Elucidating Compound Mechanism of Action and Polypharmacology with a Large-scale Perturbational Profile Compendium.

Lucas ZhongMing Hu[#1], Tal Hirschhorn[#2], Eugene Douglass[1,3], Diana Murray[1], Stephen J. Trudeau[1], Mikko Turunen[1], Sergey Pampou[1,4,5], Adina Grunn[1], Ron Realubit[1,4,5] , Albert A. Antolin[6], Alexander L.E. Wang[1], Hai Li[1,4,5], Prem Subramaniam[1], Prabhjot S Mundi[1], Charles Karan[1,4,5], Mariano Alvarez[7], Barry Honig[1,8,9,10], Brent R. Stockwell[2,4,11,12,13,16], Andrea Califano[1,4,8,9,14,15,16,17]

[1] Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA 10032.

[2] Department of Biological Sciences, Columbia University, New York, NY USA 10027.

[3] Pharmaceutical and Biomedical Sciences, University of Georgia, 250 W. Green Street, Athens, GA 30602.

[4] Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, 1130 Saint Nicholas Ave, New York, NY USA 10032.

[5] JP Sulzberger Columbia Genome Center, Columbia University Medical Center, New York, NY 10032, USA.

[6] Centre for Cancer Discovery, The Division of Cancer therapeutics, The Institute of Cancer Research, London, UK

[7] DarwinHealth, Inc. New York, NY USA 10018.

[8] Department of Medicine, Columbia University Irving Medical Center, 630 W 168th Street, New York, NY USA 10032.

[9] Department of Biochemistry & Molecular Biophysics, Columbia University Irving Medical Center, 701 W 168th Street, New York, NY USA 10032.

[10] Zuckerman Mind Brain and Behavior Institute, Columbia University, New York, NY, USA 10027

[11] Department of Chemistry, Columbia University, New York, NY USA 10027.

[12] Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY 10032

[13] Columbia University Digestive and Liver Disease Research Center, New York, NY 10032, USA

[14] Department of Biomedical Informatics, Columbia University Irving Medical Center, 622 W 168th Street, New York, NY USA 10032.

[15] Chan Zuckerberg Biohub New York, New York, NY, USA.

[16] Senior author.

[17] Lead contact.

[#] Contributed equally

Corresponding Author: Brent R. Stockwell (bs2198@columbia.edu), Andrea Califano (ac2248@cumc.columbia.edu)

Authors Disclosures: A.C. is founder, equity holder, consultant, and director of DarwinHealth Inc., a company that has licensed some of the algorithms used in this manuscript from Columbia University. M.J.A. is Chief Scientific Officer and equity holder at DarwinHealth, Inc., a company that has licensed some of the algorithms used in this manuscript from Columbia University. C.K. is a DarwinHealth consultant and shareholder. B.R.S. is an inventor on patents and patent applications involving ferroptosis; co-founded and serves as a consultant to ProJenX, Inc. and Exarta Therapeutics; holds equity in Sonata Therapeutics; serves as a consultant to Weatherwax Biotechnologies Corporation and Akin Gump Strauss Hauer & Feld LLP.

Abstract (146 words)

Drug Mechanism of Action (MoA) is generally represented as a small repertoire of tissue-independent, high-affinity binding targets. Yet, drug activity and polypharmacology are associated with a broad range of off-target and tissue-specific effector proteins. To address this challenge, we have generated a compendium of drug perturbation profiles for >700 oncology drugs in cell lines representing high-fidelity models of 23 aggressive tumor subtypes and developed an integrative computational framework for the proteome-wide assessment of drug-mediated, tissue-specific differential protein activity. This allows systematic, mechanism-based elucidation of tissue context-specific compound MoA and polypharmacology, including discovery and experimental validation of post-translationally-mediated inhibition of undruggable oncoproteins—such as MYC, CTNNB1, CENPF and UHRF1—as well as previously unknown inhibitors of MAPK, PI3K, and folate pathways, among others, as further suggested by protein structure analysis. This framework can help characterize the MoA of discovery compound libraries, thus supporting more systematic and quantitative approaches to precision oncology.

Introduction

Targeted therapy and precision medicine are predicated on the ability to accurately elucidate the mechanism of action (MoA) of small molecule compounds and to match it to disease-specific dependencies. However, the binding affinity and molecular structure of targets revealed by standard biochemical assays (*on-target*) are typically well yet not comprehensively characterized. Moreover, drug efficacy and toxicity is being increasingly linked to their complex polypharmacology, as mediated by additional low-affinity binding proteins (*off-target*) and by tissue-specific indirect effector proteins (*secondary target*) that are poorly characterized or even undetectable by traditional assays[1]. Indeed, MoA studies—including those based on recent advances in pull-down technology[2]—rely on *in vitro* affinity binding assays designed to investigate only specific protein classes, such as kinases[3], thus failing to reveal a compound's full range of high- and low-affinity binding targets and, perhaps more importantly, its tissue-specific effectors. Elucidating the proteome-wide mechanisms that determine drug's clinical efficacy and toxicity—as well as those that are pharmacologically irrelevant—has thus emerged as a critically-relevant, yet still elusive endeavor[3,4] in precision medicine, including in oncology-related applications[1,5,6].

To address this challenge, we generated a large-scale repertoire of genome-wide, tissue context-specific perturbational profiles and analyzed it using a novel MoA characterization framework based on the extensively validated VIPER algorithm, see[7-12] for instance. The resulting Pan-Cancer Assessment of Compound Effectors and Activity (PanACEA) compendium comprises >18,000 RNA-seq profiles and ~50 million drug-mediated, differential protein activity measurements, thus representing the largest genome-wide resource of its kind. Attesting to the quality and value of the PanACEA perturbational profiles, we have already shown that, on a tumor-specific basis, their use was instrumental in identifying drugs targeting patient specific

dependencies in multiple pre-clinical studies, across a wide range of aggressive human malignancies[9,10], as well as in a clinical context[13].

Perturbational profiles, characterizing the response of a tissue to one or more small molecules, have been previously leveraged to study MoA. These include both proteomic assays[14-17], for instance to elucidate signal-transduction protein targets[18], and transcriptional profiles focusing on changes in transcriptional cell state following drug perturbation[5,6,19]. Among these, proteomic assays are still expensive, labor intensive, and quite sparse. As such, they have been restricted to study only a handful of compounds of interest in specific cellular contexts, rather than large compound libraries[14,15], or to focus only on individual cell lines[16]. In addition, these assays report on protein abundance, which is an indirect measure of a protein's enzymatic or regulatory activity.

Transcriptional profiles are cheaper and more amenable to automation. However, their ability to directly inform on drug-mediated changes in protein activity or abundance is limited. In addition, their use has been limited either to small, landmark gene sets[19] or to a handful of tissues[6,20].

To transform genome-wide RNA-seq profiles from drug perturbations into accurate protein activity profiles that optimally inform on drug MoA, we leveraged the VIPER algorithm[7] . Akin to a highly multiplexed gene reporter assay, VIPER measures a protein's differential activity—in this case comparing a drug vs. vehicle control-treated sample—based on the differential expression of its tissue-specific targets (*regulon*). VIPER protein activity measurements compare favorably to antibody-based measurements[12,21] and can be used to effectively monitor >6,500 regulatory and signaling proteins. The experimental component of this study is thus based on the high-throughput generation of comprehensive RNA-seq profiles, using the PLATE-Seq technology[22], representing the response of multiple tissues to a comprehensive repertoire of clinically relevant drugs. These are then VIPER-transformed into proteomic profiles,

representing the differential activity of transcription factor/co-factor (TF/co-TF) and signaling proteins in drug vs. vehicle-control treated samples.

While large-scale perturbational profiles libraries, such as LINCS[19], have become available, they present limitations to the protein-level elucidation of drug MoA, including their suitability to VIPER-based analyses. Specifically, (a) they only report the effect on a small set (~1,000) of landmark genes, (b) perturbations are generally performed at isomolar concentrations (*e.g.*, 1μM, 5μM, etc.), thus potentially missing the relevant pharmacological window where the compound's MoA is optimally manifested; (c) perturbed cell lines are not tested as high-fidelity models of human disease; and (d) compound libraries comprise few clinically relevant drugs, such as FDA-approved and late-stage developmental compounds (*i.e.*, in phase 2 or 3 trials). A critical consequence is that the resulting reduced-representation profiles cannot be converted to protein activity profiles, using VIPER, because only 1/20th of each protein's targets is available, on average.

In contrast, PLATE-Seq provides a low-cost, fully-automated, genome-wide RNA-seq profiling technology[22] supporting VIPER analyses that are virtually indistinguishable from those based on high-depth TruSeq profiles[7]. In addition, the 23 cell lines initially represented in PanACEA—with additional cell lines being constantly added to this resource—were carefully selected to represent high-fidelity models of human cancer subtypes, based on conservation of Master Regulator (MR) proteins that are ultra-conserved across same-subtype tumors[23] (OncoMatch algorithm[24,25]). This is critical because we have shown that the MoA of drugs assessed in such high-fidelity *in vitro* models is faithfully recapitulated in PDX models *in vivo*[10]. Finally, as shown in[6,26], to prevent activation of stress or cell death pathways that inevitably confound the MoA analysis and to avoid using concentrations that are too low to engage relevant MoA mechanisms, perturbational profiles were generated at each compound's highest sublethal concentration, as assessed from experimental drug response curves (DRC).

PanACEA profiles are thus expected to effectively supporting *de novo*, tissue-specific drug MoA characterization, including: (a) canalization of different drug perturbations, resulting in activation of common response mechanisms, (b) proteome-wide assessment of drug MoA similarity, (c) elucidation of drug polypharmacology, and (d) discovery of novel inhibitors of undruggable proteins, including TFs/co-TFs previously identified as critical dependencies of molecularly distinct tumor subtypes[23].

Supporting these claims, we show that PanACEA analysis identified drugs that could inhibit undruggable proteins—identified as context-specific tumor dependencies—including CTTNNB1, UHRF1, MYC and CENPF, including via post-translational mechanisms, such as protein degradation and subcellular localization, that would elude expression-based analyses. We also used PanACEA to predict and experimentally validate novel polypharmacology as well as the MoA of poorly characterized compounds. For example, as predicted by the analysis, we show that (a) KW2499, originally characterized as a FLT3/ABL/Aurora kinase inhibitor, also potently inhibits the PIK3CA/AKT pathway, (b) biib021 (an HSP90 inhibitor[27]) and Az12419304 (a compound with unknown MoA) are effective MEK pathway inhibitors, and (c) Mgcd265, a multi-kinase inhibitor, plays a critical role in folate metabolism by inhibiting the DHFR. These results are further supported by protein structure-based predictions, using the PrePCI algorithm[28,29].

The proposed approach is fully generalizable and can be adapted to studying the MoA of novel chemical libraries in any physiologic or pathologic context, from neurological and neurodegenerative diseases[30] to targeting mechanisms associated cell state reprograming[31] and viral infection[32]. The PanACEA resource, including both RNA-seq and protein activity profiles, is freely available to the research community.

Results

PanACEA Database Assembly: PanACEA comprises n = 18,046 RNA-Seq profiles representing the genome-wide transcriptional response of 23 cell lines at 24h following treatment with >700 FDA-approved and late-stage investigational compounds, as well as vehicle control media, in duplicate. The profiling protocol was specifically designed to optimize elucidation of tissue-specific drug Mechanism of Action (MoA) and polypharmacology, see **Figure 1** for an illustrative workflow of the study methodology.

Critical differences between PanACEA and Connectivity Map (CMap)[20] or LINCS (L1000)[19] profiles include the following: (a) PanACEA comprises both genome-wide RNA-Seq profiles and VIPER-assessed differential protein activity profiles for ~6,500 regulatory and signaling proteins; (b) PanACEA was designed to support clinical applications by (*i*) selecting cell-lines representing high-fidelity models of rare and/or aggressive malignancy subtypes (**Figure 1A & Figure 1C**), and (*ii*) including a vast majority of clinically-relevant oncology drugs—including FDA-approved and late-stage experimental agents in Phase 2 and 3 clinical trials (**Table S1**); and (c) drug-concentration was chosen as the 48h, cell-line specific IC_{20}, as assessed from 10-point dose response curves (DRC) (**Figure 1B**). While this adds significant complexity and cost to the design, it is also critical to avoid confounding effects associated to stress/death pathway activation—accounting for up to 75% of the variance in existing drug-perturbation datasets—that confound MoA elucidation by masking critical targets, while artificially increasing drug MoA similarity[33,34].

The key advantage of this design is its optimal suitability to VIPER-based analyses aimed at evaluating drug-induced changes in the activity of all TF/co-TF and signaling proteins[7]. Notably, of the 23 PanACEA cell lines in PanACEA, only two (LoVo and BT20) are also included in LINCS[19]. This is important because PanACEA cell lines were selected to optimally represent the

biology of human tumors based on their ability to represent high-fidelity models of human tumors, based on either published databases or individual patients enrolled in the N of 1 study at Columbia, representing rare, aggressive adult/pediatric malignancies.

High-fidelity cell line model selection: To support their ability to survive *ex vivo*, cell lines accumulate genetic, epigenetic, and aneuploidy artifacts that are absent in their parental tumors. While poorly characterized and understood, these events can significantly affect the pathophysiologic fidelity of the model, including their response to drugs. As a result, selecting cell lines for perturbational analysis based on the either (a) the presence of specific driver mutations, (b) the activity of specific surface markers, (c) their *in vitro* growth kinetics, or (d) their morphology may not represent the best strategy. For instance, the commonly used strategy of selecting cell lines with optimal growth kinetics can bias their sensitivity to cell cycle-dependent chemotoxic agents, thus masking subtler yet more critical dependencies in slower-growing human tumors. In contrast, as shown in[23], VIPER-inferred MRs representing mechanistic determinants of cancer cell state are ultra-conserved within distinct tumor subtypes yet more rarely between tumor and cell lines presenting similar genetic alterations. For instance, among many neuroblastoma cell lines harboring a *MYCN* amplification, only three recapitulated the MR signature that was hyper-conserved in *MYCN*-amplified human tumors[11].

Based on these observations, we have developed[11,24] and extensively validated[9,10,35] the OncoMatch algorithm to evaluate *model fidelity* based on the statistical significance of MR conservation in a model/tumor pair. Specifically, OncoMatch tests for the enrichment of the top 50 MRs—*i.e.*, 25 most activated (MR↑) and 25 most inactivated (MR↓)—in a patient tumor in proteins differentially active in a cancer model (*e.g.*, cell line, PDX, organoid, etc.) and vice-versa. We have shown that this metric effectively recapitulates drug MoA—as assessed in high-

fidelity cell lines—in matched, low-passage PDXs[10]. Moreover, this metric is highly stable when the number of selected MRs varies between 10 and 200[10].

We used OncoMatch to optimize selection of high-fidelity cell lines for perturbational studies, thus making them more relevant to clinical translation (STAR Methods). The 23 PanACEA cell lines include: H1793 (lung adenocarcinoma, LUAD), LNCAP and DU145 (AR-dependent and AR-independent prostate cancer, PRAD), PANC1 and ASPC1 (basal and classical pancreatic adenocarcinoma subtypes, PAAD), NLF (bladder cancer, BLCA), HCC1143 and BT20 (triple-negative breast cancer subtype, TNBC), N87 (stomach, STAD), LoVo (colon cancer, COAD), HST-1 and KRJ1 (gastro-enteropancreatic neuroendocrine tumor, GEP-NET), IOMM (meningioma, ME), MSKOST and SAOS2 (pediatric osteosarcoma, OS), GIST (pediatric neuroblastoma, NBL), and HF2597 and U87 (proneural and mesenchymal glioblastoma subtypes, GBM), see STAR Methods for details on individual cell line selection.

Perturbational profile generation: For each cell line, cells were first cultured and expanded using the recommended culture media (STAR Methods) until a sufficient number was available to be transferred to 384-well plates for 10-point Dose Response Curve generation, using a repertoire of 90 to 451 drugs (**Table S1**). Indeed, since the collection of these profiles has been ongoing for the last 10 years, the number and diversity of compounds per cell line is variable. However, a majority of these libraries includes ~140 FDA approved antineoplastics, ~170 late-stage experimental drugs, a variable number of tool compounds from the Tocris and MicroSource collections with $EC_{20} \leq 2\mu M$—as required for optimal utilization of 96- or 384-well PLATE-Seq plates—and vehicle control media (DMSO).

Cell lines were then expanded again until a sufficient number of cells were available for dispensing into 96- or 384-well plates, with an average of 4,000 cells per well (STAR methods). Cells were allowed to rest for 24h followed by treatment with each library compound, at its cell

line-specific 48h EC_{20}, as assessed by Dose Response Curves. Cells were then harvested at 24h following treatment and enzymatically processed for library preparation using the PLATE-Seq methodology to a depth of 1M to 2M reads per well. RNA-seq profiles were then processed for quality control and normalized (STAR methods).

Data Reproducibility: To assess gene expression profile reproducibility, we leveraged the LoVo and BT20 cell lines, which were included in both PanACEA and LINCS. As shown in **Figure S1 A-D**, the differences in Pearson and Spearman correlation between technical replicates (*i.e.*, same cell line and drug), as assessed for the same set of ~1,000 LINCS landmark genes, were statistically significantly higher in PanACEA compared to LINCS, as assessed by the conservative, non-parametric Wilcoxon test. For LoVo, Wilcoxon p-values were $p = 2.75\times10^{-111}$ (Pearson) and $p = 6.77\times10^{-47}$ (Spearman), while, for BT20, they were $p = 4.71\times10^{-186}$ (Pearson) and $p = 4.39\times10^{-128}$ (Spearman). More critically, PanACEA includes whole-genome expression profiles, while LINCS measures only ~1,000 landmark genes and uses computational methods to infer the expression of all other genes. The latter may be problematic since predictions are based on the analysis of cellular populations at steady state rather than following drug perturbation.

Next, as suggested by recent publications[23], we asked whether VIPER analyses could further improve reproducibility. This analysis could not be performed on LINCS profiles since use of landmark genes decreases regulon coverage by ~20-fold. Transcriptional targets of TF/co-TF and signaling proteins were inferred by ARACNe analysis[36] of RNA-seq profiles from histology-matched human tumor cohorts (see **Table S2** for the corresponding tumor cohort list). Reproducibility was assessed by GSEA analysis of differentially expressed genes and differential activity of MR proteins, see STAR Methods. As assessed by Wilcoxon analysis, protein activity was significantly more reproducible than gene expression in all but one cell line, ranging from $p = 1.13\times10^{-13}$ in BT20 cells to $p = 1.53\times10^{-209}$ in SAOS2 cells (**Figure S2 A-W**).

IOMM, a meningioma cell line, represented the only exception (p = 0.24) (**Figure S2 X**); this is not entirely surprising because, due to lack of large-scale, high-quality meningioma cohorts in literature; the interactome was generated using profiles from 106 samples collected at Columbia (**Table S2**), a sample size barely above the minimum ARACNe requirements and much smaller than for other represented malignancies.

Consistent with the fact that several drugs may be virtually inert in a specific cellular context—*e.g.*, due to baseline inactivation of their high-affinity target—several cell lines presented violin plots with a clear bi-modal structure, including one mode corresponding to lack of drug activity (centered around NES = 0) and one corresponding to significant perturbation of cell state (centered around a higher NES value). Strong correlation of VIPER-assessed bioactivity and reproducibility was already shown in[24].

Drug Similarity Assessment: Consistent with these results—and based on prior studies showing the relevance of protein activity analysis in predicting drug sensitivity[12,23,24]—we analyzed the VIPER-based protein activity profiles to elucidate drug MoA and polypharmacology.

To assess the MoA similarity between two drugs, Rx_A and Rx_B, in the same cell line, we introduce a Drug Similarity Score ($DSS_{A\rightarrow B}$), representing the enrichment of the most differentially active proteins following treatment with Rx_A in protein differentially active following treatment with Rx_B, vs. vehicle control. The symmetric Drug Similarity Score metric ($DSS_{A,B}$) is then the average of $DSS_{A\rightarrow B}$ and $DSS_{B\rightarrow A}$ (**STAR methods**). Since drug MoA is critically tissue specific, we also define a cell line MoA similarity score ($CMSS_{C_1\rightarrow C_2}$), which assesses the similarity of the same drug MoA in two different cell lines (C_1 and C_2). This is computed as the enrichment of the most differentially active proteins, following drug vs. vehicle control treatment in C_1 cells, in proteins differentially active following C_2 cell treatment with the same drug. The

symmetric metric Cell Line MoA Similarity Score ($CMSS_{C_1,C_2}$) is then defined as the average of $CMSS_{C_1 \to C_2}$ and $CMSS_{C_2 \to C_1}$ (**STAR methods**).

Pleiotropic correction: Straightforward use of the DSS and CMSS metrics to compare drug MoA could result in poor specificity, because two completely unrelated drugs may still induce differential activity of overlapping proteins, for instance due to engagement of non-specific cell stress or cell death pathways. Using sublethal drug concentrations largely mitigates the issue yet a non-drug-specific signature still emerges as a potential confounding factor, leading to MoA specificity loss. To remove such bias, we assigned a cell line-specific weight $w_{i,j} \in [0,1]$ to each protein in the analysis—where the i index identifies a specific protein and the j index a specific cell line—representing the overall conservation of its differential activity across all perturbational profiles for that cell line. The weight is computed using a sigmoidal transformation of the absolute NES value, for a specific protein and cell line, across all drugs treatments. Proteins affected by a large number of drugs will produce lower (weighted) contribution to MoA similarity, thus mitigating non-specific confounding effects emerging (STAR methods).

Cellular Context Specificity: Previous perturbational profile-based studies either assess compound MoA in a single cell line or integrate it across all cell lines perturbed with that specific compound[37]. As previously suggested, however, ignoring the tissue-specific nature of compound MoA can confound the analysis[38]. Indeed, PanACEA profiles show that individual drugs may present MoAs that are highly conserved in a subset of cell lines yet virtually orthogonal in other subsets, including in cell lines representing the same tumor subtype.

For instance, as expected, resminostat and panobinostat—a class I/IIb and pan histone deacetylase (HDAC) inhibitor, respectively—present highly significant MoA similarity in the BT20 and HCC1143 TNBC cell lines ($p = 1.6\times10^{-9}$ and $p = 1.1\times10^{-12}$, respectively) (**Figure 2A & Figure 2B**). In contrast, pictilisib (a class I PI3K inhibitor) and neratinib (an ErbB family inhibitor)

have highly significant conserved MoA in BT20 cells ($p = 9.6\times10^{-15}$) but not in HCC1143 cells ($p = 5.3\times10^{-1}$) (**Figure 2C & Figure 2D**), both representing triple negative breast adenocarcinoma.

To further explore the tissue-specific nature of drug MoA, we assessed the enrichment of drug-pairs with the highest MoA conservation in one cell line (top 1% DSS score) in drug-pairs ranked by their DSS score in a second cell line. Surprisingly, comparison of histologically distinct cell lines, such as U87 (GBM) and H1793 (LUAD), showed strikingly significant MoA conservation across a wide range of drugs ($p = 3.5\times10^{-12}$) (**Figure 2E**). In sharp contrast, histologically related cell lines, such as HCC1143 and BT20, both TNBC-derived, presented 5 orders of magnitude lower enrichment statistics ($p = 9.6\times10^{-7}$) (**Figure 2F**).

The conclusion from these findings is that MoA analysis must directly account for tissue-specificity, independent of cell histology. Indeed, averaging MoA across all available cell lines (including histology-matched ones) diluted MoA similarity assessment and failed to reveal highly conserved, cell line subset-specific MoAs.

This is consistent with expectations, since different cell lines may harbor different mutations—potentially bypassing or abrogating drug activity—or express different gene programs—including missing critical primary or secondary targets of the drug. Yet, a comprehensive map of such cell specific differences is still missing. For instance, the estrogen receptor (ESR1), an important pharmacological target in luminal breast cancer, has low or no expression in several PanACEA cell lines. As such, conservation of ESR1 inhibitors' MoA in these cells is likely to be rather mediated by conserved off-target effects. Unfortunately, since drug targets—including both high-affinity and off targets—are only partially characterized, the cell lines in which the MoA of a drug may be conserved cannot be selected *a priori*.

To address this challenge in unbiased, data-driven fashion, we performed cluster analysis of each drug's MoA as assessed in each of the 23 available cell lines, thus revealing cell line clusters presenting virtually orthogonal inter-cluster yet high intra-cluster MoA conservation (STAR methods). As an illustrative example, consider drugs such as crizotinib, imatinib, nilotinib, tamoxifen, topotecan and vandetanib, which have rather distinct high-affinity targets (**Figure S3; Table S28).** As shown by cluster analysis, the MoA of these drugs is highly conserved in two or more distinct cell line subsets (positive DSS) yet virtually inverted between different subsets (negative DSS). The fact that analysis of 23 cell lines reveals only two to three conserved MoA cell line clusters per drug suggests that the proposed analysis is likely to converge to a small set of independent MoA per drug, thus making the problem tractable.

Drug MoA Similarity Network (MoA Network): Even though a subset of high-affinity binding targets of PanACEA drugs may be found in existing databases (*e.g.*, DrugBank[39]), their proteome-wide MoA and polypharmacology is still largely uncharacterized. In particular, DrugBank does not include potential high- and medium-affinity binding targets that cannot be revealed by traditional *in vitro* assays and completely ignores secondary, tissue specific effectors.

To systematically elucidate the tissue-specific MoA conservation of tested drugs, we leveraged the previous section's analysis to identify MoA similarities as manifested at the level of each cell line cluster. Specifically, we leveraged hierarchical clustering to identify cell line subsets (clusters) with statistically significant intra-cluster drug MoA conservation, as assessed by the CMSS metric (**Figure S3**). Since clusters with statistically significant MoA conservation for two different drugs (Rx_A and Rx_B) may have variable overlap, we only considered cluster pairs with ≥ 50% overlap (*i.e.*, Jaccard index ≥ 1/3) and with ≥ 2 cell lines each (STAR methods).

Each cluster pair where the MoA of two drugs (Rx_A and Rx_B) was significantly conserved was then represented as an edge in a global MoA Network, with an associated p-value generated by integrating the DSS p-value of each cell line in the intersection, using Stouffer's method (see STAR methods for a relevant example). Such a network is difficult to effectively visualize and is better represented in tabular form (**Table S29;** STAR methods). For visualization purposes, we generated a simplified MoA Network with a single edge between drugs with statistically significant MoA conservation in one or more cell line subsets, associated with the p-value of the most significant edge in the complete network.

MoA Network Modularity Analysis: Network modularity analysis—such as used to identify protein complexes in protein-protein interaction network[40-45], for instance—represents an effective mean to identify subset of drugs (modules) presenting high intra-module and low-extra-module MoA conservation. In addition, such analyses can be used to identify drugs presenting strong MoA conservation across two or more modules, thus revealing potential polypharmacology. For this purpose, we leveraged the ClusterONE algorithm[46], followed by a conservative statistical significance analysis (STAR methods). The analysis identified 30 distinct drug modules, associated with equally distinct pharmacologic properties (**Figure 3A & Table S3**).

To assess intra-module MoA conservation, we computed a Drug Modularity Score (DMS), and its associated p-value, representing the overall confidence that a specific drug belongs to a module. This was accomplished by integrating the p-values representing the drug MoA conservation with every other drug in the module (based on the DSS score), using Stouffer's method. To assess the global confidence in the entire module, a Module MoA Conservation Score ($MMCS$) was assessed by integrating the DMS of each drug in the module, also using Stouffer's method (STAR methods for null hypothesis models).

Based on literature analysis, a majority of the drug modules identified by the analysis were highly enriched in drugs with a common established MoA. However, several drugs were identified to also associated with modules unrelated to the drug's known MoA. This identifies drugs presenting potential polypharmacology, either directly (based on their off-target effects) or indirectly due to tissue-specific crosstalk or canalization mechanisms in a subset of the cell lines. Some key examples are shown in **Figure 3**. Obviously comprehensive validation of all novel MoA predictions would be prohibitive. As a result, we provide extensive description for some of the more critical modules (M1, M2, M3, M18 and M22) and validation of polypharmacology and/or novel MoA predictions for selected compounds in M1, M3, and M24.

In Module M1, 13 of 14 drugs represent PI3K/AKT/mTOR axis inhibitors in Drug Bank (**Figure 3B)**, while the 14th drug, KW2499, represents a multi-kinase inhibitor, whose high-affinity binding targets include FLT3, ABL, ABL-T315I and Aurora kinase[47]. MoA analysis shows that KW2499 has highly significant DSS—in the range [10^{-27}, 10^{-326}], with the other drugs in the module, critically across most cell lines (*i.e.*, 13 of 23 in the study). This cannot be discounted as an indirect effect mediated by its primary targets because none of the other FLT/ABL/AURKA inhibitors in PanACEA were found in this module (*e.g.*, ENMD-2076). To validate this polypharmacology prediction, we measured the ratio of phosphorylated p70 ribosomal S6 kinase (p-P70S6K) by total p70 ribosomal S6 kinase (t-P70S6K) after drug treatment in SAOS2 cells, via Western blot. This showed that p-70S6K levels were significantly reduced in KW2499 treated cells compared to vehicle-treated and positive control (mTOR pathway activation with 20% FBS) treated cells. Results from KW2499 treatment are consistent with temsirolimus (a known mTOR inhibitor) (**Figure 4A-C**).

Structure based analysis, based on the PrePCI algorithm[28,29], confirmed Kw2499 as a likely, novel PI3K/AKT/mTOR inhibitor. Indeed, PrePCI not only recapitulated the established KW2449 targets but also provided several high-confidence predictions (FPR < 0.001) for protein

annotated as members of PI3K/AKT/mTOR signaling pathways[48] (A in **Table S4)**, thus further supporting KW2499's polypharmacology. For instance, **Figure 4J** shows a structural model of the KW2449/INPP5D complex. INPP5D, also known as SHIP1, is the phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase-1 responsible for hydrolyzing the 5-phosphate of lipid PI(3,4,5)P3 to produce PI(3,4)P2. Both lipids are potent mitogenic signals involved in activating AKT and, thus, contribute to the regulation of PI3K/AKT signaling[49]. This suggests INPP5D as a novel, high-likelihood off-target of KW2449, whose inhibition contributes to the observed p-70S6K level reduction. The INPP5D-KW2449 complex model was constructed by rigid-body docking, as described in STAR Methods.

Briefly, PrePCI predicts protein-small molecule interactions based on the concept of Structural BLAST[50]. Specifically, protein structural alignment is used to detect evolutionary relationships between proteins that may be undetectable by protein sequence alone thus allowing non-trivial predictions for how compounds represented within experimentally determined complexes in the Protein Data Bank[51] may bind to other candidate proteins. PrePCI can also help identify binding of small molecules not represented in PDB, by leveraging PDB complexes in which related compounds are included, based on Tanimoto coefficient (TC) similarity between 2D representations of the compounds[52], *e.g.*, as reported by PubChem[53].

In **Module M2**, 13 of 14 drugs are established HDAC inhibitors (**Figure 3C**), while the 14th drug, cladribine, is an established adenosine analog that prevents the adenosine deaminase enzyme from modulating DNA processes[54]. Remarkably, its DSS with belinostat and rocilinostat—both established HDAC inhibitors—is even more significant ($p \leq 1.22\times10^{-136}$ and $p \leq 1.2\times10^{-105}$, respectively), than the DSS between Belinostat and Rocilinostat ($p \leq 5.21\times10^{-99}$), as supported by in cell lines PANC1, HSTS and GIST430. While the MoA of cladribine may suggest potential functional similarity to HDAC inhibitor-mediated chromatin remodeling, other adenosine and, more generally, purine analogs in PanACEA—*e.g.*, pentostatin, clofarabine, mercaptopurine,

and thioguanine, for instance—presented no statistically significant DSS with HDAC inhibitors, suggesting a potentially unique moonlighting function for this drug.

Figure S13 depicts a structure-based model for the complex between cladribine and HSP90AA1—a heat-shock protein included in HDAC-related pathways[48]—based on a high-confidence PrePCI prediction (FPR < 0.001). HDAC inhibitor-mediated deregulation of chaperone function has been shown to disrupt protein homeostasis and induce protein misfolding[55] suggesting a functional link between HDAC and HSP90 inhibition. **Table S4 B** illustrates that cladribine has PrePCI-predicted targets (FPR < 0.001) in common with five of the drugs in this module (entinostat, romidepsin, dacinostat, pracinostat, and abexinostat). All six compounds are predicted to target HSP90AA1.

In **Module M3**, 17 of 22 drugs represent MAPK pathway inhibitors, including inhibitors of MEK and of its upstream modulators and downstream effector kinases (**Figure 3D**). The other 5 drugs represent a quite heterogeneous mixture, including az12419304 and az12456623—two small molecule inhibitors with a yet-to-be-determined function, which may thus represent novel MAPK inhibitors—biib021 (an HSP90 inhibitor[27]), danusertib (an AURKA inhibitor[56]), and ly2835219 (a CDK4/6 inhibitor[57]). To evaluate predictions for this module, we focused on biib021 and az12419304 as illustrative examples. Specifically, we measured the ratio of phosphorylated Erk (p-Erk) by total Erk (t-Erk) in N87 cells using western blot, and observed that this ratio was significantly reduced following biib021 and az12419304 treatment, compared to DMSO control, in a dose-dependent manner (**Figure 4D-F**).

PrePCI makes many high-confidence for az12419304, az12456623, danusertib, and ly2835219 as MAPK pathway proteins (GO: 0000165) binding partners (**Table S4 C-F**). A handful of PrePCI predictions are supported by direct experimental evidence in PDB, including the danusertib-EPHB2[58] and ly2835219-CDK6[59] complexes. However, many predictions are novel,

such as the interaction between az12419304 and BRAF, whose structural model prediction is shown in **Figure 4K**.

In **Module M18**, 12 of 13 drugs represent DNA cross-linking/alkylation and topoisomerase inhibitors (**Figure 3E**). The polo kinase (PLK1) inhibitor hmn214[60] was also predicted as a highly statistically significant member of M3 (ranging from $p \leq 5.26\times10^{-42}$ with irinotecan to 1.49×10^{-107} with fluorouracil), potentially consistent with its role as a modulator of DNA damage repair mechanisms.

Module M22 (**Figure 3F**) includes four established VEGF pathway inhibitors (sunitinib, axitinib, sorafenib, and cabozantinib), as well as two drugs not previously annotated as such, including amonafide—a topoisomerase inhibitor[61] that, surprisingly, failed to cluster with other topoisomerase inhibitors in M22—and bicalutamide, a selective androgen receptor blocker[62]. Both had highly significant DSS with VEGF pathway inhibitors. Specifically, amonafide's MoA similarity with sorafenib and cabozantinib was highly significant ($p \leq 1.03\times10^{-34}$ and $p \leq 6.22\times10^{-55}$, respectively), while bicalutamide similarity with axitinib and sorafenib was even more significant ($p \leq 1.36\times10^{-72}$ and $p \leq 4.64\times10^{-92}$, respectively).

As another illustrative example, of the five drugs in **Module M24**, three (methotrexate, pemetrexed, and pralatrexate) are established dihydrofolate receptor (DHFR) inhibitors (**Figure G**), while a fourth one, floxuridine, is a thymidylate synthetase inhibitor[63], an enzyme also playing a key role in folate biosynthesis. However, Mgcd265 (also known as glesatinib) is an established multi-target kinase inhibitor[64] with no known relation to folate biosynthesis, thus establishing a potential link between these functions. To validate this prediction, we treated LoVo cells using Mgcd265 at various concentrations (0.5 μM, 1 μM and 5 μM, respectively) for 24 hours. Confirming our predictions, significant DHFR protein level reduction was observed at all three concentrations (**Figure 4G-I**). PrePCI further supports this finding by predicting that

Mgcd265 shares several targets with the other three DHFR inhibitors in M24, most notably, DHFR itself (**Table S4 G**).

Critically, modularity analysis also highlighted MoA conservation that is remarkably tissue specific. For example, Module M29 includes four DNA damage drugs (mercaptopurine, ifosfamide, streptozotocin, and procarbazine) and two steroidogenesis inhibitors (mitotane and aminoglutethimide) (**Figure S10A)**. While the two cytotoxic drug ifosfamide and mercaptopurine showed highly similar MoA across most cell lines, aminoglutethimide, mitotane, ifosfamide, procarbazine and streptozotocin formed a highly statistically significant sub-module only in three sarcoma-derived cell lines, including SAOS2 and MSKRMS (osteosarcoma) and SKNEP (Ewing sarcoma), again providing a sarcoma specific link between steroidogenesis and DNA-mediated cytotoxicity. An additional compelling example is provided by Module M13 (**Figure S10B)**, where drugs have highly distinct established MoA. While individual, statistically significant drug-drug edges are only supported by a handful of tissues, a highly interconnected sub-module, comprising five drugs (Gmx1778, Bay-117085, Cp100356, Ym201636 and Paroxetine) is identically conserved in the two TNBC cell lines (HCC1143 and BT20) suggesting a mechanism of drug effect canalization that is specific to this context. These illustrative examples show that tissue-specificity is a critical issue in drug MoA elucidation and suggest potential interesting tissue-specific biology for future investigation.

Targeting Undruggable Targets: Transcription factors and co-factors have emerged as some of the most critical tumor-essential proteins in many studies. Unfortunately, most of them are considered to be undruggable for lack of specific binding pockets[65] and lack of structural information. Indeed, for many of these proteins, only the DNA-binding domain(s) have been structurally resolved. We thus reasoned that the experimental perturbational profiles in PanACEA may offer critical insight into drugs that may inhibit these undruggable proteins,

potentially as downstream effectors of a complex pharmacologic field-effect resulting from the polypharmacology-based, tissue-specific inhibition of multiple upstream modulators.

For instance, systematic pancancer analysis of ~10,000 samples in The Cancer Genome Atlas (TCGA) elucidated Master Regulator (MR) proteins representing ultra-conserved, mechanistic determinants of tumor cell state, as well as the regulatory bottleneck responsible for canalizing the effect of patient-specific mutational landscapes[23]. Critically, MRs have been shown to be highly enriched in essential or synthetic lethal proteins, thus constituting a novel class of candidate drug targets[11,23,31,66,67]. Unfortunately, the majority of these proteins are considered to be undruggable. Critically, 37 proteins were as MR in ≥ 15 cancer subtypes each, thus establishing them as the most recurrent aberrantly activated MRs in cancer. We thus interrogated the PanACEA repository to assess whether any of them could be inhibited by one or more PanACEA drugs in tissue contexts where they had been identified as MRs. For this purpose, we first identified cell lines where these 37 proteins had above-average activity—*i.e.*, VIPER score, VS > 0, in untreated PanACEA cell lines, compared to the centroid of the CCLE and Genentech cell lines repositories[68]. We then analyzed the PanACEA perturbational profiles to identify drugs inducing statistically significant MR activity decrease in drug vs. vehicle control-treated cells. For each cell line, the top 15 drugs producing the strongest MR activity inhibition are shown as an NES heatmap (**Figure 5A)**. The three most potent, tissue-specific drugs targeting a cell line specific MRs are then recapitulated for each of the 23 PanACEA cell lines (**Figure 5B & Figure S4**).

Conclusions are two-fold: First, the top inhibitors of each MR were highly tissue-specific. This is not surprising, since the effect is likely mediated by the tissue-specific signaling and regulatory logic of the cell. Second, available drugs are only effective in targeting a specific MR in a subset of selected cell lines (**Figure 5B & Figure S4**). For example, statistically significant inhibitors of

the OVOL1 MR could only be predicted for 4 of 23 cell lines. Taken together, these data suggest that the PanACEA framework could be leveraged to identify or develop novel, tissue-specific inhibitors for virtually any protein representing a potential dependency of tumor cells or a target in non-tumor related contexts. In addition, this suggests that the antitumoral activity of some drugs may be mediated by indirect effects of critical MR proteins, rather than on their high-affinity targets.

To identify drugs representing MR inhibitors across multiple tissue contexts, thus suggesting a more direct and conserved mechanism, we generated a histogram showing the number of cell lines where the top 15 inhibitors of each MR were found to be statistically significant (**Figure 5C & Figure S5**). A table summarizing the effect of each drug on each MR, on a cell line by cell line basis, is also provided (**Table S5-27)**.

Experimental Validation of Tissue-specific, MR-targeting Drugs: In addition to mechanisms related to binding a protein's active site or to allosteric binding, protein activity modulation may result from three primary mechanisms, typically including one of the following: (a) indirect downregulation of the encoding gene's expression, (b) activation of ubiquitin-dependent proteasomal degradation, and (c) subcellular compartment sequestration (*e.g.*, cytosol, nuclear/cell membrane, etc.). Since no single assay can reveal these complementary effects, we validated predicted MR inhibitors using three complementary assays, including (a) qRT-PCR, (b) Western Blot, and (c) confocal immunohistochemistry, respectively.

Selection of Drug/MR/cell line triplets for validation was largely based on availability of key reagents, such as MR-specific antibodies[23], cell line growth culture conditions, and baseline expression of the target.

First, we show four MR-inhibitors downregulated the mRNA of the associated gene in the associated cell line context, as measured by qRT-PCR analysis, using three biological

replicates and three technical replicates (**Figure 6G**). These include (a) Bortezomib-mediated UHRF1 downregulation in H1793 cells (LUAD) (VIPER score, VS = -10.65, $p \leq 1.74\times10^{-26}$), (b) Bortezomib-mediated CENPF downregulation in HF2597 cells (GBM) (VS = -6.18, $p \leq 6.41\times10^{-10}$), (c) Pracinostat-mediated UHRF1 downregulation by in H1793 cells (LUAD) (VS = -11.21, $p \leq 3.64\times10^{-29}$), and (d) Bortezomib-mediated UHRF1 downregulation in HF2597 cells (GBM) (VS = -8.14, $p \leq 3.95\times10^{-16}$).

Four additional candidate MRs and predicted inhibitors/cell contexts were selected for Western-blot-based validation, including: (a) CTNNB1 inhibition by the DNA intercalator dactinomycin in KRJ1 cells (GEP-NET) (VS = -4.17, $p \leq 3.05\times10^{-5}$) (**Figure 6A**). β-catenin is a key effector of WNT signaling, which plays an important role in tumorigenesis[69]; (b) UHRF1 inhibition by the HDAC inhibitor pracinostat in HF2597 cells (GBM) (VS = -11.07, $p \leq 1.75\times10^{-28}$) (**Figure 6B**). UHRF1 plays an important role in the maintenance of DNA methylation in mammalian cells[70] has been identified as a potential drug target as it is essential to the survival of multiple human malignancies[71,72]; (c) MYC inhibition by the MEK1/2 inhibitor pimasertib in HCC1143 cells (TNBC) (VS = -6.2, $p \leq 5.65\times10^{-10}$) (**Figure 6C**). MYC is one of the most frequently amplified or translocated proto-oncogenes, associated with tumorigenesis in multiple of human maligancies[73], as well as a highly elusive drug target; and (d) CENPF inhibition by amsacrine in HF2597 cells (GBM) (VS = -12.8, $p \leq 1.64\times10^{-37}$) (**Figure 6D**). CENPF is a large coiled-coil protein involved in centromere maintenance during mitosis[74]. In previous work, we have shown that co-inhibition of CENPF and FOXM1 abrogates viability of malignant castration resistant prostate cancer[75]. Indeed, these two proteins are co-aberrantly activated in several of the most aggressive malignancies in TCGA[23].

Finally, four additional MR-inhibitors were validated by confocal microscopy of immunofluorescently stained cells, with statistical significance assessed by CellProfiler analysis

using the conservative Wilcoxon test[76]. These include: (a) Nuclear CENPF delocalization in U87 cells (GBM) by the topoisomerase II inhibitor amsacrine (VS = -12.31, $p \leq 8.00\times10^{-35}$) (**Figure 6E**). Nuclear vs. cytoplasm CENPF localization was significantly higher in in DMSO vs. amsacrine-treated samples ($p = 1.9\times10^{-6}$); (b) Nuclear CENPF delocalization in U87 cells (GBM) by the proteasome inhibitor bortezomib (VS = -11.44, $p \leq 2.64\times10^{-30}$) (**Figure 6E**). Nuclear vs. cytoplasm CENPF localization was significantly higher in in DMSO vs. bortezomib-treated samples ($p = 3.8\times10^{-45}$); (c) MYC sequestration to the nuclear lamina in LoVo cells (COAD) by the MEK1/2 inhibitor pimasertib (VS = -6.32, $p \leq 2.62\times10^{-10}$), consistent with prior reports of MYC activity inhibition by phosphorylation[77] (**Figure 6F**). Nuclear vs. nuclear-lamina MYC localization was significantly higher in DMSO vs. pimasertib-treated samples ($p = 5.3\times10^{-186}$); and (d) Nuclear MYC delocalization in LoVo cells (COAD) by the MEK1/2 inhibitor refametinib (VS = -8.23, $p \leq 1.87\times10^{-16}$) (**Figure 6F**). Nuclear vs. cytoplasm MYC localization was significantly higher in in DMSO vs. bortezomib-treated samples ($p = 2.8\times10^{-63}$).

Discussion

This study reports on a 10-year-long effort to create PanACEA, a collection of gene expression and protein activity profiles for the network-based elucidation of tumor-relevant drug mechanism of action and polypharmacology. For this purpose, PanACEA represents the genome/proteome-wide response of 23 cell lines, representing high-fidelity models of molecularly distinct cancer subtypes, to clinically relevant drug perturbations. Its design offers three main advantages: First, by leveraging the PLATE-Seq technology, cell response could be characterized on a genome-wide scale, both inexpensively and without substantial bias. Second, cell lines were chosen to represent clinically relevant models of human malignancies, based on OncoMatch analysis. This represents a critical aspect of this study as we have previously shown that MoA predictions based on PanACEA cell lines were recapitulated *in vivo*—in PDX models for which they were predicted as high-fidelity models by OncoMatch—for 15 of 18 tested drugs[10]. Third, drugs were carefully titrated at their highest sub-lethal concentration (EC_{20}) as assessed in each cell line, thus mitigating activation of cell stress and death pathways that may confound MoA analysis and preventing use of drugs at non bioactive concentrations. Most critically, availability of genome-wide perturbational profiles supports using the extensively validated VIPER algorithm to generate activity profiles for ~6,500 regulatory and signaling proteins, representing much more effective reporters of drug MoA, including secondary effectors and low-affinity binding targets may explain drug polypharmacology and toxicity.

VIPER assessment of differential protein activity in drug-treated cells is critical because most drugs do not affect the abundance of target proteins or the expression of encoding gene but rather their enzymatic or regulatory activity—as shown for estrogen receptor inhibitors[6,7], for instance. A critical advantage of VIPER is its high reproducibility, which has led to NY and CA Dept. of Health approval[78] of two VIPER-based algorithms, OncoTreat[24] and OncoTarget[79] for

clinical use. Indeed VIPER-based protein activity measurements compare favorably with antibody-based approaches, including at single cell level, where gene dropout severely compromises individual gene readouts[12,21]. VIPER prediction accuracy was confirmed in this study, since virtually all the drugs identified as potential context-specific Master Regulator inhibitors were experimentally validated, including via mechanisms, such as cell compartment localization, that would be difficult to assess by traditional assays.

Overall, the combination of the largest collection of genome/proteome-wide drug perturbation profiles and the associated analytical framework provides a valuable resource to facilitate the identification of drugs targeting specific proteins, mechanisms, and pathways. This investigational framework can be easily extended to the study of small molecule compounds and biologics in virtually any disease-related or physiological context. For instance, we recently shown how this framework could be used to identify drugs, such as gemcitabine, that successfully reprogrammed the tumor tropism of regulatory T cells inducing dramatic synergy with immune checkpoint inhibitors, as well as predicting that optimal concentration would be 1/10th of what is currently used clinically[31]. In the context of drug discovery, VIPER analysis of drug perturbation profiles provides a universal phenotypic assay that, once generated, can be retooled at will to report on the effect of drugs on a variety of biologically relevant protein activity signatures—especially when compared to the typical highly-specialized single protein reporter assay used in drug discovery—thus providing deep characterization of an agent's pharmacology and toxicity for approximately the same cost per well.

The value of perturbational profiles in individual high-fidelity cell lines has been demonstrated in several previous studies. For instance, we have shown that OncoTreat analysis of a small subset of PanACEA cell lines could predict drugs that induce 91% disease control rate in PDX from patients who had failed multiple lines of therapy[10]. Similarly, OncoTreat analysis of HST-S and DU145 cell lines, was instrumental in the identification of drugs inducing partial response *in*

vivo for gastroenteropancreatic neuroendocrine tumors (GEP-NETs) and castration resistant metastatic prostate cancer (mCRPC), respectively. In addition, analysis of DU145 perturbation profiles identified four drugs capable of activating Master Regulator protein modules (MR-blocks) that are inactivated in the most aggressive subtype of prostate cancer, thus inhibiting growth, immune evasion, and cell migration hallmarks[23]. Supporting PanACEA's ability to extend to disease other than cancer, it should be noted that analysis of its perturbational profiles, for lung and colon cancer cell lines, was instrumental in identifying 18 drugs predicted to invert the activity of Master Regulator regulating host cell response to SARS-CoV-2 infection, 80% of which were experimentally validated[80]. Despite these success stories, however, none of these individual studies would have supported the large-scale analysis of MoA context-specificity nor the creation of a large-scale, tissue-specific MoA similarity and polypharmacology map discussed in this study.

Taken together, the results of this study suggest that PanACEA may provide a highly flexible and valuable methodological and experimental resource to support a more quantitative application of precision medicine principles, based on the identification of small molecules that target critical, context-specific dependencies of both human malignancies, with straightforward generalization to other human diseases.

Figure Legends

Figure 1: Schematic view of PanACEA resource for the characterization of oncology drug MoAs. (A) The OncoMatch algorithm was used to select CCLE cell lines that recapitulate the MR signature of human tumor cohorts. (B) Compendium of cell lines, associated tissue of origin and number of drugs they were treated with to produce the PanACEA resource. (C) The IC_{20} of each drug in each cell line was assessed from 10-point dose response curves. (D) 23 selected cancer cell lines were treated with >700 distinct oncology drugs (in total) for 24 hours at their established IC_{20}. (E) PLATE-Seq was used to generate genome-wide RNA-Seq profiles following each drug perturbation. (F) The ARACNe and VIPER algorithms were used to compute differential protein activity in drug vs. vehicle control-treated cells. (G) Drug-drug MoA similarity was assessed and clustered to generate conserved MoA modules. (H) A functional Drug MoA Similarity Network was generated to visualize same-MoA drug modules suggesting polypharmacology or novel MoA prediction. (I) Predicted MoA, polypharmacology, and MR-specific activity was validated for selected drugs.

Figure 2: Enrichment analysis shows drug effects are cellular contexts dependent. (A) Two-tailed GSEA analysis shows high MoA similarity for resminostat and panobinostat in BT20 cells. Red and blue bars represent the 25 proteins most activated and 25 most inactivated in resminostat vs. vehicle control-treated cells. The x-axis shows proteins ranked by their differential activity in panobinostat vs. vehicle control-treated BT20 cells. (B) GSEA analysis shows high MoA conservation for resminostat and panobinostat in HCC1143 cells. (C) GSEA analysis shows unexpected high MoA conservation for pictilisib (Gdc0941) and neratinib—two drugs with very different high-affinity binding targets—in BT20 cells. (D) GSEA analysis showing low MoA conservation for pictilisib (Gdc0941) and neratinib in HCC1143 cells. (E) GSEA analysis of the 1% of drug pairs with the highest MoA similarity in U87 cells vs. all statistically significant ($p < 0.01$, Bonferroni corrected) drug pairs in H1793 cells, ranked by their p-value.

The analysis shows high drug MoA similarity conservation, despite the different cell line histology (glioblastoma vs. lung adenocarcinoma, respectively) (F) Same analysis for HCC1143 vs. BT20 cells (Both triple negative breast cancer). The analysis shows 5 orders of magnitude lower MoA similarity conservation compared to U87 vs. H1793.

Figure 3: PanACEA derived drug functional map with highlighted examples. (A) The 24 drug modules identified by MoA Network modularity analysis. Each node (circle) represents a drug, with same-color nodes indicating assignment to the same module and circle color representing the z-score of the Drug Specificity Test (*i.e.*, the confidence in drug-module assignment) (STAR Methods, **Table S3**). Modules with a clearly identifiable established MoA are labeled accordingly (*e.g.*, tubulin inhibitor module 2). Gray lines identify drug pairs with significantly MoA similarity in at least two of the 23 cell lines. Six highly representative modules, identified by a red dotted box, were selected for detailed discussion and validation in the manuscript. These are illustrated with further details in panels (B-G). Within each panel, same-color nodes indicate drugs with an established consensus MoA (*e.g.*, PIK3/AKT/mTor inhibitors). Red nodes identify drugs not previously known to share the module's consensus MoA and thus representing either novel MoA or polypharmacology candidates. A yellow rounded-edge square identifies drugs shared by multiple modules.

Figure 4: Experimental validation of novel drug MoA or polypharmacology predictions. The red circles in each panel identify the drugs selected for experimental validation. (A) PI3K/AKT/mTOR pathway inhibitor consensus module (M1): (B) Consistent with predictions, KW2449 inhibited phosphorylation of the canonical mTor substrate p70 ribosomal S6 kinase (P70S6K) in dose-dependent manner. Saos2 cells were cultured in 10% FBS, to induce mTor pathway activation, and then treated for 6 hours with vehicle control or the indicated concentration of KW2449 and temsirolimus (a known mTor inhibitor). Cell pellets were collected and lysed, and even amounts of proteins were separated by SDS-PAGE and blotted for

phospho-P70S6 (p-P70S6), total-p70S6 (t-P70S6) and GAPDH as loading control. (C) Quantification of treatment-mediated p-P70S6K abundance changes, normalized to GAPDH. (D) MEK pathway inhibitor consensus module (M3) (E) Consistent with predictions, biib021 and az12419304 inhibited phosphorylation of ERK, selected as the most downstream reporter of MAPK pathway activation. N87 cells were treated with biib021 or az12419304 for 4 hours, at the indicated concentrations. Cell pellets were collected and lysed, and even protein amounts were separated by SDS-PAGE and blotted for phospho-ERK (p-ERK), total-ERK (t-ERK) and tubulin as loading control. Trametinib (a known MEK inhibitor) was used here as positive control. (F) Quantification of treatment-mediated p-ERK abundance changes normalized to tubulin. (G) DHFR pathway inhibitor consensus module (M24). (H) Consistent with predictions, Mgcd265 induced DHFR protein abundance reduction. LoVo cells were treated with Mgcd265 for 24 hours, at the indicated concentrations. Cell pellets were collected and lysed, and even protein amounts were separated by SDS-PAGE and blotted for DHFR and tubulin (loading control). (I) Quantification of treatment-mediated DHFR abundance changes, normalized to tubulin. (J) Atomic level representation of the protein/compound interacting region based on a Kw2449-INPP5D structural model predicted by the PrePCI algorithm and refined by molecular docking analysis. (H) Atomic level representation of the protein/compound interacting region based on a Az12419304-BRAF structural model predicted by the PrePCI algorithm and refined by molecular docking analysis.

Figure 5: PanACEA-predicted inhibitors of the 37 most recurrent MR proteins in TCGA, as identified by MOMA analysis[23]. (A) MR inhibition heatmap showing the effect of the top 15 drugs inferred as potent MR inhibitors (*i.e.*, VS ≤ -3), by VIPER analysis of drug vs. vehicle control-treated cells. Rows represent the 37 MOMA-nominated MRs, while columns represent the 23 cell lines in PanACEA. For each drug/cell line pair, the heatmap shows the effect of the top 15 VIPER-inferred inhibitors (different for each heatmap), sorted left to right from the most

(dark blue) to the least potent (light blue). The inactive MRs (VIPER score <= 0 in untreated PanACEA cell lines, compared to the centroid of the CCLE and Genentech cell lines repositories) are shown in light green. (B) This panel shows up to the top three potent (i.e., VS ≤ -3) MR inhibitor predictions, ranked by their VIPER Score. Only MRs (n = 8) for which at least one potent inhibitor is identified in at least one cell line are shown. The inactive MRs are colored by the same color as in (A). (C) Bar plot showing the number of cell lines in which the 15 most recurrently inferred, potent inhibitors of each MRs in panel B are identified. Higher number of cell lines (*e.g.*, vorinostat for CENPF) suggest higher conservation of the mechanisms leading to drug-mediated MR inhibition.

Figure 6: Validation of context-specific MR inhibitor predictions. (A) Western Blot (WB) assays confirm that dactinomycin reduces CTNNB1 expression in KRJ1 cells. (B) WB assays show that pracinostat abrogates UHRF1 expression in HF2597 cells. (C) WB assays show that pimasertib reduces MYC expression in HCC1143 cells. (D) WB assays show that amsacrine abrogates CENPF expression in HF2597 cells. (E) Consistent with transcriptional activity inhibition predictions, confocal image analysis shows that nuclear localization of CENPF is significantly reduced in amsacrine and bortezomib-treated U87 cells. (F) Consistent with transcriptional activity inhibition predictions, MYC nuclear localization is significantly reduced in refametinib-treated LoVo cells and MYC trans-localizes from the nucleoplasm to the nuclear lamina in pimasertib-treated LoVo cells. (G). qRT-PCR assays in triplicate show significant gene expression reduction for genes encoding for four MR proteins whose drug-mediated transcriptional inhibition was predicted by VIPER. Error bars show ± 1 standard deviation. The chemical structure of validated drugs is shown in each panel. Scale bars in E and F, shown as a horizontal white bar represent a length of 10 μm.

Data Avaiability

Drug perturbation RNA-seq data generated from this work have been deposited to the Gene Expression Omnibus (GEO: GSE275348).

Acknowledgements

We acknowledge the High-Throughput Screening and the Genome Sequencing and Analysis Facilities at the JP Sulzberger Columbia Genome Center, the Confocal and Specialized Microscopy Shared Resource (SCMSR) managed by the Herbert Irving Comprehensive Cancer Center at Columbia University and supported in part by the NIH/NCI Cancer Center Support grant P30CA13696. This work was also supported by the NCI Cancer Target Discovery and Development (CTD2) awards U01CA272610 and U01CA217858, Center for Cancer Systems Therapeutics (CaST) award U54CA274506, and the NIH Shared Instrumentation Grants S10OD032433, S10OD012351 and S10OD021764, all to A.C. The work was also supported by NCI grant P30CA013696, NIH grants S10OD020056 and S10OD030282 and by NIH grant 5P30DK132710, all to B.R.S. ; by NIH grant R35-GM139585 to B.H.

Conflict of Interest

Authors Disclosures: A.C. is founder, equity holder, consultant, and director of DarwinHealth Inc., a company that has licensed some of the algorithms used in this manuscript from Columbia University. M.J.A. is Chief Scientific Officer and equity holder at DarwinHealth, Inc., a company that has licensed some of the algorithms used in this manuscript from Columbia University. C.K. is a DarwinHealth consultant and shareholder. B.R.S. is an inventor on patents and patent applications involving ferroptosis; co-founded and serves as a consultant to ProJenX, Inc. and

Exarta Therapeutics; holds equity in Sonata Therapeutics; serves as a consultant to Weatherwax Biotechnologies Corporation and Akin Gump Strauss Hauer & Feld LLP.

Open Access Statement

This content of this study is published under open access subject to licensing agreements. The authors allow researchers to distribute and use data from the original source provided it is distributed free of cost for non-commercial purposes. Any profit-making is strictly prohibited.

References

1 Lin, A. *et al.* Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Science translational medicine* **11**, doi:10.1126/scitranslmed.aaw8412 (2019).

2 Meissner, F., Geddes-McAlister, J., Mann, M. & Bantscheff, M. The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat Rev Drug Discov* **21**, 637-654, doi:10.1038/s41573-022-00409-3 (2022).

3 Klaeger, S. *et al.* The target landscape of clinical kinase drugs. *Science* **358**, doi:10.1126/science.aan4368 (2017).

4 Anighoro, A., Bajorath, J. & Rastelli, G. Polypharmacology: challenges and opportunities in drug discovery. *J Med Chem* **57**, 7874-7887, doi:10.1021/jm5006463 (2014).

5 Iorio, F. *et al.* Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci U S A* **107**, 14621-14626, doi:10.1073/pnas.1000138107 (2010).

6 Woo, J. H. *et al.* Elucidating Compound Mechanism of Action by Network Perturbation Analysis. *Cell* **162**, 441-451, doi:10.1016/j.cell.2015.05.056 (2015).

7 Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* **48**, 838-847, doi:10.1038/ng.3593 (2016).

8 Ding, H. *et al.* Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nature communications* **9**, 1471, doi:10.1038/s41467-018-03843-3 (2018).

9 Vasciaveo, A. *et al.* OncoLoop: A Network-Based Precision Cancer Medicine Framework. *Cancer discovery* **13**, 386-409, doi:10.1158/2159-8290.CD-22-0342 (2023).

10 Mundi, P. S. *et al.* A Transcriptome-Based Precision Oncology Platform for Patient-Therapy Alignment in a Diverse Set of Treatment-Resistant Malignancies. *Cancer discovery* **13**, 1386-1407, doi:10.1158/2159-8290.CD-22-1020 (2023).

11 Rajbhandari, P. *et al.* Cross-Cohort Analysis Identifies a TEAD4-MYCN Positive Feedback Loop as the Core Regulatory Element of High-Risk Neuroblastoma. *Cancer discovery* **8**, 582-599, doi:10.1158/2159-8290.CD-16-0861 (2018).

12 Obradovic, A. *et al.* Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell* **184**, 2988-3005 e2916, doi:10.1016/j.cell.2021.04.038 (2021).

13 Jamison, J. K. *et al.* Entinostat in patients with relapsed or refractory abdominal neuroendocrine tumors. *The oncologist*, doi:10.1093/oncolo/oyae118 (2024).

14 Blattmann, P. *et al.* Systems Pharmacology Dissection of Cholesterol Regulation Reveals Determinants of Large Pharmacodynamic Variability between Cell Lines. *Cell Syst* **5**, 604-619 e607, doi:10.1016/j.cels.2017.11.002 (2017).

15 Ruprecht, B. *et al.* A mass spectrometry-based proteome map of drug action in lung cancer cell lines. *Nat Chem Biol* **16**, 1111-1119, doi:10.1038/s41589-020-0572-3 (2020).

16 Mitchell, D. C. *et al.* A proteome-wide atlas of drug mechanism of action. *Nat Biotechnol*, doi:10.1038/s41587-022-01539-0 (2023).

17 Zecha, J. *et al.* Decrypting drug actions and protein modifications by dose- and time-resolved proteomics. *Science* **380**, 93-101, doi:10.1126/science.ade3925 (2023).
18 Lee, M. J. *et al.* Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780-794, doi:10.1016/j.cell.2012.03.031 (2012).
19 Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e1417, doi:10.1016/j.cell.2017.10.049 (2017).
20 Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929-1935, doi:313/5795/1929 [pii] 10.1126/science.1132939 (2006).
21 Vlahos, L. *et al.* Systematic, Protein Activity-based Characterization of Single Cell State. *bioRxiv* **bioRxiv 2021.05.20.445002**, doi:https://doi.org/10.1101/2021.05.20.445002 (2023).
22 Bush, E. C. *et al.* PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. *Nature communications* **8**, 105, doi:10.1038/s41467-017-00136-z (2017).
23 Paull, E. O. *et al.* A modular master regulator landscape controls cancer transcriptional identity. *Cell* **184**, 334-351 e320, doi:10.1016/j.cell.2020.11.045 (2021).
24 Alvarez, M. J. *et al.* A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat Genet* **50**, 979-989, doi:10.1038/s41588-018-0138-4 (2018).
25 Alvarez, M. J. *et al.* Reply to 'H-STS, L-STS and KRJ-I are not authentic GEPNET cell lines'. *Nat Genet* **51**, 1427-1428, doi:10.1038/s41588-019-0509-5 (2019).
26 Bansal, M. *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol* **32**, 1213-1222, doi:10.1038/nbt.3052 (2014).
27 Lundgren, K. *et al.* BIIB021, an orally available, fully synthetic small-molecule inhibitor of the heat shock protein Hsp90. *Mol Cancer Ther* **8**, 921-929, doi:10.1158/1535-7163.MCT-08-0758 (2009).
28 Trudeau, S. J. *et al.* PrePCI: A structure- and chemical similarity-informed database of predicted protein compound interactions. *Protein Sci* **32**, e4594, doi:10.1002/pro.4594 (2023).
29 Hwang, H., Dey, F., Petrey, D. & Honig, B. Structure-based prediction of ligand-protein interactions on a genome-wide scale. *Proc Natl Acad Sci U S A* **114**, 13685-13690, doi:10.1073/pnas.1705381114 (2017).
30 Shelanski, M. *et al.* A systems approach to drug discovery in Alzheimer's disease. *Neurotherapeutics* **12**, 126-131, doi:10.1007/s13311-014-0335-5 (2015).
31 Obradovic, A. *et al.* Systematic elucidation and pharmacological targeting of tumor-infiltrating regulatory T cell master regulators. *Cancer Cell* **41**, 933-949 e911, doi:10.1016/j.ccell.2023.04.003 (2023).
32 Laise, P. *et al.* The Host Cell ViroCheckpoint: Identification and Pharmacologic Targeting of Novel Mechanistic Determinants of Coronavirus-Mediated Hijacked Cell States. *bioRxiv*, doi:10.1101/2020.05.12.091256 (2020).

33 Szalai, B. *et al.* Signatures of cell death and proliferation in perturbation transcriptomics data-from confounding factor to effective prediction. *Nucleic Acids Res* **47**, 10010-10026, doi:10.1093/nar/gkz805 (2019).
34 Lukacisin, M. & Bollenbach, T. Emergent Gene Expression Responses to Drug Combinations Predict Higher-Order Drug Interactions. *Cell Syst* **9**, 423-433 e423, doi:10.1016/j.cels.2019.10.004 (2019).
35 Alvarez, M. J. *et al.* Unbiased Assessment of H-STS cells as high-fidelity models for gastro-enteropancreatic neuroendocrine tumor drug mechanism of action analysis. *bioRxiv* **677435**, doi:doi: https://doi.org/10.1101/677435 (2019).
36 Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat Genet* **37**, 382-390, doi:10.1038/ng1532 (2005).
37 Zhu, J. *et al.* Prediction of drug efficacy from transcriptional profiles with deep learning. *Nat Biotechnol* **39**, 1444-1452, doi:10.1038/s41587-021-00946-z (2021).
38 Smith, I. & Haibe-Kains, B. Similarity bias from averaging signatures from the Connectivity Map. *bioRxiv*, 2022.2001.2024.477615, doi:10.1101/2022.01.24.477615 (2022).
39 Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* **46**, D1074-D1082, doi:10.1093/nar/gkx1037 (2018).
40 Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**, 637-643, doi:10.1038/nature04670 (2006).
41 Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068-1081, doi:10.1016/j.cell.2012.08.011 (2012).
42 Wan, C. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339-344, doi:10.1038/nature14877 (2015).
43 Hu, L. Z. *et al.* EPIC: software toolkit for elution profile-based inference of protein complexes. *Nat Methods* **16**, 737-742, doi:10.1038/s41592-019-0461-4 (2019).
44 Xu, C. *et al.* Global Landscape of Native Protein Complexes in Synechocystis sp. PCC 6803. *Genomics Proteomics Bioinformatics* **20**, 715-727, doi:10.1016/j.gpb.2020.06.020 (2022).
45 Pourhaghighi, R. *et al.* BraInMap Elucidates the Macromolecular Connectivity Landscape of Mammalian Brain. *Cell Syst* **11**, 208, doi:10.1016/j.cels.2020.08.006 (2020).
46 Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* **9**, 471-472, doi:10.1038/nmeth.1938 (2012).
47 Shiotsu, Y. *et al.* KW-2449, a novel multikinase inhibitor, suppresses the growth of leukemia cells with FLT3 mutations or T315I-mutated BCR/ABL translocation. *Blood* **114**, 1607-1617, doi:10.1182/blood-2009-01-199307 (2009).
48 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).
49 Eramo, M. J. & Mitchell, C. A. Regulation of PtdIns(3,4,5)P3/Akt signalling by inositol polyphosphate 5-phosphatases. *Biochem Soc Trans* **44**, 240-252, doi:10.1042/BST20150214 (2016).
50 Dey, F., Cliff Zhang, Q., Petrey, D. & Honig, B. Toward a "structural BLAST": using structural relationships to infer function. *Protein Sci* **22**, 359-366, doi:10.1002/pro.2225 (2013).

51 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).
52 Bajusz, D., Racz, A. & Heberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* **7**, 20, doi:10.1186/s13321-015-0069-3 (2015).
53 Kim, S. *et al.* PubChem 2023 update. *Nucleic Acids Res* **51**, D1373-D1380, doi:10.1093/nar/gkac956 (2023).
54 Leist, T. P. & Weissert, R. Cladribine: mode of action and implications for treatment of multiple sclerosis. *Clin Neuropharmacol* **34**, 28-35, doi:10.1097/WNF.0b013e318204cd90 (2011).
55 Rao, R., Fiskus, W., Ganguly, S., Kambhampati, S. & Bhalla, K. N. HDAC inhibitors and chaperone function. *Adv Cancer Res* **116**, 239-262, doi:10.1016/B978-0-12-394387-3.00007-0 (2012).
56 Gontarewicz, A. *et al.* Simultaneous targeting of Aurora kinases and Bcr-Abl kinase by the small molecule inhibitor PHA-739358 is effective against imatinib-resistant BCR-ABL mutations including T315I. *Blood* **111**, 4355-4364, doi:10.1182/blood-2007-09-113175 (2008).
57 Gelbert, L. M. *et al.* Preclinical characterization of the CDK4/6 inhibitor LY2835219: in-vivo cell cycle-dependent/independent anti-tumor activities alone/in combination with gemcitabine. *Invest New Drugs* **32**, 825-837, doi:10.1007/s10637-014-0120-7 (2014).
58 Heinzlmeir, S. *et al.* Chemical Proteomics and Structural Biology Define EPHA2 Inhibition by Clinical Kinase Drugs. *ACS Chem Biol* **11**, 3400-3411, doi:10.1021/acschembio.6b00709 (2016).
59 Chen, P. *et al.* Spectrum and Degree of CDK Drug Interactions Predicts Clinical Performance. *Mol Cancer Ther* **15**, 2273-2281, doi:10.1158/1535-7163.MCT-16-0300 (2016).
60 Takagi, M. *et al.* In vivo antitumor activity of a novel sulfonamide, HMN-214, against human tumor xenografts in mice and the spectrum of cytotoxicity of its active metabolite, HMN-176. *Invest New Drugs* **21**, 387-399, doi:10.1023/a:1026282716250 (2003).
61 Hsiang, Y. H., Jiang, J. B. & Liu, L. F. Topoisomerase II-mediated DNA cleavage by amonafide and its structural analogs. *Mol Pharmacol* **36**, 371-376 (1989).
62 Furr, B. J. & Tucker, H. The preclinical development of bicalutamide: pharmacodynamics and mechanism of action. *Urology* **47**, 13-25; discussion 29-32, doi:10.1016/s0090-4295(96)80003-3 (1996).
63 Ferguson, P. J. *et al.* Antisense down-regulation of thymidylate synthase to suppress growth and enhance cytotoxicity of 5-FUdR, 5-FU and Tomudex in HeLa cells. *Br J Pharmacol* **127**, 1777-1786, doi:10.1038/sj.bjp.0702728 (1999).
64 Bradley, C. A. *et al.* Targeting c-MET in gastrointestinal tumours: rationale, opportunities and challenges. *Nat Rev Clin Oncol* **14**, 562-576, doi:10.1038/nrclinonc.2017.40 (2017).
65 Henley, M. J. & Koehler, A. N. Advances in targeting 'undruggable' transcription factors with small molecules. *Nat Rev Drug Discov* **20**, 669-688, doi:10.1038/s41573-021-00199-0 (2021).

66 Aytes, A. *et al.* Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638-651, doi:10.1016/j.ccr.2014.03.017 (2014).

67 Carro, M. S. *et al.* The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318-325, doi:nature08712 [pii] 10.1038/nature08712 (2010).

68 Klijn, C. *et al.* A comprehensive transcriptional portrait of human cancer cell lines. *Nat Biotechnol* **33**, 306-312, doi:10.1038/nbt.3080 (2015).

69 Anastas, J. N. & Moon, R. T. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer* **13**, 11-26, doi:10.1038/nrc3419 (2013).

70 Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764, doi:10.1126/science.1147939 (2007).

71 Unoki, M., Brunet, J. & Mousli, M. Drug discovery targeting epigenetic codes: the great potential of UHRF1, which links DNA methylation and histone modifications, as a drug target in cancers and toxoplasmosis. *Biochem Pharmacol* **78**, 1279-1288, doi:10.1016/j.bcp.2009.05.035 (2009).

72 Kostyrko, K. *et al.* UHRF1 is a mediator of KRAS driven oncogenesis in lung adenocarcinoma. *Nature communications* **in press** (2023).

73 McKeown, M. R. & Bradner, J. E. Therapeutic strategies to inhibit MYC. *Cold Spring Harb Perspect Med* **4**, doi:10.1101/cshperspect.a014266 (2014).

74 Feng, J., Huang, H. & Yen, T. J. CENP-F is a novel microtubule-binding protein that is essential for kinetochore attachments and affects the duration of the mitotic checkpoint delay. *Chromosoma* **115**, 320-329, doi:10.1007/s00412-006-0049-5 (2006).

75 Aytes, A. *et al.* Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638-651, doi:10.1016/j.ccr.2014.03.017 (2014).

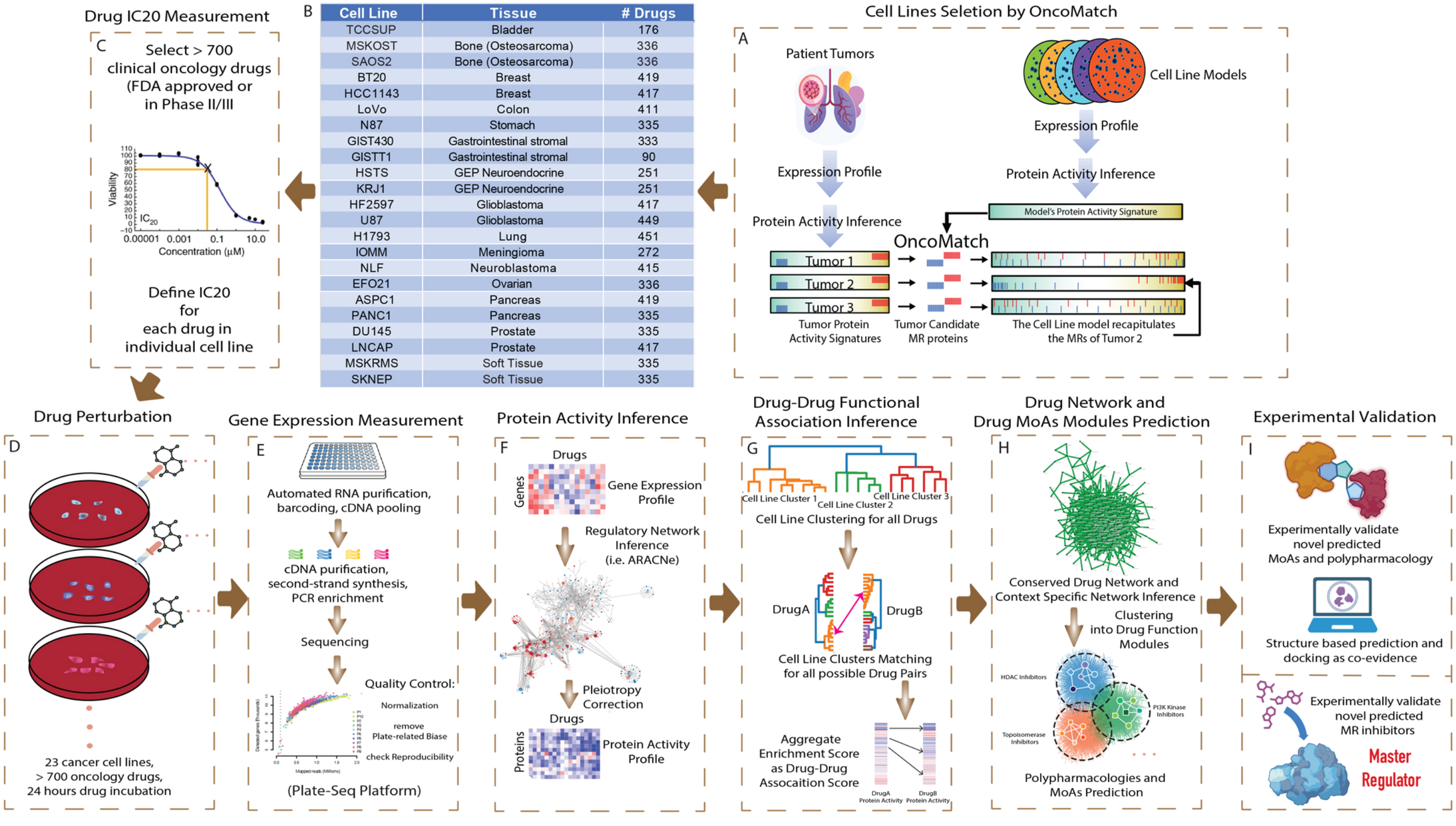
76 Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability. *Bmc Bioinformatics* **22**, 433, doi:10.1186/s12859-021-04344-9 (2021).

77 Myant, K. *et al.* Serine 62-Phosphorylated MYC Associates with Nuclear Lamins and Its Regulation by CIP2A Is Essential for Regenerative Proliferation. *Cell Rep* **12**, 1019-1031, doi:10.1016/j.celrep.2015.07.003 (2015).

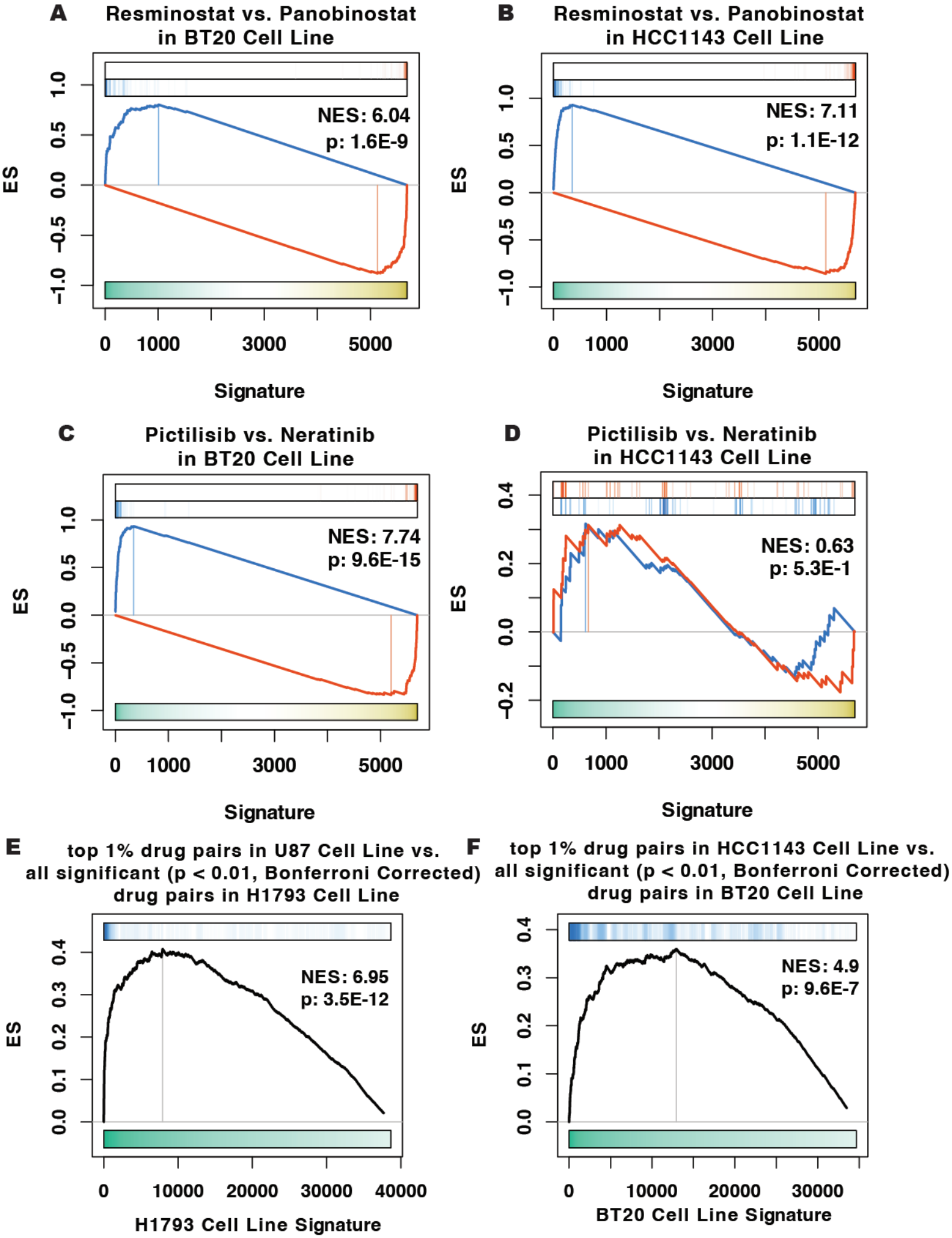
78 Alvarez, M. J. & Califano, A. *Darwin OncoTarget/OncoTreat: NY CLIA certified tests to identify effective drugs on an individual cancer patient basis from RNASeq profiles* <https://www.pathology.columbia.edu/diagnostic-specialties/personalized-genomic-medicine/oncology-testing/darwin-oncotarget-tm-oncotreat> (2018).

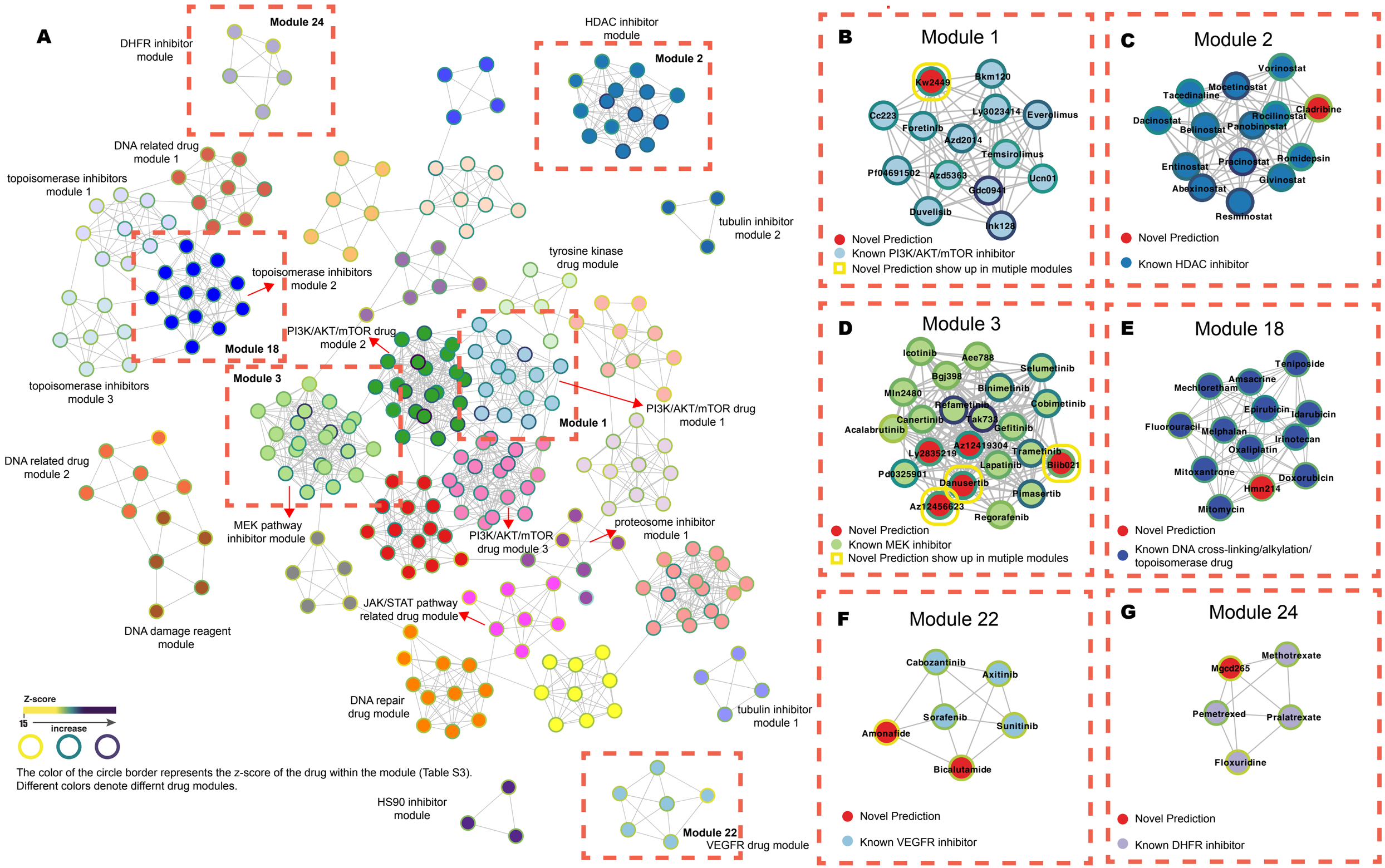
79 Zeleke, T. Z. *et al.* Network-based assessment of HDAC6 activity predicts preclinical and clinical responses to the HDAC6 inhibitor ricolinostat in breast cancer. *Nat Cancer* **4**, 257-275, doi:10.1038/s43018-022-00489-5 (2023).

80 Laise, P. *et al.* A model for network-based identification and pharmacological targeting of aberrant, replication-permissive transcriptional programs induced by viral infection. *Commun Biol* **5**, 714, doi:10.1038/s42003-022-03663-8 (2022).

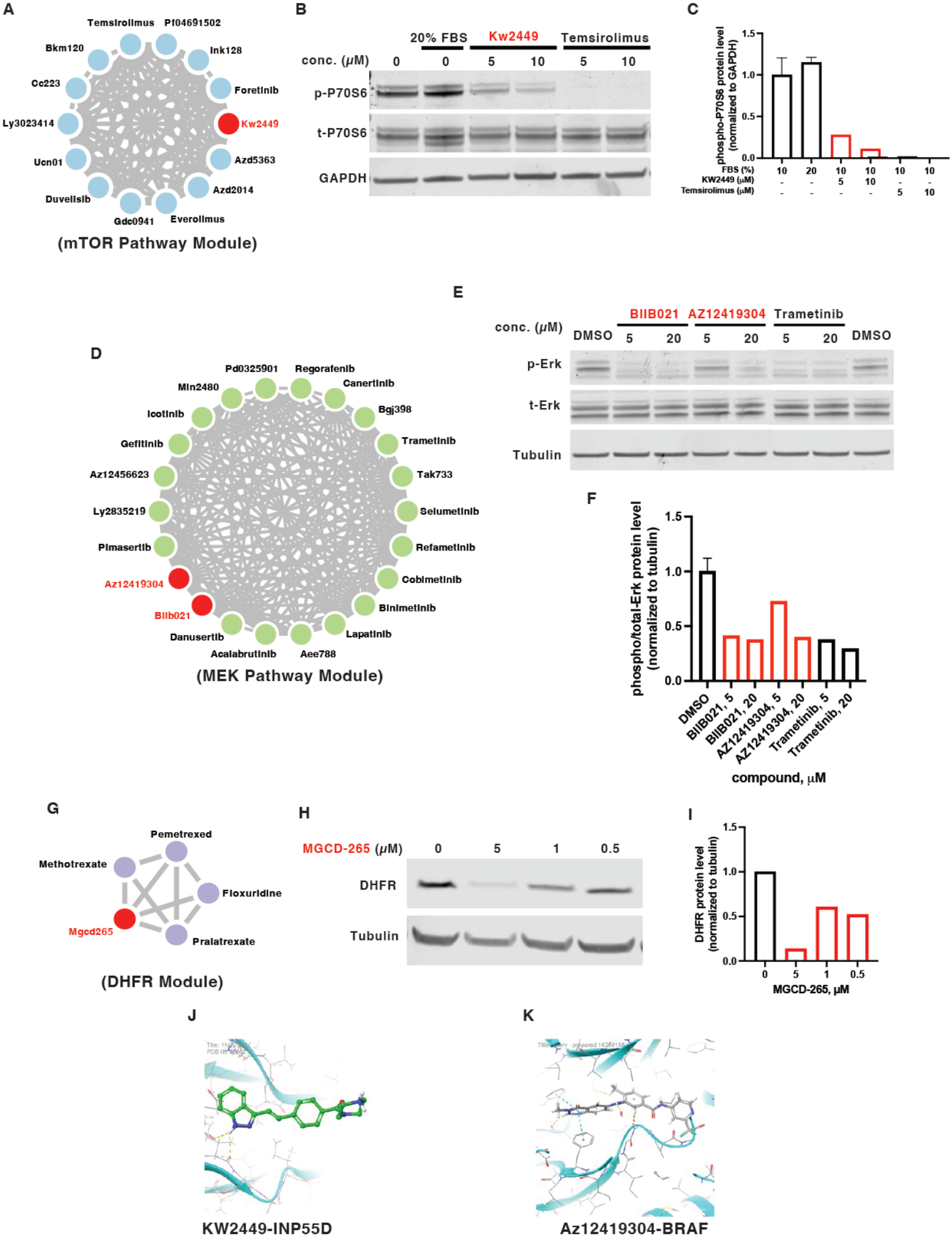


Cell Line	Tissue	# Drugs
TCCSUP	Bladder	176
MSKOST	Bone (Osteosarcoma)	336
SAOS2	Bone (Osteosarcoma)	336
BT20	Breast	419
HCC1143	Breast	417
LoVo	Colon	411
N87	Stomach	335
GIST430	Gastrointestinal stromal	333
GISTT1	Gastrointestinal stromal	90
HSTS	GEP Neuroendocrine	251
KRJ1	GEP Neuroendocrine	251
HF2597	Glioblastoma	417
U87	Glioblastoma	449
H1793	Lung	451
IOMM	Meningioma	272
NLF	Neuroblastoma	415
EFO21	Ovarian	336
ASPC1	Pancreas	419
PANC1	Pancreas	335
DU145	Prostate	335
LNCAP	Prostate	417
MSKRMS	Soft Tissue	335
SKNEP	Soft Tissue	335

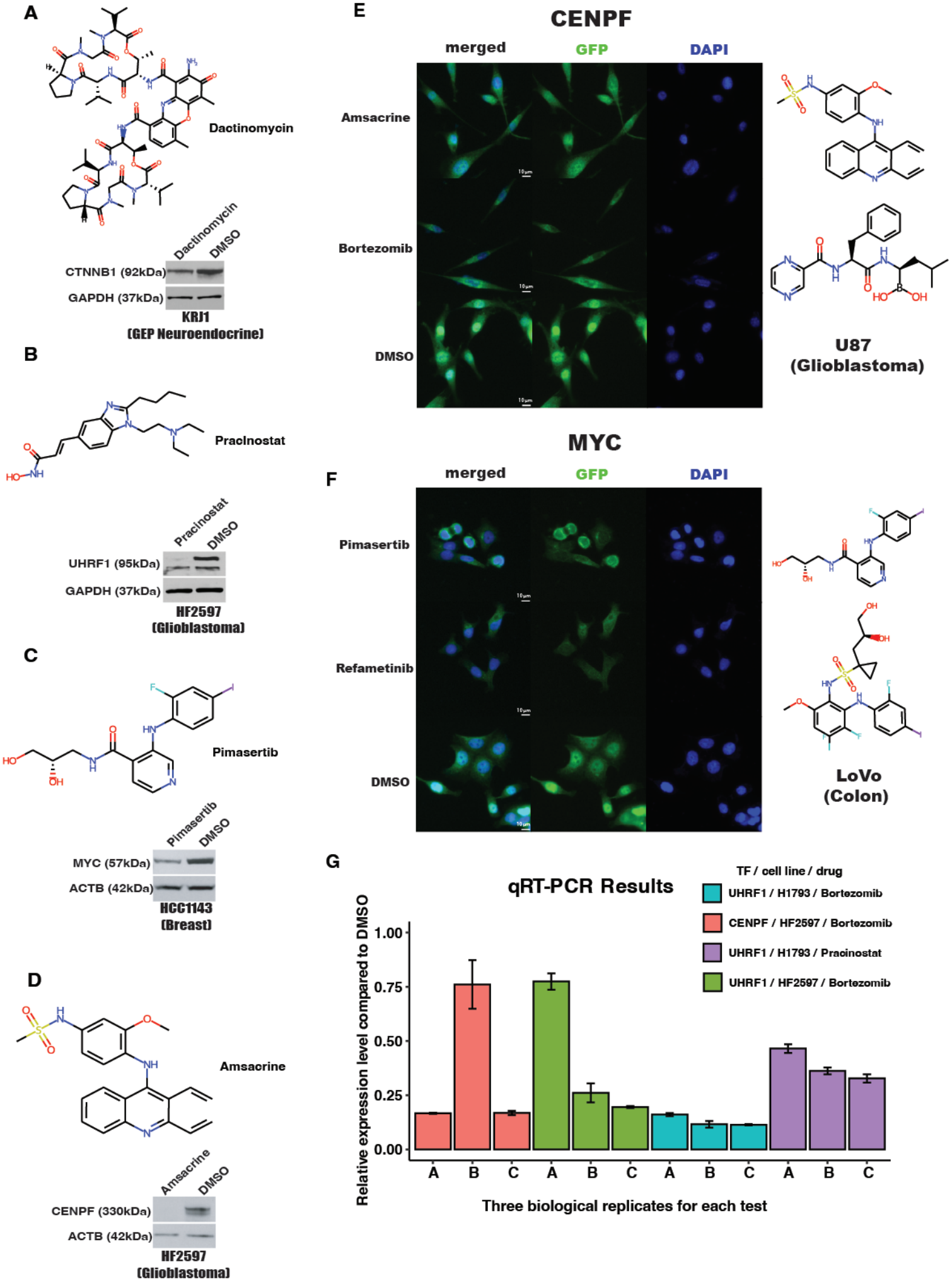

A
Resminostat vs. Panobinostat
in BT20 Cell Line
NES: 6.04
p: 1.6E-9
ES
Signature
B
Resminostat vs. Panobinostat
in HCC1143 Cell Line
NES: 7.11
p: 1.1E-12
ES
Signature
C
Pictilisib vs. Neratinib
in BT20 Cell Line
NES: 7.74
p: 9.6E-15
ES
Signature
D
Pictilisib vs. Neratinib
in HCC1143 Cell Line
NES: 0.63
p: 5.3E-1
ES
Signature
E
top 1% drug pairs in U87 Cell Line vs.
all significant (p < 0.01, Bonferroni Corrected)
drug pairs in H1793 Cell Line
NES: 6.95
p: 3.5E-12
ES
H1793 Cell Line Signature
F
top 1% drug pairs in HCC1143 Cell Line vs.
all significant (p < 0.01, Bonferroni Corrected)
drug pairs in BT20 Cell Line
NES: 4.9
p: 9.6E-7
ES
BT20 Cell Line Signature



The color of the circle border represents the z-score of the drug within the module (Table S3). Different colors denote differnt drug modules.


A
Temsirolimus
Pf04691502
Bkm120
Ink128
Cc223
Foretinib
Ly3023414
Kw2449
Ucn01
Azd5363
Duvelisib
Azd2014
Gdc0941
Everolimus
(mTOR Pathway Module)
B
20% FBS
Kw2449
Temsirolimus
conc. (μM)
0
0
5
10
5
10
p-P70S6
t-P70S6
GAPDH
C
phospho-P70S6 protein level (normalized to GAPDH)
1.5
1.0
0.5
0.0
FBS (%)
KW2449 (μM)
Temsirolimus (μM)
D
Pd0325901
Regorafenib
Min2480
Canertinib
Icotinib
Bgj398
Gefitinib
Trametinib
Az12456623
Tak733
Ly2835219
Selumetinib
Pimasertib
Refametinib
Az12419304
Cobimetinib
Binimetinib
Bilb021
Danusertib
Lapatinib
Acalabrutinib
Aee788
(MEK Pathway Module)
E
BIIB021
AZ12419304
Trametinib
conc. (μM)
DMSO
5
20
5
20
5
20
DMSO
p-Erk
t-Erk
Tubulin
F
phospho/total-Erk protein level (normalized to tubulin)
1.5
1.0
0.5
0.0
DMSO
BIIB021, 5
BIIB021, 20
AZ12419304, 5
AZ12419304, 20
Trametinib, 5
Trametinib, 20
compound, μM
G
Pemetrexed
Methotrexate
Floxuridine
Mgcd265
Pralatrexate
(DHFR Module)
H
MGCD-265 (μM)
0
5
1
0.5
DHFR
Tubulin
I
DHFR protein level (normalized to tubulin)
1.5
1.0
0.5
0.0
0
5
1
0.5
MGCD-265, μM
J
KW2449-INP55D
K
Az12419304-BRAF

A
Protein Activity Heatmap for most potent Drugs for 37 MRs across 23 Cell Lines
37 MRs
SAOS2
IOMM
N87
TCCSUP
H1793
KRJ1
EFO21
LOVO
LNCAP
NLF
ASPC1
MASKRMS
HSTS
GISTT1
GIST430
U87
HCC1143
PANC1
HF2597
DU145
MSKOST
BT20
SKNEP
B
Cell Lines
CTNNB1
MYC
STAT1
CENPF
UHRF1
OVOL1
JUP
HDAC1
C
Drugs
Number of cell lines that drug is potent


A
Dactinomycin
Dactinomycin
DMSO
CTNNB1 (92kDa)
GAPDH (37kDa)
KRJ1
(GEP Neuroendocrine)
B
Pracinostat
Pracinostat
DMSO
UHRF1 (95kDa)
GAPDH (37kDa)
HF2597
(Glioblastoma)
C
Pimasertib
Pimasertib
DMSO
MYC (57kDa)
ACTB (42kDa)
HCC1143
(Breast)
D
Amsacrine
Amsacrine
DMSO
CENPF (330kDa)
ACTB (42kDa)
HF2597
(Glioblastoma)
E
CENPF
merged
GFP
DAPI
Amsacrine
Bortezomib
DMSO
U87
(Glioblastoma)
F
MYC
merged
GFP
DAPI
Pimasertib
Refametinib
DMSO
LoVo
(Colon)
G
qRT-PCR Results
TF / cell line / drug
UHRF1 / H1793 / Bortezomib
CENPF / HF2597 / Bortezomib
UHRF1 / H1793 / Pracinostat
UHRF1 / HF2597 / Bortezomib
Relative expression level compared to DMSO
1.00
0.75
0.50
0.25
0.00
A B C A B C A B C A B C
Three biological replicates for each test