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 μ M 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 (\mathbf{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 (I) 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% CO₂. 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 \leq 3.5 for at least one cell line (AUC=7 corresponds to no compound effect), for resistant enrichments an AUC \geq 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 - $\mu_{random,n}/\sigma_{random}$. 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, <u>www.ingenuity.com</u>). 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 prepared by incubating cell pellets were 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% CO₂, 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 performed 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 μ M GSK3 β 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, 1.5×10^6 cells were seeded into T75 flasks, allowed to adhere overnight, then treated with either DMSO or 4 μ M 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 μ M 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 *8*2, 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 reanalyzed 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 S2I. 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 CTD² 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 < 3.67×10^{-9}); exclude hematopoetic cell lines (pKS < 2.34×10^{-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 < 1.63×10^{-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 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

Home	Insert Page Layout	Formulas Data Review	View						@ - =
Normal Page Page Layout Pro	e Break eview Views Screen	Ruler Formula Bar Gridlines Headings Message Bar	Zoom 100% Zoom to selection Window All	ge Freeze Panes • Unhide	View Side by Side Synchronous Scrolling Save Workspace	Switch Windows *			
Work	oook Views	Show/Hide	Zoom	Un <u>f</u> reeze P	anes	Macros			
A1		II_Line_Subset		Unlock all ro	ws and columns to scroll				
A	В	C D	E	Freeze Top	Row		J K	L	M N
Cell Line Subset	Cell_Line_Exclusion Feature	_Dataset Compound_Name	Enriched_Feature Nur	Keep the top	row visible while scrolling	Enrichment_Pvalue Chi_Squa	ared_Pvalue Square_Max_Pvalue	Log_Pvalue_Score_FDR	Qvalue Log_FDR_Qvalue
2 ALL_CCL_LINEAGE	S EXCLUDE_GENES_MU lineage	Merck60	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	through the	rest of the worksheet.	4.08E-14	5.51E-17 1.67E-27	-26.778	2.62E-24 -23.583
3 ALL_CCL_LINEAGE	S EXCLUDE_NONE lineage	Merck60	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	Ereeze Firs	t Column	2.72E-13	3.96E-15 7.40E-26	-25.131	1.26E-22 -21.9
4 ALL_CCL_LINEAGE:	S EXCLUDE_GENES_MU lineage	LY-2183240	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	Keen the first	column visible while scrollin Freeze Top	Row 1.11E-11	1.80E-11 3.22E-22	-21.492	5.06E-19 -18.296
5 ALL_CCL_UNEAGE:	S EXCLUDE_GENES_MU lineage	piperlongumine	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	through the	rest of the worksheet.	2.47E-11	3.75E-12 6.11E-22	-21.214	9.59E-19 -18.018
6 ALL_CCL_LINEAGE:	S EXCLUDE_NONE lineage	LY-2183240	HEMATOPOIETIC_AND_LYMPHOID_TISSUE			5.23E-12	1.81E-10 3.27E-20	-19.486	5.57E-17 -16.255
7 ALL_CCL_UNEAGE:	S EXCLUDE_GENES_MU lineage	tipifarnib-P2	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	209	23 sensitive	3.34E-10	1.85E-12 1.12E-19	-18.952	1.75E-16 -15.756
8 ALL_COL_LINEAGE	S EXCLUDE_NONE lineage	tipifarnib-P2	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	242	24 sensitive	8.37E-10	1.68E-10 7.00E-19	-18.155	1.19E-15 -14.924
9 ALL_CCL_LINEAGE:	S EXCLUDE_NONE lineage	piperlongumine	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	242	24 sensitive	6.33E-10	1.03E-09 1.07E-18	-17.97	1.82E-15 -14.739
10 ALL_CCL_LINEAGE	S EXCLUDE_GENES_MU lineage	SR-II-138A	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	120	19 sensitive	4.56E-13	1.61E-09 2.59E-18	-17.587	3.61E-15 -14.442
11 ALL_CCL_LINEAGE:	S EXCLUDE_NONE lineage	VER-155008	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	242	24 sensitive	1.80E-09	9.86E-12 3.26E-18	-17.487	5.55E-15 -14.256
12 ALL_COL_LINEAGE	S EXCLUDE_GENES_MU lineage	FQH1	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	100	18 sensitive	5.90E-10	2.76E-09 7.64E-18	-17.117	1.01E-14 -13.997
13 ALL_CCL_LINEAGE:	S EXCLUDE_GENES_MU lineage	VER-155008	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	209	23 sensitive	2.87E-09	1.32E-10 8.26E-18	-17.083	1.30E-14 -13.887
14 ALL_CCL_LINEAGE	S EXCLUDE_GENES_MU lineage	isoevodiamine	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	119	19 sensitive	1.76E-11	4.39E-09 1.93E-17	-16.714	2.69E-14 -13.57
15 ALL_CCL_LINEAGE	S EXCLUDE_NONE lineage	sirolimus	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	242	24 sensitive	5.64E-09	3.23E-12 3.18E-17	-16.497	5.42E-14 -13.266
16 ALL_CCL_LINEAGE:	5 EXCLUDE_GENES_MU sublinea	ige Merck60	ACUTE_MYELOID_LEUKAEMIA	209	7 sensitive	1.10E-08	1.97E-09 1.21E-16	-15.916	9.52E-14 -13.021
17 ALL_CCL_LINEAGE	S EXCLUDE_GENES_MU lineage	parbendazole	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	99	18 sensitive	1.11E-11	1.41E-08 1.98E-16	-15.704	2.59E-13 -12.586
18 ALL_CCL_LINEAGE:	5 EXCLUDE_NONE lineage	SR-II-138A	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	141	19 sensitive	5.95E-13	1.33E-08 1.77E-16	-15.752	2.92E-13 -12.535
19 ALL_CCL_LINEAGE	S EXCLUDE_GENES_MU lineage	sirolimus	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	209	23 sensitive	1.60E-08	1.04E-10 2.55E-16	-15.594	4.00E-13 -12.398
20 ALL_CCL_UNEAGE	5 EXCLUDE_NONE lineage	SB-225002	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	242	24 sensitive	2.67E-12	1.61E-08 2.60E-16	-15.585	4.43E-13 -12.354
21 ALL_COL_LINEAGE	S EXCLUDE_NONE sublinea	ige Merck60	ACUTE_MYELOID_LEUKAEMIA	242	7 sensitive	2.42E-08	4.66E-09 5.86E-16	-15.232	4.99E-13 -12.302
22 ALL_CCL_LINEAGE	5 EXCLUDE_GENES_MU lineage	SB-225002	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	209	23 sensitive	3.80E-12	2.92E-08 8.54E-16	-15.069	1.34E-12 -11.873
23 ALL_CCL_LINEAGE	S EXCLUDE_HEMATO sublinea	ige ABT-737	SMALL_CELL_CARCINOMA	218	21 sensitive	3.63E-08	4.52E-13 1.32E-15	-14.881	2.22E-12 -11.653
24 ALL_CCL_LINEAGE	5 EXCLUDE_GENES_MU lineage	JQ-1	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	83	14 sensitive	1.36E-08	5.37E-08 2.88E-15	-14.54	3.55E-12 -11.449
25 ALL_CCL_LINEAGE	S EXCLUDE_NONE lineage	BRD7137	HEMATOPOIETIC_AND_LYMPHOID_TISSUE	242	24 sensitive	2.30E-09	5.65E-08 3.19E-15	-14.496	5.43E-12 -11.265
26 LUNG	EXCLUDE MONE sublines	ide I Y.2383240	SMALL CELL CARCINOMA	91	21 centitive	3.03E-09	6 46E-08 4 17E-16	.14.38	5 61F-12 -11 261

• 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.

Home I	insert Page Layout	Formulas Data Revie	w View	-							0
From From From Access Web Text	n From Other Sources *	Refresh All - Connections	2↓ 2X X↓ Sort Filter Advanced Col	xt to Remove Dat Iumns Duplicates Validat	a Consolidate What-If Analysis	Group Ungroup	Subtotal				
Geri		Connections	30 00 100	Data	TOOIS		Jutime	J			
G13	▼ (* J× 23										
A	В	C D	E	F	G	Н		J	K L	M	N
Cell_Line_Subset	Cell_Line_Exclusion Teature	Dataset Compound_Name	Enriched_Feature	Number_of_Cell_Lines I	lumber_of_Mutant_Cell_Lines	Enrichment_Direction	Enrichment_Pvalue Chi_Squa	ed_Pvalue 👅 Square_Ma	n_PvalueLog_Pvalue_1	So(FDR_Qvalue	Log_FDR_Qval(~
2 ALL_CCL_LINEAGES	EXCLUDE_GENES_MU lineage	Merck60	HEMATOPOIETIC_AND_LYMPHOID_TISSI	UE 209	23	sensitive	4.08E-14	5.51E-17	1.67E-27 -2	26.778 2.62E-24	4 -23.583
3 ALL_CCL_LINEAGES	EXCLUDE_NONE Ineage	Merok60	HEMATOPOIETIC_AND_LYMPHOID_TISS	JE 242	26	sensitive	2.72E-13	3.962-15	7.40E-26	-26.131 1.26E-22	2 .21.9
4 ALL_CUL_LINEAGES	EXCLUDE_GENES_MU ineage	LY-2183240	HEMATOPOLETIC_AND_LYMPHOID_TISS	UE 209	23	sensitive	1.11E-11	1.80E-11	3.22E-22 -	21.492 5.06E-15	3 -18,236
6 ALL_COL_LINEAGES	EXCLUDE_GENES_WO Invage	piperiongumine	HEMATOPOLETIC_AND_LYMPHOID_TISS	UE 208	23	sensitive	2.47E-11	3.75E-12	6.IIE-22 ·	-21216 3.53E-12	3 -18.018
6 ALL_COL_LINEAGES	EXCLUDE_NUME Inteage	L1-2103240	HEMATOPOETIC_AND_LTMPHOID_TISS	UE 242	24	sensiove	5.232-12	1012-10	3.275-20 -	13.405 0.07E-1/	7 -16,200
ALL OCL LINEAGES	EXCLUDE NONE	dpramb+2	LIENATOPOIETIC_AND_LIMPHOID_TISS	203	23	sensiove	0.075 10	1005-12	7,007,10	10.002 LTOE 1	5 (D.705
ALL_COL_LINEAGES	EXCLUDE_NONE Intege	upramo-r-z	HEMATOPOIETIC_AND_LYMPHOID_TISS	UE 646	24	sensiove	0.37E-10	1005-10	1.075.49	17.07 1.03E-1	3 -14.324 E 14.729
10 ALL COL UNEAGES	EVELUDE GENES MULTINATION	CD # 129.0	HEMATOPOLETIC AND LYMPHOID TISS	UE 120	24	concitivo	6.33E-10	1012-09	2 695 10	17.67 2.615 1	E 14.442
11 ALL CCL UNEAGES	EVELUDE NONE ineage	VED.155009	MEMATOPOLETIC AND LYMPHOID TISS	UE 242	24	rentitive	1905-09	9.965.12	2.2002-10	17.497 5.555.1	5 .14.256
12 ALL CCL LINEAGES	EXCLUDE GENES MU Ineage	FOLI	HEMATOPOIETIC AND LYMPHOID TISS	UE 100	18	concitivo	5 905-10	2.765-09	7.645-18	17.107 0.00E-1	4 .13 997
12 ALL COL UNEAGES	EVCLIDE GENES MU ineage	VED. (55009	WEMATOPOIETIC AND LYMPHOID TISS	UE 209	22	rentitive	2.975.09	1225-10	9.265.19	17.092 1.20E.1	4 .12.997
14 ALL CCL LINEAGES	EXCLUDE GENES MU Ineage	iscendiamine	HEMATOPOIETIC AND LYMPHOID TISS	JE 119	19	concitive	1765-11	4 39E-09	1935-17	-16 714 2 69E-1	4 .13.57
15 ALL CCL LINEAGES	EXCLUDE NONE lineage	sioimus	HEMATOPOLETIC AND LYMPHOID TISS	IF 242	24	sensitive	5.64E-03	3235-12	3 18E-17	16.497 5.42E-1	4 -13.266
16 ALL COLLINEAGES	EXCLUDE GENES MIL sublines	ne Merck 60	ACUTE MYELOD LEUKAEMIA	209	7	concitivo	110E-08	197E-09	1215-16	-15 916 9.52E-1	4 .13.021
17 ALL COL LINEAGES	EXCLUDE GENES MU ineage	patheodazole	HEMATOPOLETIC AND LYMPHOID TISS	IF 99	18	sensitive	1115-11	141E-08	1985-46	15 704 2 59E-1	.12 586
18 ALL CCL LINEAGES	EXCLUDE NONE Ineage	SB-II-138A	HEMATOPOIETIC AND LYMPHOID TISS	UE 141	19	sensitive	5.95E-13	1.33E-08	1.77E-16	15.752 2.92E-1	3 -12 535
19 ALL CCL LINEAGES	EXCLUDE GENES MU lineage	sirolimus	HEMATOPOIETIC AND LYMPHOID TISSI	UE 209	23	sensitive	160E-08	104E-10	2.55E-16	15.594 4.00E-1	.3 -12.398
20 ALL CCL LINEAGES	EXCLUDE NONE lineage	SB-225002	HEMATOPOIETIC AND LYMPHOID TISS	JE 242	24	sensitive	2.67E-12	1.61E-08	2.60E-16	15.585 4.43E-1	3 -12.354
21 ALL CCL LINEAGES	EXCLUDE NONE sublinea	ge Merok60	ACUTE MYELOID LEUKAEMIA	242	7	sensitive	2.42E-08	4.66E-09	5.86E-16	15.232 4.99E-1	.3 -12.302
22 ALL CCL LINEAGES	EXCLUDE GENES MU lineage	SB-225002	HEMATOPOIETIC AND LYMPHOID TISSI	UE 209	23	sensitive	3.80E-12	2.92E-08	8.54E-16	15.069 1.34E-1	.2 -11.873
23 ALL CCL LINEAGES	EXCLUDE HEMATO sublinea	oe ABT-737	SMALL CELL CARCINOMA	218	21	sensitive	3.63E-08	4.52E-13	1.32E-15	-14.881 2.22E-1	.2 -11.653
24 ALL CCL LINEAGES	EXCLUDE GENES MU lineage	JQ-1	HEMATOPOIETIC AND LYMPHOID TISSI	UE 83	14	sensitive	1.36E-08	5.37E-08	2.88E-15	-14.54 3.55E-1	.2 -11.449
25 ALL CCL LINEAGES	EXCLUDE NONE lineage	BRD7137	HEMATOPOIETIC AND LYMPHOID TISSI	UE 242	24	sensitive	2.30E-09	5.65E-08	3.19E-15	14.496 5.43E-1	.2 -11.265
26 LUNG	EXCLUDE NONE sublinea	ge LY-2183240	SMALL CELL CARCINOMA	91	21	sensitive	9.03E-09	6.46E-08	4.17E-15	-14.38 5.61E-1	.2 -11.251
27 ALL COL LINEAGES	EXCLUDE GENES MILlineage	VK 4.279	HEMATOPOIETIC AND LYMPHOID TISS	IE 49	12	rensitive	14/6-09	128E-07	1635.44	13 789 1545.*	11 .10 914

• 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.

									-									
-	Home	Insert Page Layou	t Form	nulas Data	Review	View	Acrobat											🤍 – 🗖
38	A 325 A 3	EA XEA I		Conn	ections .		Clear				10	\$C3 \$C3	SEE 95 Show	w Detail				
6				C 3	2			+				먹을 🖊을						
5.00		The second se		Prope	rties	Cash	Reapply	Tauta	Data Data	Concellatory 14		Casua Managara	Hide	e Detail				
Acc	m From I	Text Sources 7 Conn	ections 1	All - See Edit Li	inks 🕺 🕺	2011	Advanced	Column	s Duplicates Validatio	consolidate w	nat-ti alveie v	Group Ungroup	Subtotal					
-		Tek Esternel Date	centons	Connection		100	at 0. Diltar	cordini	Data Ta	ala Ali	aiyaia		Notifier a					
		set External Data		Connection	s	50	it of Filter		Data it	iois			Jutime					
	G13	- () fx	23															
	A	В	С		D		E		F	G		н	1	J	K	L	м	N
1 0	Cell_Line_Subset	Cell_Line_Exclusion 💌 F	Feature_Data:	set 👕 Compound_Na	ime	 Enriched_F 	eature	- N	umber_of_Cell_Lines 👅 Nu	mber_of_Mutant_Cell_Li	nes 🕋 En	richment_Direction 🕋	Enrichment_Pvalue	Chi_Squared_Pvalue 🕋	Square_Max_Pvalue	Log_Pvalue_So(FDR_Qvalue -	Log_FDR_Qval
24	ALL_CCL_LINEA	GES EXCLUDE_GENES_1 👌	Sort A to	οZ		HEMATOR	POIETIC_AND_LYMPHOID	TISSUE	209		23 set	nsitive	4.08E-14	5.51E-17	1.67E-27	-26.778	2.62E-24	-23.583
3 /	LL_CCL_LINEA	GES EXCLUDE_NONE	1.7			HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	242		24 set	nsitive	2.72E-13	3.96E-15	7.40E-26	-25.131	1.26E-22	-21.5
4 A	LL_CCL_LINEA	BES EXCLUDE_GENES_	Sort Z to	D A		HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	209		23 set	nsitive	1.11E-11	1.80E-11	3.22E-22	-21.492	5.06E-19	-18.296
5 A	ILL_CCL_LINEA	GES EXCLUDE_GENES_f	Sort by	Color		HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	209		23 set	nsitive	2.47E-11	3.75E-12	6.11E-22	-21.214	9.59E-19	-18.015
6 A	ALL_COL_LINEA	BES EXCLUDE_NONE				HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	242		24 set	nsitive	5.23E-12	1.81E-10	3.27E-20	-19.486	5.57E-17	-16.255
7 A	ALL_CCL_LINEA	GES_EXCLUDE_GENES_f 🌾	Clear Fil	Iter From "Compou	ind_Name"	HEMATOR	POIETIC_AND_LYMPHOID	TISSUE	209		23 set	nsitive	3.34E-10	185E-12	1.12E-19	-18.952	1.75E-16	-15.756
8 /	LL_CCL_LINEA	BES EXCLUDE_NONE	Filter by			HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	242		24 set	nsitive	8.37E-10	1.68E-10	7.00E-19	-18.155	1.19E-15	-14.924
9 A	ALL_CCL_LINEA	GES EXCLUDE_NONE	T inter by	Color		HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	242		24 set	nsitive	6.33E-10	1.03E-09	1.07E-18	-17.97	182E-15	-14.735
10 A	ULL_CCL_LINEA	GES EXCLUDE_GENES_f	Text Filt	ers	•	HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	120		19 set	nsitive	4.56E-13	1.61E-09	2.59E-18	-17.587	3.61E-15	-14.442
11 A	LL_CCL_LINEA	BES EXCLUDE_NONE				HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	242		24 set	nsitive	180E-09	9.86E-12	3.26E-18	-17.487	5.55E-15	-14.256
12 A	ILL_CCL_LINEA	GES EXCLUDE_GENES_F		Select All)	-	HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	100		18 541	nsRive	5.90E-10	2.76E-09	7.64E-18	-17.117	1.01E-14	-13.997
13 8	ALL_COL_LINEA	SES EXCLUDE_GENES_F		6 beta bromoandro	osterone	HEMATUR	YUETIC_AND_LYMPHUD	TISSUE	209		2,9 set	nstrve	2.87E-09	132E-10	8.26E-18	-17.083	1.30E-14	-13.887
14 4	UL_COL_LINEA	GES EXCLUDE_GENES_F	- 21	LS, 3R-RSL-3		HEMATOR	POIETIC_AND_LYMPHOID	TISSUE	119		19 591	nskive	1.76E-11	4.39E-09	183E-17	-16./19	2.69E-14	-13.57
10 4	CL_CCL_LINEA	SES EXCLODE_NONE	2	2.4-dideoxy-DC-45-	A2	HEMATOP	"OETIC_AND_ETMPHOID	TISSUE	242		29 501	nskive	0.64E-03	3.23E-12	3.18E+17	-16.437	0.42E-14	-13.268
10 4	ALL_COL_LINEA	BES EXCLUDE_GENES_I	23	8-chloroalanine		ACOTE_M	TELOID_LEUKAEMIA	T10010	203		7 381	nsilve	1.002-00	1.572-03	1.210-10	-10.316	3.02E-14	-13.02
10 1	UL COL LINEA	SES EXCLUDE_GENES_I		Imethylfacudi		HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	39		18 501	nskive	LIE-II	1005.00	L38E-16	-15.704	2.09E-13	-12.08t
19 7	UL COL UNEA	SES EXCLUDE GENES A		ALCOCE3		HEMATOR	POIETIC_AND_LYMPHOID	TICCHE	209		22 64	nskive	1005.09	1045-10	2 555.10	-15.594	A 005-12	-12.295
20 0	UL COL UNEA	DES EVOLUDE NONE		07 202 3		HEMATOR	DOIETIC AND LYMPHOD	TICOLE	240		24	nakiye	2.675 12	1040-10	2.00E-10	10.004	4.405.10	12.054
21 0	LL CCL UNEA	SES EVOLUDE NONE		401-737		ACUTE M	VELOD LEUKAEMIA	110000	242		7 00	ncitiue	2.425.09	A 665 09	E OCE 10	16.000	4 995 12	12 205
22 6	LL CCL UNEA	SES EVELUDE GENES /	- MA	481-751		HEMATOR	POIETIC AND LYMPHOID	TISSUE	209		22 54	nzitiue	3.80E-12	2.92E.09	9.545.10	-15.069	1345.12	-11971
23 4	LL CCL UNEA	SES EXCLUDE HEMATC	- 🖌 A	AC55649		SMALL CR	FUL CARCINOMA		218		21 541	nsitiue	3.635.08	4.52E-00	1325-15	-14.881	2.22E-12	-11653
24 4	UL COL UNEA	BES EXCLUDE GENES I	🖌 A	AG14361		HEMATOR	POIETIC AND LYMPHOID	TISSUE	83		14 500	nsitive	1365-08	5.37E-08	2 88E-15	-16 56	355E-12	.11445
25 4	LL COL LINEA	BES EXCLUDE NONE	- 🖌 A	AGK-2		HEMATOR	POIETIC AND LYMPHOID	TISSUE	242		24 541	nsitive	2.30E-09	5.65E-08	3.19E-15	-14.496	543E-12	-11.265
26 L	UNG	EXCLUDE NONE	- 🗸 A	M-580		SMALL CE	ELL CARCINOMA		91		21 591	nsitive	9.03E-09	6.46E-08	4.17E-15	-14.38	5.61E-12	-11.25
27 8	LL COL LINEA	BES EXCLUDE GENES I	A	PO-866		HEMATOP	POIETIC AND LYMPHOID	TISSUE	48		13 set	nsitive	1.41E-09	1.28E-07	1.63E-14	-13,789	1.54E-11	-10.814
28 A	LL CCL LINEA	GES EXCLUDE GENES I	1.0.			ACUTE M	IYELOID LEUKAEMIA		209		7 set	nsitive	2.43E-08	1.43E-07	2.04E-14	-13.691	1.60E-11	-10.796
29 /	LL CCL LINEA	BES EXCLUDE GENES (4		•	HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	209		23 set	nsitive	1.32E-08	1.28E-07	1.64E-14	-13.785	2.58E-11	-10.585
30 A	LL_CCL_LINEA	GES EXCLUDE_NONE				HEMATOR	POIETIC_AND_LYMPHOID	TISSUE	97		14 set	nsitive	2.98E-08	1.81E-07	3.29E-14	-13.483	5.18E-11	-10.288
31 /	LL_CCL_LINEA	BES EXCLUDE NONE		OK	Cancel	ACUTE_M	IYELOID_LEUKAEMIA		242		7 501	nsitive	1.84E-07	1.37E-14	3.37E-14	-13.473	5.74E-11	-10.24
32 A	ALL_COL_LINEA	GES EXCLUDE_GENES_F				HEMATOP	POIETIC_AND_LYMPHOID	TISSUE	83		14 set	nsitive	2.18E-07	1.08E-07	4.75E-14	-13.323	5.85E-11	-10.233
33 A	LL_CCL_LINEA	GES EXCLUDE_GENES_MU :	ublineage	ABT-737		SMALL_CE	ELL_CARCINOMA		209		21 set	nsitive	1.94E-07	5.72E-10	3.76E-14	-13.424	5.91E-11	-10.225
34 L	UNG	EXCLUDE_GENES_MU s	ublineage	LY-2183240		SMALL_CE	ELL_CARCINOMA		86		21 set	nsitive	2.36E-08	2.18E-07	4.73E-14	-13.325	5.93E-11	-10.227
35 A	LL_CCL_LINEA	GES EXCLUDE_NONE s	ublineage	ABT-737		SMALL_CE	ELL_CARCINOMA		242		21 set	nsitive	1.89E-07	4.27E-10	3.57E-14	-13.447	6.08E-11	-10.216

• 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)

Home Insert Page Layout Formulas Data Review	View Acrobat						🕑 🗕 🖷 🗙
E Connections A	NZ K Clear		10 1	Show Detail			
L S Properties Z+	ZĂ V Reapply	in 📑 🗆 🛒	- 1 3 13 13	Hide Detail			
From From From Other Existing Refresh Z	Sort Filter Tex	t to Remove Data Consolid	ate What-If Group Ungroup	Subtotal			
Access Web Text Sources Connections All - 60 Edit Links	Advanced Colu	umns Duplicates Validation *	Analysis * * *				
Get External Data Connections	Sort & Filter	Data Tools	0	Dutline 😡			
	A		X				
G13 • (<i>f</i> x 23							*
A B C D	E	F G	н	- I J	К	L M	N
1 Cell Line_Subset Cell Line_Exclusion Feature_Dataset Compound_Name	Enriched_Feature	Number_of_Cell_Lines Number_of_Mutar	t_Cell_Lines T Enrichment_Direction	Enrichment_Pvalue Chi_Squared_P	ralue 🚬 Square_Max_Pvalue 🛀	Log_Pvalue_So(FDR_Qvalue	Log_FDR_Qval()
2 ALL_CCL_LINEAGES_EXCLUDE_GENES_ 2 Sort A to Z	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 209	23 sensitive	4.08E-14	5.51E-17 1.67E-27	-26.778 2.62E-2	-23.583
3 ALL_CCL_LINEAGES EXCLUDE_NONE Z Sort 7 to A	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 242	24 sensitive	2.72E-13	3.96E-15 7.40E-26	-25.131 126E-2	-21.9
ALL_CULLUNEAGES EXCLUDE_GENES_TAY SOLUTION	HEMATOPOIETIC_ANU_LYMPHOID_TISSO	E 209	23 sensitive	1.11E-11	180E-11 3.22E-22	-21.492 5.06E-1	-18.296
Sort by Color	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 209	23 sensitive	247E-II	3.75E-12 6.11E-22	-21214 3.53E-1	-18.018
ALL_COL_UNEAGES EXCLUDE_NUNE	HEMATOPOETIC_AND_LYMPHOID_TISSU	E 242	24 sensitive	2.245.10	1000 1000 1000 1000 1000	-13.405 0.07E-1	-16.200
ALL COLUMEAGES EXCLUDE MONE ALL COLUMEAGES EXCLUDE MONE	HEMATOPOLETIC_AND_LYMPHOLD_TISSU	E 205	24 concitiue	9.275.10	1605-10 7.005-19	10.352 1.752-1	10.706
ALL COLUMEAGES EVOLUEE NONE Filter by Color	HEMATOPOIETIC AND LYMPHOID TISSU	E 242	24 condition	6.072.10	1025.09 1075.19	17.97 1.925 1	14.729
10 ALL COLUMEAGES EVOLUDE GENES I	HEMATOPOLETIC AND LYMPHOID TISSU	E 676 E 120	19 capritine	4.505-10	1615.09 2.595.19	.17.597 2.615.1	.14.442
11 ALL COLUMEAGES EXCLUDE NONE	HEMATOPOIETIC AND LYMPHOID TISSU	E 242	24 sensitive	1805-09	9.88E-12 3.26E-18	.17 487 5 555-1	14 256
12 ALL CCI LINEAGES EXCLUDE GENES (HEMATOPOIETIC AND LYMPHOID TISSU	E 100	18 sensitive	5 30E-10	76F-09 764F-18	-17.117 101E-1	-13.997
13 ALL CCL LINEAGES EXCLUDE GENES 1	HEMATOPOIETIC AND LYMPHOID TISSU	E 209	23 sensitive	2.87E-09	1.32E-10 8.26E-18	-17.083 1.30E-1	-13.887
14 ALL CCL LINEAGES EXCLUDE GENES 1	HEMATOPOIETIC AND LYMPHOID TISSU	E 119	19 sensitive	1.76E-11	4.39E-09 1.93E-17	-16.714 2.69E-1	-13.57
15 ALL CCL LINEAGES EXCLUDE NONE	HEMATOPOIETIC AND LYMPHOID TISSU	E 242	24 sensitive	5.64E-09	3.23E-12 3.18E-17	-16.497 5.42E-1	-13,266
16 ALL CCL LINEAGES EXCLUDE GENES (MLN-4924	ACUTE MYELOID LEUKAEMIA	209	7 sensitive	1.10E-08	1.97E-09 1.21E-16	-15.916 9.52E-1	-13.021
17 ALL_CCL_LINEAGES_EXCLUDE_GENES_1 MST-312	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 99	18 sensitive	1.11E-11	1.41E-00 1.98E-16	-15.704 2.59E-1	-12.586
18 ALL_CCL_LINEAGES EXCLUDE_NONE N9-isopropylolomoucine	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 141	19 sensitive	5.95E-13	1.33E-08 1.77E-16	-15.752 2.92E-1	-12.535
19 ALL_CCL_LINEAGES_EXCLUDE_GENES_F	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 209	23 sensitive	1.60E-08	1.04E-10 2.55E-16	-15.594 4.00E-1	-12.398
20 ALL_CCL_LINEAGES EXCLUDE_NONE	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 242	24 sensitive	2.67E-12	1.61E-08 2.60E-16	-15.585 4.43E-1	-12.354
21 ALL_CCL_LINEAGES EXCLUDE_NONE	ACUTE_MYELOID_LEUKAEMIA	242	7 sensitive	2.42E-08	4.66E-09 5.86E-16	-15.232 4.99E-1	-12.302
22 ALL_CCL_LINEAGES_EXCLUDE_GENES_f	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 209	23 sensitive	3.80E-12	2.92E-08 8.54E-16	-15.069 1.34E-1	-11.873
23 ALL_CCL_LINEAGES EXCLUDE_HEMATC	SMALL_CELL_CARCINOMA	218	21 sensitive	3.63E-08	4.52E-13 1.32E-15	-14.881 2.22E-1	-11.653
24 ALL_CCL_LINEAGES EXCLUDE_GENES_f	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 83	14 sensitive	1.36E-08	5.37E-08 2.88E-15	-14.54 3.55E-1	-11.449
26 ALL_CCL_LINEAGES EXCLUDE_NONE	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 242	24 sensitive	2.30E-09	5.65E-08 3.19E-15	-14.496 5.43E-1	-11.265
26 LUNG EXCLUDE_NONE NSC 74859	SMALL_CELL_CARCINOMA	91	21 sensitive	9.03E-09	3.46E-08 4.17E-15	-14.38 5.61E-1	-11.251
27 ALL_CCL_LINEAGES EXCLUDE_GENES_fNSC23766	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 48	13 sensitive	1.41E-09	128E-07 163E-14	-13.789 1.54E-	1 -10.814
28 ALL_CCL_LINEAGES EXCLUDE_GENES_1	ACUTE_MYELOID_LEUKAEMIA	209	7 sensitive	2.43E-08	1.43E-07 2.04E-14	-13.691 1.60E-	1 -10.796
28 ALL_CCL_LINEAGES_EXCLUDE_GENES_1	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 209	23 sensitive	1.32E-08	1.28E-07 1.64E-14	-13.785 2.58E-	1 -10.589
30 ALL_CCL_LINEAGES EXCLUDE_NONE	HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 97	14 sensitive	2.98E-08	1.81E-07 3.29E-14	-13.483 5.18E-	1 -10.286
31 ALL_CCL_LINEAGES_EXCLUDE_NONE	ACUTE_MYELOID_LEUKAEMIA	242	7 sensitive	1.84E-07	1.37E-14 3.37E-14	-13.473 5.74E-	1 -10.241
32 ALL_CCL_LINEAGES_EXCLUDE_GENES_1	.: HEMATOPOIETIC_AND_LYMPHOID_TISSU	E 83	14 sensitive	2.18E-07	1.08E-07 4.75E-14	-13.323 5.85E-	1 -10.233
33 ALL_CUL_UNEAGES_EXCLUDE_GENES_MU_sublineage ABT-737	SMALL_CELL_CARCINOMA	209	21 sensitive	1346-07	5.72E-10 3.76E-14	-13.424 5.91E-	1 -10.228
34 LUNUi EXCLUDE_GENES_MU sublineage LY-2183240	SMALL_CELL_CANCINOMA	86	21 sensitive	2.36E-08	2.18E-07 4.73E-14	-13.325 5.93E-	1 -10.227

• 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
 - enriched_feature: BRAF
 - o feature_dataset: Onco

10000		-																						
	Home	Incert	Page	e lavout Er	rmulac	Data	Deview	View	Acrobat															
	rionic	mach	rug	c cayour i r					Acrobat															- w
25-0	A STELL	245-22	25-2			Conne	ctions			Class.		Free Free			100	¢re.	4 E 🖂		QE Show Datail					
G		Tim			2	Conne	cuons			W Clean	*				- P	(B)	2	955	- 2 Show Detail					
						Proper		Z A	U 7	& Reannly				ي 🗆					Hide Detail					
From	n From	From Fro	om Other	Existing	Refresh	1		Z Sort	Filter		Text to	Remove	Data	Consolida	te What-If	Group	Ungroup '	Subtotal	-					
Acce	ss Web	Text S	ources *	Connections	All 🔻	👓 Edit Lir	nks	A+		🐓 Advanced	Columns	Duplicate	s Validation	1*	Analysis *									
		Call Endance	al Data						Cost 9, Eilke				Data Ta				0	allin a						
		Get Extern	hai Data			onnections			Soft & Fifte	r i			Data Io	015				utiine						
			6	8																				
	L498431		- (0	f_{κ}																				
	L498431	A	• (•	f _x		c		D		E	F			G		н		1	Ŀ		к	L	м	N
1	L498431 Cell Line S	A ubset =	- Cell Line	fs: B Exclusion	- Feature	C e Dataset 🗷	Compound	D	Z Enriche	E od Feature 7	F Jumber of Ce	Lines 💌	Number of M	G stant Cell Line:	Enrichmer	H X Direction	Enrichmer	l It Pvalue	• Chi Squared Pv	alue 💌	K Square Max Pvalue 💌	L Log Pvalue Score ×	M FDB Qvalue ×	N Log FDR Qvalue
1 3743	L498431 Cel_Line_S ALL CCL	A ubset -		fx B Exclusion HEMATO	• Feature	C e_Dataset Z	Compound	D LName b	Enriche BRAF	E rd_Feature II I	F lumber_of_Ce	Lines • 1	Number_of_M	G .tant_Cell_Lines	Enrichmer	H X_Direction	• Enrichmer	 1_Pvalue [; 0.0004637	J Chi_Squared_Pv	alue 💌 8,29E-15	K Square_Mat_Pvalue 2.15E-07	L Log_Pvalue_Score - -6.6674	M FDR_Qvalue × 0.00017971	N Log_FDR_Qvalue × -3.7454
1 3743 4326	L498431	A ubset - JNEAGES JNEAGES	Cell Line EXCLUDE EXCLUDE	fx B Evolusion HEMATO FREQ SENS	 Feature Onco Onco 	C e_Dataset Z	Compound vemurafenit vemurafenit	D L Name b	Enriche BRAF BRAF	E rd_Feature 2	F lumber_of_Ce	Lines × 1 178 176	Number_of_M	G utant_Cell_Line:	Enrichmer 14 sensitive 14 sensitive	H <u>X_Direction</u>	Enrichmer	1 1 <u>. Pvalue</u> 0.0004637 0.0005181	J <u> Chi_Squared_Pv</u> 78 14	alue * 8.29E-15 3.84E-12	K Square_Mat_Pvalue 216E-07 2.68E-07	L Log_Pvalue_Score × -6.6674 -6.571	M FDR_Qvalue × 0.00017971 0.00022431	N Log_FDR_Qvalue * -3.7454 -3.6432
1 3743 4326 9768	L498431 CellLine_S ALL_CCL_ ALL_CCL_ ALL_CCL_	A ubset • INEAGES INEAGES INEAGES	Cell Line EXCLUDE EXCLUDE EXCLUDE	f _x B Exclusion HEMATO FREQ_SENS GENES_MUT_HI	Feature Onco Onco 3H Onco	C e_Dataset 🗷	Compound vemurafenit vemurafenit vemurafenit	D LName b b b	Enriche BRAF BRAF BRAF	E ed_Feature 2	F Number_of_Ce	Lines × 1 178 176 173	Number_of_M	G .tant_Cell_Lines	Enrichmer 14 sensitive 14 sensitive 12 sensitive	H X_Direction	• Enrichmer	1 t_Pvalue 0.0004637 0.0005181 0.0005964	J Chi_Squared_Pv 78 14	alue × 8.29E-15 3.84E-12 6.55E-14	K Square_Mat_Pvalue - 2.65E-07 2.63E-07 3.33E-07	L Log_Pvalue_Score - -6.6874 -6.571 -6.003	M FDR_Qvalue * 0.00017971 0.00022431 0.00074959	N Log_FDP_Qvalue * -3.7454 -3.6482 -3.1252
1 3743 4326 9768 11506	L498431 Cell Line S ALL_CCL ALL_CCL ALL_CCL ALL_CCL	A INEAGES INEAGES INEAGES INEAGES	Cell Line EXCLUDE EXCLUDE EXCLUDE EXCLUDE EXCLUDE	fx B HENOLUZION HEMATO FREQ_SENS GENES_MUT_HI NONE	 Feature Onco Onco Onco Onco 	C e_Dataset 🗷	Compound vemurafenit vemurafenit vemurafenit	D LName b b b b	2 Enriche BRAF BRAF BRAF BRAF	E ed_Feature 🗷 I	F Number_of_Ce	Lines × 1 178 176 173 202	Number_of_M	G .tant_Cell_Line:	Enrichmen Sensitive Sensitive Sensitive Sensitive Sensitive	H X_Direction	• Enrichmen	1 0.0004637 0.0005181 0.0009964 0.001046	J Chi_Squared_Pv 78 14 41 59	alue * 8.29E-15 3.84E-12 6.55E-14 7.39E-14	K Square_Mat_Pvalue _ 2.68E-07 3.33E-07 1.10E-06	L Log_Pvalue_Score • -6.6674 -6.003 -6.003 -5.9602	M FDR_Qvalue * 0.00017971 0.00022431 0.00074959 0.00092448	N Log_FDR_Qvalue * -3.7454 -3.6452 -3.1252 -3.0341
1 3743 4326 9768 11506 27107	L498431 Cell Line S ALL_CCL ALL_CCL ALL_CCL ALL_CCL ALL_CCL	A ubset • INEAGES INEAGES INEAGES INEAGES INEAGES	Cell Line EXCLUDE EXCLUDE EXCLUDE EXCLUDE EXCLUDE EXCLUDE	fx B Exolusion HEMATO FREQ_SENS GENES_MUT_HI NONE SUSPENSION	Feature Onco Onco Onco 3H Onco Onco Onco	C e_Dataset 🗷	Compound vemurafenil vemurafenil vemurafenil vemurafenil	D LName b b b b b b b	Z Enriche BRAF BRAF BRAF BRAF BRAF	E od_Feature 🗷 I	F Number_of_Ce	Lines / 178 178 176 173 202 168	Number_of_M	G stant_Cell_Line:	Enrickmer Sensitive Sensitive Sensitive Sensitive Sensitive Sensitive	H *_Direction	• Enrichmen	1 0.0004637 0.000588 0.000588 0.000588 0.000588	J Chi_Squared_Pu 78 14 15 13	alue * 8.29E-15 3.84E-12 6.55E-14 7.39E-14 2.76E-12	K Square_Max_Pvalue = 2.15E-07 2.58E-07 9.93E-07 1.10E-06 5.02E-06	L Log Pvalue_Score = -6.6571 -6.003 -6.5902 -5.239	M FDR_Qualue + 0.00072971 0.00022431 0.00092448 0.00092448	N -3,7454 -3,8492 -3,1252 -3,0241 -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 (<u>http://www.broadinstitute.org/ccle/home</u>) 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:
 - o compound_name: P-0850
 - enriched_feature: BRAF
 - cell_line_subset: ALL_CCL_LINEAGES
 - cell_line_exclusion: EXCLUDE_NONE

	Home	Insert	Page Layout Fo	rmulas Data	Review	View	Acrobat									0 - 🗉
From	From SWeb	From Fro Text So	om Other purces *	Refresh All + 00 Edit L	ections rties inks	↓ <u>2 Z</u> ↓ Sort	Filter	Text to Remov Columns Duplica	e Data Consolidate tes Validation *	What-If Analysis *	Ungroup Subtotal	Show Detail Hide Detail				
		Get Extern	al Data	Connection	\$	Sor	rt & Filter		Data Tools		Outline	5				
	6498400		fx fx													
	0450400															
	A	۹.	в	С		D	E	F	G	н	1	J	K	L	м	N
1	Cel_Line_Su), ibset 💽	B Cell_Line_Exclusion	C Feature_Dataset	Compound_N	D lame	E Enriched_Feature Z M	F Jumber_of_Cell_Lines	G Number_of_Mutant_Cell_Lines =	H Enrichment_Direction	Enrichment_Pvalue	J Chi_Squared_Pvalue	K Square_Max_Pvalue	L Log_Pvalue_Score *	M FDR_Qvalue 💌	N Log_FDR_Qvalue
1 8294	Cell_Line_Su	A Ibset ·	B Cell_Line_Exclusion EXCLUDE_NONE	C Feature_Dataset = OncoGeno	Compound_N vemurafenib	D Jame	E Enriched_Feature	F Jumber_of_Cell_Lines 202	G Number_of_Mutant_Cell_Lines -	H Enrichment_Direction	Enrichment_Pvalue 0.00053385	J Chi_Squared_Pvalue - 7.86E-0	K Square_Mat_Pvalue _ 7 3.53E-07	Log_Pvalue_Score - -6.4526	M FDR_Qvalue * 0.00059493	N Log_FDR_Qvalue - -3.2255
1 8294 11506	Cell Line Su ALL_CCL_UI ALL_CCL_UI	NEAGES	B Cell_Line_Exclusion EXCLUDE_NONE EXCLUDE_NONE	C Feature_Dataset _ OncoGeno Onco	Compound N vemurafenib vemurafenib	D Jame	E Enriched_Feature Z M BRAF BRAF	F Jumber_of_Cell_Lines _ 202 202	G Number_of_Mutant_Cell_Lines *	H Enrichment_Direction sensitive sensitive	Enrichment_Pvalue	J Chi_Squared_Pvalue 7.86E-0 7.39E-0	K Square_Max_Pvalue 7 3.53E-07 1.10E-06	Log_Pvalue_Score - -6.4526 -5.9602	M FDR_Qvalue * 0.00059493 0.00092448	N Log_FDR_Qvalue × -3.2255 -3.0341
1 8294 11506 20640	Cell_Line_Su ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI	NEAGES NEAGES NEAGES NEAGES	B Cell_Line_Exclusion EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE	C Feature_Dataset _ OncoGeno Onco TES-C	Compound_N vemurafenib vemurafenib vemurafenib	D lame	E BRAF BRAF BRAF BRAF	F Jumber_of_Cell_Lines 202 202 196	G Number_of_Mutant_Cell_Lines • 1 1 1 1	H Enrichment_Direction sensitive sensitive sensitive	Enrichment_Pvalue * 0.00059385 0.0010469 0.00067929	J Chi_Squared_Pvalue 7.86E-0 7.39E-0 4.12E-1	K Square_Mar_Pvalue * 7 3.53E-07 1 10E-06 1 4.61E-07	Log_Pvalue_Score - -6.4526 -5.9602 -6.3359	M FDR_Qvalue * 0.00059493 0.00092448 0.0019653	N Log_FDP_Qvalue * -3.2255 -3.0341 -2.7066
1 8294 11506 20640 239012	Cell_Line_Su ALL_CCL_LI ALL_CCL_LI ALL_CCL_LI ALL_CCL_LI	NEAGES NEAGES NEAGES NEAGES NEAGES	B Cell_Line_Exolution EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE	Feature_Dataset = OncoGeno Onco TES-C MUT	Compound_N vemurafenib vemurafenib vemurafenib vemurafenib	D lame	E BRAF BRAF BRAF BRAF BRAF BRAF	F Jumber_of_Cell_Lines 200 200 196 200	G Number_of_Mutant_Cell_Lines • N N 11 22	H Enrickment_Direction sensitive sensitive sensitive	Enrichment_Pvalue Enrichment_Pvalue	J Chi_Squared_Pvalue 7.86E-0 7.39E-1 4.12E-1 1.78E-01	K Square_Max_Pvalue - 7 3.53E-07 1.10E-06 1 4.61E-07 8 0.00037465	Log_Pvalue_Score - -6.4528 -6.9602 -6.3359 -3.4284	M FDR_Qvalue • 0.00059493 0.00092448 0.0019653 0.028729	N -3.2255 -3.0341 -2.7066 -1.5417
1 8294 11506 20640 239012 248645	ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI	NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES	B Cell Line Exolution EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE	C Festure_Dataset = OncoGeno Onco TES-C MUT TES	Compound_N vemurafenib vemurafenib vemurafenib vemurafenib	D Jame	E BRAF BRAF BRAF BRAF BRAF BRAF	F Jumber_of_Cell_Lines 200 200 196 200 196	G Number_of_Mutant_Cell_Lines • 1 1 1 2 2 2 2	H Enrichment_Direction sensitive sensitive sensitive sensitive sensitive	Enrichment Pvalue * 0.00059385 0.0010469 0.00057359 0.019356 0.019356	J Chi_Squared_Pvalue * 7.86E-0 7.39E-1 4.12E-1 1.78E-0 3.12E-0 3.12E-0	K Square_Mat_Pvalue * 7 3.53E-07 6 1.00E-06 1 4.61E-07 8 0.00037465 8 0.0003321	Log_Pvalue_Score - -6.4526 -5.9602 -6.3359 -3.4284 -3.4787	M FDR_Qvalue * 0.00059493 0.00092448 0.0019653 0.028729 0.030095	N Log_FDR_Qvalue + -3.2255 -3.0341 -2.7066 -1.5417 -1.5215
1 8294 11506 20640 239012 248645 255373	ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI	NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES	B Cell_Line_Exolusion EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE	C Feature_Dataset * OncoGeno Onco TES-C MUT TES TES-A	Compound_N vemurafenib vemurafenib vemurafenib vemurafenib vemurafenib	D Jame	Enriched_Feature Z N BRAF BRAF BRAF BRAF BRAF BRAF BRAF	F lumber_of_Cell_Lines 200 2002 196 2002 196 196 196	G Number_of_Mutant_Cell_Lines	H Enrickment Direction sensitive sensitive sensitive sensitive sensitive sensitive sensitive	 Enrichment_Pvalue 0.00053385 0.00067329 0.01356 0.01356 0.01356 0.01356 0.025804 	J Chi_Squared_Pvalue 7.88E-0 7.39E-0 4.02E-1 179E-0 3.12E-0 1.32E.0	Square Mat Pvalue - 7 3.53E-07 1.10E-06 1.10E-06 1 4.61E-07 3 0.00037465 0 0.00037455 0.00003211 3 0.000057267 3	L Log Pvalue Score - -6.4526 -6.3399 -3.4284 -3.4787 -3.1722	M FDR_Qvalue = 0.00059438 0.0019653 0.028729 0.030095 0.031136	N Log_FDR_Qvalue + -3.2255 -3.0341 -2.7066 -1.5215 -1.5205 -1.5067
1 8294 11506 20640 239012 248645 255373 268649	ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI	NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES	B Cell Line Erolusion EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE EXCLUDE_NONE	C Presture Dataset * OncoGeno Onco TES-C MUT TES TES-A TES-A TES-CNV	Compound_N vemurafenib vemurafenib vemurafenib vemurafenib vemurafenib vemurafenib	D lame	Eniched Feature 2 M BRAF BRAF BRAF BRAF BRAF BRAF BRAF BRAF	F Jumber_of_Cell_Lines 200 200 199 200 199 200 199 200 200 200 200 200 200 200 200 200 2	G Number of Mutant Cell Lines • 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	H Enrickment_Direction sensitive sensitive sensitive sensitive sensitive sensitive sensitive	Enrichment_Pvalue * 0.00059365 0.00059365 0.00067329 0.00067329 0.019356 0.019324 0.028945 0.028945	J Chi_Squared_Pvalue - 7.86E-0 7.39E-6 4.82E-1 179E-0 3.12E-0 1.32E-0 1.32E-0 1.32E-0 1.32E-0 1.32E-0 1.32E-0 1.32E-0	K Square Max Pvalue - 7 3.53E-07 1 10E-06 1 4.61E-07 8 0.00033218 8 0.00003257 8 0.000033382	L Log Pvalue Score • -6.4508 -6.359 -3.4264 -3.4787 -3.1722 -3.4047	M FDP_Qvalue * 0.00059493 0.0019553 0.028729 0.03095 0.03095 0.03136 0.036497	N -3.2255 -3.0341 -2.7066 -1.5417 -1.525 -1.595 -1.595 -1.597 -1.4377
1 8294 11506 20640 239012 249645 295373 288649 297738	CellLine_Su ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI ALL_CCL_UI	NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES NEAGES	B Cell Line Exclusion EXCLUDE, NONE EXCLUDE, NONE EXCLUDE, NONE EXCLUDE, NONE EXCLUDE, NONE EXCLUDE, NONE EXCLUDE, NONE EXCLUDE, NONE	C Pesture Dataset = OncoGeno Onco TES-C MUT TES TES-A TES-CNV TES-ACNV-H	Compound_N vemutafenib vemutafenib vemutafenib vemutafenib vemutafenib vemutafenib	D Jame	Enriched Feature Z h BRAF BRAF BRAF BRAF BRAF BRAF BRAF BRAF	F Jumber_of_Cell_Lines 200 2000 196 200 196 200 200 200 200 200 200 200 200 200 20	G Number_of_Matant_Cell_Lines * 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	H Enrichment Direction sensitive sensitive sensitive sensitive sensitive sensitive sensitive sensitive	 Ensistement Pvalue Ensistement Pvalue 0.0005385 0.00067323 0.0019465 0.001945 0.025394 0.02539	J Chi_Squared_Pvalue - 7.38E-0 7.38E-1 7.38E-1 7.39E-1 3.12E-0 1.32E-0 1.32E-0 8.11E-0 8.11E-0	K Square Max Pvalue 7 3.550-07 1 1000-06 1 4.610-07 8 0.00037465 0.00007257 8 0.00007257 9 0.00037852 9 0.00037852	L Log Pvalue Score - -6.4526 -6.3359 -3.4284 -3.47877 -3.47877 -3.47877 -3.47877 -3.47877 -3	M FDR_Qvalue = 0.00059493 0.00092448 0.0019653 0.028729 0.030095 0.031136 0.036497 0.038091	N Log_FDR_Qvalue + -3.2255 -3.0341 -2.7066 -1.5417 -1.5215 -1.5067 -1.4312 -1.4312

- 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 (<u>http://www.broadinstitute.org/ccle/home</u>) 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
 - o enriched_feature: CTNNB1
 - o feature_dataset: Onco

Home Insert Page Layout Fi	Data Review	view Acrobat					U - U
From From From Other Access Web Text Sources	Image: Connections ↓ Image: Properties ↓ Refresh ↓ All + ∞ Edit Links ↓	Sort	Text to Remove Data Consolidate What-If Columns Duplicates Validation ~ Analysis ~	Group Ungroup Subtotal	Detail etail		
Get External Data	Connections	Sort & Filter	Data Tools	Outline	5		
G498400 - Jx							
G498400 • Jx	С	D E	FG	н	J K	L M	N
G498400 G498400 G4	C Feature_Dataset Z Compound_Na	D E ame Z Enriched_Feature Z Nu	F G mber_of_Cell Lines × Enrichme	H I I I I I I I I I I I I I I I I I I I	J K red_Pvalue × Square_Mat_Pvalue × L	.og_Pvalue_Score 💌 FDR_Qvalue 🗵	N Log_FDR_Qualue
G498400 J× B 1 Cel_Line_Subset Cel_Line_Exclusion 174 ALL_CCL_UNEAGES EXCLUDE_SUSPENSION	C Feature_Dataset 2 Compound_Na Onco navitoslas	D E ame Z Enriched_Feature Z Nu CTNNB1	F G mber_of_Cel_Lines Vumber_of_Mutant_Cell_Lines 207 10 sensitive	H I ent_Direction × Enrichment_Pvalue × Chi_Squar e 0.00017201	J K red_Pvalue_x Square_Mar_Pvalue_x L 6.64E-06 2.96E-08	.og_Pvalue_Score × FDP_Qvalue × -7.5289 2.50E-05	N Log_FDR_Qvalue × -4.6026
G498400 A B Cel Line Subset Cel Line Exclusion IT/4 ALL_CCL_INEARES EXCLUDE SUSPENSION IS00 ALL_CCL_INEARES EXCLUDE HEMATO	C Feature_Dataset @ Compound_Na Onco navitoclas Onco navitoclas	D E ame Z Enriched Feature Z Nu CTNNB1 CTNNB1	F mber of Cel Lines V Number of Mutant Cel Lines Enrichme 207 10 sensitive 217 11 sensitive	H I Value Chi Squar e 0.00017201 e 0.0002534	J K red_Pvalue × Square_Mat_Pvalue × L 6.64E-06 2.36E-08 1.32E-05 4.688E-08	L M .og_Pvalue_Score × FDP, Qvalue × -7.5289 2.50E-05 -7.3297 3.95E-05	N Log_FDR_Qvalue * -4.6026 -4.4036
G498400 A G498400 Cel Line Subset Cel Line	C Compound_Na Droco navitoclas Droco navitoclas Droco navitoclas	D E ame Z Enriched_Feature Nu CTNNB1 CTNNB1	F mber_of_Cell_Lines > Number_of_Mutant_Cell_Lines > Enrichme 207 ID sensitive 217 III sensitive 241 III sensitive	H I Pusice Chi Squar ent_Direction * Enrichment Pusice * Chi Squar e 0.00027201 e 0.0005753	J K red_Pvalue Square_Mar_Pvalue L 6.64E-06 2.96E-00 1.02E-05 4.68E-06 4.25E-05 2.66E-07	L M -0g_Pvalue_Score × FDR_Qvalue × -7.5288 2.50E-05 -7.3297 3.95E-05 -6.5749 0.0002285	N Log_FDR_Qvalue × -4.6026 -4.4036 -3.6449
G498400 A B Cel Line Subset Cel Line Exclusion T177 ALL_CCL_UNEAGES EXCLUDE_SUSPENSION T1910 ALL_CCL_UNEAGES EXCLUDE_VENATO A455 ALL_CCL_UNEAGES EXCLUDE_VENATO T4555 ALL_CCL_UNEAGES EXCLUDE_VENATO EXCLUDE_VENATO T4555 ALL_CCL_UNEAGES T45555 ALL_CCL_UNEAGES T45555 ALL_CCL_UNEAGES T45555 ALL_CCLUEAGES T455555 ALL_CCL_UNEAGES T455555555 T45555555555555555555555555	C Compound_Na Deco navitoclar Onco navitoclar Onco navitoclar Onco navitoclar	D E ame Z Enriched_Feature X Nu CTNNB1 CTNNB1 CTNNB1 CTNNB1	F G Dutter of Autor Cell Lines E Enrichme 207 10 sensitive 217 11 sensitive 241 11 sensitive 209 9 5 sensitive	H I Christian Ch	J K red_Pvalue Square_Mar_Pvalue L 6.64E-06 2.96E-06 1.32E-05 4.68E-08 4.25E-05 2.66E-07 0.00017515 1.53E-06	L M -05 Pvalue Score V FDF Qvalue V -7,5289 2.50E-05 -7,3297 3.95E-05 -6,5749 0.0002265 -5,8485 0.00006003	N -4.6026 -4.4026 -3.8449 -3.0655
G498400 ✓ Jr G498400 ✓ CH Line Extension 174 ALL CC, LINEAGES EXCLUDE, SIPEPNION 1950 ALL CCL, LINEAGES EXCLUDE, SPERSION 1955 ALL CCL, LINEAGES EXCLUDE, SPERSIONE 1958 ALL CCL, LINEAGES EXCLUDE, SPERSIONE 1959 ALL CCL, LINEAGES EXCLUDE, SPERSIONE 1950 DE NDOWNERTIAM 1950 DE NDOWNERTIAM	C C C C C C C C C C C C C C C C C C C	D E Eature Nu CTNNB1 CTNNB1 CTNNB1 CTNNB1 CTNNB1 CTNNB1	F mer of Cel Lines S 207 H 201 H 201 H 201 H 203 S 201 H 209 S sensible 0 S sensible	H I Direction * Enrichment Prudue * CPI Squar • 0.00017201 • 0.000569 • 0.0005732 • 0.0005732	J value Square Mar Pvalue L 6.64E-06 2.96E-06 1.12E-05 4.68E-07 4.25E-05 2.68E-07 0.00017515 1.53E-06 0.039424 0.0014772	L M -7,5289 2,50E 05 -7,5289 2,50E 05 -7,5287 3,39E 05 -6,65749 0,0002265 -5,8145 0,0002685 -2,8396 0,0095056	N -4.6026 -4.4036 -3.6449 -3.6555 -1.025

- 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 (<u>http://www.broadinstitute.org/ccle/home</u>) 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
 - o feature_dataset: Onco

File Home Inset Page Layout Formulas Data Review View 🛆 🚱 🗗 🔂												
	Connections	A Clear	y 📑 📑	诗 😰 阳	Show Detail							
From From From Other Existing Access Web Text Sources T Connections	All - Se Edit Links	Sort Filter Advance	ted Columns Duplicates Validation *	Consolidate What-If Grou Analysis • •	p Ungroup Subtotal							
Get External Data	Connections	Sort & Filter	Data Tools		Outline 5							
C497392 - 🖍							^					
							4					
A B	C D	E	F G	н	I J	K L	M N 🔺					
1 cell_line_subset ~ cell_line_exclusion ~	feature_dataset 4 compound_nam	enriched_featur I number	r_of_cell_lines 🗠 number_of_mutant_cell	lines 🔄 enrichment_direction	enrichment_p_valu = chi_squared_p_valu	squared_max_p_val = log_p_value_scor	FDR_q_valu = log_q_value_sco =					
409981 LUNG EXCLUDE_FREQ_SENS	Onco neratinib	EGFR	42	4 sensitive	6.25E-05 0.04498	37 0.0020238 -2.69	938 0.06874 -1.1628					
452242 LUNG EXCLUDE_SUSPENSION	Onco neratinib	EGFR	35	4 sensitive	1.91E-05 0.0205	7 0.00042312 -3.37	736 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 (<u>http://www.broadinstitute.org/ccle/home</u>) and access the appropriate files.









5-20

number of pH2AX foci

>20









