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 ($p_{KS} < 3.67\times10^{-9}$); exclude hematopoetic cell lines ($p_{KS} < 2.34\times10^{-10}$)

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 ($p_{KS} < 1.63\times10^{-11}$) 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 $(s_i - \mu_{random,n})/\sigma_{random}$. 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 ($pKS < 3.67\times10^{-9}$); exclude hematopoetic cell lines ($pKS < 2.34\times10^{-10}$)

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 ($pKS < 1.63\times10^{-11}$) 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

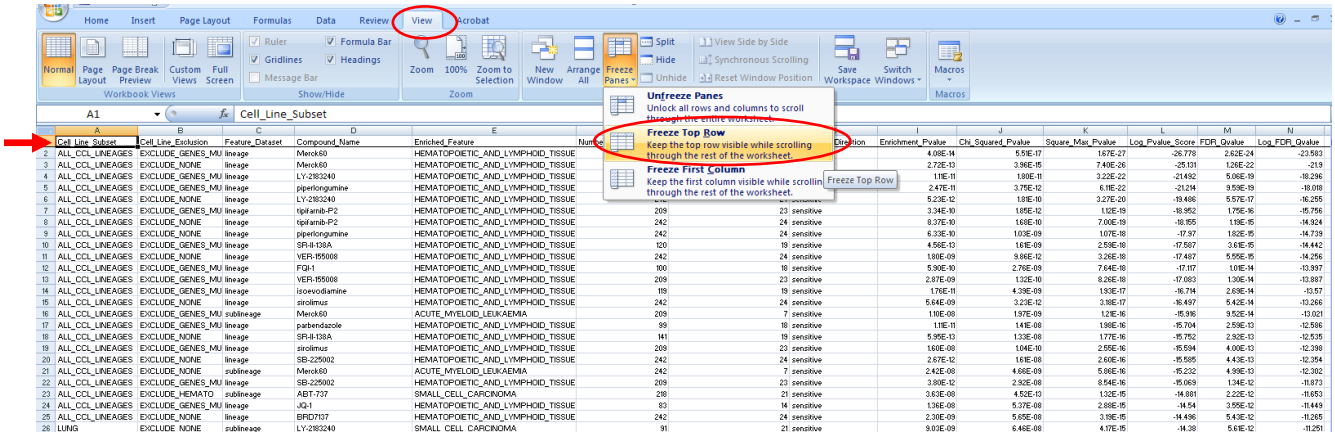
Datasets

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)

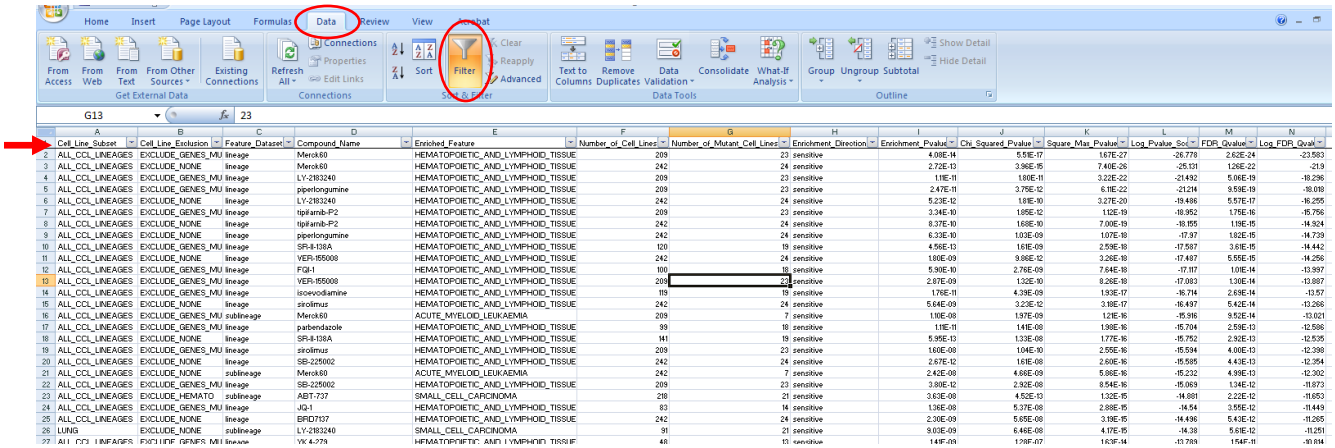
HOW TO MANIPULATE THE WORKBOOK

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

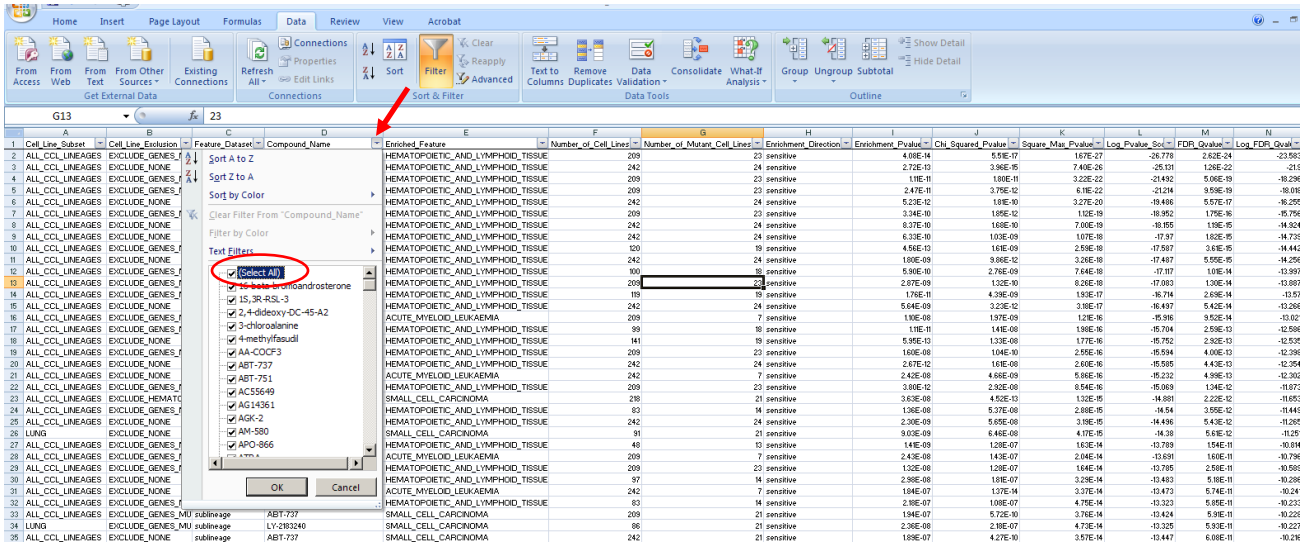
- Open the resource file with Excel. The list is currently sorted by FDR_q_value (best at the top of the list).
- Select to the "View" tab, highlight top row and under "Freeze Panes", select "Freeze Top Row". This will allow you to scroll down the list and always be able to view the column headers



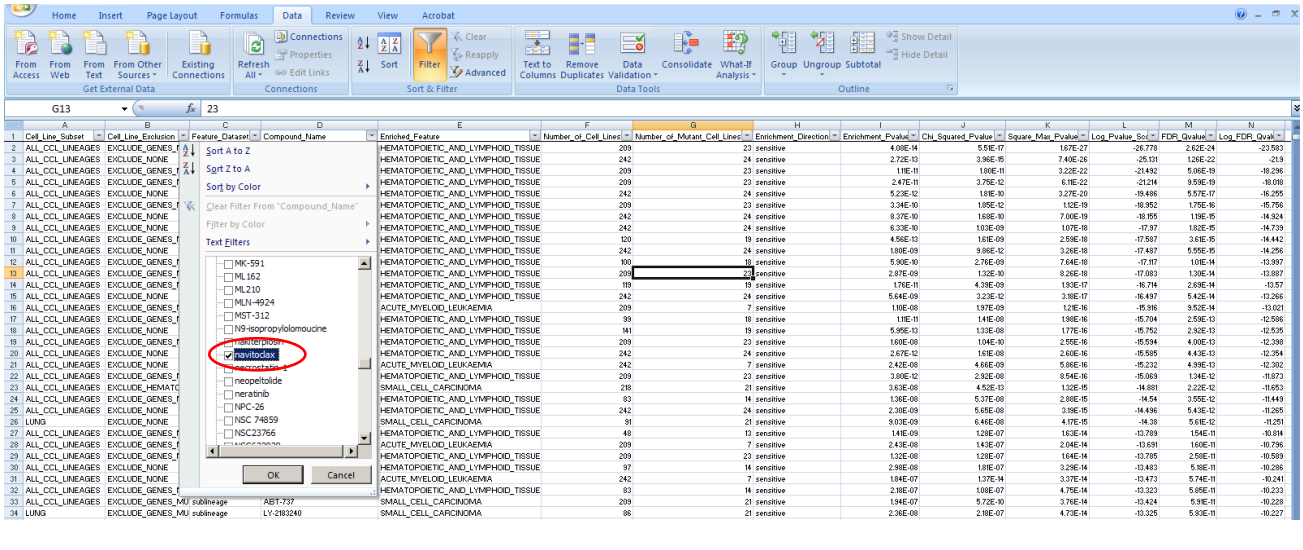
- Select the “Data” tab, highlight the top row and select “Filter”. This places drop-down menu arrows that will allow you to select features within each column.



- 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.



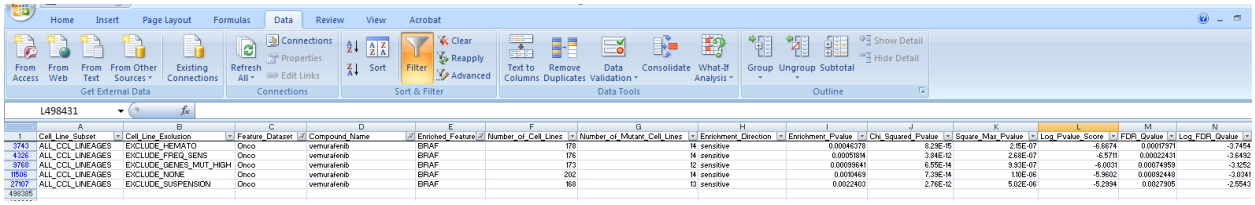
- Click on the box for “Select All” to remove the arrow and deselect all compounds. Scroll down the list to select your compounds of choice (example: navitoclax)



- 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:
    - compound_name: P-0850
    - enriched_feature: BRAF
    - feature_dataset: Onco

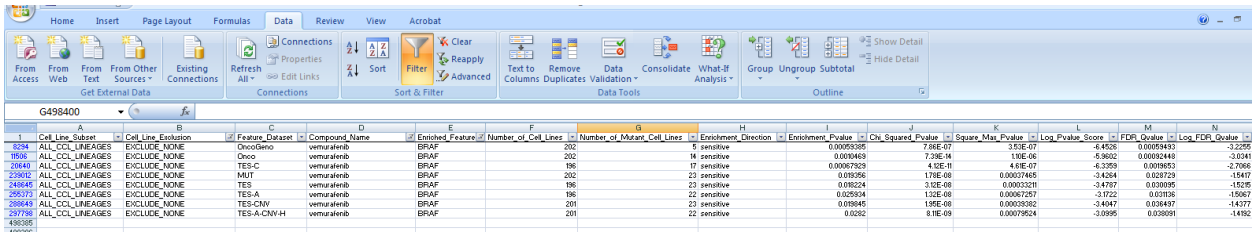


	Cell_Line_Subset	Cell_Line_Exclusion	Feature_Dataset	Compound_Name	Enriched_Feature	Number_of_Cell_Lines	Number_of_Mutant_Cell_Lines	Enrichment_Direction	Enrichment_Pvalue	Chi_Squared_Pvalue	Square_Max_Pvalue	Log_Pvalue_Score	FDR_Qvalue	Log_FDR_Qvalue
3743	ALL_CCL_LINEAGES	EXCLUDE_HEMATO	Onco	vemurafenib	BRAF	179	14	sensitive	0.00046378	8.23E-15	2.18E-07	-6.6674	0.00077971	-3.7454
4326	ALL_CCL_LINEAGES	EXCLUDE_FREQ_SENS	Onco	vemurafenib	BRAF	176	14	sensitive	0.0005184	3.84E-12	2.68E-07	-6.5711	0.00022431	-3.6492
9768	ALL_CCL_LINEAGES	EXCLUDE_GENES_MUT_HIGH	Onco	vemurafenib	BRAF	173	12	sensitive	0.00039641	6.55E-14	9.93E-07	-6.0031	0.00074959	-3.1252
11506	ALL_CCL_LINEAGES	EXCLUDE_NONE	Onco	vemurafenib	BRAF	202	14	sensitive	0.0010469	7.39E-14	1.10E-06	-5.9602	0.00092449	-3.0341
27907	ALL_CCL_LINEAGES	EXCLUDE_SUSPENSION	Onco	vemurafenib	BRAF	168	13	sensitive	0.0022403	2.76E-12	5.02E-06	-5.2994	0.0027905	-2.5543

- 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



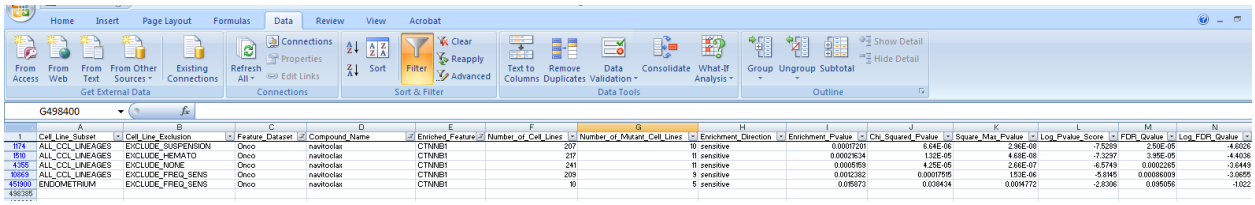
	Cell_Line_Subset	Cell_Line_Exclusion	Feature_Dataset	Compound_Name	Enriched_Feature	Number_of_Cell_Lines	Number_of_Mutant_Cell_Lines	Enrichment_Direction	Enrichment_Pvalue	Chi_Squared_Pvalue	Square_Max_Pvalue	Log_Pvalue_Score	FDR_Qvalue	Log_FDR_Qvalue
8294	ALL_CCL_LINEAGES	EXCLUDE_NONE	OncoGeno	vemurafenib	BRAF	202	5	sensitive	0.00059395	7.84E-07	3.53E-07	-6.4526	0.00084493	-3.2255
11506	ALL_CCL_LINEAGES	EXCLUDE_NONE	Onco	vemurafenib	BRAF	202	14	sensitive	0.0010469	7.39E-14	1.10E-06	-5.9602	0.00092448	-3.0341
20640	ALL_CCL_LINEAGES	EXCLUDE_NONE	TES-C	vemurafenib	BRAF	196	17	sensitive	0.00067329	4.12E-11	4.61E-07	-6.3359	0.0019653	-2.7066
239012	ALL_CCL_LINEAGES	EXCLUDE_NONE	MUT	vemurafenib	BRAF	202	23	sensitive	0.019356	1.78E-08	0.00037465	-3.4264	0.029729	-1.5417
248645	ALL_CCL_LINEAGES	EXCLUDE_NONE	TES	vemurafenib	BRAF	196	23	sensitive	0.018224	3.12E-08	0.00033211	-3.4797	0.030095	-1.5215
265373	ALL_CCL_LINEAGES	EXCLUDE_NONE	TES-A	vemurafenib	BRAF	196	22	sensitive	0.025934	1.32E-08	0.00067257	-3.1722	0.031136	-1.5067
286649	ALL_CCL_LINEAGES	EXCLUDE_NONE	TES-CNV	vemurafenib	BRAF	201	23	sensitive	0.019845	1.95E-08	0.00039382	-3.4047	0.036497	-1.4377
297798	ALL_CCL_LINEAGES	EXCLUDE_NONE	TES-A-CNV-H	vemurafenib	BRAF	201	22	sensitive	0.0282	9.11E-09	0.00079524	-3.0995	0.038891	-1.4192

- 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

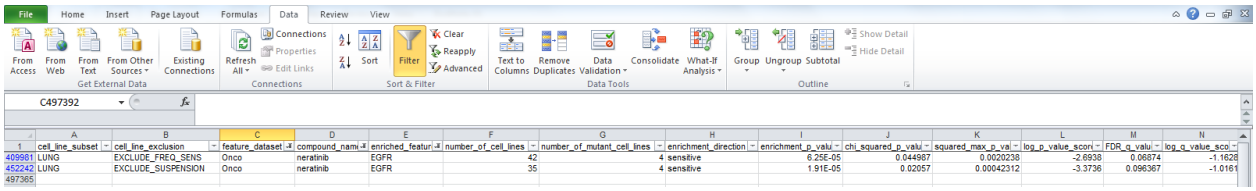


	Cell_Line_Subset	Cell_Line_Exclusion	Feature_Dataset	Compound_Name	Enriched_Feature	Number_of_Cell_Lines	Number_of_Mutant_Cell_Lines	Enrichment_Direction	Enrichment_Pvalue	Chi_Squared_Pvalue	Square_Max_Pvalue	Log_Pvalue_Score	FDR_Qvalue	Log_FDR_Qvalue
174	ALL_CCL_LINEAGES	EXCLUDE_SUSPENSION	Onco	navitoclax	CTNNB1	207	10	sensitive	0.00097201	6.64E-06	2.96E-08	-7.5289	2.59E-05	-4.5928
1510	ALL_CCL_LINEAGES	EXCLUDE_HEMATO	Onco	navitoclax	CTNNB1	217	11	sensitive	0.00021634	1.32E-05	4.68E-08	-7.3297	3.95E-05	-4.4036
4365	ALL_CCL_LINEAGES	EXCLUDE_NONE	Onco	navitoclax	CTNNB1	241	11	sensitive	0.0005159	4.25E-05	2.66E-07	-6.5749	0.0002265	-3.6449
19863	ALL_CCL_LINEAGES	EXCLUDE_FREQ_SENS	Onco	navitoclax	CTNNB1	209	9	sensitive	0.0012382	0.00017515	1.53E-06	-6.1945	0.0008089	-3.0855
451800	ENDOMETRIUM	EXCLUDE_FREQ_SENS	Onco	navitoclax	CTNNB1	10	5	sensitive	0.015973	0.038434	0.0014772	-2.8306	0.095056	-1.022
486385														

- 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.

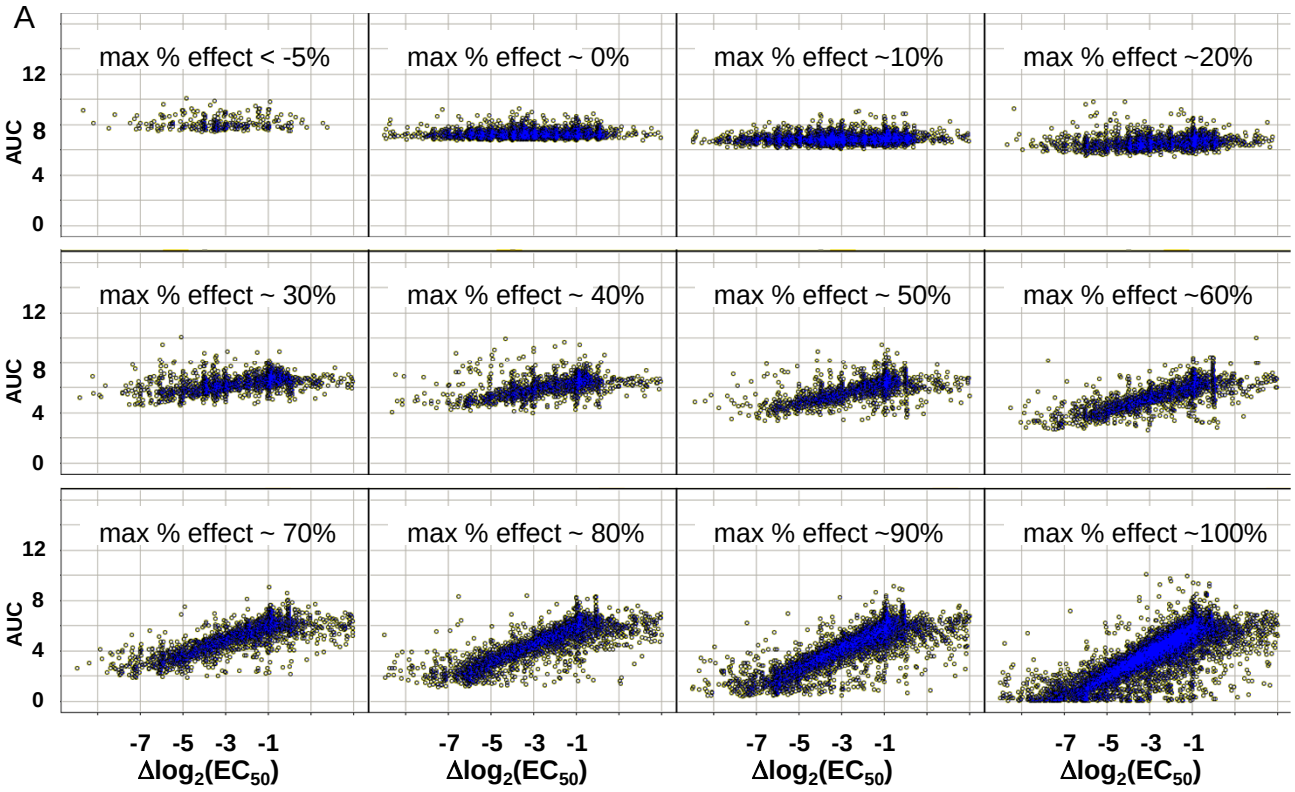
Example 4: Finding a connection within a lineage alone experiment that does not occur in an "ALL_CCL_LINEAGES" experiment.

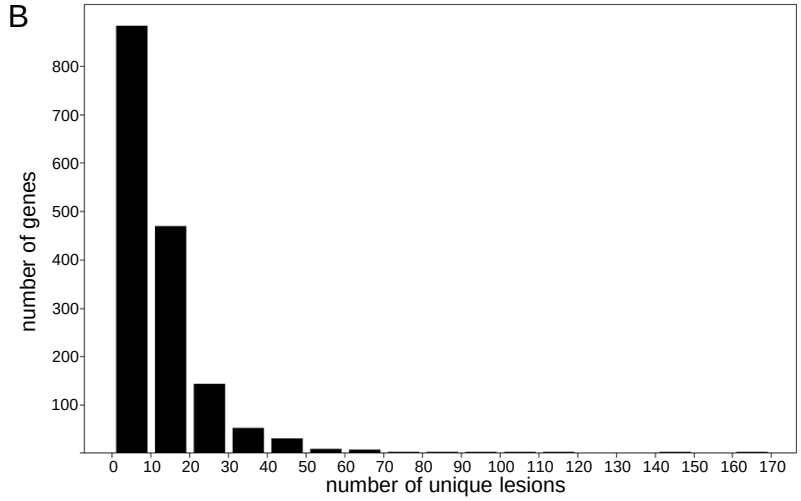
- Follow the above guidelines to filter onto:
    - compound_name: nertinib
    - enriched_feature: EGFR
    - feature_dataset: Onco

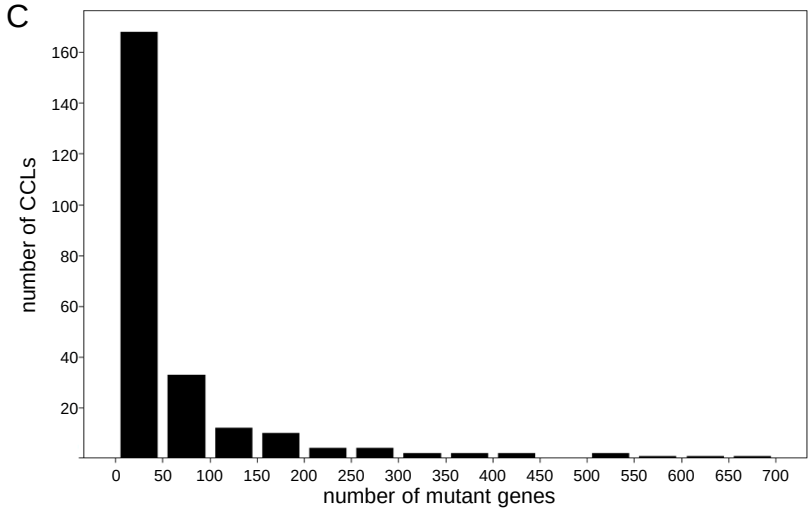


- You will observe that this correlation occurs only within in the “LUNG” lineage.
- In the “square_max_p_value” column, all values are $p < 0.05$, thus indicating statistical significance. In the “FDR_q_value” column, you can observe the change in false discovery rate. All values are $q < 0.25$.
- To 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.

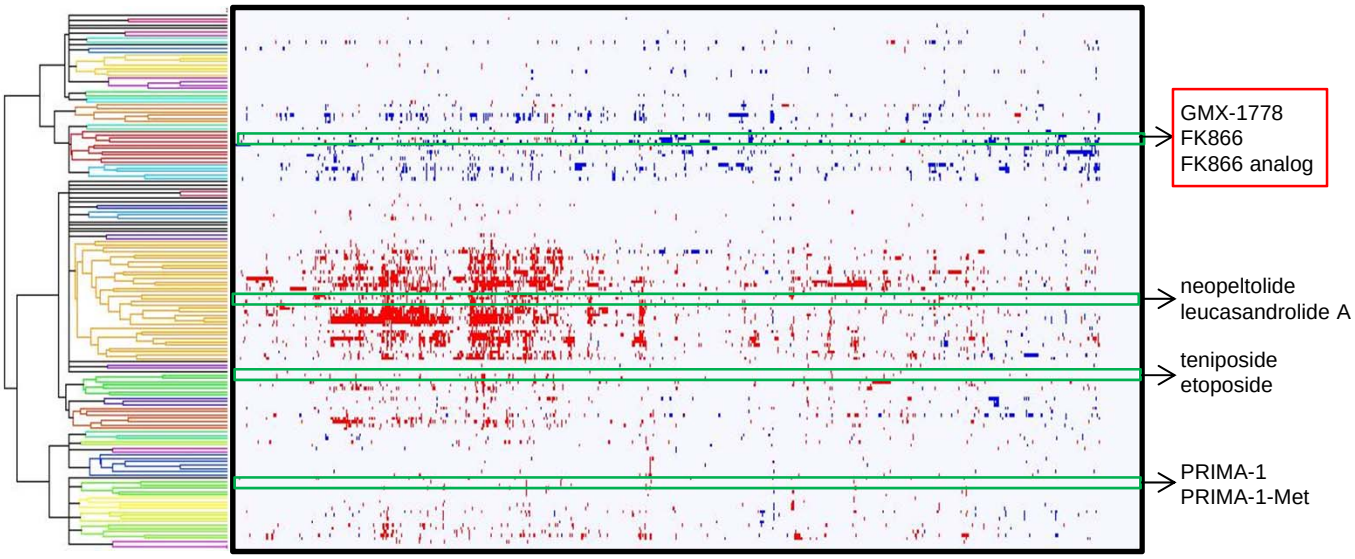
Supplemental Figure 1



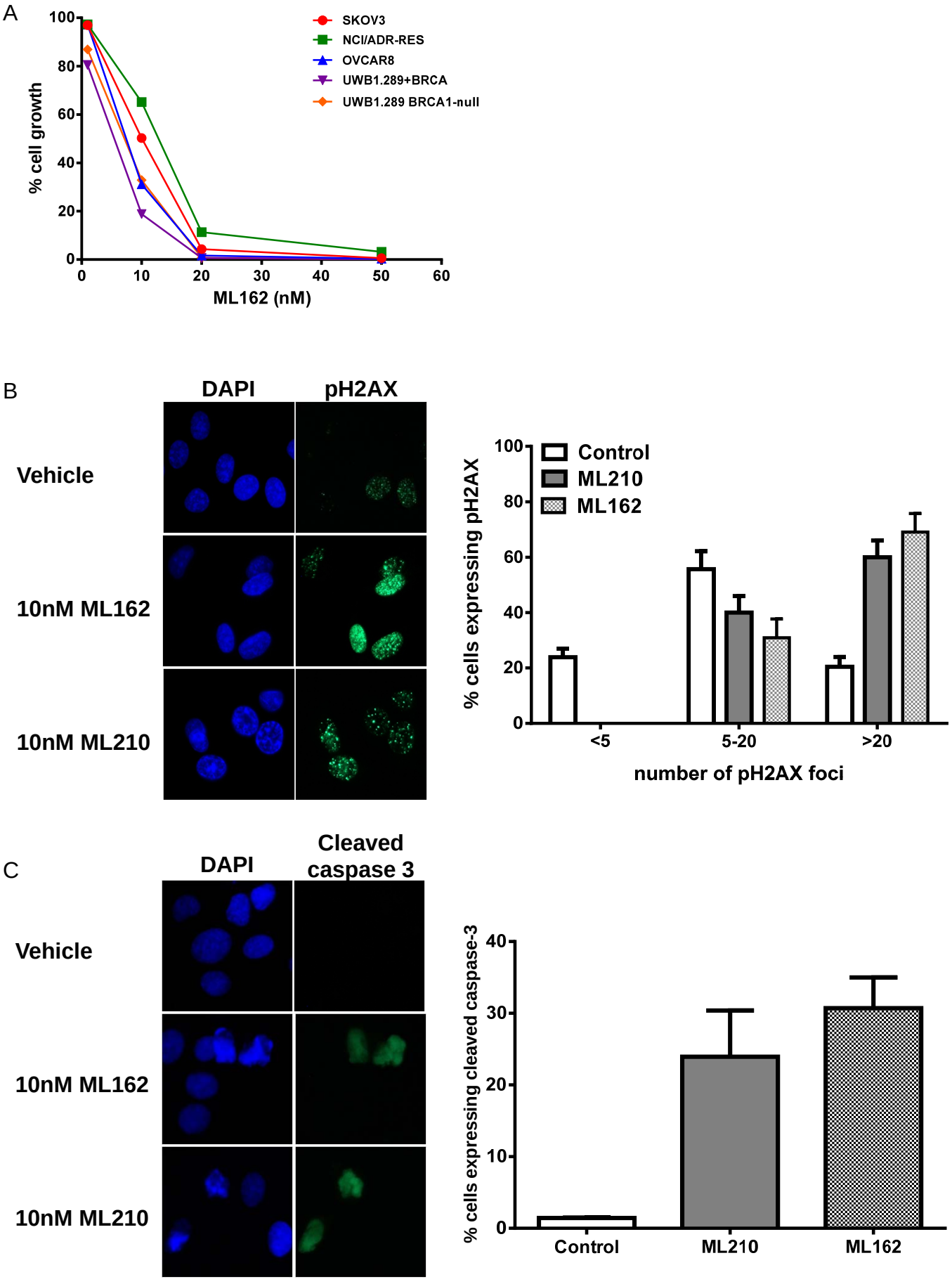




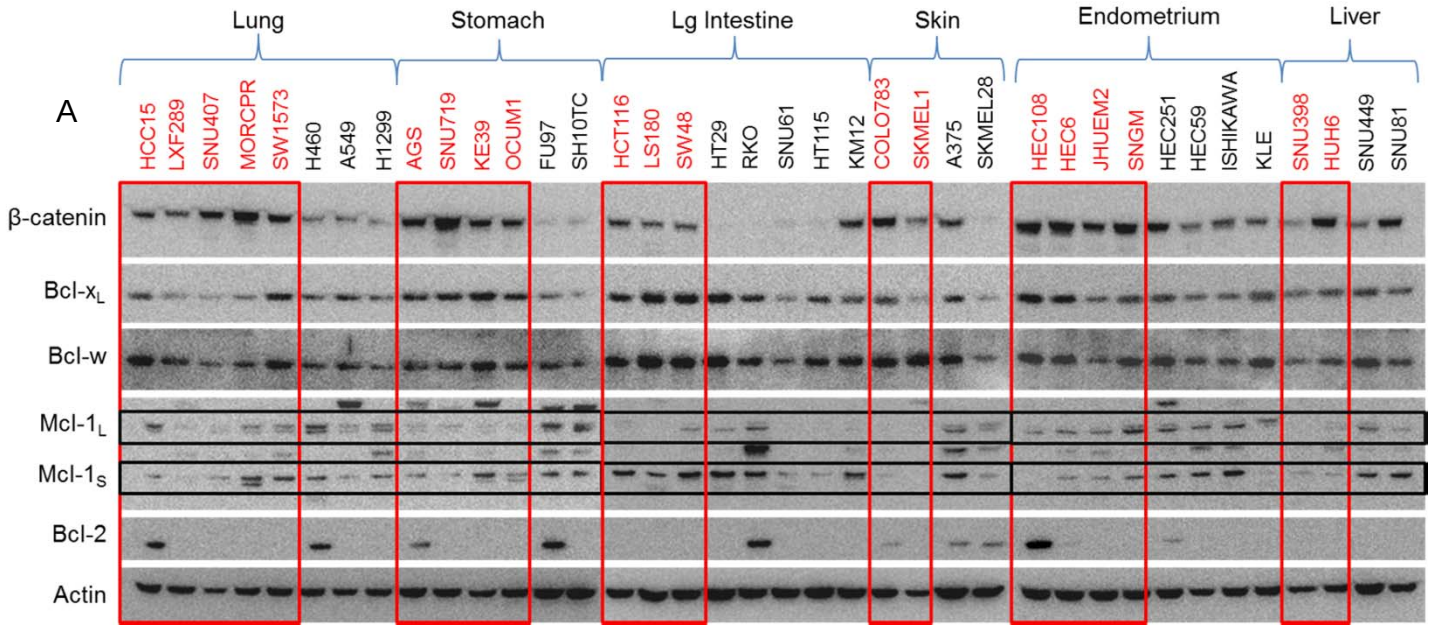
Supplemental Figure 2

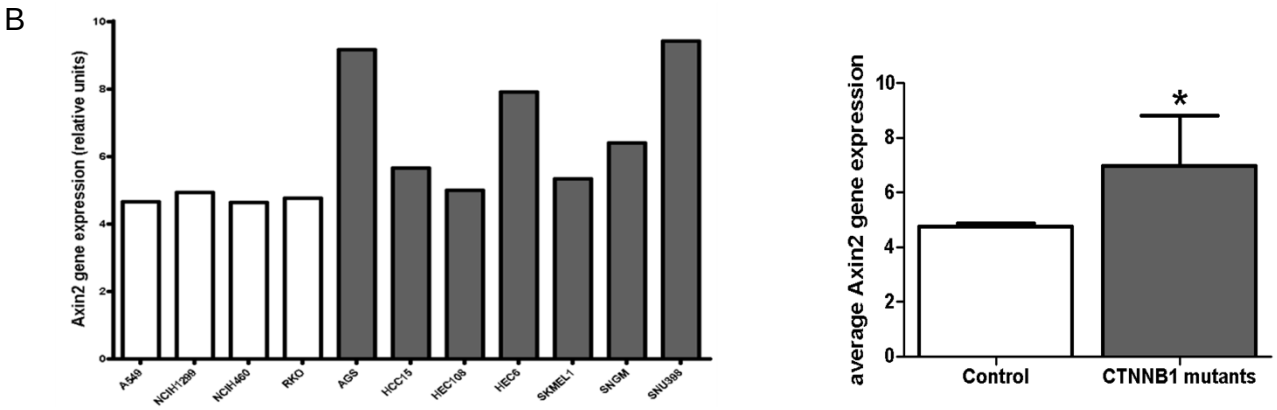


Supplemental Figure 3



Supplemental Figure 4





Supplemental Figure 5