An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story.

Accessibility

An interactive resource to identify

cancer genetic and lineage dependencies

targeted by small molecules

Amrita Basu[1]*, Nicole E. Bodycombe[1]*, Jaime H. Cheah[1]*, Edmund V. Price*, Ke Liu*, Giannina I. Schaefer*, Richard Y. Ebright*, Michelle L. Stewart*, Daisuke Ito[†], Stephanie Wang*, Abigail L. Bracha*, Ted Liefeld*, Mathias Wawer*, Joshua C. Gilbert*, Andrew J. Wilson[≠], Nicolas Stransky[*†], Gregory V. Kryukov*, Vlado Dancik*, Jordi Barretina[*†], Levi A. Garraway*, C. Suk-Yee Hon*, Benito Munoz*, Joshua A. Bittker*, Brent R. Stockwell[€], Dineo Khabele[≠], Andrew M. Stern*, Paul A. Clemons[2]*, Alykhan F. Shamji[2]*, Stuart L. Schreiber[2]*

* The Broad Institute of Harvard and MIT, Cambridge, MA 02142

[≠] Vanderbilt University School of Medicine, Nashville, TN 37232

[€] Columbia University, New York, NY 10027

[†] Present addresses: H3 Biomedicine Inc., Cambridge, MA 02139 (D.I.), Blueprint Medicines, Cambridge, MA, 02142 (N.S.), Novartis Institutes for Biomedical Research (NIBR), Cambridge, MA 02139 (J.B.)

Contact email: pclemons@broadinstitute.org; ashamji@broadinstitute.org; stuart_schreiber@harvard.edu

---

[1] These authors contributed equally

[2] Corresponding authors

Summary

The high rate of clinical response to protein kinase-targeting drugs matched to cancer patients with specific genomic alterations has prompted efforts to use cancer cell-line (CCL) profiling to identify additional biomarkers of small-molecule sensitivities. We have quantitatively measured the sensitivity of 242 genomically characterized CCLs to an Informer Set of 354 small molecules that target many nodes in cell circuitry, uncovering protein dependencies that: 1) associate with specific cancer-genomic alterations and 2) can be targeted by small molecules. We have created the Cancer Therapeutics Response Portal (www.broadinstitute.org/ctrp) to enable users to correlate genetic features to sensitivity in individual lineages and control for confounding factors of CCL profiling. We report a candidate dependency, associating activating mutations in the oncogene β-catenin with sensitivity to the Bcl2-family antagonist, navitoclax. The resource can be used to develop novel therapeutic hypotheses and accelerate discovery of drugs matched to patients by their cancer genotype and lineage.

Highlights:

- We built a portal to identify cancer genotype/compound sensitivity relationships

- We correlated genetic features of cancer cell lines to their response to compounds

- We controlled for potential confounding factors of genomic cell-line profiling

- We suggest a strategy for treating cancers with mutations in the oncogene β-catenin

Introduction

Insights into cancer genomes and advances in small-molecule science are providing a foundation for future cancer therapeutics—ones linked to genomic alterations present in patients' cancers. Several drugs that target dependencies acquired by cancers as a result of somatic mutations or translocations are yielding high clinical response rates, although beneficial responses are observed in a fraction of cancer patients and not always durable (Gonzalez de Castro et al., 2013). Current targeted drugs inhibit protein kinases encoded by driver oncogenes or their wild-type alleles directly ('oncogene dependencies'). It is not known whether similar clinical responses can result from drugs targeting non-oncogenes that become essential for cancer survival or progression in the context of specific genetic features ('oncogene-induced dependencies'). To accelerate discovery of patient-matched therapies, systematic approaches are needed to identify: 1) the dependencies cancers acquire as a result of specific genetic features, and 2) small-molecule drugs that target the dependencies.

Cancer cell-line profiling has been used to reveal patterns of small-molecule sensitivities across diverse cancer cell lines (CCLs). These efforts initially focused on relating sensitivity to CCL lineage (Shoemaker, 2006), but now increasingly relate sensitivity to genetic and epigenetic features (Barretina et al., 2012; Garnett et al., 2012; Heiser et al., 2012; Larsen et al., 2011; Sharma et al., 2010; Sun et al., 2007). This approach identified dependencies on oncogenic alleles of *EGFR* and *BRAF* that are now exploited by targeted cancer therapeutics (McDermott et al., 2007). Manifestation of genetic dependencies in a lineage-restricted manner, for example sensitivity of V600E BRAF melanoma but not colorectal cancers to BRAF-targeting vemurafenib (Prahallad et al., 2012), highlights the need to integrate genetic and lineage features in CCL profiling.

CCL profiling studies have historically been limited in the quantity, diversity, or level of characterization of CCLs and small molecules used. One of the earliest CCL profiling efforts, the NCI-60, probed a set of 59 CCLs from various lineages with now $>10^5$ diverse small molecules. While this approach has been valuable for identifying lineage-selective small-molecule sensitivities, the relatively small number of CCLs and limited genomic characterization restricted the usefulness of these data. More recent studies have aimed to address this limitation. One recent study profiled 479 CCLs with significant genomic characterization against 24 anti-cancer drugs (Barretina et al., 2012). A second study profiled 350 CCLs against 130 pre-clinical or clinical anti-cancer agents, though the genomic alterations correlated to sensitivity were limited to ~70 genes (Garnett et al., 2012). In order for genomic and lineage CCL profiling to link cancer genetic alterations systematically with potential drug-targetable dependencies, we need to obtain sensitivity measurements for extensively characterized CCLs against a larger set of small molecules that span a broad array of cell processes.

Here, we provide a resource, the Cancer Therapeutics Response Portal (CTRP; www.broadinstitute.org/ctrp), that enables researchers to analyze relationships between genetic and lineage features of cancer and small-molecule sensitivity. We profiled the sensitivity of 242 CCLs to an Informer Set of small molecules with well-annotated targets and activities that collectively modulate a broader range of cellular processes than is currently being investigated in cancer drug discovery. We correlated the compound-sensitivity measurements of CCLs to their genomic alterations, identifying significant correlations involving 60% of the compounds tested and suggesting candidate dependencies on their targets. We intend the CTRP to be a living resource, incorporating new data over time

involving additional CCLs and compound treatments (single agent and combination), and new analyses linking sensitivity to additional types of cellular features.

Results

Creating an interactive resource

Profiling the sensitivity of CCLs to an Informer Set of small-molecule probes. The two main considerations for inclusion of small molecules in the Informer Set were high selectivity for their targets (e.g., rapamycin (Brown et al., 1995)), and/or collective targeting of many distinct nodes in cell circuitry. Compounds having different structures but targeting the same protein (e.g., cyclosporin A and tacrolimus targeting calcineurin (Liu et al., 1991)), and compounds having differential selectivity towards distinct members of a protein family (e.g., histone deacetylases (Pan et al., 2012)), were included to validate that genetic feature/sensitivity correlations can be attributed to dependency on a defined protein target. Compounds in clinical development, with strong selectivity data, or with pharmacokinetic data, were prioritized to enable rapid drug development. The current Informer Set comprises 35 FDA-approved drugs, 54 clinical candidates, and 266 probes (~30% prepared for this project by synthesis, **Table S1**).

The 242 CCLs (**Table S2j**), chosen to align with lineages studied by The Cancer Genome Atlas and in published genome-wide RNAi screens (Cheung et al., 2011), are a subset of the Cancer Cell Line Encyclopedia collection of ~1,000 genetically characterized CCLs. Data regarding gene expression, amplifications/deletions, somatic mutations in 1,645 cancer genes, and lineage/histological subtypes are freely available (www.broadinstitute.org/ccle). Each CCL was grown in its preferred media, plated at a density optimized during assay development

6

(**Table S2j**) and treated with compound at 8 concentrations for 72h. Sensitivity was assayed using CellTiter-Glo to measure cellular ATP levels as a surrogate for cell number and growth. The area under percent-viability curves (AUC) was computed as a metric of sensitivity (**Extended Experimental Procedures**) since AUC reflects both relative potency and total level of inhibition observed for a compound across CCLs (**Table S2g** and **Figure S1**).

Analysis of sensitivity data. The ability of genomic CCL profiling to identify clinically relevant biomarkers of drug response depends on the ability of CCLs to model tumor responses, which cannot be confirmed without patient-response data to the same perturbations. To evaluate the performance of CCLs in this study, we analyzed distributions of AUCs across all compounds to identify trends among various subpopulations (**Figure 1A**). While most CCLs respond differentially across our Informer Set, we observed that CCLs within specific tissue lineages and suspension CCLs were often more sensitive to many compounds tested (**Figure 1B**). These observations motivated us to perform analyses of AUC distributions that include all CCLs, as well analyses that exclude specific context-dependent subsets of CCLs to control for potential "confounding factors" (**Figure 1C** and **Figure S1**; **Extended Experimental Procedures**).

We performed statistics-based enrichment analyses that combined rank-based and parametric tests (**Experimental Procedures**) to identify genetic alterations and cellular features that are significantly enriched among sensitive (AUC < 3.5) or unresponsive (AUC > 5.5) CCLs. Analyses were performed for each compound across all CCLs and relevant subsets. These correlations are available as a table for download (**Table S2 and User Guide S1**), and those exceeding a specific threshold of statistical significance are visualized in the CTRP (**Extended Experimental Procedures**).

Querying the CTRP resource

Validating known dependencies. The resource identified several known mutation/sensitivity relationships, such as the increased sensitivity of *BRAF*-mutant CCLs to P-0850, an analog of the FDA-approved BRAF-V600E inhibitor, vemurafenib (Smalley, 2010) (**Figure 2A**). Inspecting P-0850-unresponsive V600E CCLs identified additional features previously associated with resistance to vemurafenib: 1) the unresponsive colorectal CCL, RKO, is reported to produce high levels of hepatocyte growth factor, which activates CRAF via MET in an autocrine fashion and circumvents dependence on BRAF (Corcoran et al., 2011; Straussman et al., 2012); and 2) the unresponsive CCL, SKMEL28, contains an activating mutation in *EGFR*, and enhanced EGFR signaling has been linked to resistance in vemurafenib-treated colon cancers (Prahallad et al., 2012). These data show the resource can identify candidate resistance mechanisms that may suggest rational combination therapies.

The resource identified increased sensitivity of *NRAS*- and *KRAS*-mutant CCLs to the MEK1/2 inhibitor, selumetinib, which has shown preliminary moderate activity in clinical trials with *KRAS*-mutant patients (Janne et al., 2012; National Cancer Institute, 2000; Yoon et al., 2011) (**Figure 2A**). Several mutant CCLs are unresponsive to selumetinib, suggesting that *KRAS/NRAS* may be one of several factors determining the response. Analysis of unresponsive outliers may reveal additional features modifying the response to selumetinib.

In some cases, genetic features correlated better with small-molecule sensitivity in the context of specific lineages. We observed that EGFR-mutant lung CCLs were highly sensitive to

neratinib, a dual ERBB2/EGFR inhibitor (**Figure 2A**) (Arteaga, 2006), currently in Phase II trials for advanced non-small cell lung cancer.

Mining for new dependencies. The CTRP suggests dependencies involving oncogenes for which targeted therapies are lacking. We observed that CCLs with *MYC* mutations, including those interfering with MYC protein degradation (Vervoorts et al., 2006), had increased sensitivity to (-)-gallocatechin-3-monogallate (GCG), a green tea-derived natural product (**Figure 2B**). Previous studies report that treatment of digestive tract-derived CCLs or mouse tumor models with epigallocatechin-3-monogallate, a GCG analog, led to decreased *MYC* expression (Ju et al., 2005; Ran et al., 2005). We also observed that mutations in *MYC* and, to a lesser degree, amplifications of *MYC* (**Figure 2B**) correlated with sensitivity to SB-225002, an inhibitor of a chemokine receptor (CXCR2) implicated in promoting oncogene-induced senescence (Acosta et al., 2008). Though the relationship of MYC and SB-225002-targeted biology is not understood, the correlation of sensitivity to SB-225002 with two different types of genomic alterations in *MYC* supports a potential connection.

We also identified small molecules with strong potency against CCLs of a specific lineage. Although they display a range of sensitivity, ovarian CCLs were among the most sensitive to two probes (ML210 and RSL3; **Figure 3A**) identified for their ability to kill oncogenically engineered cell lines selectively, BJeLR (HRasG12V, SV40 large T and small T antigens) and DRD (HRasG12V, hTERT, SV40 small T oncoprotein, dominant negative p53, cyclin D1, and mutant CDK4), relative to untransformed controls (Weiwer et al., 2012). Mutations in *HRAS* did not correlate with sensitivity to RSL3 or ML210 in CCLs used in our study. We confirmed the potency of ML210 against five ovarian CCLs (IC_{50} ~10 nM), including three not previously

profiled, using sulforhodamine B to detect cellular protein content as an assay for cytotoxicity (Skehan et al., 1990) (**Figure 3B**). Treatment of SKOV3 cells with ML210 also increased expression of the DNA-damage marker phospho-H2AX (**Figure S3B**) and levels of cleaved caspase-3 (**Figure S3C**), a marker of apoptosis, suggesting that ML210 is cytotoxic. Similar results were obtained with ML162, another probe identified in the phenotypic screen that yielded ML210 (**Figures S3A,B,C**).

RSL3 and related compounds are thought to induce cell death via ferroptosis (Dixon et al., 2012), though they appear to promote markers of apoptosis in this context. It is possible that the sensitive CCLs, whether of ovarian or other lineages, have features in common with the engineered cells described above that render them sensitive to ferroptosis modulators. Testing these compounds in more CCLs and performing multi-feature correlation analyses may help uncover these features.

Global analyses. We determined whether studying gene/compound connections as sets rather than individually could yield insights about dependencies. We limited these analyses to connections having greater statistical significance than those included in **Table S2** (**Extended Experimental Procedures**) to ensure clustering experiments were not dominated by relatively weaker connections. Hierarchical clustering of compounds based on their profile of connections to genetic features (**Table S3i**) yielded several clusters of compounds that share similar mechanisms of action, including: PRIMA-1 and PRIMA-1-Met (re-activators of mutant p53 signaling), FK866 and GMX-1778 (NAMPT inhibitors), neopeltolide and leucascandrolide A (modulators of respiration), and teniposide and etoposide (topoisomerase inhibitors) (**Figure S2**).

We also analyzed the frequency with which mutated genes correlate with sensitivity or unresponsiveness to different compounds (**Table S3**). We found several genes (*STK11*, *EGFR*, *BRAF*) correlated with unresponsiveness to many compounds. *EGFR*-mutated CCLs were unresponsive to different compounds with the same mechanism of action (NAMPT inhibition; **Figure 2C**). For example, the top-ranked gene, *STK11*, has been implicated in resistance to docetaxel in a murine model of KRAS-mutant lung cancer (Chen et al., 2012). Similarly, retrospective clinical analyses indicate patients with *BRAF*-mutant metastatic colon cancers tend to be non-responsive to EGFR-targeted therapy (Sartore-Bianchi et al., 2009). Our CCL data suggest a possibility that mutations driving certain cancers may lead to unresponsiveness to a wide range of small molecules.

CCLs with activating mutations in β-catenin are more sensitive to navitoclax

The resource suggests oncogene-induced dependencies involving oncogenic alleles of the transcription factor β-catenin (*CTNNB1*) and alterations in genes encoding proteins that regulate β-catenin stability. *CTNNB1* is mutated in several cancer types, yet no targeted treatment has been identified. Activating mutations in the *CTNNB1* degradation box (amino acids 32-45) are known to interfere with its phosphorylation and proteasomal degradation, leading to aberrant increases in protein levels (Sparks et al., 1998).

We found that *CTNNB1*-mutant CCLs were among those most sensitive to navitoclax (**Figure 4A**), an inhibitor of anti-apoptotic BCL2 family members (BCL-xL, BCL2, BCL-w, but not MCL-1 or BFL-1/A1) previously studied in clinical trials (Gandhi et al., 2011). In studying other proteins regulating β-catenin degradation (APC, AXIN1, CSNK1A1, GSK3B, βTRC), we found that alterations in *AXIN1* and *CSNK1A1* also correlate with sensitivity to navitoclax (**Figure**

4A). Collectively, these functionally related alterations account for 37% of the CCLs most sensitive to navitoclax, suggesting alterations increasing β-catenin levels or activity may create a dependency of cancer cells on BCL2 family members for survival (**Figure 4B**). Our results are consistent with a recent study showing the level of β-catenin pathway activity in CCLs correlates with sensitivity to knockdown of *BCL2L1* (encodes BCL-xL) (Rosenbluh et al., 2012). Increased β-catenin activity has also been linked to enhanced expression of *BCL2* and *BCL2L1* (Kaga et al., 2006; Rosenbluh et al., 2012) and to suppression of BAX-mediated apoptosis (Wang et al., 2009). We do not observe a correlation between *CTNNB1* mutation and BCL-2, BCL-xL, or BCL-W protein levels (**Figure S4A**). We note that in some lineages, unresponsive CCLs lacking *CTNNB1* mutations have increased MCL-1 protein levels, which is reported to confer resistance to navitoclax (Tahir et al., 2010). We also observed that correlation between *MCL1* gene expression and unresponsiveness to navitoclax ranked highly (top 2%) compared to all other genes (**Table S5**; **Extended Experimental Procedures**).

Analytical and experimental confirmation in CCLs. We analyzed the sensitivity data using an orthogonal analytical approach, elastic-net regression (Zhu and Hastie, 2004), that aims to identify a parsimonious model that best predicts response to navitoclax. This analysis used both somatic mutation and copy-number data for each gene as candidate predictive features. Consistent with enrichment analysis, mutation of *CTNNB1* is among the top-ranked features in predicting sensitivity to navitoclax (**Figure 4C and Table S4**). To assess model performance using ten-fold cross validation, we calculated a root-mean-squared error between the predicted and observed sensitivities and compared this value to the day-to-day variability of AUCs in our profiling data (AUC=0.98, 90th percentile). Our estimated prediction error (1.45) is greater but comparable to the biological replicate variability of AUCs. Ingenuity Pathway Analysis (IPA)

(Jimenez-Marin et al., 2009) of the full list of predictive features revealed that a majority of the genes in our model are directly linked via interactions annotated in the Ingenuity knowledge base. Our highest-scoring IPA network ($p=10^{-38}$) identified β-catenin as a centrally connected node in the network, and links it to destruction complex members CSNK1A1 and APC, as well as anti-apoptotic BCL2, a target of navitoclax (**Figure 4D**).

To confirm the relative sensitivity of *CTNNB1*-mutant CCLs observed in the large-scale profiling data, we re-tested navitoclax in a subset of CCLs. We first examined a panel of lineage-matched non-mutant and *CTNNB1*-mutant CCLs and confirmed that mutant CCLs had increased CTNNB1 protein levels (**Figure S4A**) (Sparks et al., 1998) and *AXIN2* expression levels (**Figure S4B**) (Jho et al., 2002). We then re-tested the sensitivity of seven *CTNNB1*-mutant and four non-mutant CCLs to 72-hr treatment with navitoclax (CellTiterGlo); the AUCs were similar to those in our profiling data (**Figure 5A**). We also found that the CCLs most sensitive to navitoclax elicited the largest increase in caspase 3/7 activation, as measured by Caspase-Glo, indicating that loss of viability resulted from apoptosis (**Figure 5B**) (Tse et al., 2008). We also tested five previously untested CCLs with *CTNNB1* mutations in the degradation box and their sensitivity to navitoclax was similar to the original *CTNNB1*-mutant lines (**Figure 5C**). These data support our hypothesis that mutations in *CTNNB1* and alterations in its destruction complex are biomarkers for sensitivity to navitoclax.

Our data suggest that *CTNNB1* mutations that increase β-catenin protein levels sensitize cells to navitoclax. We reasoned that small molecules that increase β-catenin protein levels may also sensitize cells to navitoclax in the absence of *CTNNB1* mutations. To explore this hypothesis, we tested CCLs for differential changes in β-catenin levels following a 3d treatment with CHIR-99021, a GSK3β inhibitor that prevents phosphorylation and degradation of β-

catenin (Bennett et al., 2002). We identified four CCLs for further testing: RKO and HT29, which lack *CTNNB1* mutations and showed significant CHIR-99021-induced increase in β-catenin (**Figure 6A**); HEC59, which lacks *CTNNB1* mutations and showed little change in β-catenin in response to CHIR-99021; and SW48, which contains a mutant GSK3β phosphorylation site in *CTNNB1* (S33Y) and did not increase β-catenin in response to CHIR-99021 (**Figure 6D**).

Each CCL was pre-treated with CHIR-99021 and then co-treated with both CHIR-99021 and navitoclax. ATP levels were measured as a surrogate for cell viability. The IC_{50} of both RKO and HT29 response to navitoclax shifted 4-to-8-fold lower (**Figures 6B and 6C**) after pre-treatment with CHIR-99021 and was more pronounced than after only co-treatment (**Figure S5A**). Neither CHIR-99021 pre-treatment nor co-treatment significantly increased sensitivity to navitoclax in HEC59 (**Figure 6E, Figure S5B**) or in SW48 (**Figure 6F, Figure S5C**). These data suggest that increasing β-catenin levels may correlate with increased sensitivity to navitoclax. We have not determined whether increasing protein levels leads to increased activity, which we observed also correlates to navitoclax sensitivity. Of the 242 CCLs profiled in our study, 41 were tested previously for β-catenin activity in a TCF4 reporter assay (Rosenbluh et al., 2012). Of those 41 CCLs, 8 were sensitive and the remaining 33 unresponsive to navitoclax in our assay. Of the 8 sensitive CCLs, 6 were active in the reporter assay, while only 7 of the 33 unresponsive CCLs were considered active in the reporter assay ($p < 0.05$; data not shown). Thus, compounds that increase β-catenin protein levels may also increase β-catenin activity, rendering them sensitive to navitoclax.

Discussion

Genomic and lineage CCL profiling offers an approach to identify cancer dependencies that are targetable with small molecules, and suggest combinations of compounds that mitigate drug resistance. The Cancer Therapeutic Response Portal (CTRP) suggests candidate dependencies associated with common and medically significant oncogenes. The first version of the CTRP resulted from profiling an Informer Set of small molecules, many of which target non-altered proteins that work in partnership with oncogenes. Exploiting oncogene-induced dependencies contrasts to a related approach based on targeting cell-biological 'hallmarks' common to cancers (Hanahan and Weinberg, 2011) without linking these 'non-oncogene addictions' to specific genomic alterations (Luo et al., 2009). For example, navitoclax has been tested in phase-I/II clinical trials for small-cell lung cancer (Gandhi et al., 2011); however, our data suggest that navitoclax might best be targeted to patients harboring *CTNNB1* mutations, which are present in colorectal, hepatocellular, gastric, and endometrial cancers. We observe that CTNNB1-mutant CCLs are sensitive to navitoclax in several lineages, though more strongly in some (e.g., gastric) than others. The same selectivity was not observed for ABT-199, a BCL-2 specific inhibitor (Souers et al., 2013) (data not shown), suggesting that inhibition of other BCL-2 family members underlies the differential response. Consistently, Rosenbluh *et al.* recently showed that knockdown of *BCL2L1* (BCL-xL) in β-catenin-active CCLs impairs proliferation (Rosenbluh et al., 2012), implicating BCL-xL as a relevant target for navitoclax in *CTNNB1*-mutant cancers.

Profiling data for single agents may also suggest drug combinations to prevent or overcome drug resistance. By studying the response of BRAF-V600E-mutant CCLs to V600E inhibition, we show how outlier cell lines unresponsive to a small molecule in an otherwise sensitive cohort can reveal additional features that correlate with and confer resistance, in this case

upregulation of HGF. Combined treatment with a MET inhibitor to block HGF signaling was sufficient to sensitize these cells to BRAF-V600E inhibition (Corcoran et al., 2011; Straussman et al., 2012). These observations also suggest that correlating small-molecule response to groups of features rather than individual ones may yield biomarkers with greater predictive accuracy.

Since the same oncogene may give rise to different dependencies in different cancer types (e.g., BRAF in melanoma vs. colorectal), the CTRP has been built to allow users to identify dependencies in all CCLs or only in specific lineages. For example, we find that *KRAS* mutations correlate significantly with sensitivity to navitoclax among colorectal CCLs, but not among all CCLs (data not shown). Interestingly, Corcoran *et al.* recently showed that navitoclax synergizes with selumetinib to kill *KRAS*-mutant CCLs in several lineages, but most strongly and consistently in colorectal CCLs (Corcoran et al., 2013).

Corcoran *et al.* also highlight an important lesson for interpreting CCL profiling data. The authors attribute a lack of efficacy of selumitinib as a single agent in *KRAS*-mutant tumors to the fact that it is largely cytostatic rather than cytotoxic. Combination with navitoclax, which activates apoptosis, was required for induction of cell death. We note that most CCL profiling data, gathered using a readout for cell growth or proliferation rather than death, may identify gene/sensitivity relationships involving cytostasis; indeed, while *KRAS*-mutant lines are among those most affected by selumetinib in our study, the compound only leads to partial inhibition of ATP levels, suggestive of cell growth inhibition. Corcoran *et al.* describe a screening approach for how cytostatic compounds with selectivity for specific cancer genotypes might be exploited in combination strategies to achieve greater efficacy. It also suggests the importance of

considering level of inhibition in analyses of existing data, and motivates incorporation of scaleable assays for cell death in future data collection.

The CTRP currently associates small-molecule sensitivity with individual features. In some cases, multiple features associated with sensitivity co-occur in the same CCLs, making it challenging to interpret if an associated feature is causal. For example, we observe that hematopoietic/lymphoid CCLs are more sensitive than those from other lineages to many compounds, including the BRD4 inhibitor JQ-1. *MYC*-mutant CCLs are among those most sensitive to JQ-1, but they are also frequently of hematopoietic/lymphoid origin, making it difficult to assess whether the genetic or lineage feature is the key determinant of sensitivity. While in this example a larger set of mutant lines will be needed to study *MYC* mutations separately within hematopoetic/lymphoid and solid tumor CCLs, the CTRP has been built to allow users to perform this analysis in general; for example, *MYC* amplification associates with sensitivity to SB-225002 whether all CCLs are analyzed or only those from solid cancers.

While CCLs have a long history as models for human cancer, their use in large-scale genomic CCL profiling has emerged more recently. Decisions associated with selection of CCLs, growth conditions, data collection (e.g., assay choice), data filtering (e.g., for possible confounding CCLs), data analysis, and formulation of questions in controlled computational experiments may contribute to differences in results and interpretation of existing profiling studies. For example, this study correlated *CTNNB1* mutations with sensitivity to navitoclax, while Garnett *et al.* correlated sensitivity to navitoclax with *NOTCH1* mutations. However, the portal from Garnett *et al.* suggests that *CTNNB1* mutations correlate with sensitivity to TW-37, a pan-Bcl-2 family inhibitor. Despite such differences, both studies identify several similar mutation/sensitivity connections including both established (e.g., *KRAS-NRAS*/selumetinib,

BRAF/V600E inhibitors) and novel associations (e.g., *CDKN2A*/GW-843682X, a PLK1 inhibitor), as well as features associated with unresponsiveness to compounds (e.g., *TP53*/nutlin-3). Encouragingly, there are no examples of targeted cancer drugs today that were not predicted by previous genomic CCL profiling studies.

The CTRP is available online (www.broadinstitute.org/ctrp) and the primary sensitivity data underlying the resource can be downloaded from the NCI-CTD[2] data portal (http://ctd2.nci.nih.gov). Genetic feature data for CCLs tested can be downloaded from the Broad/Novartis CCLE portal (http://www.broadinstitute.org/ccle).

The CTRP is expected to evolve to include additional data and analyses as they become available. We expect associations identified with the current dataset to change in strength as new lines are examined, and entirely new associations to be uncovered. We are extending this approach to test new probes and drugs, including compounds with novel physical and biological properties, or rationally selected combinations of compounds, across a larger set of CCLs. We are also undertaking systematic analyses to correlate sensitivity to combinations of cellular features, as well as other types of features, including gene expression, signatures of pathway activity (Liberzon et al., 2011), and activity of master regulators inferred computationally (Lefebvre et al., 2010). Further CCL annotations, such as metabolic, proteomic, and epigenetic profiles, will enable additional types of predictive biomarkers to be identified. Our hope is that the cancer biology community will use the CTRP to identify hypotheses for deeper investigation and to accelerate discovery of patient-targeted therapies with better treatment outcomes.

Experimental Procedures

Cancer Cell Line (CCL) Profiling

Frozen cells were obtained from the Broad Institute Biological Samples Platform or ATCC[TM]. CCLs were grown in their specified medium at 37°C/5% CO_2. Media were replaced every 2 days. Each CCL was tested for mycoplasma infection (Takara PCR Mycoplasma Detection Set). A list of all CCLs and media conditions is provided (**Table S2)** and resides on the NCI-CTD[2] data portal (http://ctd2.nci.nih.gov).

Cells were plated at a density optimized during assay development (**Extended Experimental Procedures**) in 384-well opaque, white assay plates and incubated overnight at 37°C/5% CO_2. Compound stocks were plated in 384-well format in 8-pt, 2-fold concentration ranges defined by literature review. Compounds were pin-transferred (CyBio Vario) into duplicate assay plates and incubated for 72h. ATP levels were measured using CellTiter-Glo as a surrogate for cell viability.

Assembling the Informer Set

354 small molecules that perturb targets and processes on which cancer cells may become dependent were identified by careful evaluation of the probe-development literature including seminars, journals, NIH Molecular Libraries Initiative Probe Reports, and patents. ~30% of the Informer Set was accessed through organic synthesis. A list of all compounds, with annotated targets and structures, is provided (**Table S1)** and resides on the NCI-CTD[2] data portal (http://ctd2.nci.nih.gov).

Data processing

At each compound concentration, we computed a percent-viability score relative to the effect observed for vehicle-control (DMSO) treatment of the same CCL. Concentration-response curves using percent-viability scores were fit using cubic splines and areas under percent-viability curves (AUC) were computed used as a measure of sensitivity for subsequent analyses (**Extended Experimental Procedures**).

Genetic Data

Our analyses use publicly available annotations of CCLs, including: gene expression (Affymetrix GeneChip Human Genome U133 Plus 2.0 Array), copy number (Affymetrix Genome-Wide Human SNP Array 6.0), and mutation status from massively parallel sequencing of >1,600 genes and from mass spectrometric genotyping (OncoMap) for 492 mutations in 33 oncogenes/tumor suppressors (Barretina et al., 2012). To illustrate the genetic diversity of the CCLs, we report frequency distributions of the number of mutant genes across the number of CCLs, and the number of unique lesions for each gene (**Figure S1**).

Enrichment and regression analysis

For each compound, profiling across CCLs yielded a ranked list of sensitivities (AUCs) that could be analyzed for genetic features correlating with the response. For each compound, we used a sorting-based enrichment scoring algorithm (Cormen et al., 2000) to measure how genetic features distribute across the ranked list of sensitivities, followed by a chi-squared test of homogeneity to account for compound potency. The maximum (worst) of the p-values from these two tests was used in subsequent analysis to correct for multiple hypothesis tests, resulting in false-discovery rate (FDR) q-values (Benjamini and Hochberg, 1995). We applied a cutoff of q<0.25 in **Table S2**, and a more stringent cutoff in the CTRP. For elastic-net

regression analysis, we normalized copy number variation, mutation, and lineage features using a z-score (standard normal distribution, with μ=0 and σ=1) for each feature. Elastic net was implemented using Matlab, Python, & R using a core algorithm component from the original authors (Zhu and Hastie, 2004) (**Extended Experimental Procedures**)**.**

Global analyses of the resource

Global analysis was performed on the subset of connections most robust relative to potential confounding factors. When multiple datasets suggested the same compound-gene connection, the best-scoring connection was retained. Frequency, sum of scores, and average scores for every gene and compound were computed (**Table S3**). Statistical significance of the number of overlapping genes and compounds (hypergeometric distribution) and a hierarchical clustering of compounds using enrichment scores was performed (**Table S3; Extended Experimental Procedures)**.

Confirming sensitivity of ovarian CCLs to ML210 and ML162

CCLs were plated in 384-well plates at 2000 cells per well in their preferred media and treated with 4 concentrations of ML210 and ML162. Cell number and growth were assayed after 72h treatment using an SRB assay. Results from assays were confirmed using 6 replicates at each compound concentration in each of 3 runs (**Extended Experimental Procedures**).

Confirming association of *CTNNB1* mutation and sensitivity to navitoclax

Four navitoclax-resistant control CCLs and 7 *CTNNB1*-mutant CCLs were seeded into 384-well plates as during profiling experiments. Caspase 3/7 activity was measured using Caspase-Glo (Promega) after 1.5-h incubation. ATP levels were measured 72h after treatment. Results were confirmed using 8 replicates at each compound concentration in each

of 3 runs. Five additional *CTNNB1*-mutant CCLs were also assayed for sensitivity to navitoclax in a separate run for further comparison (**Extended Experimental Procedures**).

Induction of β-catenin protein levels and sensitivity to navitoclax

Four CCLs were pre-treated with either DMSO or 4 μM GSK3β inhibitor CHIR-99021, the maximum concentration that did not cause reduction in ATP levels after 3d continuous treatment, as measured by CellTiter-Glo. Cell samples were collected from untreated cells and after 3d of DMSO- or CHIR-99021 treatment, and β-catenin protein levels were assayed by Western blotting (**Extended Experimental Procedures**).

For pre-treatment experiments, cells were plated overnight, treated with either DMSO or 4 μM CHIR-99021 for 72h, and seeded into 384-well plates with media supplemented with DMSO or CHIR-99021. Cells were incubated overnight, treated with navitoclax in a 12-pt, 2-fold dilution series for 72h, assayed for viability using CellTiter-Glo. All experiments were performed in 8 replicates in each of 2-3 runs. Cells were simultaneously plated for co-treatment experiments under similar conditions (**Extended Experimental Procedures**).

Compound Management team for handling the Informer Set, D.-K. Jang, A. Li, and M. Reich for supporting web portal development, and the Biological Samples Platform for providing CCLs. The project was enabled by the Broad Institute Chemical Biology Program and Platform. L.A.G. is a consultant for and equity holder in Foundation Medicine, Inc. and received sponsored research from Novartis, Inc. The authors are grateful for the leadership of the CTD[2] Network by Daniela Gerhard (Director, Office of Cancer Genomics, NCI). S.L.S. is an Investigator at the Howard Hughes Medical Institute.

Author contributions

Author contributions: C.S.H., B.M., A.M.S., P.A.C., A.F.S., and S.L.S. designed research; A.B., N.E.B., J.H.C., E.V.P., K.L., G.I.S., R.Y.E., M.L.S., D.I., S.W., A.L.B., A.J.W., D.K. and P.A.C. performed research; A.B., N.E.B., J.C.G., T.L., M.W., N.S., G.V.K., J.B., V.D., L.A.G., C.S.H., B.M., and P.A.C. contributed new reagents/analytic tools; A.B., N.E.B, J.H.C., E.V.P., K.L., R.Y.E., G.I.S., M.L.S., D.I., S.W., A.L.B., J.C.G., A.M.S., P.A.C., A.F.S., and S.L.S analyzed data; and A.B., N.E.B, J.H.C., E.V.P., R.Y.E, M.L.S., A.M.S., P.A.C., A.F.S., and S.L.S. wrote the paper.

References

Acosta, J.C., O'Loghlen, A., Banito, A., Guijarro, M.V., Augert, A., Raguz, S., Fumagalli, M., Da Costa, M., Brown, C., Popov, N*., et al.* (2008). Chemokine signaling via the CXCR2 receptor reinforces senescence. Cell *133*, 1006-1018.

Arteaga, C.L. (2006). EGF receptor mutations in lung cancer: from humans to mice and maybe back to humans. Cancer Cell *9*, 421-423.

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 *483*, 603-607.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) *57*, 289-300.

Bennett, C.N., Ross, S.E., Longo, K.A., Bajnok, L., Hemati, N., Johnson, K.W., Harrison, S.D., and MacDougald, O.A. (2002). Regulation of Wnt signaling during adipogenesis. The Journal of biological chemistry *277*, 30998-31004.

Brown, E.J., Beal, P.A., Keith, C.T., Chen, J., Shin, T.B., and Schreiber, S.L. (1995). Control of p70 s6 kinase by kinase activity of FRAP in vivo. Nature *377*, 441-446.

Chen, Z., Cheng, K., Walton, Z., Wang, Y., Ebi, H., Shimamura, T., Liu, Y., Tupper, T., Ouyang, J., Li, J.*, et al.* (2012). A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response. Nature *483*, 613-617.

Cheung, H.W., Cowley, G.S., Weir, B.A., Boehm, J.S., Rusin, S., Scott, J.A., East, A., Ali, L.D., Lizotte, P.H., Wong, T.C.*, et al.* (2011). Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc Natl Acad Sci U S A *108*, 12372-12377.

Corcoran, R.B., Cheng, K.A., Hata, A.N., Faber, A.C., Ebi, H., Coffee, E.M., Greninger, P., Brown, R.D., Godfrey, J.T., Cohoon, T.J.*, et al.* (2013). Synthetic lethal interaction of combined BCL-XL and MEK inhibition promotes tumor regressions in KRAS mutant cancer models. Cancer Cell *23*, 121-128.

Corcoran, R.B., Settleman, J., and Engelman, J.A. (2011). Potential therapeutic strategies to overcome acquired resistance to BRAF or MEK inhibitors in BRAF mutant cancers. Oncotarget *2*, 336-346.

Cormen, T., Dehne, F., Fraigniaud, P., and Matias, Y. (2000). ACM Symposium on Parallel Algorithms and Architectures - Guest editors' foreword. Theor Comput Syst *33*, 335-335.

Dixon, S.J., Lemberg, K.M., Lamprecht, M.R., Skouta, R., Zaitsev, E.M., Gleason, C.E., Patel, D.N., Bauer, A.J., Cantley, A.M., Yang, W.S.*, et al.* (2012). Ferroptosis: an iron-dependent form of nonapoptotic cell death. Cell *149*, 1060-1072.

Gandhi, L., Camidge, D.R., Ribeiro de Oliveira, M., Bonomi, P., Gandara, D., Khaira, D., Hann, C.L., McKeegan, E.M., Litvinovich, E., Hemken, P.M.*, et al.* (2011). Phase I study of Navitoclax (ABT-263), a novel Bcl-2 family inhibitor, in patients with small-cell lung cancer and other solid tumors. J Clin Oncol *29*, 909-916.

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.

Gonzalez de Castro, D., Clarke, P.A., Al-Lazikani, B., and Workman, P. (2013). Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance. Clin Pharmacol Ther *93*, 252-259.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Heiser, L.M., Sadanandam, A., Kuo, W.L., Benz, S.C., Goldstein, T.C., Ng, S., Gibb, W.J., Wang, N.J., Ziyad, S., Tong, F.*, et al.* (2012). Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc Natl Acad Sci U S A *109*, 2724-2729.

Janne, P.A., Shaw, A.T., Pereira, J.R., Jeannin, G., Vansteenkiste, J., Barrios, C.H., Franke, F.A., Grinsted, L., Smith, P.D., Zazulina, V.*, et al.* (2012). Phase II double-blind, randomized study of selumetinib (SEL) plus docetaxel (DOC) versus DOC plus placebo as second-line treatment for advanced KRAS mutant non-small cell lung cancer (NSCLC). ASCO Meeting Abstracts *30*, 7503.

Jho, E.H., Zhang, T., Domon, C., Joo, C.K., Freund, J.N., and Costantini, F. (2002). Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Molecular and cellular biology *22*, 1172-1183.

Jimenez-Marin, A., Collado-Romero, M., Ramirez-Boo, M., Arce, C., and Garrido, J. (2009). Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. BMC Proceedings *3*, S6.

Ju, J., Hong, J., Zhou, J.N., Pan, Z., Bose, M., Liao, J., Yang, G.Y., Liu, Y.Y., Hou, Z., Lin, Y.*, et al.* (2005). Inhibition of intestinal tumorigenesis in Apcmin/+ mice by (-)-epigallocatechin-3-gallate, the major catechin in green tea. Cancer research *65*, 10623-10631.

Kaga, S., Zhan, L., Altaf, E., and Maulik, N. (2006). Glycogen synthase kinase-3beta/beta-catenin promotes angiogenic and anti-apoptotic signaling through the induction of VEGF, Bcl-2 and survivin expression in rat ischemic preconditioned myocardium. Journal of molecular and cellular cardiology *40*, 138-147.

Larsen, J.E., Cascone, T., Gerber, D.E., Heymach, J.V., and Minna, J.D. (2011). Targeted therapies for lung cancer: clinical experience and novel agents. Cancer J *17*, 512-527.

Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C.*, et al.* (2010). A human B-cell interactome identifies

MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol Syst Biol *6*, 377.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739-1740.

Liu, J., Farmer, J.D., Jr., Lane, W.S., Friedman, J., Weissman, I., and Schreiber, S.L. (1991). Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP- FK506 complexes. Cell *66*, 807-815.

Luo, J., Emanuele, M.J., Li, D., Creighton, C.J., Schlabach, M.R., Westbrook, T.F., Wong, K.K., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell *137*, 835-848.

McDermott, U., Sharma, S.V., Dowell, L., Greninger, P., Montagut, C., Lamb, J., Archibald, H., Raudales, R., Tam, A., Lee, D.*, et al.* (2007). Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. Proc Natl Acad Sci U S A *104*, 19936-19941.

National Cancer Institute (2000). NLM Identifier:NCT01229150. Randomized Phase II Study of AZD6244 MEK-Inhibitor With Erlotinib in KRAS Wild Type and KRAS Mutant Advanced Non-Small Cell Lung Cancer. (ClinicalTrials.gov [Internet]. Bethesda (MD): National LIbrary of Medicine (US). 2000-[cited 2012Jun18].).

Pan, H., Cao, J., and Xu, W. (2012). Selective histone deacetylase inhibitors. Anticancer Agents Med Chem *12*, 247-270.

Prahallad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., Beijersbergen, R.L., Bardelli, A., and Bernards, R. (2012). Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. Nature *483*, 100-103.

Ran, Z.H., Zou, J., and Xiao, S.D. (2005). Experimental study on anti-neoplastic activity of epigallocatechin-3-gallate to digestive tract carcinomas. Chin Med J (Engl) *118*, 1330-1337.

Rosenbluh, J., Nijhawan, D., Cox, A.G., Li, X., Neal, J.T., Schafer, E.J., Zack, T.I., Wang, X., Tsherniak, A., Schinzel, A.C*., et al.* (2012). beta-Catenin-driven cancers require a YAP1 transcriptional complex for survival and tumorigenesis. Cell *151*, 1457-1473.

Sartore-Bianchi, A., Di Nicolantonio, F., Nichelatti, M., Molinari, F., De Dosso, S., Saletti, P., Martini, M., Cipani, T., Marrapese, G., Mazzucchelli, L*., et al.* (2009). Multi-determinants analysis of molecular alterations for predicting clinical benefit to EGFR-targeted monoclonal antibodies in colorectal cancer. PLoS ONE *4*, e7287.

Sharma, S.V., Haber, D.A., and Settleman, J. (2010). Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. Nature reviews Cancer *10*, 241-253.

Shoemaker, R.H. (2006). The NCI60 human tumour cell line anticancer drug screen. Nature reviews Cancer *6*, 813-823.

Skehan, P., Storeng, R., Scudiero, D., Monks, A., McMahon, J., Vistica, D., Warren, J.T., Bokesch, H., Kenney, S., and Boyd, M.R. (1990). New colorimetric cytotoxicity assay for anticancer-drug screening. Journal of the National Cancer Institute *82*, 1107-1112.

Smalley, K.S. (2010). PLX-4032, a small-molecule B-Raf inhibitor for the potential treatment of malignant melanoma. Curr Opin Investig Drugs *11*, 699-706.

Souers, A.J., Leverson, J.D., Boghaert, E.R., Ackler, S.L., Catron, N.D., Chen, J., Dayton, B.D., Ding, H., Enschede, S.H., Fairbrother, W.J*., et al.* (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nature medicine *19*, 202-208.

Sparks, A.B., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998). Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. Cancer research *58*, 1130-1134.

Straussman, R., Morikawa, T., Shee, K., Barzily-Rokni, M., Qian, Z.R., Du, J., Davis, A., Mongare, M.M., Gould, J., Frederick, D.T*., et al.* (2012). Tumor microenvironment contributes to innate RAF-inhibitor resistance through HGF secretion. Nature *in press.*

Sun, S., Schiller, J.H., Spinola, M., and Minna, J.D. (2007). New molecularly targeted therapies for lung cancer. The Journal of clinical investigation *117*, 2740-2750.

Tahir, S.K., Wass, J., Joseph, M.K., Devanarayan, V., Hessler, P., Zhang, H., Elmore, S.W., Kroeger, P.E., Tse, C., Rosenberg, S.H*., et al.* (2010). Identification of expression signatures predictive of sensitivity to the Bcl-2 family member inhibitor ABT-263 in small cell lung carcinoma and leukemia/lymphoma cell lines. Molecular cancer therapeutics *9*, 545-557.

Tse, C., Shoemaker, A.R., Adickes, J., Anderson, M.G., Chen, J., Jin, S., Johnson, E.F., Marsh, K.C., Mitten, M.J., Nimmer, P*., et al.* (2008). ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. Cancer research *68*, 3421-3428.

Vervoorts, J., Luscher-Firzlaff, J., and Luscher, B. (2006). The ins and outs of MYC regulation by posttranslational mechanisms. The Journal of biological chemistry *281*, 34725-34729.

Wang, Z., Havasi, A., Gall, J.M., Mao, H., Schwartz, J.H., and Borkan, S.C. (2009). Beta-catenin promotes survival of renal epithelial cells by inhibiting Bax. J Am Soc Nephrol *20*, 1919-1928.

Weiwer, M., Bittker, J.A., Lewis, T.A., Shimada, K., Yang, W.S., MacPherson, L., Dandapani, S., Palmer, M., Stockwell, B.R., Schreiber, S.L*., et al.* (2012). Development of small-molecule probes that selectively kill cells induced to express mutant RAS. Bioorganic & medicinal chemistry letters *22*, 1822-1826.

Yoon, J., Koo, K.H., and Choi, K.Y. (2011). MEK1/2 inhibitors AS703026 and AZD6244 may be potential therapies for KRAS mutated colorectal cancer that is resistant to EGFR monoclonal antibody therapy. Cancer research *71*, 445-453.

Zhu, J., and Hastie, T. (2004). Classification of gene microarrays by penalized logistic regression. Biostatistics *5*, 427-443.

Figure Legends

Figure 1. Response of CCLs to Informer Set. Sensitivity of 242 CCLs to small-molecule probes/drugs was assessed at dose (CellTiterGlo) and areas under the concentration-response curve (AUC) were computed. Data are shown as box plots indicating distributions of AUC values for each compound **(A)** and a heatmap of AUC values (scale represents AUC values ranging between 1 (sensitive; red) and 6 (unresponsive; blue)) **(B)** for single CCLs (columns) treated with single compounds (rows). Missing numerical values in heatmap were imputed using a k-nearest neighbors approach. AUC distributions were analyzed by incorporating context-dependent exclusions **(C)** of cell lines (grey bars represent excluded cell lines). See also Figure S1 and Table S1.

Figure 2. Genetic dependencies targeted by small molecules. The distribution of CCL response (AUC values) to compound treatment is represented as a heatmap denoting sensitivity (red) or unresponsiveness (blue) aligned with genomic alterations for corresponding CCLs (gray bars). The resource identified known clinically drug-targeted genetic dependencies (**A**) and known drug-resistance mechanisms (BRAF V600E outlier cell lines: *RKO; #SKMEL28). The resource also suggests dependencies with both mutation and copy number

variation in *MYC* (**B**). Global analysis of the resource showed *EGFR*-mutated CCLs are unresponsive to NAMPT inhibitors (**C**). CNV-H: high-copy number (≥8 copies), TES: all targeted-exome sequencing mutant calls, TES-A: targeted-exome sequencing, non-neutral missense mutations; Onco: Oncomap mutant calls, MUT: any mutation call. See also Figure S2, Table S2, User Guide S1, and Table S3.

Figure 3. Lineage dependencies targeted by small molecules. Ovarian CCLs are highly sensitive to ML210 and RSL3 (**A**). An expanded panel of ovarian CCLs showed sensivity to ML210 (IC$_{50}$ of ~10 nM) independent of the BRCA1 status of the CCLs (**B**). See also Figure S3.

Figure 4. Mutations in β-catenin associate with sensitivity to navitoclax. Activating mutations in β-catenin (*CTNNB1*) or mutations in members of its destruction complex (*AXIN1*; *CSNK1A1*) correlate with sensitivity to navitoclax (**A**). Previous studies have linked the Wnt/β-catenin pathway to expression of Bcl2 family members (**B**). An elastic-net regression model (black circles: observed; red crosses: predicted; weighted root-mean-squared error: 1.45) predicts AUC sensitivity values across CCLs treated with navitoclax (**C**). Heatmap depicts model features (rows; e.g., mutation, copy number) sorted by descending weight (black bars) across all CCLs tested (columns). Scale represents range of normalized values between -3 and 3 (red: relative higher copy number, presence of mutation; blue: relative lower copy number). All model features (**Table S4**) were input to Ingenuity Pathway Analysis (Jimenez-Marin et al., 2009) and the highest-scoring network (**D**) contains β-catenin as a central node (p=10^{-38}). The network contains members of the β-catenin pathway present in the regression

model (brown), other genes present in the model (dark grey), and molecular interactions with non-regression-model features (light grey). See also Table S4 and Table S5.

Figure 5. Confirmation experiments for navitoclax/β-catenin. Response to navitoclax observed in large-scale profiling was confirmed in 3 independent experiments (**A**) with the 7 most sensitive *CTNNB1*-mutant CCLs (gray bars) and 4 control CCLs lacking mutations in *CTNNB1* (white bars). In parallel, caspase 3/7 activation after navitoclax treatment was measured (**B**), showing that loss of viability was due to induction of apotosis. The response of previously untested *CTNNB1*-mutant CCLs (red bars) to navitoclax was measured using the same conditions from our initial profiling experiments (**C**). Data are represented as mean +/- SD. See also Figure S4.

Figure 6. Small-molecule induction of β-catenin levels and sensitivity to navitoclax. Treatment with the GSK3β inhibitor CHIR-99021 led to increased levels of β-catenin in RKO and HT29 (non-mutant) cells (**A**) and relatively little change in HEC59 (non-mutant) and SW48 (S33Y CTNNB1 mutant) cells (**D**). Sensitivity to navitoclax was assessed after pre-treatment with CHIR-99021 (red) and compared to DMSO-pretreated controls (black). RKO (**B**) and HT29 (**C**) cells, which had increased levels of β-catenin, showed 4-fold increase in sensitivity to navitoclax, while HEC59 (**E**) and SW48 (**F**) cells, which had unchanged levels of β-catenin, demonstrate no significant change in sensitivity. Data are represented as mean +/- SD. See also Figure S5.

Figure 1
Click here to download Figure: Figure1.pdf
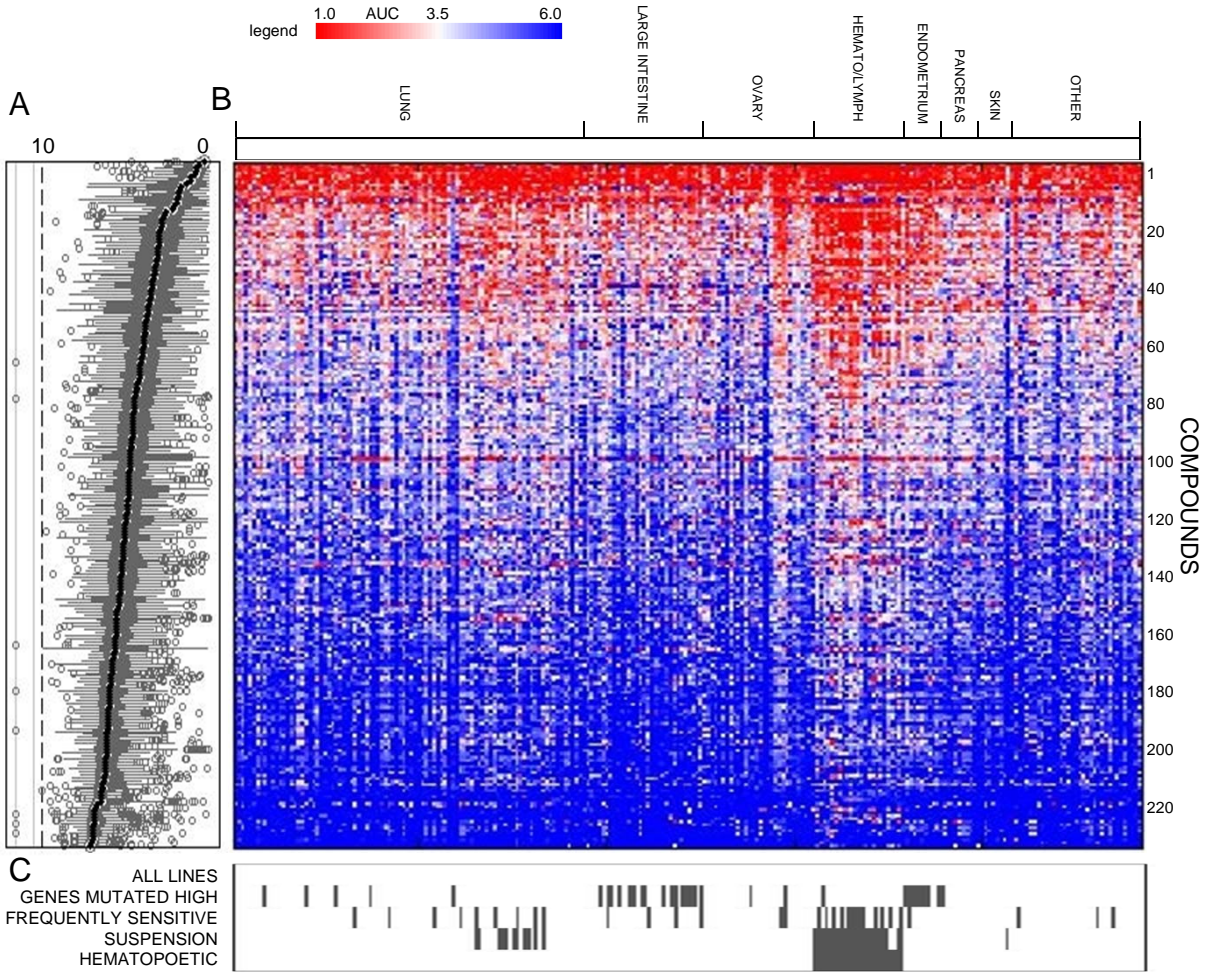
Figure 2
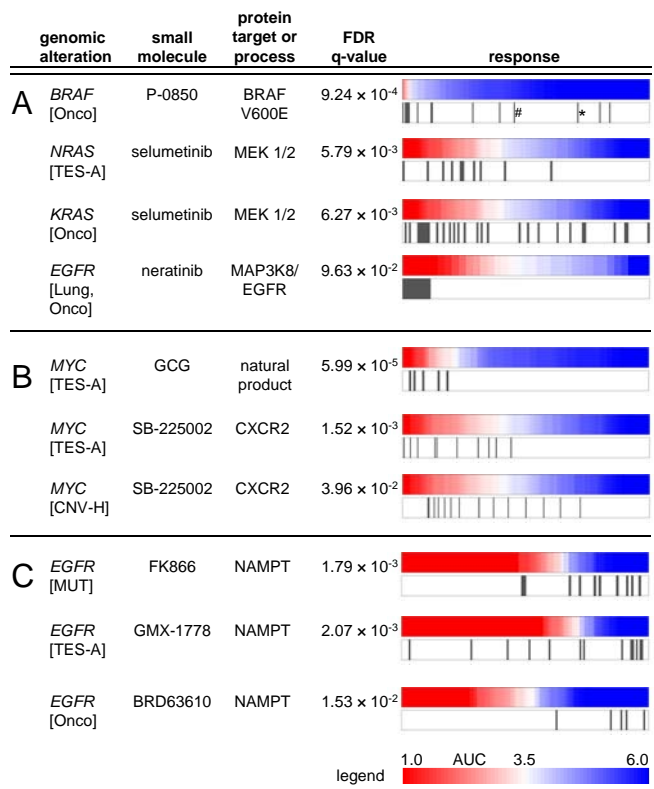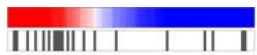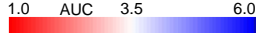
genomic alteration	small molecule	protein target or process	FDR q-value	response
A BRAF [Onco]	P-0850	BRAF V600E	9.24×10^{-4}	
NRAS [TES-A]	selumetinib	MEK 1/2	5.79×10^{-3}	
KRAS [Onco]	selumetinib	MEK 1/2	6.27×10^{-3}	
EGFR [Lung, Onco]	neratinib	MAP3K8/ EGFR	9.63×10^{-2}	
B MYC [TES-A]	GCG	natural product	5.99×10^{-5}	
MYC [TES-A]	SB-225002	CXCR2	1.52×10^{-3}	
MYC [CNV-H]	SB-225002	CXCR2	3.96×10^{-2}	
C EGFR [MUT]	FK866	NAMPT	1.79×10^{-3}	
EGFR [TES-A]	GMX-1778	NAMPT	2.07×10^{-3}	
EGFR [Onco]	BRD63610	NAMPT	1.53×10^{-2}	

1.0   AUC   3.5   6.0
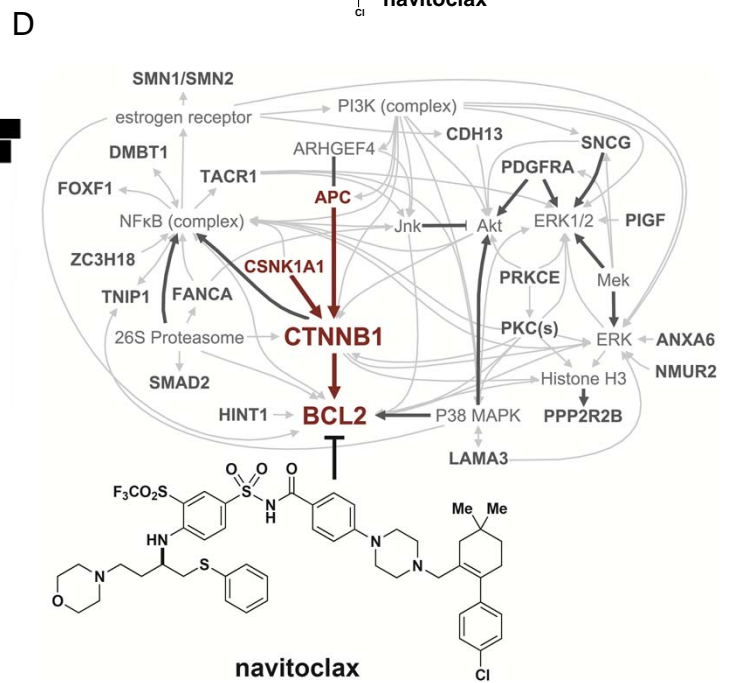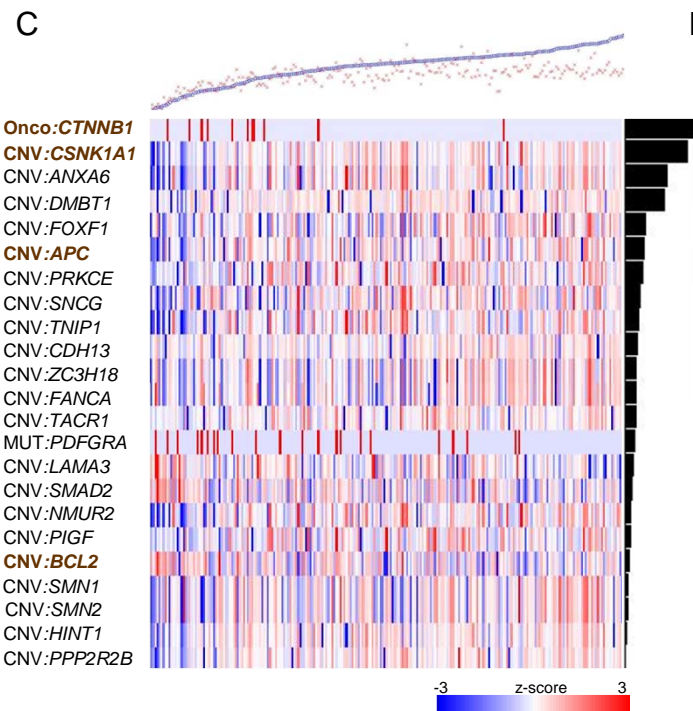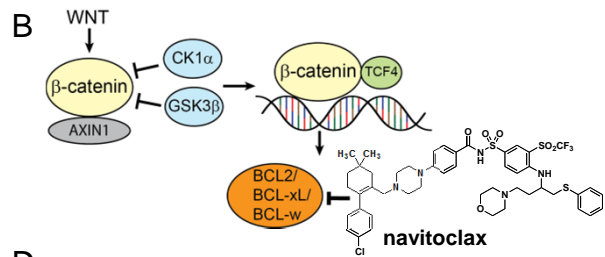
legend
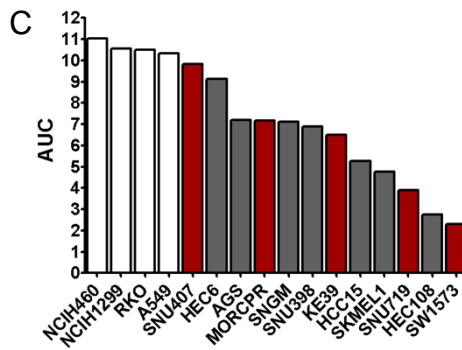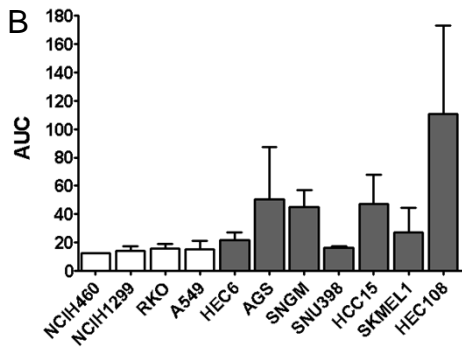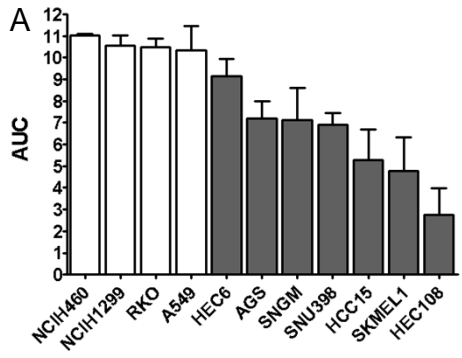
Figure 3

Figure 4

Figure 6
Click here to download Figure: Figure6.pdf
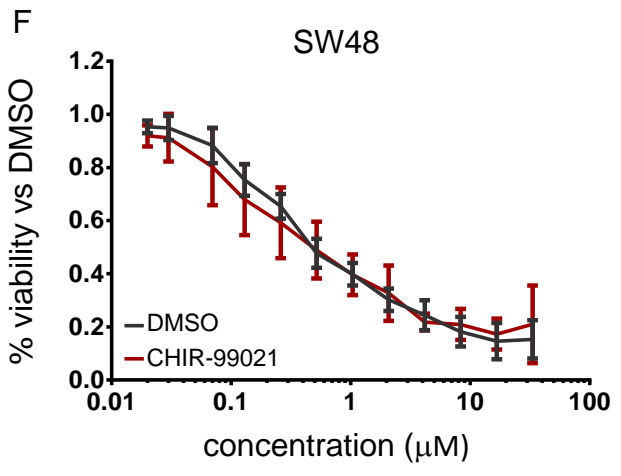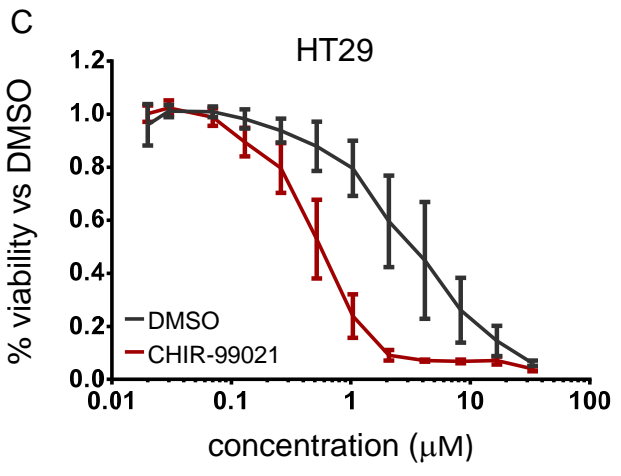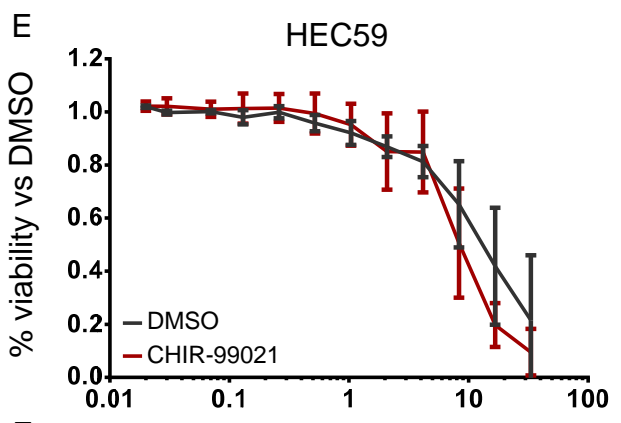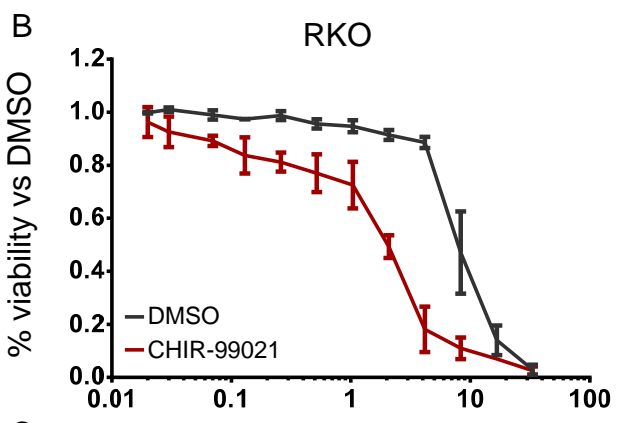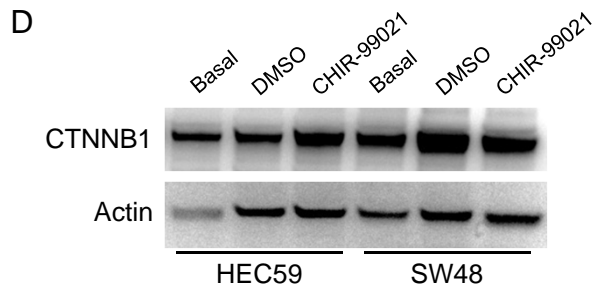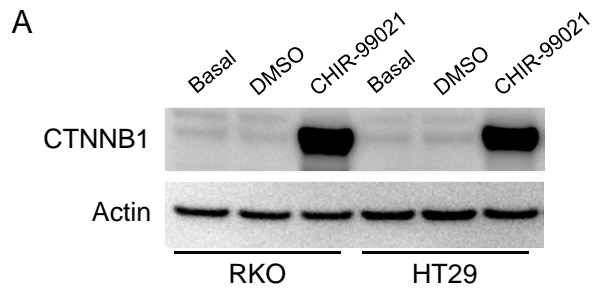
Supplemental Figure and Table Legends

Figure S1, related to Figure 1. **Properties of CCL profiling and genetic data.** Area under concentration-response curves (AUC) accounts for both EC_{50} and strength of effect (**A**). At low percent effect (i.e., when cell viability is relatively unaffected by compound treatment), AUC is essentially independent of relative EC_{50}. In contrast, as percent effect increases, the dependence of AUC on EC_{50} (as judged by the slope of their correlation) increases such that at 100% effect, changes in AUC are equivalent to changes in $log(EC_{50})$ (slope=1). Data presented represent a summary of 37,592 curve-fits (74.1% of all experiments in this study) for which the EC_{50} estimate was greater than 1/8 of the lowest concentration tested and less than 8X the highest concentration tested. Relative EC_{50}s were computed relative to the highest concentration tested for each compound, and strengths of effect were binned into groups centered on the indicated values for trellis display**.** Distributions of unique lesions (**B**) and frequencies of genes mutated in CCLs tested (**C**). The median CCL has mutations in 75 genes (5 percent of total genes sequenced). A large fraction of genes has several unique lesions.

Figure S2, related to Figure 2. **Properties of global connections**. Dendrogram of all compounds used in the global analysis (using cosine distance in complete-linkage analysis); boxed cluster is described in the main text.

Figure S3, related to Figure 3. Identification of lineage dependencies targeted by small molecules. Enrichment analysis revealed that ovarian CCLs are more sensitive to

two small molecule probes, ML210 and RSL3 than other lineages. An expanded panel of ovarian CCLs showed to sensivity to ML162, an analogous compound to ML210 identified in the same phenotypic screen, (IC50 of ~10 nM) and is independent of the BRCA1 status of the CCLs (**A**). In the SKOV3 line, treatment with ML210 and ML162 elicited increased expression of both pH2AX (**B**) and cleaved caspase-3 (**C**), showing that both ML210 and ML162 induced cytotoxicity in these ovarian CCLs.

Figure S4, related to Figure 5. Increased protein levels and activity in *CTNNB1*-mutant CCLs. CTNNB1, Bcl-2, Bcl-xL, Bcl-w, and Mcl-1 protein levels were measured by Western blot in a panel of *CTNNB1* mutant cell lines (red) and navitoclax-unresponsive control cell lines lacking mutations in *CTNNB1* (black) across six different lineages. Actin was used to control for differences in protein loading (**A**). Expression levels for *AXIN2* for a subset of lines (4 navitoclax-unresponsive control CCLs and 7 *CTNNB1*-mutant CCLs) were obtained from the CCLE portal (www.broadinstitute.org/ccle) and averaged across experimental groups (white, control CCLs; grey, *CTNNB1* mutant CCLs) (**B**).

Figure S5, related to Figure 6. Co-treatment of CCLs with CHIR-99021 is insufficient to induce sensitivity to navitoclax. RKO (**A**), HEC59 (**B**), and SW48 (**C**) CCLs were co-treated with either DMSO (grey) or 4 µM CHIR-99021 (red) and navitoclax in a 12-pt, 2-fold dilution series. Cell viability was measured using Cell Titer-Glo as a surrogate for cell viability.

Table S1, related to Figure 1. The CTRP Informer Set. The Informer Set is a collection of 354 small-molecule probes and drugs that selectively target distinct nodes in cell circuitry.

Table S2, related to Figure 2. Table of enrichments underlying the CTRP. The introduction (**a**) details all contents of the table and gives a summary of how data were produced. The summary table (**b**) is a pivoted table of enrichment p-values < 0.05 and FDR q-values < 0.25 for each genetic feature (i.e., sensitive [red] or unresponsive [blue]) for 177 compounds using all cancer cell lines; more than 2 mutant cell lines were present in each enrichment. The lineage summary (**c**) is a pivoted table of enrichment p-values < 0.05 and FDR q-values < 0.25 for each genetic feature and enrichment direction (i.e., sensitive [red] or unresponsive [blue]) for 177 compounds using cancer cell line subsets from individual lineages; more than 2 mutant cell lines were present in each enrichment. The full table (**d**) is an un-pivoted table of enrichment scores from all cell line subsets, cell line exclusions, and genetic feature datasets with enrichment p-values < 0.05 and FDR q-values < 0.25 for 203 compounds. The spearman correlations between basal gene expression (**e**) and gene copy number (**f**) and sensitivity values of cell lines for each compound with the z-score of each correlation are reported. The sensitivity values (**g**) are calculated area-under-dose curve (AUC) values for each cell line and compound. AUCs < 3.5 are considered sensitive to compound treatment. AUCs > 5.5 are considered unresponsive to compound treatment. AUC values were used as input for all enrichment analyses. The viability scores (**h**) are percent viability values for each cancer cell line and compound for every concentration point tested. The compound information table (**i**)

contains contextual compound information and annotations. The cell line information table (**j**) contains contextual cancer cell line information and annotations. The media composition table (**k**) contains basal media names and short descriptions of media additives used in cancer cell line profiling experiments. The media components table (**l**) contains basal media names and lists all media components and concentrations.

User Guide S1, related to Figure 2. A how-to guide for using the CTRP Resource . A guide to understanding the CTRP table of enrichments (Table S2).

Table S3, related to Figure 2. Global analysis of the CTRP. We calculated the frequency, sum of scores, and average scores, for each gene and compound individually in both sensitive and unresponsive directions (**a-d**). We also computed the number of overlapping genes and compounds and their significance (by hypergeometric distribution), for each pair of compounds and genes, respectively (**e-h**). We performed complete-linkage clustering analysis on the compounds using a cosine similarity distance based on the presence or absence of a connection between each compound and gene (binary calls) (**i**). We report all non-zero genes that were associated with each cluster of compounds in (**i**) and their respective weights (**j**).

Table S4, related to Figure 4. Elastic Net Regression Features. The complete list of predictive features for navitoclax, using elastic net regression, with weights for each feature.

Table S5, related to Figure 4. Direct correlation between *MCL1* gene expression and navitoclax sensitivity. .We calculated the Spearman and (rank) and Pearson correlation coefficients between *MCL1* gene expression and sensitivity to navitoclax across all lineage-controlled experiments and CCL subpopulation experiments. The table lists the gene ranking of *MCL1* in comparison to all other genes for which expression data are available (18,893 genes), Spearman and Pearson correlation coefficients, and results of permutation tests (n=16,384) perfomed by randomizing CLLs labels, allowing calculation of nominal p-values for each CCL subpopulation experiment. Highlighted nominal p-values are significant (p<0.05) after considering Bonferonni correction.

Extended Experimental Procedures

Assay development

Measuring response of each cancer cell line (CCL) at various densities to treatment with staurosporine resulted in the determination of their optimized plating density. The Z' factor at each concentration point was calculated and compared between each cellular density to determine the largest dynamic detection window. Briefly, adherent cells or suspension cells were plated in a range of 500-2000 cells/well or 500-5000 cells/well, respectively, in 384-well opaque, white plates and incubated overnight at 37°C/5% CO_2. Cells were treated with staurosporine starting at a high concentration of 3.3µM, in a 16-pt, 1.67-fold dilution series, 16 replicates/concentration, for 72 h. Sensitivity was assayed using CellTiter-Glo (Promega), which measures cellular ATP levels as a

surrogate for cell number and growth, according to the manufacturer's protocol, with one change. The solution was diluted 1 part CellTiter-Glo to 2 parts PBS before a 1:1 addition to the volume on the plate. Luminescence was measured using a PerkinElmer Envision. Sensitivity summary scores were computed as areas under concentration-response curves (AUCs), as described below. While using different cell densities across CCLs could potentially introduce variability in the AUCs, we determined that this variability was of the same magnitude as day-to-day variability in AUCs and favored performing the assays at densities with more robust signal detection.

Data processing

At each concentration of compound, we compute a compound-performance score (D-score) that expresses effect size as a weighted average of differences between treatment and control, and statistical significance by estimating the likelihood that an observed effect size is different than effects expected for mock-treatment (DMSO) in the assay. Any number of replicates of a compound treatment across plates and days can be combined by the method of maximum likelihood into such weighted average and its uncertainty; the ratio of the difference to the uncertainty is the D-score, a normalized value for each compound in an assay. By computing the weighted average on log-transformed small-molecule sensitivity data, we obtain an appropriately weighted average of ratios (i.e., weighted fold-change) of compound-treated to mock-treated wells, which after re-exponentiation we use as the percent-viability score. Concentration-response curves using percent-viability scores were fit using smooth cubic splines for multivariate data from the MATLAB curve-fitting tool box. The resulting

areas-under-curve (AUCs) were used as a measure of sensitivity and used in subsequent enrichment and regression analyses. Pipeline Pilot protocols and MATLAB scripts and functions used in data-processing and analysis are available on request.

Enrichment and regression analysis

Our primary sorting-based enrichment-scoring algorithm (Cormen et al., 2000) results in p-values that reveal the enrichment of genetic alterations relative to ranked sensitivities measured for a single compound across many cell lines. For each compound, each genetic feature receives an p-value that corresponds to the likelihood of seeing that pattern of alterations (or stronger) enriched among the ranked sensitivities by chance. A drawback of using a non-parametric test alone to rank feature-compound pairs is that these scores do not take the relative or absolute potency of the compound into account. Accordingly, we initially observed apparently significant connections to compounds whose sensitivity distributions do not exhibit patterns of sensitivity in which we were interested.

We addressed these issues first by filtering out cases where the range of sensitivities was undesirable. Three filtering criteria were used: (1) for sensitive enrichments an AUC≤3.5 for at least one cell line (AUC=7 corresponds to no compound effect), for resistant enrichments an AUC≥5.5 or higher for at least one cell line; (2) a minimum difference of 3 between the lowest AUC and the highest AUC; (3) and at least one cell line with the genetic alteration under consideration. Next, to produce a list of candidate cancer dependencies that have statistical significance with the desired compound sensitivity performance, we performed a parametric chi-squared test of homogeneity to

account for the absolute potency of each compound in relation to the distribution of genetic alterations.

To obtain a significance score using both non-parametric and parametric tests, we squared the maximum (worst case) of their two p-values for subsequent multi-test correction and ranking. We applied the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) to control for multiple hypothesis testing within each family of hypotheses (different genetic or lineage features sharing a measured AUC distribution), resulting in q-values (adjusted squared maximum p-values). In **Table S2**, we applied the procedure at a false-discovery rate (FDR) cutoff of q<0.25, while in the CTRP, we used a more stringent cutoff.

For elastic-net regression analysis, we normalized copy-number, hybrid-capture, Oncomap, and lineage data (~24K features), using a z-score (standard normal distribution, with μ=0 and σ=1) for each feature. Following preprocessing, we used two different methods to examine the link between the genetic features and cancer cell line sensitivities. First, we used an elastic net regression approach (Zou and Hastie, 2005) to predict AUCs for cell lines treated with selected compounds. In comparison to other regression methods, elastic-net regression is typically used for high-dimensional data to achieve an optimal balance between sparse and highly correlated input features. The algorithm adjusts the values of two parameters, λ and α, that control the relative optimization for model parsimony (i.e., lasso regression) and covariance-strengthened models (i.e., ridge regression) recording the root-mean-squared (RMS) error of its prediction per iteration. Elastic net was implemented using MATLAB, Python, & R (Lang and Ihaka, 2008) using a core algorithm component from the original authors (Zou and

Hastie, 2005).

In order to pre-select input features that are highly correlated with our sensitivities, we calculated the Spearman correlation between each feature and sensitivity (AUC) across all cell lines. The minimum value of the RMS error (or weighted RMS error, see below) was used to determine the best elastic net model parameters during 10-fold cross validation. Elastic net was combined with an optimization algorithm that progressively searched for the minimum RMS error of elastic net predictions among runs with different numbers of input features as determined by the Spearman-ranking. In addition to input optimization, bootstrapping (random sampling of cell lines *with* replacement) (Paparoditis and Politis, 2005) was used to select the highest frequency features that enter into a final model. We imposed a constraint on the number of bootstrapped features entering the final run, where the number of features should not exceed the number of observed responses. In all elastic net steps (including optimization, cross-validation, and bootstrapping), a novel weighting method was used to adjust the magnitude with which sensitive versus resistant cell lines received weights, where the sensitive or resistant cell lines obtained a weight proportional to exponentiating the absolute value of the standard score of the responses. This weighting scheme forces elastic net to be more careful in predicting the extreme sensitive (or resistant) cell lines. The regression model we display (**Figure 4C**) applies a greater magnitude of weight to the sensitive cell lines. The final model produces a list of genes, each with a weighting coefficient, and the importance of each feature to the overall model used to predict the pattern of sensitivity for cell lines treated with a selected compound. Pipeline Pilot protocols and MATLAB scripts and functions used in data-processing and analysis are available on request.

CCL Sub-population Analysis

We wondered whether all lineage-based sub-populations and excluded sub-populations actually required consideration. We checked each lineage and sublineage separately, and combinations of lineage and exclusions for qualification in our analysis (**User Guide S1**). To assess the extent to which certain sub-populations of cell lines might have non-compound-specific sensitivity characteristics that confound our enrichment analyses, we undertook a two-stage strategy to (1) identify such sub-populations independently of their specific genetic lesions, and (2) characterize the consequences to enrichment analyses of analyzing various sub-populations separately versus together.

Identification of potentially confounding sub-populations

For qualitative (categorical) exclusions, we performed K-S tests to ask whether the exclusion of growth conditions (adherent, suspension, mixed) or individual lineages had a significant effect on AUC distributions. The following two exclusions were found to be significant: exclude suspension cell lines ($p_{KS} < 3.67 \times 10^{-9}$); exclude hematopoetic cell lines ($p_{KS} < 2.34 \times 10^{-10}$)

To find an appropriate threshold to identify cell lines with a statistically large number of mutations, we assumed that this subset of cell lines would have a different distribution of sensitivities than all other cell lines. We used a collection of i two-distribution Komolgorov-Smirnov (K-S) tests, subject to Bonferroni correction, to evaluate the difference between distributions of cell line sensitivities between the i most-frequently-mutated cell lines and the remainder. Cell lines that harbored a fraction

of mutations greater than the most significant K-S test ($p_{KS} < 1.63 \times 10^{-11}$) were those we termed 'genes mut high', of which there are 33 CCLs. A similar analysis was attempted to identify cell lines that were frequently sensitive to many compounds, but many K-S test significance values exceeded machine precision. We chose cell lines that were sensitive to more than 25% of the compounds as those we termed 'frequently sensitive', of which there are 32 CCLs. We therefore exclude these subsets of cell lines as part of our enrichment analyses.

For lineage and sublineage feature analysis, we checked whether each lineage or sublineage had more than 3 lines present in the dataset to qualify for inclusion. For lineage and exclusion combination analysis, we included combinations where there were at least three cell lines that passed one of the four exclusion criteria and were present in the specified lineage or sublineage (see User Guide S1 for additional details). Sub-lineages tested within lung are: adenocarcinoma, broncheoalveolar carcinoma, large cell carcinoma, non-small cell carcinoma, small-cell carcinoma, and squamous cell carcinoma; and within hematopoetic are: acute myeloid leukemia, diffuse large B-cell lymphoma, plasma cell myeloma.

Global Analysis

Overall, the list of enrichments contains 397,270 connections, which includes lineage/sub-lineage connections, and exclusions across all datasets. Pairs of scores were considered in the absence or presence of each exclusion for each compound-gene connection. We conceptualized these score-pairs as falling into interpretable regions of a scatterplot that indicate whether connections were improved, preserved, or diminished

upon running an exclusion experiment. We counted score pairs for each region defined in the scatter plot, and compared these values to a randomly permuted score matrix. Since excluding suspension or hematopoietic cell lines is independent of our sensitivity data, while our other exclusions make use of it, we elected to use only these two categorical exclusions to qualify connections for global analysis. For connections improved or preserved when both suspension and hematopoietic lines were excluded, we kept the best-scoring connection. When either exclusion diminished the score, we kept the diminished score, and if either exclusion resulted in an insignificant score, we removed the connection. We also did not include connections with contradictory scores between exclusions or between an exclusion experiment and the primary analysis. When multiple datasets suggested the same compound-gene connection, we kept the best-scoring connection. We observed that this set of connections was dominated (in number) by relatively weaker q-value scores. To eliminate weaker connections, we excluded connections with fewer than 3 mutations, fewer than 3 sensitive mutants or unresponsive mutants, and connections where more than half the cell lines tested were mutant cell lines. Using the filters described, we began with 108,635 candidate compound-gene connections for global analysis. We determined the optimal threshold for a more stringent q-value to retain the strongest gene-compound connections and protect against type I errors. Empirically, we varied the q-value between 0 and 0.25, and observed the fraction of remaining connections. At q-value cutoffs of less than 0.025, we observed small changes in the number of connections (i.e., achieved relative stability in the connection list). Thus, we used a cutoff of q<0.025, and used the remaining 16,667 qualified distinct compound-gene connections for further analyses.

We calculated the frequency, sum of scores, and average scores for each gene and compound individually in both sensitive and unresponsive directions (**Table S3a-d**). We also computed the number of overlapping genes and compounds and their significance (by hypergeometric distribution), for each pair of compounds and genes, respectively (**Table S3e-h**). Finally, we performed complete-linkage clustering analysis on the compounds using a cosine similarity distance based on the presence or absence of a connection between each compound and gene (binary calls) (**Table S3i, Figure S2**). In each cluster, for each gene where the compound connection score with the gene is non-zero, a weight was calculated proportional to the fraction of compounds to which the gene connected. For example, if the gene connected to all compounds in the cluster, then that particular gene obtained a weight of 1. If the gene connected to half of the compounds in the cluster, that gene obtained a weight of 0.5. This weight per gene was multiplied by the mean q-score of the gene across all the compounds in the cluster, and summed. To calculate how significant this score was, we computed a random score for a cluster of size n over 100 iterations. The score reported (**Table S3i**) is (s_i − $\mu_{random,n}$)/σ_{random}. We also report all non-zero genes that were associated with each reported cluster of compounds and their respective weights (**Table S3j**) For the lineage-specific analysis, we had 46,175 total qualified connections for global analysis that involved more than 2 mutants, more than 2 examples (across all datasets), which resulted in a total of 12,518 distinct gene-compound connections.

Ingenuity Pathway Analysis

The networks, functional analyses, and pathway connections were generated through

the use of Ingenuity Pathway Analysis (IPA; Ingenuity Systems, www.ingenuity.com). IPA was performed using all gene features (HUGO symbols) from the elastic-net regression analysis. The IPA p-value indicates the likelihood that the assembly of a set of focus genes in a network could be explained by random chance alone. The Ingenuity pathway was redrawn, maintaining all direct and indirect connections, excluding feedback loops, and including only interactions that were experimentally observed or predicted with high confidence.

Confirming sensitivity of ovarian CCLs to ML210 and ML162

Five ovarian CCLs, SKOV3, OVCAR8, NCI/ADR-RES, UWB1.289 (BRCA1 null), and UWB1.289+BRCA (stably expressed BRCA1-WT) CCLs were seeded into 384-well plates at 2000 cells per well in their own preferred media and treated with four concentrations of ML210 and ML162. Cell viability was assayed after 72h of treatment using an SRB assay as previously described (Skehan et al., 1990) with minor modifications for seeding at the chosen density. Absorbance was measured at 510 nm using a Spectramax M5 spectrophotometer (Molecular Devices). Results from assays were confirmed using six replicates at each compound concentration in a single run and three runs were performed.

Immunofluorescence

SKOV3 ovarian cancer cells were seeded on coverslips in 6-well culture plates (at approximately 100,000 cells/well). After overnight culture, the cells were treated for 24h with 10 µM compound or DMSO as a control. Cells were then fixed, permeabilized, and

stained with anti-phospho H2AX (Ser139) (pH2AX) (Millipore Corporation) or cleaved caspase 3 (Cell Signaling Technology), as per established protocols (Wilson et al., 2003; Wilson et al., 2011). Binding of each primary antibody was detected with Alexa Fluor anti-rabbit and IgG 488 secondary antibody (Invitrogen). DAPI (Invitrogen) was used to stain the nuclei. Images were acquired and analyzed as previously described. A minimum of 50 cells were counted in 3 independent fields in each experiment. Results are from three independent experiments with *p<0.05 by the Student's t test.

Western Blots

A panel of 20 *CTNNB1* mutant and 18 navitoclax-unresponsive non-mutant CCLs across six different lineages were chosen for examination of CTNNB1 and Bcl2-family protein levels. Whole-cell lysates were prepared by incubating cell pellets in radio-immunoprecipitation assay buffer (RIPA buffer; Pierce) with protease inhibitors (Roche) for 10m. After clarification by centrifugation, protein concentrations were determined by BCA protein assay (Pierce). 50 µg of total protein was boiled with LDS buffer for 10m and separated by electrophoresis on NuPAGE 4-12% Bis-Tris polyacrylamide gels (Invitrogen). Proteins were transferred to nitrocellulose membranes using the Invitrogen iBlot system, blocked for 1h in 5% milk in Tris buffered-saline + 0.1% Tween (TBST), probed with antibodies against CTNNB1, Bcl-2, Bcl-xL, Bcl-w, Mcl-1(Cell Signaling Technologies, 1:1000) and actin (Sigma, 1:15000) overnight at 4°C. Blots were washed with TBST and incubated with HRP-linked secondary antibodies (Cell Signaling Technologies, 1:1000) in 5% milk in TBST for 1h at room temperature. Blots were washed with TBST, developed using SuperSignal® West Pico Chemiluminescent

Substrate (Pierce), and detected on a Kodak image station. Band intensities were determined using Carestream Molecular Imaging Software, and CTNNB1 intensity was normalized to GAPDH (**Figure S4A**). Values were averaged across control CCLs and *CTNNB1*-mutant lines and plotted.

Confirming association of *CTNNB1* mutation and sensitivity to navitoclax

Four navitoclax-resistant control cell lines (A549, H1299, H460, RKO) and seven previously profiled *CTNNB1*-mutant CCLs (AGS, HCC15, HEC108, HEC6, SKMEL1, SNGM, SNU398) were seeded into 384-well plates at densities and media conditions described previously. Cells were incubated overnight at 37°C/5% CO_2, then treated with 12-point, 2-fold dilutions of either navitoclax or DMSO control for 6h. Caspase 3/7 activity was measured as follows: Caspase-Glo (Promega), diluted 1:3 from the original stock, was added to each well, incubated 1.5h, and luminescence was measured. ATP levels were measured 72h after treatment. For both assays, results from all cells lines were confirmed using eight replicates at each compound concentration within a single run. The data shown are averages of three runs. Five new *CTNNB1*-mutant CCLs (SNU407, SNU719, MORCPR, KE39 and SW1573) were also assessed for sensitivity to navitoclax in a single run and the AUCs compared to the previously confirmed control and *CTNNB1*-mutant lines.

AXIN2 gene expression

Gene-expression data for *AXIN2* and *GAPDH* were obtained for the subset of CCLs used in the confirmation experiment from the Broad CCLE portal. Individual

AXIN2/*GAPDH* ratios (**Figure S4B**) and averaged values across control CCLs and *CTNNB1*-mutant lines were plotted.

Correlation between *MCL1* gene expression and navitoclax sensitivity. Spearman (rank) and Pearson correlation coefficients were calculated between *MCL1* gene expression and sensitivity to navitoclax across all lineage-controlled and CCL subpopulation experiments. The rank of *MCL1* in comparison to all other genes for which expression data are available (18,893 genes), Spearman and Pearson correlation coefficients, and results of permutation tests are presented (**Table S5**). Permutation tests (n=16,384) were perfomed by randomizing CLLs labels, allowing calculation of nominal p-values for each CCL subpopulation.

Small molecule induction of β-catenin protein levels and sensitivity to navitoclax

RKO, HT29, HEC59, or SW48 CCLs were pre-treated with either DMSO or 4 µM GSK3β inhibitor CHIR-99021 for 3 days. β-catenin protein levels were assessed by Western blotting and normalized to actin levels, as described above.

For pre-treatment experiments, 1.5 x 10^6 cells were seeded into T75 flasks, allowed to adhere overnight, then treated with either DMSO or 4 µM CHIR-99021. After 72 hours, the cells were seeded into 384-well plates at previously determined densities and with media supplemented the same pre-treatment compound. Cells were incubated overnight, then treated with navitoclax in a 12-pt, 2-fold dilution series. After another 72 hours, cell viability was assessed using Cell Titer-Glo. Results from all cell lines were performed eight replicates at each compound concentration in a single run and two to

three runs were performed for each cell line.

For co-treatment experiments, cells were seeded into 384-well plates at previously determined densities and incubated overnight. The cells were treated with either DMSO or 4 µM CHIR-99021 just prior to treatment with navitoclax in a 12-pt, 2-fold dilution series. After another 72 hours, cell viability was assessed using Cell Titer-Glo. Results from all cell lines were performed in eight replicates at each compound concentration in a single run and two to three runs were performed for each cell line.

Dose-response curves were generated by averaging the ATP levels of the replicates within a single run and comparing them to the average ATP levels of the wells containing DMSO. For RKO, n=2 runs and for HT29, HEC59 and SW48, n=3 runs.

References

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) *57*, 289-300.

Cormen, T., Dehne, F., Fraigniaud, P., and Matias, Y. (2000). ACM Symposium on Parallel Algorithms and Architectures - Guest editors' foreword. Theor Comput Syst *33*, 335-335.

Lang, D.T., and Ihaka, R. (2008). The Future of Statistical Computing Comment. Technometrics *50*, 443-446.

Paparoditis, E., and Politis, D.N. (2005). Bootstrap hypothesis testing in regression models. Statistics & Probability Letters *74*, 356-365.

Skehan, P., Storeng, R., Scudiero, D., Monks, A., McMahon, J., Vistica, D., Warren, J.T., Bokesch, H., Kenney, S., and Boyd, M.R. (1990). New colorimetric cytotoxicity assay for anticancer-drug screening. Journal of the National Cancer Institute *82*, 1107-1112.

Wilson, A.J., Arango, D., Mariadason, J.M., Heerdt, B.G., and Augenlicht, L.H. (2003). TR3/Nur77 in colon cancer cell apoptosis. Cancer research *63*, 5401-5407.

Wilson, A.J., Holson, E., Wagner, F., Zhang, Y.L., Fass, D.M., Haggarty, S.J., Bhaskara, S., Hiebert, S.W., Schreiber, S.L., and Khabele, D. (2011). The DNA damage mark pH2AX differentiates the cytotoxic effects of small molecule HDAC inhibitors in ovarian cancer cells. Cancer Biol Ther *12*, 484-493.

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) *67*, 301-320.

Supplemental Information Inventory

Figure S1: **Properties of CCL profiling and genetic data,** related to Figure 1. Area under concentration-response curves (AUC) account for both EC_{50} and strength of effect of compounds. Distributions of unique lesions and frequencies of genes mutated summarize the genomic features of CCLs tested.

Table S1: **The CTRP Informer Set**, related to Figure 1. The Informer Set is a collection of 354 small-molecule probes and drugs that selectively target distinct nodes in cell circuitry.

Figure S2: **Properties of global connections**, related to Figure 2. Hierarchical clustering of compounds based on their profiles of enrichment connections with genetic features illustrates relationships between compounds tested.

Table S2: **Table of enrichments underlying the CTRP**, related to Figure 2. This table contains enrichment scores from all cell line subsets, cell line exclusions, and genetic feature datasets, as well as, correlations between basal gene expression and copy number to compound sensitivity, and contextual compound and CCL information.

Table S3: **Global analysis of the CTRP**, related to Figure 2. Compounds that share similar mechanisms of action cluster together based on their connections to genetic features. We also analyzed the frequency with which mutated genes correlate with sensitivity or unresponsiveness to different compounds.

User Guide S1: **A how-to guide for using the CTRP Resource**, related to Figure 2. A guide to understanding the CTRP table of enrichments (TableS2).

Figure S3: **Identification of lineage dependencies targeted by small molecules**, related to Figure 3. ML210 and RSL3 induce cytotoxicity in ovarian CCLs.

Table S4: **Elastic net regression features for predicting sensitivity of CCLs to navitoclax treatment**, related to Figure 4. The complete list of predictive features for CCL sensitivity to navitoclax, using elastic net regression, with weights for each feature.

Table S5: **Correlation between *MCL1* gene expression and navitoclax sensitivity**, related to Figure 4. The Spearman and Pearson correlation coefficients between *MCL1* gene expression and CCL sensitivity to navitoclax across all lineage-controlled experiments and CCL subpopulation experiments.

Figure S4: **Increased β-catenin protein levels and activity in *CTNNB1*-mutant CCLs**, related to Figure 5. CTNNB1-mutant CCLs show increased β-catenin protein levels and increased *AXIN2* levels, a downstream target of β-catenin.

Figure S5: **Co-treatment of CCLs with CHIR-99021 is insufficient to induce sensitivity to navitoclax**, related to Figure 6. Pre-treatment of navitoclax-unresponsive, wild-type CCLs with CHIR-99021 leads to increase sensitivity to navitoclax, however, co-treatment does not.

USER GUIDE OVERVIEW. Our project uses genomic cancer cell-line profiling to identify, as systematically as possible, the dependencies that: 1) specific genomic alterations impart on human cancers, and 2) can be targeted with small molecules. Towards this goal, we have measured the sensitivity of a large panel of genetically characterized cancer cell lines to an Informer Set of small-molecule probes and drugs that have selective interactions with their targets, and that collectively modulate many distinct nodes in cancer cell circuitry. Using one approach for data analysis (enrichment analysis), we correlated the sensitivity measurements to the genetic features of the cell lines in order to identify dependencies conferred by specific genotypes. We have assembled these statistically significant correlations into an Excel workbook of cancer genetic dependencies to serve as a hypothesis-generating resource for the cancer biology community. A subset of the correlations found in the workbook can be visualized on the Cancer Therapeutics Response Portal (CTRP, www.broadinstitute.org/ctrp). We have made available all primary data such that it can be re-analyzed to yield further hypotheses as additional computational approaches and deeper genetic and epigenetic characterization of the cancer cell lines become available. Our hope is that the insights mined from the resource, first based on cell-line models of cancer and then substantiated in more complex environments, will yield clinically relevant predictions of how patients will respond to novel types of targeted therapies and accelerate the discovery of new genetically matched medicines.

SUMMARY OF THE WORKBOOK OF STATISTICALLY SIGNIFICANT CORRELATIONS. Table S2 contains the raw data and enrichment analyses associated with compounds tested in profiling experiments that produced enrichment p-values less than 0.05 and false-discovery-rate (FDR) q-values less than 0.25. This workbook contains the following data tables:

Table S2b. **summary workbook**: pivoted table of enrichment p-values less than 0.05 and FDR q-values less than 0.25 for each genetic feature (i.e. sensitive [red] or unresponsive [blue]) for 176 compounds using all cancer cell lines; more than 2 mutant cell lines were present in each enrichment

Table S2c. **lineage summary**: pivoted table of enrichment p-values less than 0.05 and FDR q-values less than 0.25 for each genetic feature and enrichment direction (i.e. sensitive [red] or unresponsive [blue]) for 176 compounds using cancer cell line subsets from individual lineages; more than 2 mutant cell lines were present in each enrichment

Table S2d. **full workbook**: unpivoted table of enrichment scores from all cell-line subsets, cell-line exclusions, and genetic feature datasets with enrichment p-values less than 0.05 and FDR q-values less than 0.25 for 203 compounds

Table S2e. **expression correlation**: a table of spearman correlations between basal gene expression and sensitivity values of cell lines for each compound with the z-score of each correlation (z-scores were calculated by generating a randomly permuted correlation distribution); a z-score cutoff of 2.333 (i.e., p-value equal to 0.01) was applied

Table S2f. **copy number correlation**: a table of spearman correlations between gene copy number and sensitivity values of cell lines for each compound with the z-score of each correlation (z-scores were calculated by generating a randomly permuted correlation distribution); a z-score cutoff of 2.333 (i.e., p-value equal to 0.01) was applied

Table S2g. **sensitivity values**: calculated area-under-dose curve (AUC) values for each cancer cell line and compound. AUCs < 3.5 are considered sensitive to compound treatment. AUCs > 5.5 are considered unresponsive to compound treatment. AUC values were used as input for all enrichment analyses.

Table S2h. **viability scores**: percent viability values for each cancer cell line treated with compound for every dose point tested

Table S2i. **compound information**: contextual compound information and annotation

Table S2j. **cell line information**: contextual cancer cell line information and annotation

Table S2k. **media composition**: basal media names and short description of additives

Table S2l. **media components**: basal media names and list of all media components and concentration

NOTE. This workbook is based on analysis of raw data gathered in duplicate and normalized for analysis. The data have not been confirmed in follow-up experiments. We hope this resource is useful to you in generating hypotheses about your compound or gene of interest to guide future experiments.

Small-molecule profiling of genetically characterized human cancer cell lines was performed in 384-well plate format and sensitivity to compounds was assessed using the Cell Titer-Glo assay (Promega), which measures cellular ATP levels as a surrogate for cell viability. All data were generated by the Chemical Biology Program, The Broad Institute, 7 Cambridge Center, Cambridge, MA 02142. Genetic characterization of cancer cell lines were accessed from the Broad/Novartis Cancer Cell Line Encyclopedia portal: http://www.broadinstitute.org/ccle/home

The public repository for this data is located on the NCIs CTD[2] data portal: http://ctd2.nci.nih.gov/

COLUMN HEADER LABELS

summary and lineage summary
enriched_feature = genetic feature (gene symbol, cell lineage, or gene symbol combinations)
columns C – FU = compound names

full workbook
cell_line_subset = measurements from a subset of cell lines (by lineage) used in enrichment calculation
cell_line_exclusion = measurements from a subset of cell lines that were excluded from enrichment
feature_dataset = database from which genetic feature was obtained (see Description of computational experiments>Datasets below for full descriptions of each dataset label)
compound_name = name of compound
enriched_feature = genetic feature that correlates to enrichment_direction
number_of_cell_lines = number of cell lines with genetic information for specific enrichment
number_of_mutant_cell_lines = number of cell lines with alterations in this Enriched_Feature
enrichment_direction = direction of enrichment (i.e. sensitive or unresponsive)
enrichment_p_value = probability of genetically altered cell lines for this feature enriched by chance

chi_squared_p_value = chi-squared test for homogeneity (enrichment given the potency of compound)

square_max_p_value = maximum of the enrichment_p_value or chi_squared_p_value squared

log_p_value_score = signed negative log (base 10) of square_max_p_value

FDR_q_value = false-discovery rate (FDR) used to correct square_max_p_value for multiple hypotheses

log_q_value_score = signed negative log (base 10) of FDR_q_value

expression correlation
compound_name = name of compound

gene_symbol = HUGO approved symbol for a given gene

spearman_correlation = correlation coefficient of spearman rank correlation

correlation_zscore = z-score of spearman_correlation using the mean and standard deviation of a randomly permuted correlation distribution

cpy number correlation
same columns as expression correlation

sensitivity values
compound_name = name of compound

cell_line_name = name of cancer cell line

area_under_dose_curve = the calculated area under the spline fit (MatLab) dose curve (1 unit equivalent to 2-fold difference in apparent EC50)

viability scores
compound_name = name of compound (INN was used when possible)

cell_line_name = name of cancer cell line

compound_concentration_(uM) = final assay compound concentration

percent_viability = weighted percent-viability with error propagation; shared by all replicates of a compound+concentration+cell line

compound information
compound_name = name of compound (INN was used when possible)

compound_status = status of small-molecule as it relates to treatment of patients

starting_assay_concentration_(uM) = starting top concentration of compound; each compound was profiled in 8-point two-fold dose

affected_process = area of biology relating to cancer that the compound is reported to modulate

target_or_activity_of_compound = protein or biological process that the compound is reported to target (not an exhaustive list)

gene_name_of_protein_target = HUGO gene symbol for the gene of the targeted protein

percent_compound_purity = percent compound purity as determined by LC/MS

compound_SMILES = text string representation of compound structure

cell line information
cell_line_name = name of cancer cell line

cell_line_synonym = list of known cancer cell line names that might also be used to refer to a particular cell line (not an exhaustive list)

cell_line_lineage = tissue of origin (broad)

cell_line_sublineage = histology type (specific)

growth_mode = growth behavior of cancer cell lines in culture
cells_per_well = number of cells plated into 384-well plate for Cell Titer-Glo assay
culture_media = media name for media and additives used in cell culture

media composition
culture_media = media name for media and additives used in cell culture
media_composition = text description of base media and components added

media components
culture_media = media name for media and additives used in cell culture
media_component = name of media component (either in base media or an added component)
component_amount_(mM) = concentration of media component

DESCRIPTION OF COMPUTATIONAL EXPERIMENTS

We wondered whether all lineage-based sub-populations and excluded sub-populations actually required consideration. We checked each lineage and sublineage separately, and combinations of lineage and exclusions for qualification in our analysis (see below). To assess the extent to which certain sub-populations of cell lines might have non-compound-specific sensitivity characteristics that confound our enrichment analyses, we undertook a two-stage strategy to (1) identify such sub-populations independently of their specific genetic lesions, and (2) characterize the consequences to enrichment analyses of analyzing various sub-populations separately versus together.

Identification of potentially confounding sub-populations

For qualitative (categorical) exclusions, we performed K-S tests to ask whether the exclusion of growth conditions (adherent, suspension, mixed) or individual lineages had a significant effect on AUC distributions. The following two exclusions were found to be significant: exclude suspension cell lines ($pKS < 3.67 \times 10^{-9}$); exclude hematopoetic cell lines ($pKS < 2.34 \times 10^{-10}$)

To find an appropriate threshold to identify cell lines with a statistically large number of mutations, we assumed that this subset of cell lines would have a different distribution of sensitivities than all other cell lines. We used a collection of i two-distribution Komolgorov-Smirnov (K-S) tests, subject to Bonferroni correction, to evaluate the difference between distributions of cell line sensitivities between the i most-frequently-mutated cell lines and the remainder. Cell lines that harbored a fraction of mutations greater than the most significant K-S test ($pKS < 1.63 \times 10^{-11}$) were those we termed 'genes mut high', of which there are 33 CCLs. A similar analysis was attempted to identify cell lines that were frequently sensitive to many compounds, but many K-S test significance values exceeded machine precision. We chose cell lines that were sensitive to more than 25% of the compounds as those we termed 'frequently sensitive', of which there are 32 CCLs. We therefore exclude these subsets of cell lines as part of our enrichment analyses.

For lineage and sublineage feature analysis, we checked whether each lineage or sublineage had more than 3 lines present in the dataset to qualify for inclusion. For lineage and exclusion combination analysis, we included combinations where there were at least three cell lines that passed one of the four exclusion criteria and were present in the specified lineage or sublineage (see User Guide S1 for additional details). Sub-lineages tested within lung are: adenocarcinoma, broncheoalveolar carcinoma, large cell carcinoma, non-small cell carcinoma, small-cell carcinoma, and squamous cell carcinoma; and within hematopoetic are: acute myeloid leukemia, diffuse large B-cell lymphoma, plasma cell myeloma.

The following are the lineage/exclusion combinations that qualified:

    ALL_CCL_LINEAGES__EXCLUDE_NONE
    ALL_CCL_LINEAGES__EXCLUDE_GENES_MUT_HIGH
    ALL_CCL_LINEAGES__EXCLUDE_FREQ_SENS
    ALL_CCL_LINEAGES__EXCLUDE_SUSPENSION
    ALL_CCL_LINEAGES__EXCLUDE_HEMATO
    CENTRAL_NERVOUS_SYSTEM__EXCLUDE_NONE
    ENDOMETRIUM__EXCLUDE_NONE
    ENDOMETRIUM__EXCLUDE_FREQ_SENS
    HEMATOPOIETIC_AND_LYMPHOID_TISSUE__EXCLUDE_NONE
    HEMATOPOIETIC_AND_LYMPHOID_TISSUE__EXCLUDE_GENES_MUT_HIGH
    HEMATOPOIETIC_AND_LYMPHOID_TISSUE__EXCLUDE_FREQ_SENS
    LARGE_INTESTINE__EXCLUDE_NONE
    LARGE_INTESTINE__EXCLUDE_GENES_MUT_HIGH
    LARGE_INTESTINE__EXCLUDE_FREQ_SENS
    LIVER__EXCLUDE_NONE
    LUNG__EXCLUDE_NONE
    LUNG__EXCLUDE_GENES_MUT_HIGH
    LUNG__EXCLUDE_FREQ_SENS
    LUNG__EXCLUDE_SUSPENSION
    OESOPHAGUS__EXCLUDE_NONE
    OVARY__EXCLUDE_NONE
    OVARY__EXCLUDE_GENES_MUT_HIGH
    OVARY__EXCLUDE_FREQ_SENS
    PANCREAS__EXCLUDE_NONE
    PLEURA__EXCLUDE_NONE
    SKIN__EXCLUDE_NONE
    SKIN__EXCLUDE_SUSPENSION
    SOFT_TISSUE__EXCLUDE_NONE
    STOMACH__EXCLUDE_NONE
    STOMACH__EXCLUDE_FREQ_SENS
    URINARY_TRACT__EXCLUDE_NONE
    ACUTE_MYELOID_LEUKEMIA__EXCLUDE_NONE
    ADENOCARCINOMA__EXCLUDE_NONE
    ADENOCARCINOMA__EXCLUDE_GENES_MUT_HIGH
    ADENOCARCINOMA__EXCLUDE_FREQ_SENS
    BRONCHIOLOALVEOLAR_ADENOCARCINOMA__EXCLUDE_NONE
    BRONCHIOLOALVEOLAR_ADENOCARCINOMA__EXCLUDE_GENES_MUT_HIGH
    DIFFUSE_LARGE_B_CELL_LYMPHOMA__EXCLUDE_NONE
    DIFFUSE_LARGE_B_CELL_LYMPHOMA__EXCLUDE_FREQ_SENS
    LARGE_CELL_CARCINOMA__EXCLUDE_NONE
    LARGE_CELL_CARCINOMA__EXCLUDE_FREQ_SENS
    NON_SMALL_CELL_CARCINOMA__EXCLUDE_NONE
    NON_SMALL_CELL_CARCINOMA__EXCLUDE_GENES_MUT_HIGH
    NON_SMALL_CELL_CARCINOMA__EXCLUDE_FREQ_SENS
    PLASMA_CELL_MYELOMA__EXCLUDE_NONE
    PLASMA_CELL_MYELOMA__EXCLUDE_FREQ_SENS
    SMALL_CELL_CARCINOMA__EXCLUDE_NONE
    SMALL_CELL_CARCINOMA__EXCLUDE_FREQ_SENS
    SMALL_CELL_CARCINOMA__EXCLUDE_SUSPENSION

SQUAMOUS_CELL_CARCINOMA__EXCLUDE_NONE

(1) For each gene-compound-dataset combination, we report enrichment scores, which of the lineage subpopulations, or exclusions, if any, it came from, and which of the extrema exclusions were applied, if any. The scores are defined above.
(2) For each significant enrichment, meeting a square_max_p_value<0.05 and FDR_q_value<0.25 threshold, we report it along with the experimental variables that produced it.
(3) The above was computed for ALL COMPOUNDS across all DATASETS and EXPERIMENTS

Datasets
lineage: cancer cell lineages
sublineage: cancer cell lineage subtypes
CNV: all copy-number mutant calls
CNV-L: low-copy calls only   (0 copies)
CNV-H: high-copy calls only (>= 8 copies)
TES: all targeted exome sequencing hybrid capture mutant calls
TES-CNV: union of TES calls and CNV calls
TES-A: targeted exome sequencing hybrid capture putative activating mutations (non-neutral missense mutations, in-frame shifts)
TES-A-CNV-H: union of TES-A calls and CNV-H calls
TES-C: targeted exome sequencing hybrid capture putative activating mutations within 3 amino acids of sites reported by the Sanger Institute's COSMIC database (this is a subset of TES-A)
TES-L: targeted exome sequencing hybrid capture putative loss-of-function mutations (nonsense mutations, indels, out-of-frame shifts)
TES-L-CNV-L: union of TES-L calls and CNV-L calls
Onco: all Oncomap mutant calls
MUT: union of TES and Onco mutation calls
OncoGeno: all cell line genotypes (combination of ALL mutant gene calls from Oncomap for each cell line)

HOW TO MANIPULATE THE WORKBOOK

Instructions are for used with Excel 2007 (may differ slightly with other versions)

- Open the resource file with Excel. The list is currently sorted by FDR_q_value (best at the top of the list).
- Select to the "View" tab, highlight top row and under "Freeze Panes", select "Freeze Top Row". This will allow you to scroll down the list and always be able to view the column headers
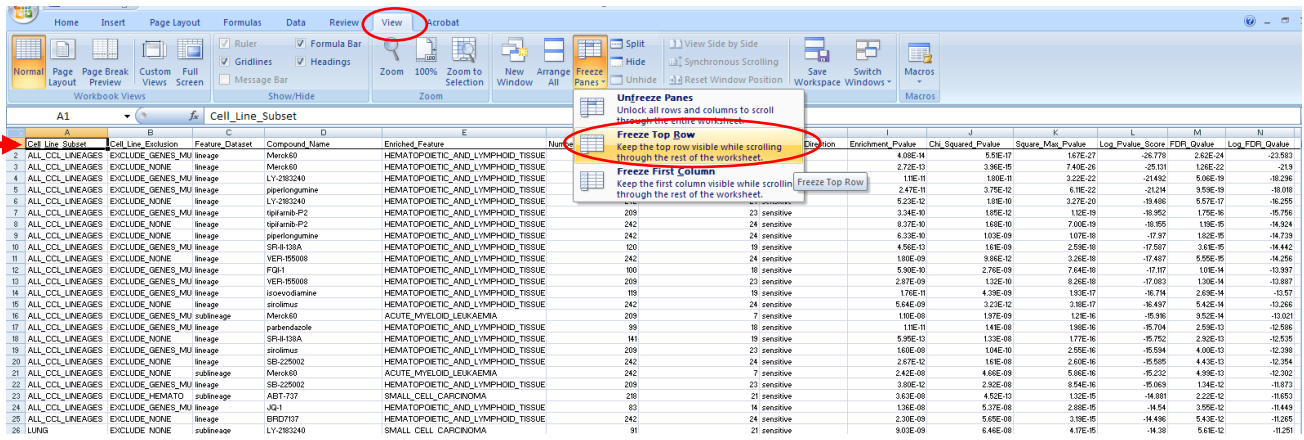
- Select the "Data" tab, highlight the top row and select "Filter". This places drop-down menu arrows that will allow you to select features within each column.
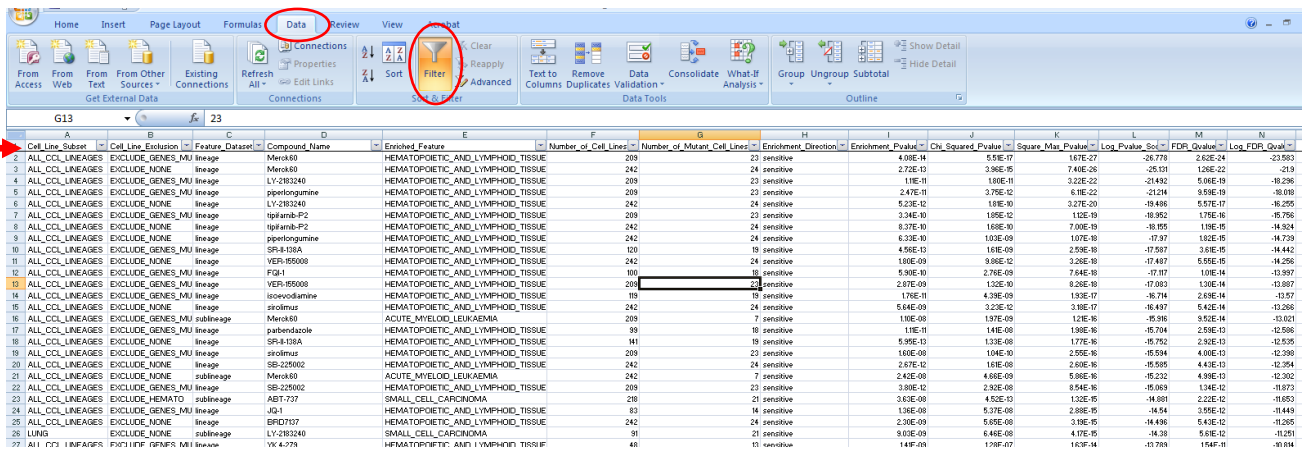
- Clicking an arrow within a given column leads to a drop down menu with all available choices within that column. For instance: under the column "compound_name", the drop-down menu shows a list of compounds. You can focus in on a compound of interest by de-selecting all of the others, and selecting only one. This will generate a list of all enrichments involving that compound.

- Click on the box for "Select All" to remove the arrow and deselect all compounds. Scroll down the list to select your compounds of choice (example: navitoclax)

- This now gives you a list of all experiments, data sets and features that correlate to your compound of interest. Use the filter option in other columns to fine-tune to your interests (see examples below).

Example 1: Filter on a single feature, a single compound, and a single dataset and compare scores across experiments

- Follow the above guidelines to filter on:
  - o compound_name: P-0850
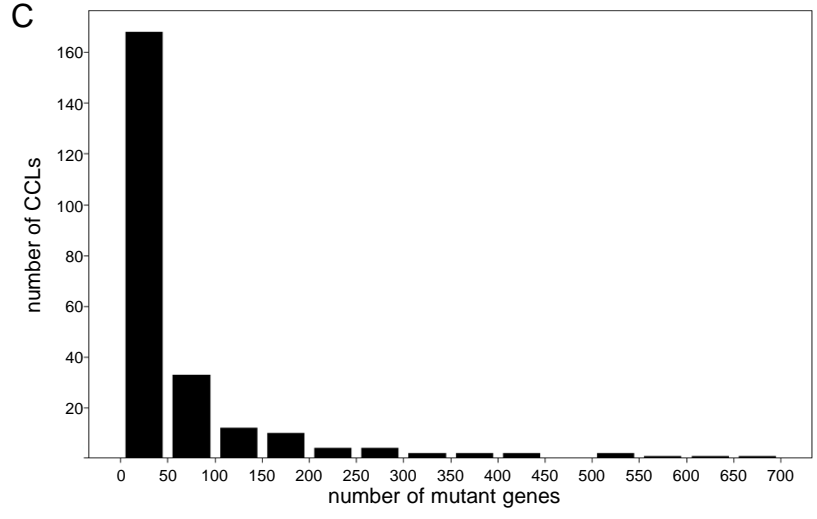 - o enriched_feature: BRAF
  - o feature_dataset: Onco

- In the "cell_line_subset" column, you will observe that this correlation persists across almost all exclusions within the ALL_CCL_LINEAGES category.
- In the "number_of_cell_lines" and "number_of_mutant_cell_lines" columns, you will observe that the numbers change based on the exclusion (e.g., 12 mutants in the EXCLUDE_GENES_MUT_HIGH vs 14 mutants in the EXCLUDE_NONE).
- In the "square_max_p_value" column, you can observe the change in probability score depending on the exclusion. All values are $p < 0.05$, thus indicating statistical significance. In the "FDR_q_value" column, you can observe the change in false discovery rate. All values are $q < 0.25$.
- Conclusion: mutations in BRAF correlate to sensitivity to P-0850 (a vermurafenib analog) treatment regardless of the experimental exclusions.
- To determine the exact nature of the mutation called within a dataset, go to the Broad/Novartis Cancer Cell Line Encyclopedia portal (http://www.broadinstitute.org/ccle/home) and access the appropriate files.

Example 2: Filter on a single gene, a single compound, a single experiment and compare scores across datasets

- Follow the above guidelines to filter on:
    - compound_name: P-0850
    - enriched_feature: BRAF
    - cell_line_subset: ALL_CCL_LINEAGES
    - cell_line_exclusion: EXCLUDE_NONE

- In the "feature_dataset" column, you will observe that this correlation persists across a number of datasets.
- In the "number_of_cell_lines" and "number_of_mutant_cell_lines" columns, you will observe that the numbers change based on the dataset.
- In the "square_max_p_value" column, you can observe the change in probability score depending on the exclusion. All values are $p < 0.05$, thus indicating statistical significance, however, as the datasets get broader, the probability scores increase suggesting that more specific mutation calls better correlated to compound sensitivity. In the "FDR_q_value" column, you can observe the change in false discovery rate. All values are $q < 0.25$.

- Conclusion: mutations in BRAF correlate to sensitivity to P-0850 treatment. As both Onco mutations and TES-C mutations are at the top of the list, it is highly likely that mutations commonly found in human tumors best correlate to sensitivity to P-0850.
- To determine the exact nature of the mutation called within a dataset, go to the Broad-Novartis Cancer Cell Line Encyclopedia portal (http://www.broadinstitute.org/ccle/home) and access the appropriate files.

Example 3: Filter on a single gene, a single compound, a single dataset and compare scores across experiments. In some cases, experiments will not be present in the list, due to the thresholds applied (square_max_p_value < 0.05, "FDR_q_value")

- Follow the above guidelines to filter onto:
    - compound_name: navitoclax
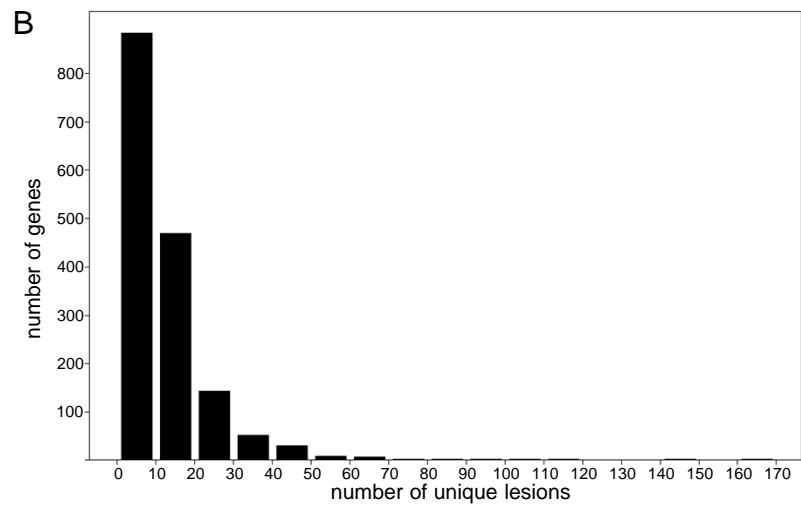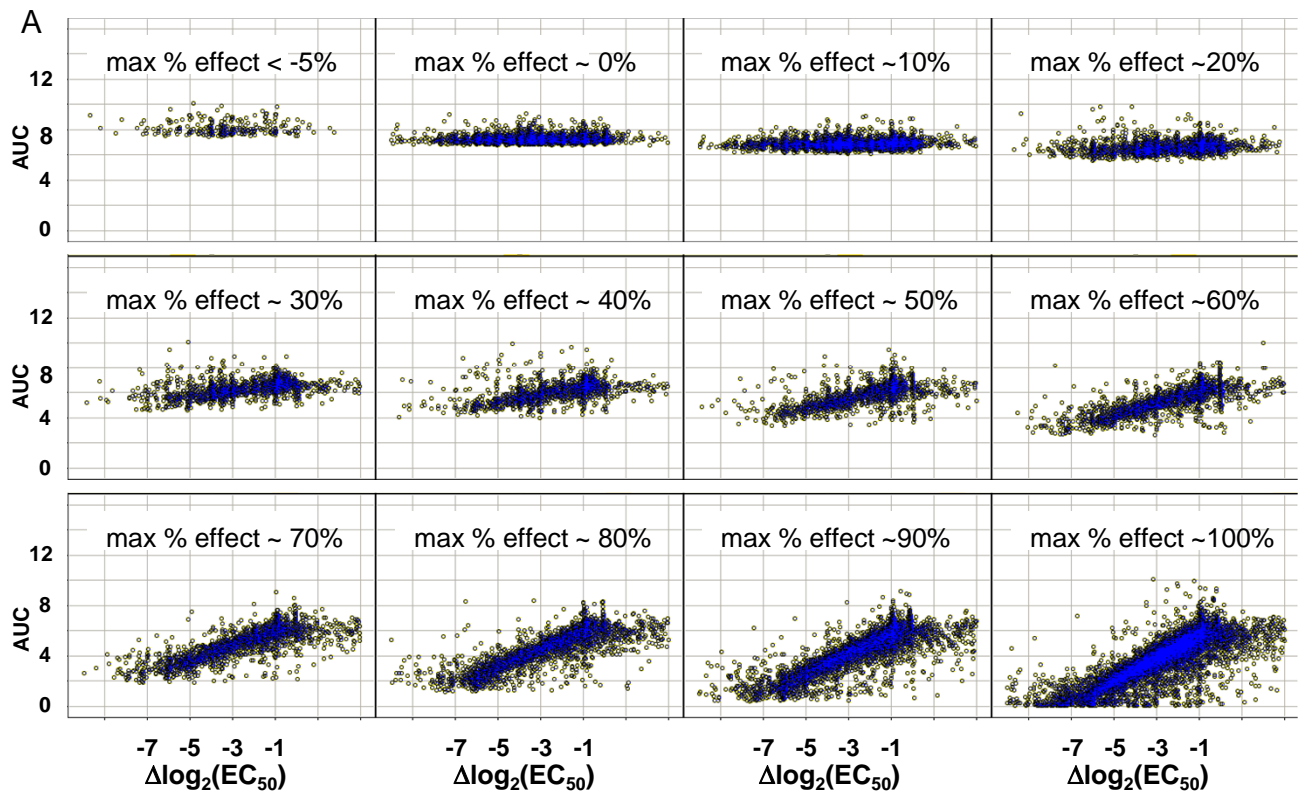   - enriched_feature: CTNNB1
    - feature_dataset: Onco

- In the "cell_line_subset" column, you will observe that this correlation persists across all "cell_line_exclusion" catagories within ALL_CCL_LINEAGES, except EXCLUDE_GENES_MUT_HIGH.
- In the "number_of_cell_lines" and "number_of_mutant_cell_lines" columns, you will observe the numbers of cell lines and mutants are very similar, indicating there is not much different between the exclusions.
- In the "square_max_p_value" column, you can observe the change in probability score is not highly affected, also confirming there is not much different between the experiments. All values are p < 0.05, thus indicating statistical significance. In the "FDR_q_value" column, you can observe the change in false discovery rate. All values are q < 0.25.
- Many of the GENES_MUT_HIGH cell lines are within endometrial and large intestine lineage, which correlates with the majority of the CTNNB1 mutants.
- To determine the exact nature of the mutation called within a dataset, go to the Broad/Novartis Cancer Cell Line Encyclopedia portal (http://www.broadinstitute.org/ccle/home) and access the appropriate files.
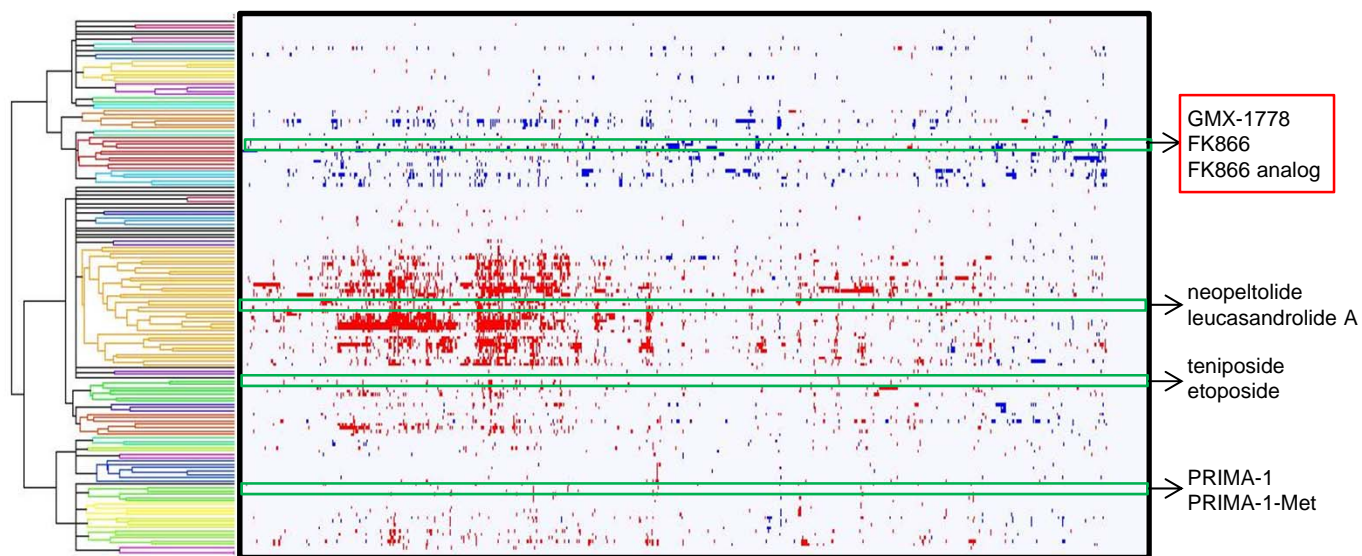
Example 4: Finding a connection within a lineage alone experiment that does not occur in an "ALL_CCL_LINEAGES" experiment.
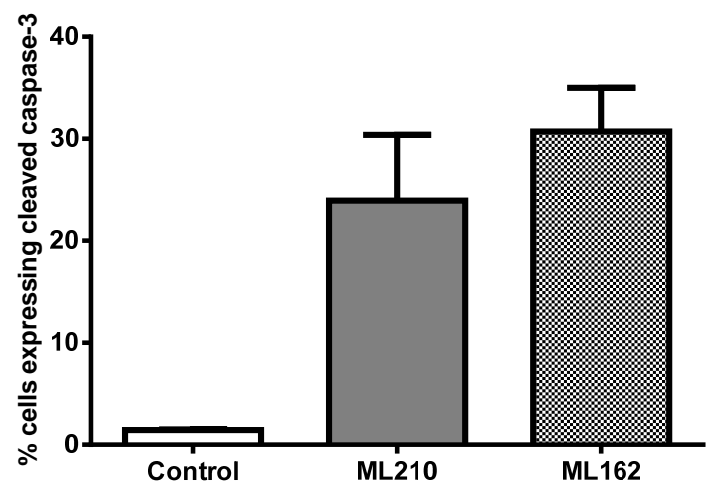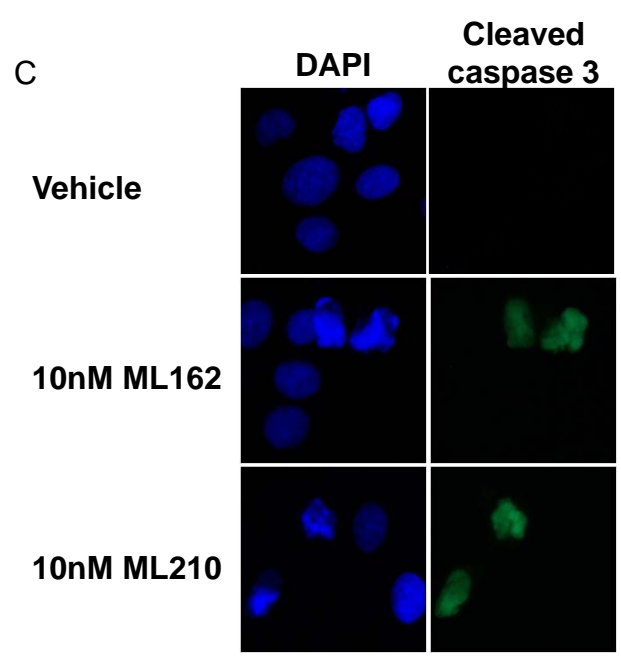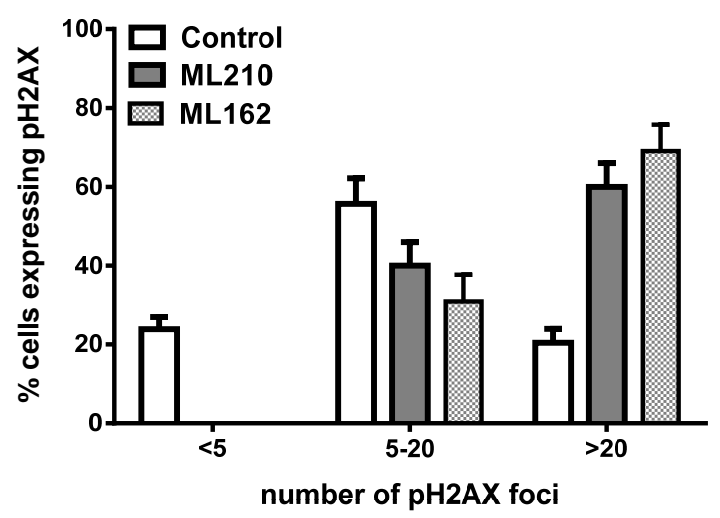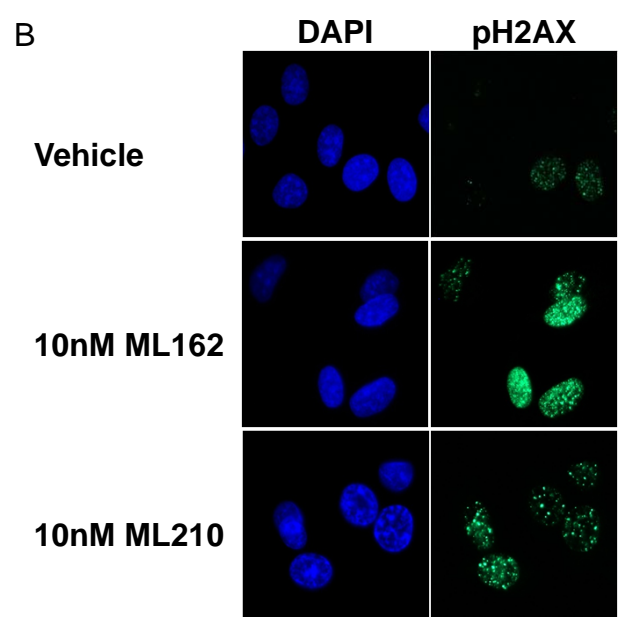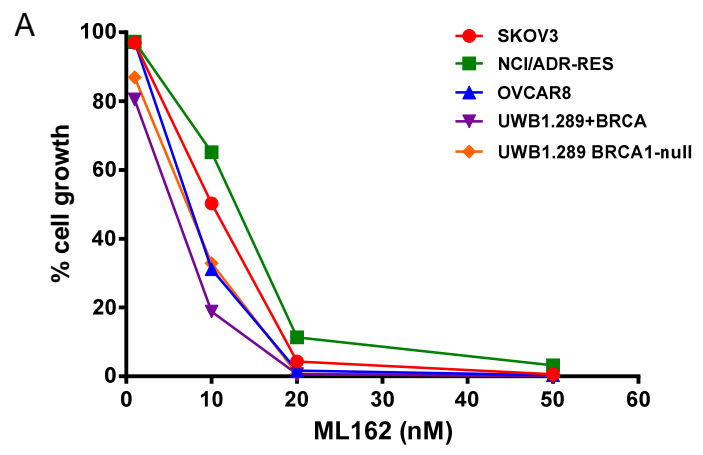
- Follow the above guidelines to filter onto:
    - compound_name: nertinib
    - enriched_feature: EGFR
    - feature_dataset: Onco

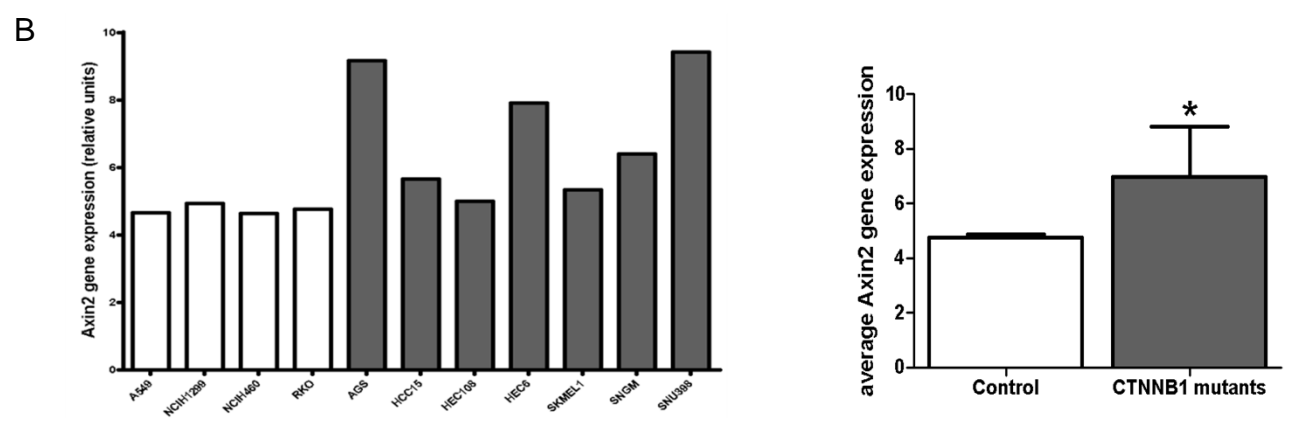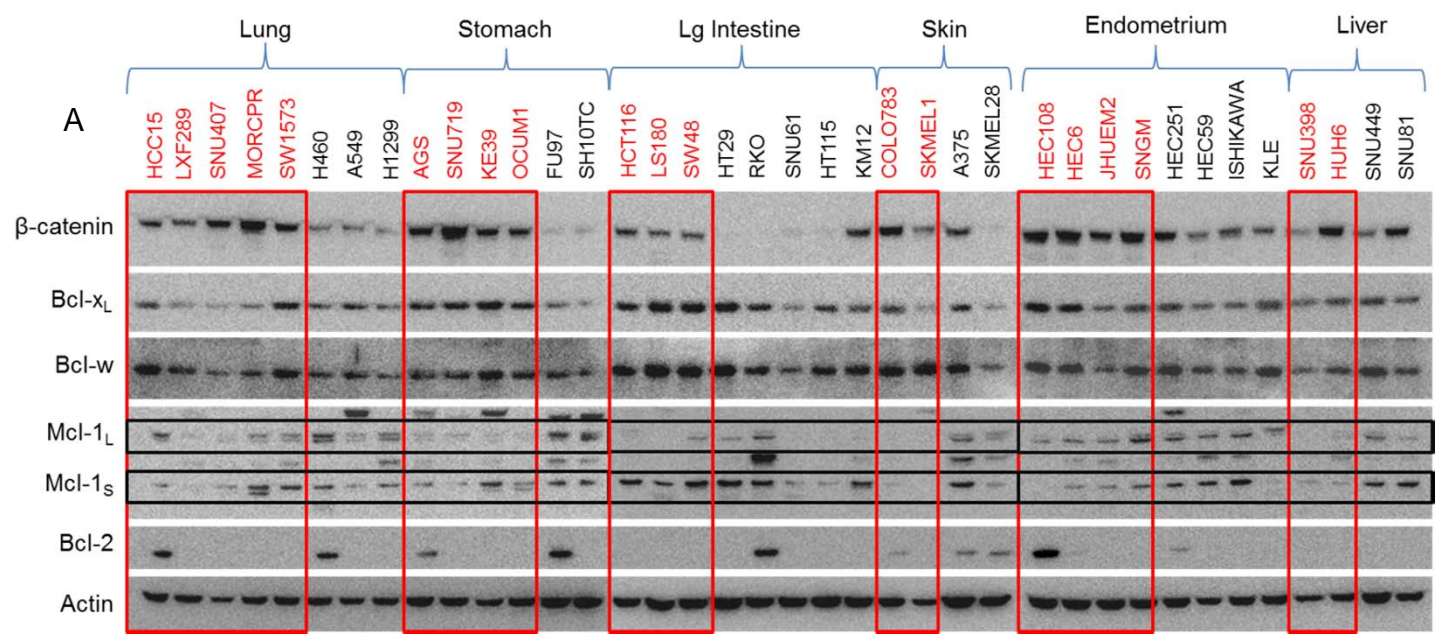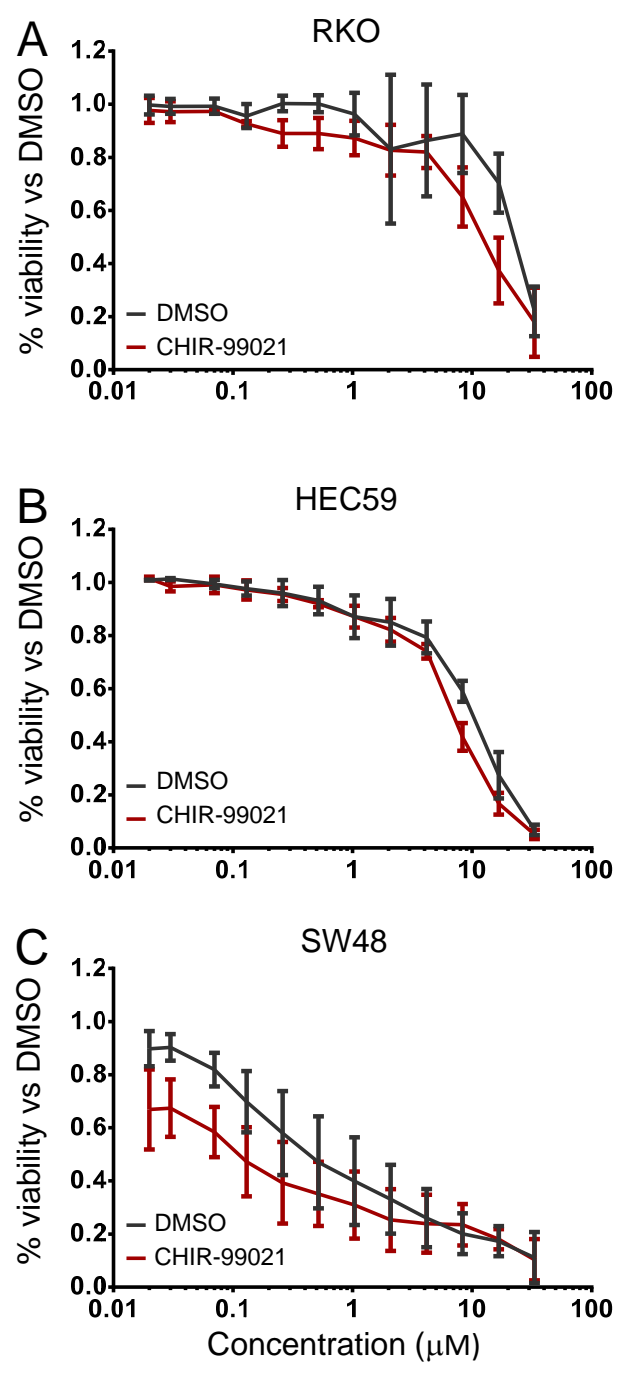|	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	cell_line_subset	cell_line_exclusion	feature_dataset	compound_name	enriched_feature	number_of_cell_lines	number_of_mutant_cell_lines	enrichment_direction	enrichment_p_value	chi_squared_p_value	squared_max_p_value	log_p_value_score	FDR_q_value	log_q_value_score
409981	LUNG	EXCLUDE_FREQ_SENS	Onco	neratinib	EGFR	42	4	sensitive	6.25E-05	0.044987	0.0020238	-2.6938	0.06874	-1.1628
452242	LUNG	EXCLUDE_SUSPENSION	Onco	neratinib	EGFR	35	4	sensitive	1.91E-05	0.02057	0.00042312	-3.3736	0.096367	-1.0161
497365														

- You will observe that this correlation occurs only within in the "LUNG" lineage.
- In the "square_max_p_value" column, all values are $p < 0.05$, thus indicating statistical significance. In the "FDR_q_value" column, you can observe the change in false discovery rate. All values are q < 0.25.
- To determine the exact nature of the mutation called within a dataset, go to the Broad/Novartis Cancer Cell Line Encyclopedia portal (http://www.broadinstitute.org/ccle/home) and access the appropriate files.

Supplemental Figure 1

GMX-1778
FK866
FK866 analog

neopeltolide
leucasandrolide A

teniposide
etoposide

PRIMA-1
PRIMA-1-Met

Supplemental Figure 3

Supplemental Figure 4

Supplemental Figure 5

eTOC

We measured the sensitivity of genomically characterized CCLs to an Informer Set of small molecules that targets many nodes in cell circuitry, uncovering protein dependencies that: 1) associate with specific lineages or cancer-genomic alterations and 2) can be targeted by small molecules. We created a Cancer Therapeutics Response Portal (www.broadinstitute.org/ctrp) to enable users to correlate CCL features to sensitivity and develop novel therapeutic hypotheses to accelerate discovery of drugs matched to patients by their cancer genotype.