis, the AUC Z-scores ($z = \frac{x-\mu}{\sigma}$) were calculated based on AUCs for each drug across all screened cell lines. We calculated the median

values of perturbed pathway scores (Δ pathway scores). The AUC difference between a specific drug group and other drugs was assessed by unpaired Student's t-test, and the significant ones were identified with the cut-off of FDR < 0.01.

To identify the differential protein markers of drug sensitivity and resistance, cell lines were classified as sensitive or resistant to a specific drug based on the consensus call of CTRPv2, GDSC2, and in-house datasets. Baseline levels (p0) and protein response levels (Δp) with a significant difference between sensitive and resistant cell lines were identified by unpaired Student's t-test. Pathway-level scores (Li et al., 2017) were similarly analyzed.

To assess the predictive power of perturbed RPPA data across cell lines, perturbation data of the same cell line treated with the same compound at different dosages or time points were averaged using mean values. For univariate analysis, seven compounds had both RPPA perturbation profiling and drug sensitivity data in ≥5 cell lines based on GDSC2 data (see Figure 5A for the sample size information of each drug). For each compound, the baseline levels (p0) and protein response levels (Δp) of each antibody were tested for associations with drug sensitivity (IC50 or AUC score) in univariate linear models. The joint markers (Δp|p0) were defined as the predictions of linear regression models, including both baseline and protein response for specific antibodies. Predictive markers were selected by Pearson correlation at a significance level of p = 0.05. We then developed a multivariate model to predict drug sensitivity using the leave-one-out approach and three sets of protein profiles: baseline level (p0), post-treatment level (p1), and joint profile (p0 and p1). For the machine-learning-based analysis, to provide sufficient statistical power for feature selection and cross validation, we examined compounds with both RPPA perturbation data and drug sensitivity data in ≥10 cell lines. Only lapatinib (drug data from Daemen et al., n = 13 cell lines.) and GSK690693 (drug data from GDSC2, n = 10 cell lines) were able to generate statistically meaningful models for assessment. In each round of cross-validation, one cell line was left out as the validation set. In the training set, candidate markers were first selected with a univariate correlation test at a significance level of p = 0.1, with a maximum size of 20. The model was then built by Elastic Net based on candidate markers in the training set and was applied to the validation set. In time-series experiments, we performed cross-validation for each time point using the three sets of protein profiles mentioned above. The performance was evaluated using the mean squared error (MSE) of prediction in the validation set.

Construction of a Drug-Protein Connectivity Map

The association of each drug-protein pair was assessed by testing the difference of protein expression between baseline (p0) and post-treatment level (p1) based on the paired t-test across cell lines. For each drug-drug pair, we used Goodman-Kruskal's γ to calculate the associations, as described in the comparison between RPPA-based protein response and L1000-based mRNA response data. The significantly correlated drug-protein and drug-drug pairs (FDR < 0.1) were used to construct a global drug-protein connectivity map. In the connectivity map, proteins were grouped and colored by their related protein functional pathways, and drugs were grouped and colored by their targeted genes or pathways. For each drug, the network densities were calculated for the two subsets of RPPA proteins: (i) proteins significantly differentially expressed between p0 and p1 (perturbed RPPA proteins), and (ii) other RPPA proteins (neutral proteins). The network density D of a protein subset with size N was defined as a ratio of the number of protein-protein interactions (E) to the number of all possible protein pairs ($E_{max} = \binom{N}{2}$), i.e., $D = E / E_{max}$. Protein-protein interaction information was obtained from the STRING database (Szklarczyk et al., 2019). A paired Wilcoxon test was performed to assess the difference of the network densities between the perturbed and other proteins of all the drugs. For Figures S4 and S5, the examples of drug-centered connectivity maps were generated separately with colored edges (red: upregulated in post-treatment; blue: downregulated in post-treatment). The edge widths were proportional to the differential expression between baseline (p0) and post-treatment levels (p1). All network views were generated by the Rcy3 library and Cytoscape (Otasek et al., 2019; Shannon et al., 2003).

Prediction and Validation of Drug Combinations

Each drug-centered connectivity map was extracted from the full connectivity map (Figure 6). The perturbed protein markers were grouped into their associated protein pathways, e.g., apoptosis and RAS/MAPK. The baseline and perturbed pathway scores were calculated based on the weighted average of baseline (obtained from CCLE) and perturbed protein levels, respectively. The perturbed pathway scores were used to infer the pathway responses to each drug. The baseline pathway scores were used to perform correlation analysis between pathways and drug sensitivity data, which were further used to infer drug resistance pathways. Focusing on the resistance pathways, drugs were predicted as combination candidates when we found (i) positive correlations (FDR < 0.1) between AUC and downregulated pathways or (ii) negative correlations (FDR < 0.1) between AUC and upregulated pathways. The top 10 combinations for each drug were reported based on the rank of their correlation coefficients. The literature and clinical evidence were obtained from PubMed (https://pubmed.ncbi.nlm.nih.gov) and clinical trial database (https://clinicaltrials.gov). The sensitivity data (AUC) of MK-2206 (AKTi), selumetinib (MEKi), and their combination were obtained from CTRPv2.

Data Portal Development

All RPPA and drug sensitivity data accompanying the pre-calculated analytic results were stored in a CouchDB database. We generated all the analytic results in R. We implemented a user-friendly and interactive web interface in JavaScript. Specifically, tabular

Cancer Cell
Article

CellPress

results were generated by DataTables, box and scatter plots were generated by HighCharts, and interactive network views were implemented by Cytoscape.js library.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using R (version 3.6.2). To assess the correlation between two continuous variables, Pearson or Spearman rank correlation test was used; to compare two groups within the RPPA platform, Wilcoxon test or Student's t-test was used; to compare multiple groups with the RPPA platform, analysis of variance (ANOVA) was used; to compare the associations between RPPA and L1000 platforms, Goodman-Kruskal's gamma (γ) test was used; for the machine-learning models for drug sensitivity prediction, elastic net regression was used. Detailed descriptions of statistical tests were provided in the Method Details section and in the respective figure legends.