Chemistry-First Approach for Nomination of Personalized Treatment in Lung Cancer

Graphical Abstract

Chemistry-first target nomination

200,000 chemicals

100 deeply annotated NSCLC lines

Automated discovery analytics

Translational correlates

200,000 chemicals

Therapeutic triads

NR3C1

GLUT8

SW036310

NR3C1

Glucocorticoids

NOTCH2 mutation

SW157765

KRAS/KEAP1 co-mutation

Cilia

TTC21B mutation

https://pops.biohpc.swmed.edu/pops

Authors
Elizabeth A. McMillan, Myung-Jeom Ryu, Caroline H. Diep, ..., John D. Minna, Hyun Seok Kim, Michael A. White

Correspondence
john.minna@utsouthwestern.edu (J.D.M.), hsfkim@yuhs.ac (H.S.K.), michael.white@utsouthwestern.edu (M.A.W.)

In Brief

Highlights
- A chemistry-first approach for druggable target identification for lung cancer
- Mapping the associations between chemicals and genetic lesions in lung cancer
- Matching chemicals with diverse patient-specific cancer-promoting mechanisms
- Validating the effect of targeting chemically addressable mechanisms in NSCLC cells

https://doi.org/10.1016/j.cell.2018.03.028
Chemistry-First Approach for Nomination of Personalized Treatment in Lung Cancer

Elizabeth A. McMillan,1 Myung-Jeom Ryu,2,14 Caroline H. Diep,1,14 Saurabh Mendiratta,1 Jean R. Clemenceau,1 Rachel M. Vaden,1 Ju-Hwa Kim,2 Takashi Motoyaji,3 Kyle R. Covington,4 Michael Peyton,8 Kenneth Huffman,5 Xiaofeng Wu,1 Luc Girard,4 Yeojin Sung,2 Pei-Hsuan Chen,6 Prema L. Mallipeddi,2 Joo Young Lee,2 Jordan Hanson,7 Sukesh Voruganti,2 Yunku Yu,8 Sunho Park,6 Jessica Sudderth,6 Christopher DeSevo,1 Donna M. Muzny,4 HarshaVardhan Doddapaneni,4 Adi Gazdar,12 Richard A. Gibbs,4 Tae-Hyun Hwang,8 John V. Heymach,9 Ignacio Wistuba,10 Kevin R. Coombes,11 Noelle S. Williams,7 David A. Wheeler,4 John B. MacMillan,7 Ralph J. Deberardinis,6 Michael G. Roth,7 Bruce A. Posner,7 John D. Minna,5,* Hyun Seok Kim,2,* and Michael A. White1,13,*

1Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
2Severance Biomedical Science Institute, Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, Korea
3Biomolecular Research Laboratories, Pharmaceutical Research Division, Takeda Pharmaceutical Company, Ltd., Fujisawa, Kanagawa, Japan
4Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
5Hamon Center for Therapeutic Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
6Children’s Research Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
7Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
8Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
9Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
10Translational Molecular Pathology, MD Anderson Cancer Center, Houston, TX 77030, USA
11Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA
12Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
13Lead Contact
14These authors contributed equally
*Correspondence: john.minna@utsouthwestern.edu (J.D.M.), hsfkim@yuhs.ac (H.S.K.), michael.white@utsouthwestern.edu (M.A.W.)
https://doi.org/10.1016/j.cell.2018.03.028

SUMMARY

Diversity in the genetic lesions that cause cancer is extreme. In consequence, a pressing challenge is the development of drugs that target patient-specific disease mechanisms. To address this challenge, we employed a chemistry-first discovery paradigm for de novo identification of druggable targets linked to robust patient selection hypotheses. In particular, a 200,000 compound diversity-oriented chemical library was profiled across a heavily annotated test-bed of >100 cellular models representative of the diverse and characteristic somatic lesions for lung cancer. This approach led to the delineation of 171 chemical-genetic associations, shedding light on the targetability of mechanistic vulnerabilities corresponding to a range of oncogenotypes present in patient populations lacking effective therapy. Chemically addressable additions to ciliogenesis in TTC21B mutants and GLUT8-dependent serine biosynthesis in KRAS/KEAP1 double mutants are prominent examples. These observations indicate a wealth of actionable opportunities within the complex molecular etiology of cancer.

INTRODUCTION

The future of cancer treatment lies in a personalization of medicine, where each patient’s treatment regime is tailored to the genetic diversity of their tumors. Accomplishing this requires a “therapeutic triad,” where appropriate context-specific intervention targets, tightly linked to response biomarkers, are coupled to agents to engage these targets. To date, this has been best realized in disease harboring a druggable oncogenic driver. However, many of the more prevalent and most lethal cancers do not present with this opportunity. Non-small cell lung cancer (NSCLC) is a leading cause of cancer-related death in the United States and is a highly heterogeneous disease. A clinically relevant contributor to this disease heterogeneity is the diversity of molecular etiology associated with individual NSCLC tumors. Specifically, lung squamous carcinoma (LUSC) and lung adenocarcinoma (LUAD) represent the second and the third most highly mutated tissue subtypes in The Cancer Genome Atlas (TCGA), with a mean non-synonymous mutation burden of ~250 mutations/tumor. This greatly increases the challenge of understanding the molecular drivers underpinning a patient’s disease, knowledge that is usually the starting point for hypothesis driven design of new therapeutic approaches. However, a large mutational burden also increases the probability that NSCLCs will contain vulnerabilities, not found in normal cells, which might be exploited therapeutically. The problem is how to identify and engage such vulnerabilities. Here, we employ...
a chemistry-driven de novo discovery strategy tailored for coincident delivery of preclinical therapeutic triads.

RESULTS

A Tiered Screening and Analytic Strategy for Chemistry-First Target Discovery

To generate an experimental testbed reflective of the molecular and mechanistic heterogeneity of lung cancer, we assembled a panel of 96 NSCLC (mainly LUAD) cell lines and 4 immortalized human bronchial epithelial cell lines (HBECs) (Data S1). Reasonable concordance of the phenotypic variation of this panel with human tumors was evaluated using legacy whole-genome transcript array data (Figures 1A and S1A). High-resolution molecular characterization was then carried out by whole-exome sequencing (WES) (Data S2), RNA sequencing (RNA-seq) (Data S3), tiled SNP arrays (Data S4), reverse phase protein array (RPPA) (Data S5), and heavy carbon tracing from glucose and glutamine into selected metabolites. For 34 NCLC cell lines, matching B cell lines from the same patients were sequenced, allowing for robust discrimination of somatic versus germline variation (Data S2). For the remaining cell lines, we developed a computational pipeline leveraging somatic alleles detected in the matched pairs and public datasets to filter probable germline variation (Figures 1B, S1B, and S1C; STAR Methods). We noted high concordance between transcript profiles from RNA-seq and hybridization arrays that were performed years apart, providing confidence in the accuracy and stability of cell line provenance (Figure S1D).

A deterministic clustering method, affinity propagation clustering (APC), (Frey and Dueck, 2007) produced more than 15 distinct phenotypic groups as defined by gene expression profiles (Figure S1E). Based on this, we devised a tiered high-throughput screening strategy to screen 202,103 chemicals across 12 cell lines representing overall phenotypic diversity of the panel (Figure S1F). Filters were implemented at each tier to enrich for small molecules that could selectively target the phenotypic variation across the NSCLC cell line panel (Figures S1F–S1I; STAR Methods). This culminated in 208 compounds and an additional 14 chemicals with known mechanism of action, which were tested for efficacy across the complete panel of 100 NSCLC cell lines at 12 doses in triplicate in two independent runs. This set was evaluated for potency and selectivity using both area under the dose response curve (AUC) and the effective inflection point while AUC values reflect magnitude of response. As both metrics provide complementary information, we employed AUC and ED50 values in all subsequent association analyses (Data S6). Furthermore, we examined potential correlations of chemical sensitivities (AUC and ED50 values) with expression of 1 of 9 known drug metabolism enzymes (LC/MS)-based approaches (Figures 2B–2E and S2E–S2J). Six compounds displayed accelerated metabolism, evident by loss of parent compound, selectively in the sensitive cell lines (Figures 2B–2E and S2E–S2J), 6 compounds displayed accelerated metabolism, evident by loss of parent compound, selectively in the sensitive cell lines (Figures 2B–2E, S2E, and S2F). We further examined SW157765 as a member of a cluster of compounds associated with activity in cells with high expression of the cytochrome p450 family collection of compounds, based on activity across the cell line panel, produced at least 38 distinct clusters (Figure S1K). A ranking of this collection, which we refer to as a “precision oncology probe set” (POPS), by potency and activity revealed robust selectivity profiles (median fold changes from 2–90,000, Figures 1C and S1I).

To begin to pursue chemical/genetic associations, we first parsed the cell line panel based on similarity of chemical response (Figure 1D) and each quantitative molecular feature set (Figures S1M–S1P). We overlaid annotations from feature-specific clusters onto those derived from chemical sensitivity, observing unimpressive correspondence with any feature set (Figures 1E, 1F, and S1Q–S1T). This suggests global molecular diversity cannot account for the observed selective chemical responses. This observation, together with similar previous observations from our group and others, led us to pursue sparse feature selection for finding robust chemical/genetic associations (Eskiocak et al., 2017; Garnett et al., 2012; Iorio et al., 2016; Kim et al., 2013). For this, we employed a combination of regularized machine learning (elastic net) and probability-based metrics (scanning Kolmogorov-Smirnov [KS]) to isolate features from each molecular dataset predicting sensitivity to each chemical. As compelling proof of concept, these methods linked high ALK expression as predicting sensitivity to the ALK inhibitor crizotinib (Figure 1G) and EGFR mutations and amplifications to predict sensitivity to the EGFR inhibitor erlotinib (Figure 1H). As expected, the EGFR mutant, erlotinib-sensitive cell lines have mutations in the kinase domain known to affect EGFR function. Notably, among the EGFR mutant non-responders, 2 cell lines harbored preexisting T790M mutations (Figure S1U), a known cell-autonomous adaptive mechanism promoting inhibitor resistance not detected by related chemical profiling efforts (Figure S1V). Together, these observations indicate that clinically relevant associations are discoverable within this experimental schema.

High-Throughput Discovery of Pharmacological Liabilities among Chemicals with Robust Selective Sensitivity Profiles

From the elastic net feature discovery approach, we noted a cohort of chemicals (12/221) for which sharp and selective sensitivity profiles (median deltas of >60-fold) were associated with expression of 1 of 9 known drug metabolism enzymes (Figures 2A and S2A). The compounds are structurally diverse and target distinct groups of cell lines (Figures S2B–S2D), prompting the attractive possibility that pharmacological liabilities could be detected and flagged early in chemistry-first target discovery screening cascades (i.e., for these chemicals, selective sensitivity may be due to selective production of a toxic metabolite). To test this, we first assessed chemical stability of each compound in groups of sensitive and resistant cell lines using liquid chromatography-mass spectrometry (LC/MS)-based approaches (Figures 2B–2E and S2E–S2I), 6 compounds displayed accelerated metabolism, evident by loss of parent compound, selectively in the sensitive cell lines (Figures 2B–2E, S2E, and S2F). We further examined SW157765 as a member of a cluster of compounds associated with activity in cells with high expression of the cytochrome p450 family.
member, CYP4F11. Notably, the CYP4F family inhibitor, HET0016, reversed metabolism of the compound (Figure 2F), and CRISPR-mediated knockout of CYP4F11 reversed toxicity in otherwise sensitive cell lines (Figures 2G and S2K–S2M). We next predicted the activity of SW001286 and SW126788, within a panel of 26 previously untested NSCLC cells, using the weighted elastic net models. Expression of the carboxylesterases CES1 and CES1P1 was cleanly predictive of SW126788 sensitivity in this test set (Figure S2N). However, prediction accuracy of SW001286 sensitivities was lower due to 4 unanticipated non-responders (Figure S2O). Like SW157756, HET0016 rescued SW001286 toxicity in two sensitive cell lines (Figure 2H). Thus, the metabolic products of SW001286 in CYP4F11-expressing cells may not be behaving as a general toxin(s), but rather may be targeting a selective vulnerability in sensitive cell lines. In line with this possibility, a manual examination of the KS test output indicated mutations in LKB1 as an additional marker that, together with CYP4F11 expression, would indicate potential sensitivity to SW001286.

Figure 1. Genomic Characterization and Chemical Sensitivities of NSCLC Cell Line Panel

(A) p values (Pearson) comparing tumors (MDACC, orange; TCGA, purple) and cell lines colored by source.
(B) Number of mutations called in the matched (red) and unmatched (blue) subsets post-filtering.
(C) NSCLC sensitivity (ED50) to POPS rank ordered by row. Red dashes, “cherry-picked.”
(D) APC clustering by similarity of POPS ED50 responses. Nodes are colored according to cluster membership.
(E) APC clustering by similarity of POPS ED50 responses (as in D). Nodes are colored according to cluster membership defined by RNA-seq-based APC (Figure S1E).
(F) APC clustering across all datasets. Cell lines are ordered according to cluster membership in chemical APC. Each cell line is colored according to cluster membership in the indicated datasets (Figures S1M–S1P). Cell lines absent from a dataset are colored in white.
(G and H) Predictive mRNA expression signatures specifying (G) crizotinib and (H) erlotinib sensitivity. Rank-ordered sensitivity values are indicated as heatmap (top row) with corresponding features plotted below.

All experiments performed in triplicate, unless otherwise indicated. Values are means. Error bars plotted as ± 1 SD. *p < 0.05; **p < 0.01. See also Data S2, S3, S4, and S5.
In response to metabolic stress, LKB1 activates AMPK to suppress anabolic and activate catabolic pathways to maintain energy and redox homeostasis via inhibition of ACC1 (Jeon et al., 2012). A mechanistic connection between SW001286 sensitivity and loss-of-function LKB1 mutations was supported by the observation that depletion of ACC1 (Figure 2J), or addition of a ROS scavenger (Figure S2Q), partially rescued SW001286-sensitivity.

Chemicals correlating with expression of CYP4F11 represent the largest observed “prodrug” class (Figure 2A). The P450 class of enzymes, of which CYP4F11 is a member, can oxidize a variety of substrates, the scope of which has not been fully
characterized but some chemical transformations tend to recur thematically within the class. One such transformation, the demethylation of aryl methoxy groups, is predicted to occur with two of the proposed pro-drug chemicals (Figure S2R). The remaining 3 compounds share a common furan-substituted alkene functional group that can likely be engaged by xenometabolitic enzymes. Furthermore, the internal alkene is a reasonable site for enzymatic oxidation and protein conjugation. Consistent with the possibility that protein/small molecule conjugation might be at play, 72 hr ED_{50}s were similar following either transient or sustained exposure to SW157765 (Figure S2S). In an effort to identify structural components of the molecule required for biological activity, we designed and synthesized a series of analogs of SW157765, of which analog 500-01 (Figure S2T) was found to be completely inert when tested for viability in a NSCLC cell line panel (Figure S2U). Interestingly, 500-01 differs from the parent molecule in only the hydrogenation of its internal alkene, suggesting SW157765 requires this functional group for biological activity. Considering that alkene and other sites of unsaturation can be transformed into points of conjugation with larger biomolecules, the discovery of SW157765a active moiety converged with our hypothesis that the molecule can behave as a covalent modifier. We next sought to identify labile metabolites of SW157765 that might be susceptible to protein conjugation, with specific attention given to metabolite species whose modifications appeared at the double bond. Mass spectrometry-based evaluation of metabolites produced by H2122 cells treated with either SW157765 or 500-1 for 8 hr identified an oxidized metabolite, unique to SW157765 (Figures S2V–S2Y), potentially containing an epoxide at the site of the internal alkene (Figure S2Z). Considering that strained epoxide ring systems are subject to facile protein adduction, we propose this epoxide metabolite is the active covalent ligand. Efforts to chemically synthesize the metabolite revealed it to be too unstable to produce in quantities needed for in vitro testing, perhaps underpinning the semi-transient nature and robust reactivity profile of the molecule in a biological setting.

As a final example of high-throughput detection of pharmacological liabilities, we noted that high expression of the multi-drug resistance transporter, ABCG2, predicts resistance to the CDK7 inhibitor THZ1 (Figure 2K). RNAi-mediated depletion of ABCG2 sensitized resistant cells to THZ1, suggesting it is an ABCG2 substrate (Figure 2L).

**NOTCH2 Mutations Predict Cellular Sensitivity to Glucocorticoids**

From within the Prestwick library, we noted a cluster of 5 glucocorticoid (GC) receptor agonists with highly correlated selective activity profiles and a strong association with mutations in NOTCH2 (Figures 3A and 3B). NOTCH2 has been implicated as a tumor suppressor in some settings. For example, NOTCH2 and NOTCH1 expression are oppositely correlated with prognosis in colorectal cancer, where low expression of NOTCH2 and high expression of NOTCH1 is predictive of poor patient outcome (Chu et al., 2011). Additionally, in experimental models of lung cancer, NOTCH2 loss, but not NOTCH1 loss, promotes aggressive disease (Baumgart et al., 2015). The “dispersed” pattern of NOTCH2 alleles detected among the NSCLC cell lines is reminiscent of loss-of-function alterations typically associated with a tumor suppressor (Figure S3A), and the glucocorticoid-sensitive cell lines harboring these mutations display downregulation of Notch pathway genes as compared to wild-type counterparts. (Figure 3C).

Depletion of the ubiquitously expressed GC receptor, NR3C1, was sufficient to reduce cellular sensitivity to GC exposure, suggesting the selective toxicity phenotype is receptor-dependent (Figure 3D). Intriguingly, several studies have linked GC response to Notch pathway activity. For example, activation of Notch signaling is associated with GC resistance in T-ALL (Inaba and Pui, 2010) and gamma-secretase inhibitors, which block the activation of Notch, restore sensitivity to GC. Moreover, a mutually antagonistic relationship exists between Notch effector, HES1, and NR3C1, in which each represses transcription of the other (Real et al., 2009; Revollo et al., 2013). Consistent with these observations, we found significantly higher basal expression levels of NR3C1 mRNA in NOTCH2 mutant, GC-responsive cell lines (Figure 3E). Transcription of NR3C1 itself is responsive to GC induction, and we observed significant induction of NR3C1 in response to GC stimulation in these cells (Figure 3F). These observations indicate that GC-responsive cells are primed to propagate an NR3C1 signal through a GC-dependent positive feedback amplification loop.

We found the selective efficacy of GCs was preserved in 3D spheroid models of lung cancer (Figure 3G). Thus, we sought to understand the mechanism by which differential activity of Notch signaling may specify sensitivity to GCs. HES1 is a general transcriptional repressor that has been described to occupy the promoters of GC-inducible genes and acts as a master negative regulator of GC response (Revollo et al., 2013). Similarly, we found that GC exposure selectively reduced cellular HES1 protein levels in GC-sensitive NSCLC cells (Figures 3H and S3C). GC exposure was cytostatic, resulting in a selective G1/S arrest (Figure 3I). GCs suppress inflammation through transcriptional activation of anti-inflammatory genes and direct inhibition of nuclear factor κB (NF-κB) and activator protein 1 (AP-1). A well-known target of both pathways is cyclin D1, which was selectively reduced in sensitive NSCLC cells exposed to GC (Figure 3I). Finally, stable overexpression of HES1 from a GC-independent CMV promoter was sufficient to rescue GC-induced cell-cycle arrest (Figures 3J, S3E, and S3F). We therefore suspect that NOTCH2 mutations, in NSCLC cells, result in reduced Notch signaling and higher basal NR3C1 expression, priming cells to respond to GC with G1 cell-cycle arrest (Figure S3G). While GC therapy is not commonly used in therapeutic doses to treat patients with lung cancer, 5.9% of LUAD tumors and 5.1% of LUSCs in the TCGA have mutations or deletions in Notch2, corresponding to thousands of patients a year that could be treated with a FDA-approved therapy.

**Ready Detection of a Biologically Diverse Array of Chemical/Genetic Associations**

To enrich for robust chemical/genetic associations that enable productive new target pursuit, we established a strict inclusion criteria threshold for automated reporting of predictive biomarker hypotheses from the elastic net (STAR Methods). Receiver operator characteristics and odds ratios
Figure 3. Glucocorticoid Sensitivity Is Predicted by NOTCH2 Mutations

(A) 2-way hierarchical cluster of 5 glucocorticoid AUC values across 100 NSCLC cell lines.

(B) Rank-ordered ED50 values for methylprednisone are indicated as heatmap (top row). Mutation status for NOTCH2 is shown below.

(C) CDF plot comparing ranked mRNA expression of genes in the indicated gene set (Z-scores) in GC sensitive (blue) and resistant cell lines (orange, KS test p = 0.003).

(D) siRNA depletion of NR3C1 or a non-targeting control (NC) followed by 72 hr hydrocortisone treatment (3 μM) in H2073 cells.

(E) Log2 mRNA expression of NR3C1 in GC-responsive and non-responsive cell lines (Illumina BeadArray).

(F) Changes in NR3C1 mRNA 72 hr post-GC treatment (5 μM) in 2 sensitive and resistant cell lines, normalized to untreated cells.

(G) Dose-response curves of cell lines grown in 3D spheroid models in response to methylprednisone. Cell lines that were sensitive (blue) or resistant (orange) to GCs in standard 2D cell-culture were evaluated in 3D. (n = 8/dose).

(H and I) Changes in (H) HES1 and (I) Cyclin D1 protein expression 72 hr post-GC treatment (5 μM) in sensitive and resistant cells.

(J) Flow cytometric histograms for H1993 cells transfected with HES1-pCMV-AC-GFP and treated 72 hr post-GC treatment (5 μM). The propidium iodide signal of cells gated by GFP fluorescence is graphed. Nocodazole (100 ng/mL) was added 48 hr post-treatment to force accumulation of proliferating cells in G2/M over the course of the next 24 hr.

See also Figure S3.
Figure 4. Biological Diversity among Robust Chemical/Genetic Associations

(A) ED_{50} values of the testing set cell lines predicted to be sensitive (blue) and resistant (orange) to each indicated chemical. Dashes indicate class means. Red font indicates chemicals with successfully predicted selectivity profiles (KS test p values < 0.05).

(B) ED_{50} values of cell lines grown in 3D spheroid format in response to the indicated chemicals. Cell lines that were sensitive (blue) or resistant (orange) to each chemical in standard 2D cell-culture were evaluated. Dashes indicate class means. Chemicals for which 2D selectivity is preserved in 3D (KS-test p < 0.05) are highlighted in red.

(C) CDF plot comparing SW036310 sensitivity (AUC) of TTC21B mutant (red) to wild-type cell lines (blue). (scanning KS p < 0.0002).

(D) Dose-response curves of cell lines grown in 3D spheroid models in response to SW036310. Cell lines that were sensitive (blue) or resistant (orange) to SW036310 in standard 2D cell-culture were evaluated (n = 8/dose).

(E) SW036310 sensitivity (AUC) plotted as a function of Ciliobrevin sensitivity (AUC). Pearson R = 0.88; p = 0.0041.

(F) SW140154 sensitivity (ED_{50}) plotted as a function of SW151511 sensitivity (ED_{50}). Pearson R = 0.62, p = 0.00036.

(G) Predictive mRNA expression signature specifies sensitivity to SW140154 and SW151511. Rank ordered ED_{50} values are indicated as a heatmap (top row). Log2 FPKM values are plotted below.

(legend continued on next page)
were calculated as confidence metrics. Finally, predicted responder population frequencies were evaluated using the TCGA LUAD cohort. To enable open access for community-based hypothesis testing, we integrated the final results and all associated quantitative data into a searchable web-based GUI (Data S6) (https://pops.biohpc.swmed.edu).

From the output, we selected 26 predicted chemical/genetic associations to experimentally evaluate for a distrbutive assessment of reliability and biological diversity. The weighted elastic net models derived from the training set (100 cell line panel) were applied to a distinct panel of 33 previously untreated NSCLC cell lines (Data S1; test set). For 21 of the 26 chemicals, at least one cell line in the test set was predicted to be sensitive. 9 of these were validated by empirical testing (Figures 4A and S4B). In addition, 13 chemicals were selected for evaluation of conservation of selective activity in spheroid assays, 9 of which validated. (Figures 4A and S4A). 4 chemicals passing one or both criteria were selected for additional functional characterization (SW036310, SW151511, SW140154, SW208097).

The automated scanning KS analysis indicated that mutations in TTC21B correspond to sensitivity to the benzothiazole-containing small molecule SW036310 (Figure 4C), in which selective efficacy was preserved in spheroid assays (Figure 4D). TTC21B (aka, IFITI139B) is the only known protein to act solely as a retrograde transport motor for primary cilia. Somatic mutations in TTC21B have not been characterized in the setting of cancer, however, loss of function TTC21B mutations upregulate cilia-dependent processes in mice (Tran et al., 2008) and are causal mutations in human developmental diseases driven by primary cilia dysfunction (ciliopathies) (Davis et al., 2011). Gain-of-function primary cilia growth and signaling occurs upon loss of TTC21B activity including activation of sonic hedgehog and potentially other processes known to be regulated by primary cilia, including NF-κB, VHL, and transforming growth factor β (TGF-β) signaling pathways. Indeed, whole genome transcript profiles indicated that gene signatures associated with activation of these pathways were selectively enriched (Figure S4B) and primary cilia were selectively detectable (Figure S4C) in TTC21B mutant, SW036310-sensitive cell lines (Figure S4B). Given these associations, we suspect that SW036310 may perturb a target(s) associated with primary cilia biology that supports survival of TTC21B mutant cells. Consistently, SW036310 sensitivity almost perfectly correlated sensitivity to the cytoplasmic dynein inhibitor, cilobiorein, known to disrupt primary cilia by perturbing anterograde trafficking to that organelle (Figure 4E).

From a distinct biological context, we examined two chemicals, with anti-correlated activity profiles (Figure 4F), corresponding to expression of positive and negative modulators of innate immune signaling (Figure 4G). Using the derived weighted sum elastic net model, we found SW140154 sensitivity was accurately predicted outside the training set by a combination of high expression of the negative regulator of Toll-like receptor signaling (TLR) pathway, SARM1, and low expression of the cytokine receptor, IL18R1 (Figures 4H and 4D) while SW151511 sensitivity could be predicted by high expression of the positive regulator of the TLR pathway, PELI2 (Figures 4H and 4E). We compared cell lines on opposing ends of the sensitivity spectrum and noted that high expression of TLR pathway genes was associated with sensitivity to SW151511 and resistance to SW140154 (Figure 4I). Sensitivity to SW151511 (Figures 4B and 4J), but not SW140154 (Figure 4B and 4F), was recapitulated in spheroid culture models. We therefore selected SW151511 for examination of global gene expression responses to chemical challenge. 2 sensitive and 2 resistant cell lines were treated with SW151511 for 24 hr prior to transcript profiling. We found significant chemically induced expression changes associated with the host defense response (Figure S4G). Notably, this signature was elevated above base-line upon compound exposure, suggesting amplification of a maladaptive innate-immune signaling program may represent a conditional vulnerability in cell lines response to SW151511.

Finally, low nanomolar sensitivity to SW208097 was predicted and validated to correspond to co-occurring mutations in TP53 and KEAP1 (Figures 4K and 4L). This is notable, as the molecule is a well-tolerated investigational drug (GSK923295) targeting the mitotic motor protein CENPE. GSK923295 has dose proportional pharmacokinetics in humans and a low number of grade 3/4 adverse events, however, responder populations have not been identified. Co-occurring TP53 and KEAP1 mutations are detected in 9.6% of LUAD in the TCGA, which extrapolates to ~17,000 patients/year potentially harboring GSK923295-responsive disease.

**A Chemically Addressable Vulnerability of KRAS/KEAP1 NSCLC Cells to Perturbation of SLC2A8**

KRAS mutant lung cancers are common, aggressive, and difficult to manage in the clinic. Therefore, we chose this class for in-depth pursuit of tool compound/target/biomarker triads. We previously reported the overarching phenotypic diversity among KRAS mutant lung cancer cell lines is essentially equivalent to the global phenotypic variation found across all characterized NSCLC cell lines (Kim et al., 2016). Consistent with this, we noted KRAS mutant NSCLC lines distributed across the majority of APC similarity clusters, defined by RNA-seq, within the larger NSCLC cell line panel (Figure 5A). This genomic mRNA expression diversity was mirrored by a diversity of sensitivity of KRAS mutant cell lines to the POPs collection (Figures 5B and 5C).

(H) Cell line sensitivities outside the training set were predicted based on associated elastic net models of SW140154 and SW151511. Boxplot represents AUC values for each prediction class (orange, predicted resistant; blue, predicted sensitive).

(I) SW151511 responsive cells show enrichment of KEGG TLR Signaling compared to SW140154 non-responsive cell lines. (GSEA ES = 0.39).

(J) Dose-response curves of cell lines grown in 3D spheroid models in response to SW151511. Cell lines that were sensitive (blue) or resistant (orange) to SW151511 in standard 2D cell-culture were evaluated. (n = 8/dose).

(K) CDF plot comparing GSK-923295 sensitivity (ED50) of cell lines with co-occurring mutations in TP53 and KEAP1 (red) to wild-type (blue) (scanning KS p < 0.0002).

(L) Dose-response curves for cell lines outside the training set predicted to be sensitive (blue) and resistant (orange) to GSK-923295. See also Figure S4 and Data S6.
Figure 5. Chemical Response Associations among KRAS Mutant NSCLC Lines
(A and B) Cell lines clustered (APC) according to (A) similarity of ED₅₀ responses to the “POPS” and (B) similarity of gene expression (RNA-seq). Cell lines (nodes) are colored according to KRAS status (blue, KRAS mutant; red, KRAS WT).
(C) 2-way hierarchical cluster of KRAS mutant cells according to response to POPs (ED₅₀). Red dash, cherry-picked dataset.
(D) CDF plot comparing SW157765 sensitivity (AUC) of cell lines with co-occurring mutations in KRAS and KEAP1 (red) to wild-type (blue; p < 0.0002; KS test).
(E) Dose-response curves of 5 cell lines not included in the original training panel in response to SW157765 (blue, KEAP1/KRAS mutant; orange, WT).

(legend continued on next page)
However, automated scanning KS analyses returned 4 chemical associations with KRAS mutant subtypes that passed p value thresholds (p < 2E−4). These subtypes were defined by co-occurring mutations in KEAP1, NUP214, PTEN, and TACC2B (Figures S5 A and 5D). Among these, the association of KRAS/KEAP1 double mutant cell lines with sensitivity to SW157765 (Figure 5D) was verifiable in a test set of NSCLC lines distinct from the training set (Figure 5E), and selective efficacy was preserved in spheroid models of lung cancer (Figures 5F and S5B). KEAP1 is a major regulator of the NRF2 antioxidant response. Under normal physiological conditions, NRF2 is constantly ubiquitinated in the cytoplasm by the CUL3/KEAP1 E3 ligase/substrate adaptor complex. Upon stress, KEAP1 inactivation facilitates NRF2 nuclear translocation and consequent activation of the NRF2-dependent anti-oxidant and cytoprotective transcriptional responses. Deleterious mutations or deletions in KEAP1 are present in ~19% of LUADs and ~12% of LUSCs corresponding to constitutive NRF2 activity. Co-occurring mutations in KEAP1 and KRAS are present in ~6% of LUADs (http://www.tcga.org), significantly more than expected by chance (p = 0.007), suggesting they are under positive selective pressure during disease development. There were, however, a few KEAP1 wild-type cell lines that were responsive to SW157765 (Figure 5G). These included the only cell line in the panel that harbors a KEAP1 homozygous deletion (Figure S5C) resulting in undetectable KEAP1 mRNA (Figure 5G). These also included cell lines (2/2) with amino acid substitutions in the degron domain of NRF2 (Figure 5G), corresponding to hotspot NRF2 mutations detected in LUAD tumors lacking functional KEAP1 degradation motifs producing constitutively activated variants (Figure S5D). We did not detect additional variants in known NRF2 pathway genes among the 8 remaining sensitive cell lines that were KRAS/KEAP1 wild-type. However, these cell lines had a significant up-regulation of an NRF2-dependent gene expression signature (p < 2.2E−16) (Figure S5E) and can be predicted to have high NRF2 pathway activity despite the absence of discernable NRF2-related lesions. High expression of this NRF2 gene signature (Figure S5F) was predictive of sensitivity to SW157765 when applied to cell lines outside the training set (Figure S4H).

SW157765 is found as a member of the “prodrug” compounds in which high expression of CYP4F11 is predictive of, and required for, cellular response (Figures 2A and 2G). Of note, CYP4F11 is a candidate NRF2 target gene whose expression is upregulated in NRF2-dependent NSCLC (Goldstein et al., 2016). Given this association, we assessed NRF2-dependent regulation of CYP4F11 in an SW157765-sensitive cell line and found that small interfering RNA (siRNA)-mediated depletion of NRF2 resulted in depletion of CYP4F11 (Figures S1 and S5G) and reduction of sensitivity to SW157765 (Figure 5J). This suggests that NRF2 pathway activity leads to selective production of a toxic SW157765 metabolite. However, we also noted that HCC44 is an SW157765-resistant cell line with high expression of CYP4F11 and corresponding SW157765 metabolism (Figure 2E), suggesting CYP4F11-dependent modification of SW157765 is not sufficient to account for chemical sensitivity. Consistent with this, we found siRNA-mediated depletion of KRAS completely rescued cellular sensitivity to SW157765 even though metabolism of the compound was unaffected (Figures S5K and S5L). Thus, KRAS and NRF2 pathway activation combine to produce a selective cellular vulnerability to SW157765 intervention.

To help identify the nature of this vulnerability, we employed an arrayed genome-scale affinity-selection/mass spectrometry screening strategy to identify SW157765 interacting proteins from among a panel of ~14,000 candidates (Figure S6A). The non-canonical glucose transporter, GLUT8 (SLC2A8), was the sole hit, with an estimated Kd = 200 nM (Figure 6A). Parental SW157765 was used as a substrate for the binding assay rather than the CYP4F11-dependent oxidized metabolite (Figure 2Z), and the higher docking score (7.9 versus 7.6). Notably, docking studies to assess binding of SW157765 to GLUT8 in comparison to the predicted metabolite (Figure S2Z). A homology model GLUT8 was developed from homologous crystal structure of GLUT1 (Kapoor et al., 2016). A crystal structure of GLUT8 has not been published, however, GLUT1 has a 48% sequence similarity with GLUT8, and most of the critical residues for glucose transport are conserved between the two proteins. SW157765 and the epoxide metabolite were predicted to dock on top of one another, except for a deviation near the epoxide region. (Figure S6B). While both chemicals were predicted to interact with residues on GLUT8 (Figure S6B), the epoxide constrains the furan ring, resulting in a significant shift of this ring (55°) relative to the parent molecule, producing stronger predicted interactions with Trp433 and a higher docking score (−7.9 versus −7.6). Notably, docking studies with the inactive analog, 500-1, indicate it fails to achieve
**Figure 6. SW157765-Sensitive Cell Lines Define a KRAS Mechanistic Subtype Addicted to GLUT8-Mediated Glucose Transport**

(A) LC/MS binding signals for ~14,000 proteins tested for the ability to bind to SW157765.

(B) siRNA-mediated depletion of GLUT8 (blue) or a non-targeting control (gray, NC) in SW157765 sensitive and resistant cell lines.

(C) Cellular accumulation of fluorescently labeled 2DG in sensitive and resistant cell lines 72 hr post-SW157765 treatment (purple, 1 μM; blue, 5 μM) or DMSO treatment.

(D) Cellular accumulation of fluorescently labeled 2DG in H647 (KEAP1, KRAS mutant) cells post-siRNA depletion of GLUT1 and GLUT8.

(E) Incorporation of [13C6] into serine (SerM3) and glycine carbons (GlyM2) in SW157765 sensitive (n = 12) and resistant (n = 57) cell lines at 24 hr post-label incubation.

(F and G) Relative viability (Z scores) of SW157765 sensitive and resistant cells in response to (F) siATF4 (G) and siPHGDH.

(H) Incorporation of [13C6] into serine (SerM2) in H647 cells 24 hr post-SW157765 or DMSO treatment.

(I) Protein expression of PHGDH in KEAP1, KRAS mutant, SW157765-sensitive cell lines (A549, H460, and H647) and KEAP1, KRAS mutant unanticipated non-responders (DFCI024, HCC44, H2030, HCC4019).

(J) Dose-response curves of HCC44 cells and HCC44 cells stably expressing either PHGDH or PHGDHV490M in response to SW157765.

(K) SW157765 response (AUC) in CYP4F11 and PHGDH-positive breast cancer cell lines (blue) compared to CYP4F11- and PHGDH-negative cell lines (orange).

See also Figure S6.
a binding pose similar to SW157765, with less H-bonding interactions and a relatively poor docking score of only ~6.8. In aggregate, these analyses are consistent with interaction of both SW157765 and its oxidized metabolite with GLUT8. Enhanced GLUT8 thermal stability in cells treated with SW157765, but not 500-1, provided orthogonal evidence for this interaction (Figure S6C).

GLUT8 is a member of the class III glucose transporters thought to mainly participate in translocation of glucose across the blastocyst membrane (Carayannopoulos et al., 2000). A role for GLUT8 in cancer has not been well studied, although it has been found to be significantly upregulated in endometrial cancer and in multiple myeloma relative to normal tissue. While the dominant glucose transporter in tumor cells is thought to be GLUT1, upregulated class III glucose transporters may support higher energy demands in some cases (Schmidt et al., 2009). Supporting this notion, glucose uptake and viability of a subset of multiple myeloma cell lines was found to be dependent on the continued expression of GLUT8 but not GLUT1 (McBrayer et al., 2012). Notably, we found SW157765-sensitive NSCLC cell lines were also selectively sensitive to glucose deprivation (Figure S6D) and to GLUT8 depletion (Figure 6B). Furthermore, SW157765 selectively inhibited fluorescent 2-deoxyglucose (2DG) uptake in SW157765-sensitive cells in a dose-dependent manner (Figure 6C). In contrast, GLUT1 depletion (Figures S6E and S6F) had no effect on 2DG uptake (Figure 6D) or viability (Figure S6G). These observations are consistent with action of SW157765 at the level of GLUT8 inhibition and a selective dependence of KRAS/KEAP1 mutant cells on GLUT8 for glucose consumption.

Uniformly labeled \(^{13} \text{C}_6\) glucose is metabolized via the glycolytic cycle to 3-phosphoglycerate (3PG), which can enter the serine biosynthetic pathway where it is converted in a series of steps to serine, which is subsequently cleaved to produce glycine and a one-carbon intermediate that can enter the folate cycle to ultimately result in the production of purines and thymidines (Figure S6H). Of relevance to this study, high NRF2 activity was recently demonstrated to promote serine/glycine biosynthesis, in some NSCLC cell types, through ATP4-dependent expression of rate-limiting serine biosynthetic enzymes (PHGDH, PSAT1, PSPH, SHMT1, and SHMT2). De novo serine biosynthesis is upregulated in subsets of lung cancer, breast cancer, glioma, and melanoma, presumably to support glutathione and nucleotide production, and can be required for tumor cell survival. When we examined carbon flux from uniformly labeled \(^{13} \text{C}_6\) glucose into serine (SerM3) and glycine (GlyM2) mass isotopomers across 63 NSCLC cell lines, we found significant enrichment of heavy carbons in the SW157765 sensitive cell lines (Figure 6E), corresponding to significantly higher expression of serine biosynthetic pathway genes (Figure S6I) and selective consequences on cell survival upon depletion of the ATF4 transcription factor (Figure 6F) and PHGDH (Figure 6G), the enzyme that catalyzes the first committed step in the serine biosynthetic pathway. Taken together, these findings indicate a dependence of KRAS/KEAP1 mutant NSCLC cells on consumption of glucose to support serine biosynthetic pathway activity.

These cumulative observations led us to consider that SW157765 may be acting to reduce carbon flux through the serine biosynthetic pathway, leading to selective targeting of cancer cells dependent on this pathway. To test this, we pretreated H647 (KRAS/KEAP1 mutant) with SW157765 for 24 hr, an interval where we do not observe significant induction of cell death (Figure S6J), followed by exposure to glucose (\(^{13} \text{C}_2\)) (Figure S6O). Heavy carbon labeling of serine reached steady state after 2 hr (SerM2), which was reduced 5-fold by SW157765 (Figure 6H). Iteration of this approach within a panel of NSCLC cell lines indicated basal carbon flux through the serine biosynthetic pathway was higher in SW157765-sensitive cell lines, as expected, and exposure to SW157765 selectively reduced serine labeling in these cell lines (Figure 6I). Notably, carbon flux from glucose through the pentose phosphate pathway (PPP; LacM1) or the citric acid cycle (TCA; CitM2) (Figures S6K, S6L, and S6O) was not affected by SW157765. Both KRAS and NFR2 are known to shunt glucose toward the PPP. Thus, the selective consequences of SW157765 exposure on serine/glycine metabolism suggests a dominant routing of available glucose to the PPP and TCA. GLUT8 may therefore support supplementary glucose consumption to an extent that provides for serine/glycine biosynthetic demands of the KRAS/KEAP1 mutant cellular context.

While KEAP1 and KRAS mutation status was robustly predictive of sensitivity to SW157765, we noted 4 unanticipated nonresponders (DFCI024, HCC44, H2030, HCC4019) with co-occurring mutations in KEAP1 and KRAS and high expression of CYP4F11 (Figure S6M). Importantly, PHGDH was greatly reduced or absent in all 4 cell lines (Figure 6J). Additionally, cell line H2030 is completely missing mRNA expression of PSAT1 (Figures S6H and S6M). These cells may be resistant to SW157765 due to a pre-existing adaptation that reduces contributions of glycolysis to serine/glycine biosynthesis. To test this, we stably expressed either full-length PHGDH or a hypomorphic mutant (PHGDH<sup>HOM</sup>) (Tabatabaei et al., 2009) in HCC44 cells under the control of a doxycycline-inducible promoter (Figure S6N). Overexpression of PHGDH, but not PHGDH<sup>HOM</sup>, sensitized HCC44 cells to SW157765 (Figure 6K), suggesting that PHGDH re-established carbon flux into serine/glycine production with consequent reliance on GLUT8 to maintain sufficient glucose consumption.

In summary, we have shown co-occurring mutations in KEAP1 and KRAS define a vulnerability to continued function of GLUT8. Inhibition of GLUT8 is associated with a reduction of glucose intake leading to a selective shunting of glucose from serine biosynthesis. We found overexpression of wild-type PHGDH can re-sensitize HCC44 cells to SW157765. Perhaps most intriguingly, re-introduction of PHGDH also can sensitize cells to GLUT8 inhibition. These findings suggest that shunting cellular consumption of glucose to serine biosynthesis generates a dependency on GLUT8 that can be selectively targeted with SW157765. To potentially help assess the generality of this relationship, we profiled SW157765 for toxicity in a panel of 27 breast cancer cell lines with publically available genomics data (Barretina et al., 2012). Importantly, we found that copy number-driven amplification of PHGDH expression together with...
high expression of CYP4F11 significantly corresponded to SW157765 sensitivity (Figure 6L).

DISCUSSION

The chemistry-first target nomination approach employed here was designed to leverage large-scale uncharacterized chemical diversity as a de novo discovery tool unconstrained by any preconceived notions of mechanistic relationships. Of the 202,103 chemicals employed, <1% are associated with known or suspected modes of action. However, computer-automated rediscovery of current precision medicine relationships from within the collection of known compounds authenticated the screening cascade. This included the association of erlotinib-sensitivity with EGFR mutation status and the association of crizotinib-sensitivity with EML4-ALK translocations. Likewise, the pipeline returned novel and robust repurposing hypotheses and biomarkers for clinically available compounds that currently lack patient selection hypotheses.

While the above observations credentialed the rigor of the experimental and data analytics pipeline, the identification of novel chemical/genetic relationships was the primary objective. To that end, 171 compounds were linked to genomic features within a 95% confidence interval by the elastic net. These chemical/genetic relationships spanned a strikingly diverse array of biological processes including selective vulnerabilities associated with host defense pathway activation, ciliogenesis, and nuclear hormone signaling. Furthermore, pharmacological relationships that were a consequence of selective chemical clearance and/or selective chemical metabolism were readily detectable.

To pursue candidate “therapeutic triads,” we focused on KRAS mutant adenocarcinoma. When considered as a single class, these cell models displayed diverse and discordant response profiles to the chemical collection employed. However, significant chemical/genetic associations were detected upon segmentation of cell models according to mutations in additional genes that co-occurred with KRAS mutations at a reasonable frequency. This is in accordance with accumulating evidence that KRAS mutant lung cancers parse across multiple distinct mechanistic subtypes. Most notably, KRAS/KEAP1 double mutant NSCLC cells were selectively sensitive to the benzothiazole, SW157765, due to the convergent consequences of dual KRAS and NRF2 modulation of metabolic and xenobiotic gene regulatory programs. GLUT8 was identified as a mechanistic target of SW157765 and is selectively required to support the diversion of glucose to serine biosynthesis in this genetic background. Modulation of these regulatory programs by orthogonal means was sufficient to modulate SW157765 responsiveness. We note that lineage-restricted biomarker discovery was key to identifying the KRAS/KEAP1 synthetic chemical relationship. Of note, parallel analysis within a large cohort of breast cancer cell lines returned sensitivity-associated biomarkers indicative of a conserved biological mode-of-action for SW155765, but there was no relationship with KRAS/KEAP1 mutation status, an oncogenotype that is exceedingly rare in breast.

We find that the application of a large diversity-oriented chemical collection, within a carefully delineated phenotypic screening convention, can uncover a compelling diversity of heretofore unappreciated target opportunities within the seeming cacophony of molecular etiology of lung cancer. Importantly, these target opportunities are, by nature of the discovery paradigm, associated with precision medicine strategies and pharmacologically addressable. Furthermore, it is clear that chemical vulnerabilities can be revealed that are linked to recurrent mutations in lung cancer patients that are not currently “actionable.” Thus, we argue that many undeveloped avenues remain open for productive pursuit of tumor-intrinsc precision medicine.

STAR METHODS

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

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Cell Lines
- **METHODS DETAILS**
  - Chemical libraries
  - Spheroid assays
  - mRNA isolation and microarray
  - Carbon tracing and metabolic flux analysis
  - Glucose deprivation assays
  - qPCR
  - Thermal stability shift assay
  - Targeted siRNA and plasmid DNA transfection
  - CRISPR-mediated cell line editing
  - Immunoblot analysis
  - 96 well dose response assays
  - Flow cytometry analysis
  - In vitro determination of compound stability with human tumor cells
  - SW157765 Metabolite Identification
  - Glucose utilization
  - Binder selection assays
  - DNA/RNA extraction for sequencing
  - Genomic characterization
  - Germline variant filtering
  - Small molecule cytotoxicity assays
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - In-silico Molecular Docking
  - Normalization of drug response data and calculation of ED50 and AUC values
  - Scanning Kolmogorov-Smirnov statistic
  - Elastic Net Regression
  - Cell line pathway activity analysis
  - Gene Set Enrichment Analysis
  - Sensitivity Prediction
  - ROC Curve Analysis
  - Affinity propagation clustering
  - NRF2 Signature
  - Comparison of cell lines to tumors
  - Peg plots
  - Other Statistical Analyses
- **DATA AND SOFTWARE AVAILABILITY**
SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and six data files and can be found with this article online at https://doi.org/10.1016/j.cell.2018.03.028.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (CA197177, CA176284, CA70907, CA142543), CPrIT (RP120732, RP110708, RP110708), Robert Welch Foundation (I-1144), the Korea Health Technology R&D project through the Korea Health Industry Development Institute (HI14C1324), and the National R&D Program for Cancer Control (1421008), funded by the Ministry of Health & Welfare, Republic of Korea. E.A.M. was supported by NIH training grant 5T32GM8203-27. C.D. and R.M.V were supported by CPrIT training grant RP141090, and R.M.V was supported by NIH training grant 5T32CA124334-09. We would like to thank Hanspeter Niederstrasser, Melissa McCoy, Shuguang Screening Core, and Anwu Zhou in the UT Southwestern High-throughput Screening Core for their support of the large-scale screening and dose-response experiments described herein.

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS

The authors have no competing interests related to this work. M.A.W. is currently an employee of Pfizer Inc., and T.M. was an employee of Takeda Oncology. Currently, Hanspeter Niederstrasser is employed by a pharmaceutical company. The authors have no competing interests related to this work.

REFERENCES


McBrayer, S.K., Cheng, J.C., Singhal, S., Krett, N.L., Rosen, S.T., and Shammu- 
nase (PHGDH) are distributed throughout the protein and result in altered enzyme kinetics. Hum. Mutat. 30, 749–756.
# STAR★METHODS

## KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUT13</td>
<td>Abcam</td>
<td>ab190163</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Cell Signaling Technology</td>
<td>5174; RRID: AB_10622025</td>
</tr>
<tr>
<td>B-actin</td>
<td>Santa Cruz Biotechnology</td>
<td>sc-4778</td>
</tr>
<tr>
<td>LaminB1</td>
<td>Cell Signaling Technology</td>
<td>12586; RRID: AB_2650517</td>
</tr>
<tr>
<td>KRAS</td>
<td>Santa Cruz Biotechnology</td>
<td>F234; RRID: AB_627865</td>
</tr>
<tr>
<td>CYP4F11</td>
<td>Santa Cruz Biotechnology</td>
<td>sc-53619; RRID: AB_2088410</td>
</tr>
<tr>
<td>HES1</td>
<td>Cell Signaling Technology</td>
<td>11988</td>
</tr>
<tr>
<td>CyclinD1</td>
<td>Cell Signaling Technology</td>
<td>2922; RRID: AB_2228523</td>
</tr>
<tr>
<td>PHGDH</td>
<td>Cell Signaling Technology</td>
<td>66350</td>
</tr>
<tr>
<td>NRF2</td>
<td>Invitrogen</td>
<td>710574; RRID: AB_2532742</td>
</tr>
<tr>
<td>B-tubulin</td>
<td>Cell Signaling Technology</td>
<td>2128; RRID: AB_823664</td>
</tr>
<tr>
<td>Cas9</td>
<td>Abcam</td>
<td>ab191468; RRID: AB_2693235</td>
</tr>
<tr>
<td>GLUT8</td>
<td>Abcam</td>
<td>ab99132; RRID: AB_10672063</td>
</tr>
<tr>
<td>GLUT1</td>
<td>Abcam</td>
<td>ab652; RRID: AB_305540</td>
</tr>
<tr>
<td>NR3C1</td>
<td>Cell Signaling</td>
<td>12041; RRID: AB_2631286</td>
</tr>
<tr>
<td><strong>Biological Samples: Cell Lines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THLE2</td>
<td>ATCC</td>
<td>CRL2706</td>
</tr>
<tr>
<td>THLE3</td>
<td>ATCC</td>
<td>CRL11233</td>
</tr>
<tr>
<td>BT-20</td>
<td>ATCC</td>
<td>HTB19</td>
</tr>
<tr>
<td>BT-474</td>
<td>ATCC</td>
<td>HTB20</td>
</tr>
<tr>
<td>BT-483</td>
<td>ATCC</td>
<td>HTB121</td>
</tr>
<tr>
<td>BT-549</td>
<td>ATCC</td>
<td>HTB122</td>
</tr>
<tr>
<td>HCC1143</td>
<td>ATCC</td>
<td>CRL2321</td>
</tr>
<tr>
<td>HCC1428</td>
<td>ATCC</td>
<td>CRL2327</td>
</tr>
<tr>
<td>HCC1187</td>
<td>ATCC</td>
<td>CRL2322</td>
</tr>
<tr>
<td>HCC1395</td>
<td>ATCC</td>
<td>CRL2324</td>
</tr>
<tr>
<td>HCC1419</td>
<td>ATCC</td>
<td>CRL2326</td>
</tr>
<tr>
<td>HCC1500</td>
<td>ATCC</td>
<td>CRL2329</td>
</tr>
<tr>
<td>HCC1569</td>
<td>ATCC</td>
<td>CRL2330</td>
</tr>
<tr>
<td>HCC1806</td>
<td>ATCC</td>
<td>CRL2335</td>
</tr>
<tr>
<td>HCC1937</td>
<td>ATCC</td>
<td>CRL2336</td>
</tr>
<tr>
<td>HCC202</td>
<td>ATCC</td>
<td>CRL2316</td>
</tr>
<tr>
<td>HCC70</td>
<td>ATCC</td>
<td>CRL2315</td>
</tr>
<tr>
<td>Hs-578-T</td>
<td>ATCC</td>
<td>HTB126</td>
</tr>
<tr>
<td>MDA-MB-157</td>
<td>ATCC</td>
<td>HTB24</td>
</tr>
<tr>
<td>MDA-MB-231</td>
<td>ATCC</td>
<td>HTB26</td>
</tr>
<tr>
<td>MDA-MB-436</td>
<td>ATCC</td>
<td>HTB130</td>
</tr>
<tr>
<td>MDA-MB-453</td>
<td>ATCC</td>
<td>HTB131</td>
</tr>
<tr>
<td>MDA-MB-468</td>
<td>ATCC</td>
<td>HTB132</td>
</tr>
<tr>
<td>MCF7</td>
<td>ATCC</td>
<td>HTB22</td>
</tr>
<tr>
<td>MDA-MB-361</td>
<td>ATCC</td>
<td>HTB27</td>
</tr>
<tr>
<td>SK-BR-3</td>
<td>ATCC</td>
<td>HTB30</td>
</tr>
<tr>
<td>T47D</td>
<td>ATCC</td>
<td>HTB133</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>UACC-812</td>
<td>ATCC</td>
<td>CRL1897</td>
</tr>
<tr>
<td>ZR-75-1</td>
<td>ATCC</td>
<td>CRL1500</td>
</tr>
<tr>
<td>NSCLC cell line collection</td>
<td>this paper</td>
<td>described in Data S1</td>
</tr>
</tbody>
</table>

### Chemicals, Peptides, and Recombinant Proteins

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSK923295</td>
<td>CalBioChem</td>
<td>532372</td>
</tr>
<tr>
<td>HET0016</td>
<td>Santa Cruz Biotechnology</td>
<td>sc-200673</td>
</tr>
<tr>
<td>nocodazole</td>
<td>Sigma Aldrich</td>
<td>M1404-G02</td>
</tr>
<tr>
<td>NAC</td>
<td>Sigma Aldrich</td>
<td>A7250</td>
</tr>
<tr>
<td>POPS</td>
<td>this paper</td>
<td>described in Data S6</td>
</tr>
</tbody>
</table>

### Deposited Data

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSCLC lung microarray</td>
<td>Kim et al., 2016</td>
<td>GEO: GSE32026</td>
</tr>
<tr>
<td>SW151511 treated microarray</td>
<td>this paper</td>
<td>GEO: GSE104757</td>
</tr>
<tr>
<td>NSCLC sequencing results</td>
<td>this paper</td>
<td>dbGAP: 28061</td>
</tr>
</tbody>
</table>

### Oligonucleotides

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLUT1 taqman probe</td>
<td>Applied Biosciences</td>
<td>Hs00892681_m1</td>
</tr>
<tr>
<td>GLUT8 taqman probe</td>
<td>Applied Biosciences</td>
<td>Hs00205863_m1</td>
</tr>
<tr>
<td>NR3C1 taqman probe</td>
<td>Applied Biosciences</td>
<td>Hs00353740_m1</td>
</tr>
<tr>
<td>CYP4F11 sgRNA sequence (CACCAGAGGCAGGCG CGGAGTGTGTC)</td>
<td>Integrated DNA Technologies</td>
<td>custom oligo</td>
</tr>
<tr>
<td>siACC1 (GAAAGCAGGCUAUCAGUGA-CAGCAA ACCUUGAUCUUCUCCAUAU CGAUAUAGGUUGUUGU)</td>
<td>Dharmacon</td>
<td>D004551-01, D004551-02, D004551-03, D004551-04</td>
</tr>
<tr>
<td>siABC2 (GCAGAUGCCUUCUUGUUA-UUGC AAUCUUGUUAUUGUCAAGAUGAUGAUGGUUGU)</td>
<td>Dharmacon</td>
<td>D009924-01, D009924-02, D009924-03, D009924-04</td>
</tr>
<tr>
<td>siACUC1 (CACUCUCAUAGGACUCUGA-CGCAGGA UCAGAUGCCUUUUAGGACAGAAGAAAGAGA)</td>
<td>Ambion</td>
<td>4390824</td>
</tr>
<tr>
<td>siNR2F2 (GGAUGACCUACGGAGCA-CAAAAAG AGGUUGCGCUAAGAUCAGAAGAAACGAUA)</td>
<td>Mission</td>
<td>SASI_Hs01_00182394, SASI_Hs01_00182393, SASI_Hs02_00341015</td>
</tr>
<tr>
<td>siKras (CGAUAUGAUCCAAAGAAA-UAGGAC UCUGAACUGUAAUGAUGAUGUUAUUGUGCUAAGAAGAAACGAUA)</td>
<td>Dharmacon</td>
<td>D-005069-01, D-005069-01, D-005069-01, D-005069-01</td>
</tr>
<tr>
<td>siNC (UCAGCCCCUCUGUAUAUGUAUGAUGUUGUUGCUAAGAAGAAACGAUA)</td>
<td>Dharmacon</td>
<td>D-001210-01, D-001210-02, D-001210-03, D-001210-04</td>
</tr>
<tr>
<td>siGLUT8 (UUGAAGAGGCCAUGUUGAUGUUAGUCUGACUGAUGUUGUUGLUUAGAUGUUGUUGCUAAGAAGAAACGAUA)</td>
<td>Dharmacon</td>
<td>D007520-01, D007520-02, D007520-03, D007520-04</td>
</tr>
<tr>
<td>siGLUT1 (UUGAAGAGGCCAUGUUGAUGUUAGUCUGAUGUUGUUGCUAAGAAGAAACGAUA)</td>
<td>Dharmacon</td>
<td>D007509-01, D007509-02, D007509-03, D007509-04</td>
</tr>
</tbody>
</table>

### Recombinant DNA

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHGDH overexpression vector</td>
<td>Locasale et al., 2011, courtesy of Matt Vander Heiden</td>
<td>NA</td>
</tr>
<tr>
<td>pLvx-Tight-Puro</td>
<td>Clontech</td>
<td>632163</td>
</tr>
<tr>
<td>HE51-pCMV6-AC-GFP</td>
<td>Origene</td>
<td>R9211709</td>
</tr>
<tr>
<td>lenti-cas9 blast</td>
<td>Addgene</td>
<td>52962</td>
</tr>
<tr>
<td>lenti-guide puro</td>
<td>Addgene</td>
<td>52963</td>
</tr>
<tr>
<td>psPax2</td>
<td>Addgene</td>
<td>12260</td>
</tr>
<tr>
<td>pMD2-VSV-G</td>
<td>Addgene</td>
<td>12259</td>
</tr>
</tbody>
</table>

(Continued on next page)
### REAGENT or RESOURCE SOURCE IDENTIFIER

<table>
<thead>
<tr>
<th>Reagent/Resource</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHGDH-V490M overexpression vector</td>
<td>Locasale et al., 2011, courtesy of Matt Vander Heiden</td>
<td>NA</td>
</tr>
<tr>
<td>pcDNA3.1</td>
<td>invitrogen</td>
<td>V79020</td>
</tr>
</tbody>
</table>

### Software and Algorithms

<table>
<thead>
<tr>
<th>Software and Algorithms</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>lumi</td>
<td>R Biodonductor</td>
<td><a href="https://bioconductor.org/">https://bioconductor.org/</a></td>
</tr>
<tr>
<td>glmnet</td>
<td>R Biodonductor</td>
<td><a href="https://bioconductor.org/">https://bioconductor.org/</a></td>
</tr>
<tr>
<td>Gene Set Enrichment Analysis</td>
<td>Subramanian et al., 2005</td>
<td><a href="http://software.broadinstitute.org/gsea/index.jsp">http://software.broadinstitute.org/gsea/index.jsp</a></td>
</tr>
<tr>
<td>FlowJo 7.6.5.</td>
<td>FlowJo</td>
<td><a href="https://www.flowjo.com/">https://www.flowjo.com/</a></td>
</tr>
<tr>
<td>Prism nonlinear software program</td>
<td>Graphpad</td>
<td><a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a></td>
</tr>
<tr>
<td>DNAcopy</td>
<td>Bioconductor</td>
<td><a href="https://bioconductor.org/">https://bioconductor.org/</a></td>
</tr>
<tr>
<td>FastQC</td>
<td>Babraham Bioinformatics Institute</td>
<td><a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a></td>
</tr>
<tr>
<td>STAR V.2.4.2</td>
<td>Dobin et al., 2013</td>
<td><a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a></td>
</tr>
<tr>
<td>RSeQC</td>
<td>Wang et al., 2012</td>
<td><a href="http://rseqc.sourceforge.net/">http://rseqc.sourceforge.net/</a></td>
</tr>
<tr>
<td>Cufflinks</td>
<td>Trapnell et al., 2010</td>
<td><a href="http://cole-trapnell-lab.github.io/cufflinks/">http://cole-trapnell-lab.github.io/cufflinks/</a></td>
</tr>
<tr>
<td>drc</td>
<td>Bioconductor</td>
<td><a href="https://bioconductor.org/">https://bioconductor.org/</a></td>
</tr>
<tr>
<td>cgdsr</td>
<td>Bioconductor</td>
<td><a href="https://bioconductor.org/">https://bioconductor.org/</a></td>
</tr>
<tr>
<td>ROCR</td>
<td>Bioconductor</td>
<td><a href="https://bioconductor.org/">https://bioconductor.org/</a></td>
</tr>
<tr>
<td>apc</td>
<td>Bioconductor</td>
<td><a href="https://bioconductor.org/">https://bioconductor.org/</a></td>
</tr>
<tr>
<td>cytoscape</td>
<td>Shannon et al., 2003</td>
<td><a href="http://www.cytoscape.org/">http://www.cytoscape.org/</a></td>
</tr>
<tr>
<td>R Version 3.4.0</td>
<td>R</td>
<td><a href="https://www.r-project.org/">https://www.r-project.org/</a></td>
</tr>
<tr>
<td>Resources website for POPs</td>
<td>this paper</td>
<td><a href="https://pops.biohpc.swmed.edu">https://pops.biohpc.swmed.edu</a></td>
</tr>
<tr>
<td>BD FACSDiva software version 8.0.1</td>
<td>BD Biosciences</td>
<td><a href="http://wwwbdbiosciences.com/">http://wwwbdbiosciences.com/</a></td>
</tr>
<tr>
<td>SARNavigator</td>
<td>Tripos, Inc</td>
<td><a href="http://www.tripossoftware.com/">http://www.tripossoftware.com/</a></td>
</tr>
<tr>
<td>Protein Preparation Wizard</td>
<td>Schrodinger</td>
<td><a href="https://www.schrodinger.com/protein-preparation-wizard">https://www.schrodinger.com/protein-preparation-wizard</a></td>
</tr>
<tr>
<td>Glide V6.5</td>
<td>Schrodinger</td>
<td><a href="https://www.schrodinger.com/glide">https://www.schrodinger.com/glide</a></td>
</tr>
<tr>
<td>Maestro</td>
<td>Schrodinger</td>
<td><a href="https://www.schrodinger.com/maestro">https://www.schrodinger.com/maestro</a></td>
</tr>
</tbody>
</table>

### Other

<table>
<thead>
<tr>
<th>Other</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powerplex 1.2 DNA fingerprinting kit</td>
<td>Promega</td>
<td>DC2705</td>
</tr>
<tr>
<td>Mycoplasma testing kit</td>
<td>Intron Bio</td>
<td>25235</td>
</tr>
<tr>
<td>96-well U-bottom low adherence plates</td>
<td>NuncIon Sphera, Thermo Scientific</td>
<td>174925</td>
</tr>
<tr>
<td>CellTiter-Glo® 3D cell viability assay</td>
<td>Promega</td>
<td>G7573</td>
</tr>
<tr>
<td>miRNeasy</td>
<td>QIAGEN</td>
<td>217004</td>
</tr>
<tr>
<td>qScript cDNA SuperMix</td>
<td>Quanta Biosciences</td>
<td>95048</td>
</tr>
<tr>
<td>taqman universal master mix II</td>
<td>Applied Biosciences</td>
<td>4440038</td>
</tr>
<tr>
<td>Glucose Uptake Essay Kit</td>
<td>Abcam</td>
<td>ab204702</td>
</tr>
<tr>
<td>Lipopectamine-2000</td>
<td>Invitrogen</td>
<td>11668027</td>
</tr>
<tr>
<td>Hoechst-33342</td>
<td>Invitrogen</td>
<td>H3570</td>
</tr>
<tr>
<td>RNAiMax</td>
<td>Invitrogen</td>
<td>13778-150</td>
</tr>
<tr>
<td>FuGENE 6</td>
<td>Promega</td>
<td>E2691</td>
</tr>
<tr>
<td>Illumina Human1M-Duo DNA Analysis BeadChip</td>
<td>Illumina</td>
<td>WG-311</td>
</tr>
</tbody>
</table>
CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Michael A. White (michael.white@utsouthwestern.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

Most NSCLC lines used in this study were part of the NCI and HCC (Hamon Cancer Center at UT Southwestern) series of cell lines, with the exception of THLE-2, THLE-3, A427, A549, Calu.1, Calu.6 (American Type Culture Collection; ATCC), Cal.12T (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH; DSMZ), DFCI.024, DFCI.032 (Dana Farber Cancer Institute, courtesy of Pasi Jänne), EKVX, Hop62 (NCI-60 panel), PC9 (Johns Hopkins University School of Medicine, courtesy of Bert Vogelstein). Cell lines from these collections were cultured in RPMI 1640 (GIBCO, 2.05mM L-glutamine) supplemented with 5% FBS (GIBCO) and 1% penicillin/streptomycin (GIBCO). Normal bronchiole epithelia-derived cell lines (Ramirez et al., 2004) were grown in ACL4 (RPMI 1640 supplemented with 0.02 mg/ml insulin, 0.01 mg/ml transferrin, 25 nM sodium selenite, 50 nM hydrocortisone, 10 mM HEPES, 1 ng/ml EGF, 0.01 mM ethanolamine, 0.01 mM O-phosphorylethanolamine, 0.1 mM triiodothyronine, 2 mg/ml BSA, 0.5 mM sodium pyruvate) with 2% FBS and 1% penicillin/streptomycin. Normal liver lines, THLE-2 and THLE-3, were grown in the Bronchial Epithelial Cell Growth Medium (Lonza, CC-3170) supplemented with 10% FBS and 1% penicillin/streptomycin. All cell lines were maintained in a humidified environment in the presence of 5% CO2 at 37°C. All cell lines were DNA fingerprinted (Powerplex 1.2 Kit, Promega) and mycoplasma free (myco kit, Boca Scientific). All chemicals beginning with the prefix SW are from the UT Southwestern Chemical Library. THZ1 was obtained from Calbiochem, ciliobrevin from Tocris, GSK923295 from SelleckChem, HET-0016 from Santa Cruz Biotechnology, nocardazole from Sigma-Aldrich.

METHODS DETAILS

Chemical libraries

The UT Southwestern chemical screening library (~200,000 small molecules; Figure S1G) is composed of 75,000 compounds purchased from ChemBridge Corporation, 100,000 compounds purchased from Chemical Diversity Labs and 22,000 compounds purchased from ComGenex, 1200 purchased from TimTek, 1100 from Prestwick, 2,500 compounds from UTSW chemistry labs, and the 450 compounds of the NIH clinical collection. The TimTek compounds are "natural product-like" synthetic compounds and the Prestwick compounds are off-patent drugs. The NIH clinical collection is composed of compounds that have been tested in phase I clinical trials. The library is free of commercial ties and is wholly owned by UT Southwestern. The compounds from commercial sources were selected to reflect the diversity available in the larger sets of compounds offered by each company and were screened for desirable, drug-like qualities and chemical diversity, using filtering software running in the CheD and SARNavigator programs (Tripos). The compounds in the library satisfy a relaxed set of Lipinsky’s rules, with 99% having a molecular weight less than 550 (average 250-300).

Spheroid assays

Cell lines were trypsinized, counted, and plated into 96-well U-bottom low adherence plates (Nunclon Sphera, Thermo Scientific). Cells were inoculated between 500-4,000 cells per well depending on growth rate. Spheroids were allowed to form over 48 hr, drug was added, and the plates incubated for an additional 96 hr. Luminescence assays were performed using CellTiter-Glo 3D cell viability assay (Promega) according to the manufactures instructions. The plates were read on a BMG Labtech FLUOstar® Optima.

RNA isolation and microarray

All cells were seeded in 6-well plates at 300,000 cells/well in 2 mL standard culture media (RPMI, 5%FBS, penicillin/streptomycin) and allowed to adhere overnight. The media was discarded and replaced with 1.5 mL treatment media containing either 0.1% DMSO vehicle control, or 10 µM of SW compound. After 24 hr of treatment, total RNA was harvested using the miRNeasy QIAEN kit according to the manufacturer’s instructions. All samples were submitted for microarray analysis at the UT Southwestern Microarray Core using an Illumina Human-HT-12 v4 Expression BeadChip. Raw intensity values were background corrected and quantile normalized using the lumi package in R. Using a minimal expression cutoff of 7, we eliminated from the analysis genes that were not expressed before or after compound treatment. For each gene, normalized values were converted to a log2 to score to indicate the fold change with compound treatment with the following equation:

$$\frac{x_{\text{comp}}}{x_{\text{DMSO}}}$$
where \( x_{\text{DMSO}} \) and \( x_{\text{comp}} \) is the normalized expression value of gene \( x \) with DMSO and 10 \( \mu \)M of SW compound treatment, respectively. The data were deposited into the Gene Expression Omnibus (GEO) under the accession number GSE104757.

Carbon tracing and metabolic flux analysis

All labeling experiments performed with cells plated at a density of 200,000 cells per 60 mm diameter dishes and grown for 48 hr. Afterward, media was removed and cells were rinsed with PBS prior to treatment with SW157765 for either 6 or 24 hr. Media was then removed and cells were rinsed with PBS prior to treatment with SW157765 in media containing glucose-free RPMI supplemented with 5% FBS and \( ^{13}C \) glucose. For the 6 hr compound treatment, \( ^{13}C \) media mixture was added for 2 hr. For the 24 hr compound treatment, cells were rinsed with phosphate-buffered saline, replenished with \( ^{13}C \) labeling medium with SW157765 and cultured for time points ranging from 0 to 2 hr as indicated at the end of the 24 hr. To determine basal metabolomics flux in the panel of 61 cell lines, cells were pre-treated with media containing either \( ^{13}C \) labeled glucose or glutamine for either 6 or 24 hr. Labeled cells were briefly rinsed with cold saline, pelleted at cold 50% methanol, lysed through at least 3 freeze-thaw cycles, and then centrifuged to remove debris. The supernatants were evaporated to dryness methoximated and derivatized by tert-butyl dimethylsilylation. One mL of the derivatized material was injected onto an Agilent 6970 gas chromatograph equipped with a fused silica capillary GC column (30 m length, 0.25 mm diameter) and networked to either an Agilent 5973 or 5975 Mass Selective Detector. Retention times of all metabolites of interest were validated using pure standards. The measured distribution of mass isotopomers was corrected for natural abundance of \( ^{13}C \).

Glucose deprivation assays

Cells were cultured in either complete media (RPMI supplemented with 5% FBS) or media lacking glucose for 5 days. Viability was determined as relative content of DNA at day 5 in reduced media as compared to cells grown in complete medium.

qPCR

Cells were plated at a density of 250,000 cells/well in 6 well format and allowed to incubate overnight. The cells were then washed with PBS twice prior to RNA extraction with the RNeasy Mini Kit (QIAGEN) following the manufacturer’s recommended protocol. 100 ng to 1ug of total RNA was mixed with qScript cDNA SuperMix for cDNA synthesis (Quanta Biosciences) or taqman universal master mix II (Applied Biosciences). Taqman gene expression probes (Applied Biosciences) for GLUT1, GLUT8 and NR3C1, were used for real-time qPCR amplification on a Light Cycler 480 II Real-Time PCR System (Roche). The cycling program was 95°C for 10 min, 95°C for 15 s, and 60°C for 40 cycles. Each sample was run in triplicate, normalized to the Cy5 standard probe, and analyzed by the comparative \( C_t \) method.

Thermal stability shift assay

3E6 cells were cultured in 75 cm² flasks for overnight growth. Cells were treated with RPMI media supplemented with 5% FBS containing either 0.1% DMSO or 1 \( \mu \)M SW157765 for 24 hr. After treatment, cells were detached with trypsin, collected by centrifugation, resuspended in PBS, and cell suspensions of 500,000 cells/tube were transferred into 8-well 0.2 mL PCR tubes and heated for 3 min.

For transfection in 96 well format, 0.1–1 \( \mu \)L siRNA (10 \( \mu \)M) of siRNA in 25 \( \mu \)L of serum-free RPMI was mixed with either 0.2 or 0.4 \( \mu \)L of RNAmax (Invitrogen) in 25 \( \mu \)L serum-free RPMI. Following a 15 min incubation, the siRNA-lipid mixture was transferred to a 96 well plate followed by plating of cells at a concentration ranging from 3000 cells/well to 5000 cells/well (depending on cellular growth rate) in 100 \( \mu \)L media. Optimal concentration of siRNA was determined by titering amounts from 0.1 to 1 \( \mu \)L per well and selecting the maximal concentration for which no death is observed with non-targeting control. Consequences on cell viability were determined 48-96 hr post-incubation. Experiments involving chemical treatment involved 48 hr pretreatment with siRNA followed by chemical treatment for 72 hr at the indicated doses. CellTiter-Glo (promega) assays were performed using 15 \( \mu \)L regent/well followed by a 10 min incubation. Luminescence was quantified with an Envision plate reader (PerkinElmer). siRNA data for siATF4 and siPHGDH (Figures 6F and 6G) was curated from a prior study (Kim et al., 2016).

For immunoblot and qPCR analyses, a 6 well plate was prepared containing mixture of 250 \( \mu \)L siRNA (Dharmacon, 10 \( \mu \)M siRNA in 240 \( \mu \)L serum free media) and 250 \( \mu \)L RNAiMax (Invitrogen, 6 \( \mu \)L RNAiMax in 244 \( \mu \)L serum free media) per well, pre-incubated for 15 min at room temperature. Cells were then plated at a final concentration of 250,000 cells/well. After 48-96 hr of transfection, cells were lysed and subjected to immunoblot or qPCR analyses.

Stable PHGDH expressing cell lines were created by transducing HCC44 cells with the pLvx-Tight-Puro (Clontech) tetracycline-inducible vector containing the human PHGDH complementary DNA fragment (kindly provided by Matthew G. Vander Heiden)
(Locasale et al., 2011). Cell colonies were selected and maintained with 0.5 μg/mL of puromycin and 0.5 mg/mL of G418 sulfate. To induce PHGDH expression, cells were pretreated with 1 μg/mL doxycycline for 24 hr prior to SW157765 treatment.

To create stable HES1 overexpressing cells, H1993 cells were seeded at 3 × 10⁶ cells/well in 6-well plates 24 hr prior to transfection. The cells were transiently transfected with the 2ug of HES1-pCMV6-AC-GFP expressing plasmid using 8ul/well of Lipofectamine-2000 (Invitrogen) according to manufacturer's instruction. At 24hrs post-transfection, 5 μM hydrocortisone or EtOH vehicle was treated to the culture medium and incubated for 72hrs. Nocodazole (300ng/ml) or DMSO vehicle was added at 48hrs post-treatment of hydrocortisone. Nocodazole treated cells were used as positive control. For cell cycle analysis, the cells were trypsinized, centrifuged at 1200rpm and stained with the cell-permeable DNA dye Hoechst-33342 (10μg/ml, Invitrogen) for 30 min at 37°C. After incubation, the stained cells were washed and resuspended with cold PBS. The DNA content of GFP-positive or negative with Hoechst positive cells were determined using FACS with UV and 488 nm lasers (LSR Fortessa, BD FACSDiva software version 8.0.1, firmware version 1.4, BD biosciences). Data were analyzed using FlowJo 7.6.5.

**CRISPR-mediated cell line editing**

CRISPR knockout cells were prepared using the two-vector system (Sanjana et al., 2014). 293T cells were cultured to 90% confluence. A mixture of 0.4 μg transfer plasmid (lenti-cas9 blast or lenti-guide puro; Addgene), 0.87 μg psPax2 (Addgene), and 1 μg pMD2-VSV-G (Addgene) were diluted to a total of 50 μL in Opti-MEM media and added to a mixture of 21 μL FuGENE 6 (Promega) in 129 μL Opti-mem after a 10 min incubation period. The mixture was allowed to sit for 20 min after which it was added dropwise to 293T cells. Fresh RPMI 5% media was added 24 hr later and 48 hr post-transfection, target cells were transduced with virus. This processes was repeated and clones were selected in 10 μg/mL blasticydin. CAS9 expression was confirmed with western blots. CAS9 expressing cells were then transduced with lenti-guide puro constructs using the same protocol. Clones were selected in puromycin and knockouts were confirmed immunoblotting. sgRNA constructs were designed according to established protocol at http://CRISPR.mit.edu and cloned into the lenti-guide puro lentiviral expression vector. The sequences are as follows: CYP4F11 CACCGAAGCGCGGCAGTTGTCAT

**Immunoblot analysis**

Cells were plated in 6 well format for at a density of 150,000 cells/well and allowed to incubate overnight. Cells treated with 5 μM GC were allowed to incubate 72 hr prior to collection. Cells were either lysed in RIPA buffer (Sigma-Aldrich) with 1X protease inhibitor (GenDEPOT) and phosphatase inhibitor (Thermo Scientific) cocktails or in 50nM Tris (pH 6.8), 2% SDS and 10% glycerol. Total 10 μg of lysates were loaded and electrophoresed on 4%-15% gradient SDS-PAGE gel (Bio-Rad) and transferred to a PVDF membrane using the Trans-blot turbo transfer system (Bio-Rad). After blocking with 5% nonfat dry milk in PBST (1X PBS, 0.1% Tween-20), membranes were probed overnight with primary antibodies diluted at either 1:500 or 1:1000 at 4°C. After washing and incubation with secondary antibody, protein signals were visualized with the Enhanced Chemiluminescence Western Blot Detection Solution (Thermo Scientific) or Supersignal West Pico Chemiluminescence Western Blot Detection Solution (Thermo Scientific). Western blotting was performed using anti-GAPDH. Whole cell lystate loading controls were either GAPDH or β-actin. Nuclear loading controls were Lamin B1. Glut13 was used as a loading control for thermal stability shift assays. Antibodies were purchased as follows: β-actin, Kras, CYP4F11 and HRP-conjugated anti-mouse or rabbit IgG antibody (Santa Cruz Biotechnology), HES1, Cyclin D1, GR and PHGDH (Cell Signaling Technology), NRF2 (Invitrogen), β-tubulin, GLUT1, GLUT8, GLUT13 and Cas9 (Abcam).

**96 well dose response assays**

To determine cytotoxicity of the small molecule compounds, NSCLC cells and HBECs were plated at a densities ranging from 3,000 of 5,000 cells per well in white tissue-culture-treated 96-well clear bottom plate (Corning), with the seeding density for each cell line based on growth rate. After culturing the cells in assay plates for 24 hr, compounds were added to each plate at the indicated doses (3 replicates per dose per cell line). After an incubation of 96 hr, 15 μL of CellTiter-Glo reagent (Promega) was added to each well and mixed. Plates were incubated for 15 min at room temperature and luminescence was determined for each well using a SpectraMax Paradigm plate reader (Molecular devices).

**Flow cytometry analysis**

For DNA content analysis, cells were seeded at density of 1.5 × 10⁶ per well in 6-well plate and after 24 hr in cell culture, 3 μM hydrocortisone or DMSO vehicle was added to medium. Nocodazole at 100 ng/ml or DMSO was added 72 hr after cell seeding. Twenty-four hr post-nocodazole/DMSO treatment, cells were collected by trypsinization, resuspended in 1 mL of ice-cold PBS-F (1 x PBS, 2% FBS), followed by drop-wise addition of 10 mL ice-cold 70% ethanol. Following overnight incubation at 4°C, cells were washed twice with PBTA (1x PBS, 1% BSA, 0.1% Tween-20), stained with propidium iodide (Sigma) containing RNase A at 37°C for 30 min. Fluorescence of the PI-stained cells was measured using a FACSCalibur (BD Biosciences) and analyzed with FlowJo software (BD Bioscience).

**In vitro determination of compound stability with human tumor cells**

Cell lines were plated at a density of either 2000 (H2122, A549, HCC95, HCC44, H1792, H460, H322, HCC1171, H920, HCC2108, H226, H647, H2086, HCC4011) or 4000 (DFCi.032, HCC3051, H3255, H1395, H2073, H1437, HCC2814, HCC515, H596, H3122)
cells per well in 96 well plates. After overnight adherence, media was removed and replaced with fresh media containing either 100 nM (SW027951, SW098382, SW126788, SW153609, SW157765, and SW159580) or 200 nM (SW103675, SW115205, and SW167255) compound. Experiments using the CYP4A and 4F inhibitor HET0016 used 50 nM SW157765 in combination with 100 nM HET0016 added after overnight cell culture. siRNA experiments involved 48 hr pretreatment with siRNA’s targeting KRAS prior to compound addition. At varying times post compound addition, media and cells were removed using trypsin and the cells were broken open and the lysate preclearned of protein by the addition of a two-fold volume of methanol containing 0.2% formic acid, 2 mM NH₄ acetate and 100 ng/ml internal standard (IS = n-benzylbenzamide or tolbutamide) followed by vigorous vortexing and centrifugation at 16,000 x g for 5 min. In experiments involving the compound SW153609, proteins were pre-cleared by addition of a two-fold volume of methanol containing 2 mM NH₄ formate and 100 ng/mg of IS. The supernatant was analyzed by LC-MS/MS and centrifugation at 140,000 g for 10 min. As shown in the figure, the amount of compound lost in media only by 24 hr equivalent to that lost in the presence of both sensitive and resistant cell lines. For quantitation of parent compound and metabolites, H2122 cells were incubated with 10 μM SW157765 or 500-1 for 8 hr and then washed and harvested as described. Samples were analyzed by LC-MS/MS on a Sciex 4000QTRAP using LightSight Software for information dependent acquisition followed by an enhanced product ion scan (IDA-EPI). Transitions associated with Phase I metabolism were evaluated.

Glucose utilization
Glucose uptake was evaluated utilizing the Glucose Uptake Assay Kit (Abcam). Briefly, 6,000 cells were plated in 96-well plates in RPMI plus 5% FBS. Twenty-four hr later, cells were pretreated with either SW157765 (1 or 5 μM, final) or equal volume vehicle (ethanol) in RPMI plus 5% FBS for 6 hr. In experiments involving siRNA, GLUT8 or GLUT1 was transfected as described and allowed to incubate for 48 hr. Media was removed and cells were washed three times with DBPS. Afterward, 0.9 mM 2-deoxyglucose (2-DG) was prepared in glucose-free RPMI plus 5% FBS and then added to each well. Plates were returned to a 37°C incubator with 5% CO₂ for 2 hr. Afterward, media was removed, cells were washed with DPBS three times to remove exogenous 2-DG and detection of uptake was determined using manufacturer’s recommended protocol.

Analysis of 2-DG uptake was performed as follows: First, fluorometric values were calculated based on the 2-deoxyglucose-6-phosphate standard curve. Next, cell count and viability was determined by the CellTiter-Glo Luminescent Assay in a separate 96-well plate that was cultured and treated in parallel to the 2-DG treated plates. Reported relative fluorescent 2-DG uptake was calculated by normalizing the fluorescent values (i.e., 2-DG) to the luminescent values (cell number).

Binder selection assays
Protein preparation
In order to prepare human protein library, the CMV promoter-based vectors encoding a cDNA fragment of human gene (~14,000) was transiently transfected into FreeStyle293 cells (Invitrogen) using 293fectin (Invitrogen) according to the manufacturer’s instructions. After 3 days of incubation, cells were disrupted in a bead beater and the cell disruptions were frozen at –80°C until use.

To prepare the membrane-associated human GLUT8, the pcDNA3.1 vector (Invitrogen) encoding a cDNA fragment of human GLUT8 was transiently transfected into FreeStyle293 cells. After 3 days of incubation, cells were washed with phosphate-buffered saline (PBS) and suspended in buffer A, which contained 50 mM HEPES (pH 7.4), 1 mM EDTA. The cells were homogenized and centrifuged at 950 × g for 10 min at 4°C, after which the supernatant was recovered. Total membrane fractions were isolated by ultracentrifugation at 140,000 × g for 60 min at 4°C, and resuspended in buffer A. Membrane fractions were frozen at –80°C until use.
**Binder selection technology**

Arrayed affinity-selection/mass spectrometry binding assays were employed to identify binding proteins for test compounds using a human protein library with one protein per well in 96 well plate format. The binding assay was performed at 4°C using the human protein library or membranes. Proteins were dissolved in assay buffer (25 mM Tris–HCl (pH 7.4), 137 mM NaCl, 2.7 mM MgCl₂, 0.005% Tween-20, 1 mM DTT). Compounds were added to protein solutions (final DMSO concentration of less than 1%) and the mixtures were incubated for 2 hr. Nonspecific binding was defined using mock-transfected membranes. The reaction mixtures were applied to a size-exclusion chromatography to separate bound compounds from free compounds. Then the bound compounds were quantified by liquid chromatography-mass spectrometry (API5000 LC/MS/MS system) and the calculated Kd value was determined using Prism nonlinear software (Graph Pad Software, CA, USA).

**DNA/RNA extraction for sequencing**

Prior to sequencing, all cell lines were DNA-fingerprinted (PowerPlex 1.2 Kit; Promega) and found to be mycoplasma-free (e-MycO Kit; BocScietific). DNA for exome or genome sequencing was purified from frozen cell line pellets using DNeasy reagents and protocols with QIAcube robot (QIAGEN). DNA spectra were quantitated using spectrophotometer (Nanodrop) and samples diluted with nuclease free water (Ambion). Cell lines were grown to approximately 70%–80% confluence, washed 2X with PBS and directly lysed from culture flasks using RLT buffer (QIAGEN). Lysates were snap frozen and stored at –80°C. RNA was purified from lysates using RNeasy kit and QIAcube robot (QIAGEN).

**Genomic characterization**

**SNP Arrays**

Whole-genome single nucleotide polymorphism (SNP) array profiling was done using the Illumina Human1M-Duo DNA Analysis BeadChip (Illumina). Cell line DNA was hybridized according to manufacturer instructions. Processing was first performed using Illumina BeadStudio to generate the ‘Log R Ratio’ which measures the relative probe intensity compared with normal diploid controls. The package DNAcopy in the R statistical software environment was then used to segment the data. Final copy number variation was interpreted as the log₂ segmented copy number values.

**RNAseq and Whole Exome Sequencing**

FastQC (Babraham Bioinformatics Institute) was used to check the sequencing quality, and high-quality reads were mapped to human reference genome (hg19) along with the gene annotation data (genecode v19) from Genecode database using STAR (v2.4.2). RSeQC was applied for assessing RNA sample quality Gene-level expression was reported in fragments per kilobase per million reads (FPKM) by Cufflinks.

**Illumina BeadChip Microarray**

Raw Illumina HumanWG-6 v3.0 BeadChip files were obtained from the Gene Expression Omnibus using accession number GEO: GSE32026 and normalized as described previously (Kim et al., 2016). Briefly, Data were background-corrected using the ‘MBCB’ package in R, which provides a model-based background correction method similar to an RMA correction with affymetrix controls. Data were then quantity-normalized to produce equivalent expression distributions among cell lines.

**Germline variant filtering**

The UTSW-92 panel of the cell lines corresponded to those in which we have tumor DNA but corresponding matched non-tumorigenic DNA is not available. These correspond to 68 lines from the ‘training set’ of cell lines and 24 lines from the ‘testing set’ of cell lines. For these, we developed a pipeline to filter out the most probable germline mutations and enrich for somatically acquired mutations. Reads were aligned as described to the hg19 reference and filtered for non-synonymous lesions (missense, non-sense, splice site mutations) (mean of 5,049 mutations/cell). We next removed any site that was annotated as corresponding to a germline mutation in the matched dataset (mean of 1,248 mutations/cell). Using publically available datasets such as the thousand genome project (TGP) as an exclusion criteria or the catalog of somatic mutations in cancer (COSMIC) as an inclusion criteria may aid in enriching for somatic mutations. We removed variants (defined based on genomic position) that were found in > 12% of the TGP (TGP filter) and where the difference in the UTSW panel frequency and the TGP frequency was < 1.8% (allele difference filter). We also removed, on a gene-level basis, genes that were highly mutated (mutated at any site in > 40% of cell lines) in the UTSW panel (mutation any site filter), but present at a low frequency (< 13%) in COSMIC (Cosmic filter) and in the UTSW-34 matched panel (< 20%) (UTSW-34 filter). This resulted in a final mean mutation count of 718 mutations/cell. We developed a strategy to find a data driven way select optimal filter cutoffs from these datasets. We selected 12 evenly distributed values for the TGP filter between 0.02% and 20%, for the allele difference filter between –10% and 10%, for the mutated any-site filter between 1.8% and 80%, for the Cosmic filter between 0.13% and 20%, (log₁₀ scale), and for the UTSW-34 filter between 2.9% and 50%. Selecting all possible combinations of these filters resulted in 248,832 possible combinations. For each filter combination, we can plot the number of mutations that pass the filters (Figure S1C), with the strictest filter combination resulting in the fewest variant being annotated as ‘somatic’ and the most lenient resulting in the most variants being included. To select the optimal filter combination in a data-driven way, we fit a cubic function to the plot of filter index (x values) versus number of mutations included at each filter index (y axis) and selected...
the value on the plot which results in the minimized second derivative for each cell line. Figure S1C indicates the mean selected filter value across the cell line panel (solid line) with 95% confidence intervals indicated (dashed line). Resulting variants are described in Data S2.

**Small molecule cytotoxicity assays**

The UTSW chemical library and screening assay format was described previously (Kim et al., 2013) and is described above. Our chemical library, consisting of ~200,000 chemicals (Figures S1B and S1G), was initially screened at a single dose (2.5 μM) in single well for each compound against a panel of 12 NSCLC cell lines. Toxicity data were converted to an activity score according to the following equation

\[
AC = -1 \times \left(100 - \frac{x}{\text{median}(x_{\text{control}})} \times 100\right)
\]

so that an activity score indicates percent kill relative to on-board DMSO controls. We subsequently converted activity scores to z-scores for each chemical across the 12 cell line panel and selected chemicals with \(z \leq -3\) in at least one cell line, resulting in 15,483 chemicals (single dose cohort). These chemicals were then re-screened in triplicate against the same 12 NSCLC cell lines along with as an immortalized human bronchial epithelial cell line (HBEC30KT) at the screening dose of 2.5 μM (confirmation dataset). From this dataset, we used two criteria to select chemicals for further follow-up. We first filtered for chemicals with a bimodal pattern of response from our panel of cell lines. Specifically, we selected chemicals with > 40% toxicity to a subset of cell lines and < 20% toxicity to the remaining NSCLC’s and HBEC30KT. As determined in downstream dose-response studies, compounds that met this criteria typically displayed IC₅₀’s in the range of our screening dose or lower for a subset of the NSCLC lines and IC₅₀ values > 10 μM in the remaining cell lines in the panel and the HBEC30KT cell lines. In terms of chemical selectivity, we expect this selection to result in compounds with at least a ½ log difference in response between sensitive and resistant cell lines. We also used a selection method to capture potent chemicals with more of a continuous distribution of cytotoxicity in our 12 cell line panel. For each compound, the responses of the cell lines were ranked from most sensitive to least. The difference (Δn) in response between each pair of ranked cell line activities for each compound was calculated. The S-score is the maximum difference (Δnmax) between two cell lines’ responses in the ranked list of responses to the compound. The two cell line responses that define the S-score therefore demarcate a boundary between sensitive and resistant response groups in the ranked list of responses for each compound. We selected chemicals for follow-up to be those with the S-score > 40%, while enforcing the criteria that the chemical not be toxic to HBEC30KT (≤ 20% observed toxicity). These chemicals were subjected to chemistry review that removed compounds with known or suspected promiscuous (off target) behavior based on historical screening data, structural alerts, and PAINS substructures. Following resupply (1 – 5 mg of powder per compound) and analytical quality control for identity and purity (LC/MS), 447 compounds were assayed in a multi-dose format (12 point dose-response curves in ½ log dilutions with the doses ranging from 50 pM to 50 μM) against the same panel of 12 NSCLC cell lines plus the HBEC30KT cell line. Each compound was assayed twice in this format and the dose-response curves compared. In cases where experimental replicates differed by more than 3-fold, we performed a third dose-response experiment and averaged the two experimental replicates that were in closest agreement. We used the same unimodal (S-score) method to select a total of 208 chemicals to be screened across the entire panel of 100 cell lines. In this case, we ranked-ordered average log₁₀(IC₅₀) values for each compound and applied a threshold of 0.5 log units for the S-score. Chemical structures for all analyzed chemicals are available at https://pops.biohpc.swmed.edu. Smile strings are described in Data S6.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

**In-silico Molecular Docking**

The GLUT8 structural model was developed from the inward, open crystal structure of human GLUT1 (hGLUT1, PDBID 5EQG). The hGLUT1 structure has 48% sequence similarity to GLUT8, and all except one of ligand binding residues are conserved between the two proteins. The sequence alignment used for this comparison was based on Long and Cheeseman (2015) and Deng et al. (2015) to ensure correct alignment of sequentially conserved residues. For unresolved loop region residues, the GLUT3 crystal structure (4ZW9) was used as template (Deng et al., 2015). The model GLUT8 structure was evaluated for proper side-chain orientation versus hGLUT1 crystal structure and was then optimized for docking using Protein Preparation Wizard. This included addition of hydrogens and stepwise energy minimization using OPLS3 force field. Sitemap was used to predict ligand-binding sites on GLUT8. A grid box was defined using the top ranked site that overlaps with glucose binding site and inhibitor binding residues. Conformations of SW157765 and the epoxy metabolite were generated in MOE2016.08 (Chemical Computing Group, QC, Canada). Multiple conformations were docked against glucose-binding site of the GLUT8 model using an SP scoring function in Glide version 6.5 (Schrodinger, LLC). In these docking experiments, the protein was treated as rigid and the liarvargand was allowed to be flexible. The docking results were analyzed visually using MOE2016.08 and Maestro (Schrödinger, LLC, New York, NY). The top-ranked docked pose exhibited H-bond interactions with Asn312 and van der Waals interaction with Trp433 and Trp457 consistent with co-crystal structures (Deng et al., 2015; Kapoor et al., 2016). SW157765 and epoxide metabolite exhibited similar docked poses expect for the double bond/epoxide region.
Normalization of drug response data and calculation of ED$_{50}$ and AUC values

Chemical response for each cell line was converted to an activity score as described above. We found normalizing to the median of the two lowest doses, as opposed to on-board DMSO controls, minimized plate effects and resulted in a better curve fit and a more accurate description of sensitivity. We used the drc package in R to fit a standard 4 parameter log-logistic fit to the data and discover ED$_{50}$ values. As imputed ED$_{50}$ values have shown to be problematic in re-tests of large drug screening datasets, we do not impute values. Rather, if the imputed ED$_{50}$ value is greater than the top tested dose (50 μM), we assign an ED$_{50}$ of the top dose. Additionally, to correct for low ED$_{50}$ values being assigned to chemicals in which the response is shallow, we assign an ED$_{50}$ value of the top tested dose if the chemical does not result in at least 30% reduction in CTG values.

We calculated AUC values by determining area under the curve of the log fitted hill equation through standard integral analysis. For many of the compounds, a large proportion of the dose range is completely innocuous for all cell lines tested. To increase the dynamic range with AUC values, we found for each compound the proportion of the curve in which there is a response across all cell lines, and eliminated the data for the lower doses. Each compound was tested in 2 separate runs, with three replicates per run. To automatically detect the best, most reproducible data, we select the two replicates from each run with the best concordance between calculated ED$_{50}$ values. The assigned ED$_{50}$ or AUC value is then the mean between the filtered runs.

Scanning Kolmogorov-Smirnov statistic

A modification to a Kolmogorov-Smirnov statistic, which we term a scanning ranked KS test, was used to determine which mutations alone or co-occurring combinations of mutation combinations can best predict a selective sensitivity to each unknown compound. In addition to single mutations, we also annotated a ‘RAS_Class’ metaclass in which we assigned a cell line a value of ‘1’ if it contained a KRAS, NRAS, HRAS, PIK3C1, or BRAF mutation or pairwise combinations of co-occurring mutations were first binarized (1 = mutated; 0 = wild-type), resulting in 446,435 combinations in which at least 5 cell lines contained the mutation combination. For each chemical, we reasoned that if a mutation combination is conferring a selective sensitivity, then the ED$_{50}$ or AUC values for cell lines that are mutated will be lower than those that are wild-type. To determine the degree to which the ED$_{50}$/AUC values for cells that are mutated are located toward the bottom of the ranked list sensitivity values, and thus lower than the background distribution, the following equation was used:

$$u = \max_{j=1}^{n} \left[ \frac{1}{n} \sum_{t=1}^{n} \left( \frac{M_{jt} - V(j)}{n} \right) \right]$$

where $v(j)$ is the position of each gene in the gene set in the ordered list of genes, $t$ is the total number of cell lines with the mutation combination, and $n$ is the total number of cell lines assayed ($n = 100$).

To determine a p value, 5000 permutations of randomized sorting of ED$_{50}$/AUC values of size $t$ was performed, and $u_{\text{random}}$ was calculated. The resulting p value was determined to be:

$$p = \frac{\# \text{ instances } u_{\text{random}} > u}{\# \text{ total permutation}}$$

p < 0.002 indicates that, out of 5000 permutations, no random value was less than the calculated distance, $u$. This process was repeated for each of the mutation combinations for each chemical using both AUC and ED$_{50}$ values as a sensitivity metric.

Elastic Net Regression

In order to assign predictive biomarkers to each chemical, we used a penalized linear regression model, the elastic net. We considered each dataset individually and separately as input into the elastic net. Candidate predictive features were selected from normalized measures of gene expression (illumina V3 BeadChip, RNaseq), copy number (Snp 6.0 arrays), protein expression (RPPA), metabolomics flux analysis and binary measures of gene mutational statuses (Whole Exome Sequencing). mRNA measures are likely to be the most variable across laboratories are subject to fluctuations from an in-vitro to an in-vivo system, and are the most difficult to assess in patient samples. Thus, mRNA measures were only considered inputs into the elastic net if there was at least a two-fold difference between lowest and highest expressing cells, and were prioritized for experimental follow-up based on maximal fold expression differences. The elastic net assigns biomarkers to a response vector of activity scores by solving a basic linear regression problem as follows:

Let $X \in \mathbb{R}^{n \times p}$ be the matrix of predictive features where $n$ is the number of cell lines included in the training dataset and $p$ is the number of features, and let $y \in \mathbb{R}^{n}$ be the vector of binary sensitivity values for the same cell line panel. Columns of the predictive features matrix and $y$ were normalized to have a mean of zero and a standard deviation of 1. The elastic net attempts to find which weighted linear combination of the columns of the predictive features matrix can best approximate $y$, or it solves the following equation for $w$:

$$\arg \min_{w} \left[ y - Xw \right]^{2}$$
The elastic net solves the above by enforcing a penalty to the solution that makes the solution both unique and sparse so that only the features that best approximate $y$ are left with non-zero weight values. It does this by combining L1-norm and L2-norm regularization parameters so that the elastic net formulation to the above problem is given by:

$$\text{argmin}_w \left[ \| y - Xw \|^2_2 + \lambda \left( \alpha \| w \|^2_1 + (1 - \alpha) \| w \|^2_2 \right) \right]$$

where $\lambda$, $\alpha$, are two adjustable parameters such that lambda controls the degree of the overall penalty and $\alpha$ controls the degree to which the L1 norm and L2 norm constraints are applied so that when $\alpha = 0$, only the L1 penalty is applied and when $\alpha = 1$, only the L2 penalty is applied. In order to determine the optimal values of alpha and lambda to use in the model, we did 100 iterations of 10-fold cross-validation where, in each iteration, the cells were randomly re-sampled into different groups. The values of alpha and lambda were chosen to be those that resulted in the minimum mean squared error for each fold. We then subjected the data to a series of 100 bootstrap permutations in which the cell lines were sampled with replacement, and features were assigned to each bootstrapped dataset. Features with a higher frequency of representation across bootstrapped permutations correspond to those with the highest confidence of being associated with drug response as opposed as being assigned due to overfitting of the model. Features were assigned to a chemical that were present in >70% of permutations. Higher weighted features in the elastic net correspond to those with better mathematical predictive capacity and lowly weighted features may be assigned due to model overfitting. Cutoff values were assigned for weight values to select the most predictive features while still maintaining sparcity in the model. As weight magnitudes and distributions vary depending on input feature set and chemical sensitivity distributions, cutoffs were selected to be tailored to each chemical/feature set combination by selecting weight values ± 2 standard deviations from the mean. Weight distributions follow a Gaussian distribution, so this cutoff contributes to the top 5% of the most predictive features being assigned to each chemical.

As input into the elastic net, we used both AUC and ED$_{50}$ values as measures of sensitivities. Additionally, we found that log$_{10}$ transformation of the sensitivity vector could better identify exceptional responders to a compound in some instances, thus we used both the linear and log transformed sensitivity measure as input to the algorithm. The elastic net was run using the glmnet package in R.

**Cell line pathway activity analysis**

To calculate pathways that were downregulated relative to the background distribution on an individual cell line basis, we used a modification of a Kolmogorov Smirnov test. Gene pathways were curated from the Broad msigdb. We first converted RNaseq data to z-scores with the following equation

$$z_{i,j} = \frac{x_{i,j} - \text{mean}(x_i)}{\text{sd}(x_i)}$$

where $z_{i,j}$ is the z-score for gene $i$ in cell line $j$. Then for a cell line, we converted z-scores to a ranked list, where a value of 1 indicates the highest z-score in that cell line.

For a pathway, to determine the degree to which the values in a set are located toward the top of a ranked list, and thus upregulated relative to background, the following equation was used:

$$u = \max_{j=1}^{T} \left[ \frac{1}{T} - \frac{V(j)}{n} \right]$$

and to determine the degree to which a set is downregulated relative to background, the following equation was used:

$$u = \max_{j=1}^{T} \left[ \frac{V(j)}{n} - \frac{(j-1)}{T} \right]$$

where $v(j)$ is the position of each gene in the gene set in the ordered list of genes, $T$ is the total number of genes in the gene set, and $n$ is the total number of genes assayed in the array.

To determine a p value, 5000 permutations of randomized sorting of genes of the same set size was performed, and $u_{\text{random}}$ was calculated. The resulting p value was determined to be:

$$p = \frac{\# \text{ instances } u_{\text{random}} > u}{\# \text{ total permutation}}$$

**Gene Set Enrichment Analysis**

Cell lines were dichotomized based on sensitivities to SW140154 and SW151511. We selected cell lines that were sensitive (ED$_{50}$ < 10 µM) to SW151511 and resistant to SW140154 (ED$_{50}$ > 20 µM) and compared them to cells resistant to SW151511 (ED$_{50}$ > 20 µM) and sensitive to SW140154 (ED$_{50}$ < 10 µM) using a GSEA analysis (Subramanian et al., 2005). For SW036310, we compared sensitive (ED$_{50}$ < 2 µM), TTC218 mutant cell lines to resistant lines (ED$_{50}$ > 40 µM) with a GSEA analysis. Enrichment plots for the top gene sets were re-plotted using R statistical software.
**Sensitivity Prediction**

Sensitivity outside the training set was predicted as described previously (Kim et al., 2013). Twenty-six NSCLC cell lines not included in the original training set were subjected to RNaseq and high throughput 384 well format cytotoxicity assays as described above in response to 21 chemicals in which at least one cell line is predicted to be sensitive. Log2 transformed FPKM values were converted to z-scores, and sensitivities were predicted according to the following equation

$$s_i = \sum_{i=1}^{n} w_i x_{ij}$$

where $w_i$ is the weight determined from the elastic net for feature $i$, and $x_{ij}$ is the normalized expression value of feature $i$ in line $j$ and $n$ is the number of features selected for a chemical as described above. The range of $s_i$ values predicts the degree of sensitivity where a high value of $s_i$ predicts resistant and a low value of $s_i$ predicts sensitive. Sensitivity to SW001286 and SW126788 was predicted based on RNaseq based expression of CYP4F11 and CES1/CES1P1, respectively. Sensitivity to SW151511 and SW140154 was predicted based on expression of PEL12 and SARM1/IL18R1. Additionally, we predicted activity of the KEGG TLR Pathway on a single-cell line basis using the protocol described above. Cells predicted to be sensitive to SW151511 and resistant to SW140154 were confirmed to have high TLR pathway expression levels.

To assess predictions, a two sample, one-sided KS test was used to determine if activity scores across the 12-point dose range in the predicted sensitive classes were significantly lower than those in the predicted resistant classes.

**ROC Curve Analysis**

For each feature set, we associated biomarkers to a sensitivity vector and predicted sensitivities on the original training panel according to the procedures outlined above. We used a cutoff of $s_j = 0$ to binarize cell lines into predicted sensitive and resistant classes. For each chemical in our dataset, we manually selected ED$_{50}$ and AUC values above which a cell line is considered resistant and below which a cell line is considered sensitive. The ROCR package in R was then used to calculate specificity (100 – false positive rate) and sensitivity (true positive rate) and plot the values. As input to the ROCR package, ‘true positives’ were considered to be those whose predictions were correct (sensitive cells predicted to be sensitive and resistant cells predicted to be resistant). Area under the ROC curve was calculated with and $p$ value was calculated to test the hypothesis that the area under the ROC curve is different from 0.5 (random) using the ROC R package.

**Affinity propagation clustering**

Affinity propagation clustering was performed as described (Kim et al., 2016) using Pearson distance as a similarity metric. Cell lines that were present in the intersection between the POPs and each dataset were clustered for similarity. Cell lines in the RPPA dataset (65 cell lines) were clustered according to 154 unique features, in the metabolomic flux (67 cell lines) analysis according to 84 unique features, and in the chemical perturbagen dataset (100 cell lines) according to 221 features. We first filtered features in the illumina BeadChip (90 cell lines), RNaseq (99 cell lines), and SNP 6.0 arrays (63 cell lines) by selecting the top 20% of the most highly variant features. RNaseq and illumina BeadChip features were further reduced by selecting features that were above a minimal expression cutoff in at least one cell line (RNaseq FPKM value of 1 and illumina BeadChip value of 6), resulting in 5075 and 5047 features, respectively. The SNP array was clustered according to a final set of 3583 features. Networks were visualized with cytoscape with edges defined according to the procedure above and edge lengths drawn proportional to pearson distance using the built-in spring embedding algorithm.

**NRF2 Signature**

We curated publicly available datasets to identify genes with NRF2 binding sites through ChiP-seq analysis (Chorley et al., 2012; Hirotsu et al., 2012; Malhotra et al., 2010). To identify a context-independent set of NRF2 regulated genes, we selected genes that were found to have NRF2 binding sites in all three datasets. Recent work has also annotated a non-small cell lung cancer specific set of genes that are upregulated in cells with gain of function mutations in the NRF2 pathway (Goldstein et al., 2016). We also included these genes in our NRF2 signature, resulting in a total of 40 genes. Cell lines in our panel were binarized into two categories. SW157765 sensitive cell lines were defined to have an AUC < 400 and an ED$_{50}$ value < 1 μM while resistant cells had an ED$_{50}$ > 30 μM. A KS test was used to determine if sensitive cells had significantly higher expression of NRF2 signature genes (two sample, one sided KS test) using the R stats package.

**Comparison of cell lines to tumors**

LUAD and LUSC RNaseq V2 RSEM normalized expression values were downloaded from the TCGA (https://cancergenome.nih.gov/) (519 LUAD tumors and 504 LUSC tumors). For each dataset, we selected the top 20% of the most highly variant genes. 509 genes represented the intersection between all three datasets. RNaseq V2 RSEM gene expression values of the 509 genes in BRCA tumors (1100 tumors) and MESO tumors (87 tumors) was downloaded using the CGDSR package in R. Though the gene signature was defined using a NSCLC dataset, every gene in the signature is expressed in at least one tumor in the MESO and BRCA dataset. Using
the 509 genes, we used a Pearson correlation to compare each tumor in the TCGA with each cell line in our dataset with the stats package in R. p values for the correlation are plotted (Figures 1A and S1A).

**Peg plots**
TCGA mutation data for the LUAD and LUSC subtypes was retrieved using the cgdsr package in R. Somatic mutations characterized as either ‘missense’ or ‘nonsense’ were plotted according to amino acid position. Non-synonymous mutations in the UTSW cell line panel for the same gene were plotted on the same scale. Domain information was obtained from the PFAM database from the following website http://pfam.xfam.org/.

**Other Statistical Analyses**
Hierarchical clustering, Pearson correlations, two sample t tests, and density calculations were performed using the stats package in R.

**DATA AND SOFTWARE AVAILABILITY**
The accession number for the microarray data in response to SW151511 reported in this paper is GEO: GSE104757. The study ID number for the raw sequencing data reported in this paper is available at dbGap: 28061.
Supplemental Figures

A. Cell alterations with hg19 reference (mean = 15,226 muts/cell)

B. 1. Remove synonymous alterations (mean = 5,049 muts/cell)

B. 2. Remove germline alterations found in matched tumor/normal (mean = 1,248 muts/cell)

B. 4. Remove alterations found in > 12% of TCGA and where (allele frequency - TCGA frequency) < 1.8% (mean = 811 muts/cell)

B. 5. Remove genes mutated (any site) > 40% AND mutated < 13% in COSMIC AND mutated < 20% in UTSW matched pairs (TAN) (mean = 718 muts/cell)

C. (continued on next page)
Figure S1. (continued).

(legend on next page)
A) Cell lines were compared to tumors through similarity of a gene expression signature, defined to be the set of the 509 most highly variant (top 20% of variance values) genes in the cell line and tumor datasets. Boxplot comparing the pairwise p values (Pearson Correlations) between lung cancer cell lines and lung cancer tumors (blue) and between lung cancer cell lines and breast cancer tumors (pink) corresponding to Figure 1A. p = 0.05 is indicated as a dashed red line.

B) For the 92 cell lines with no corresponding matched normal DNA, the series of filters shown were applied in a stepwise fashion to remove probable germline variants and retain the most likely somatic mutations. The mean number of remaining mutations per cell line after each filtering step is indicated in each box, with a final mean estimated somatic mutation burden of 718 mutations per cell line. Optimal filter values were defined according to the protocol described in (C) and described in detail in the STAR Methods. TGP = thousand genome project; COSMIC = catalog of somatic mutations in cancer.

C) A data-driven metric was applied to discover the optimal filter cutoff values applied in boxes 4-5 in (B). 12 evenly distributed filter values were selected between pre-defined ranges (.02% - 20%) for the TGP filter (B, box 4), for the allele difference filter (B, box4; allele frequency – TGP frequency) (~10% - 10%), for the mutated (any site) filter (B, box5; 1.8% - 80%), for the cosmic filter (B, box5; 0.13% - 20%) and for the UTSW matched pair filter (B, box5; 2.9% - 50%). Selecting all possible combinations of these filter values resulted in a total of 248,832 filter combination values. For each filter value, the number of mutations that pass each filter is plotted. Each cell line in the unmatched dataset is plotted as a black line. A cubic function was fit to each black curve, and the optimal filter value for each cell line was selected to be the value where the second derivative is minimized. An overall filter value was selected to be the mean across the cell lines (solid red line). The red dashed line indicates the selected filter cutoff with 95% confidence range indicated as the dashed lines.

D) Pearson correlations were calculated based on similarity of gene signature expression values of the same panel of cell lines assessed by an Illumina V3 BeadArray dataset and an RNaq dataset. Gene signatures were defined to be the set of expressed genes (illumina expression value ≥ 3 and RNaq FPKM > 1) in at least one cell line that are among the most highly variant (top 20%). UPQMA of the R values are shown, where the diagonal indicates cell line self-similarity between both datasets.

E) APC of NSCLC cell lines clustered according to similarity of a RNaseq derived gene expression. Clusters are drawn with cytoscape with edges proportional to Pearson distances. Nodes are colored according to APC-defined cluster membership. The 12 cell lines screened with the entire 200,000 compound library are highlighted in green.

F) UTSW screening approach. The entire 200K (G) chemical library was screened at a single dose (2.5 μM) in singlicate across a panel of 12 cell lines defined to be representative of overall phenotypic diversity. (E), 15,000 molecules with variable response profiles were re-screened in triplicate at 2.5 μM. 900 chemicals with reasonable bi-modal (H) or 317 chemicals with unimodal (I) response patterns were selected and filtered by chemistry review. Fresh material was resupplied and subjected to analytical quality control and purity (LC/MS). 447 chemicals were re-assayed in a multi-dose format (12 point dose responses) against 12 cell lines in duplicate. Variable response profiles were selected, resulting in 208 chemicals that were screened together with 14 ‘cherry picked’ chemicals with known mechanism across the 100 cell line panel using 12 doses (1/2 log dilutions from 50 pM to 50 μM) in triplicate, twice.

(G) The UTSW chemical library consists of 202,103 chemicals composed of 450 chemicals from the NIH clinical library, 1,120 from Prestwick, 941 from TimTek, 2,500 from the UTSW proprietary library, 21,668 from ComGenex, 75,424 from ChemBridge, and 100,000 from ChemDiv labs.

(H) Density distribution of the ED$_{50}$ values in the 100 cell line panel of a chemical that was selected with (G) the bimodal selection method and (H) a chemical selected with the unimodal selection method.

(I) For each chemical, a Pearson correlation was calculated to represent similarity between calculated ED$_{50}$ and AUC responses across the 100 cell line panel. Red dashes indicate 14 manually curated chemicals with known mechanisms of action.

(K) APC of 222 chemicals clustered based on ED$_{50}$ values across the 100 cell line panel. Clusters are drawn with cytoscape with edges proportional to Pearson distances. Nodes are colored according to APC-defined cluster membership.

(U) Lollipop plot of $EGFR$ mutations within the 100 cell line panel as compared to tumor samples in TCGA LUAD and LUSC. Mutations are ordered based on annotated amino acid position along protein length. Top panel indicates the frequency of non-synonymous mutations found in TCGA LUAD’s (blue) and LUSC’s (red); bottom panel the frequency in the UTSW cell line panel (blue = erlotinib sensitive cell line; orange = erlotinib resistant cell line). The domains in the $EGFR$ protein as annotated in PFAM are diagrammed below.

(V) Comparison with perturbations shared with (iorio et al., 2016) and (Seashore-Ludlow et al., 2015), To the protein level, we found limited commonality among the compounds evaluated in both studies. Out of 676 chemicals in the union between Iorio et al. (2016) and Seashore-Ludlow et al. (2015), 105 were also present in our primary screen. The majority of these chemicals did not pass our initial filtering criteria for selectivity of cellular response and only 26 chemicals made it to the confirmation cohort (F). None of these passed our unimodal or bi-modal filtering criteria required for selection to be tested across the full panel of 100 NSCLC cell lines. This was mainly a consequence of either broad toxicity or no detectable activity at 2.5 μM. Among the compounds with known mechanism of action that we inserted in the dose-response studies across the 100 cell lines (quality control “cherry-pick” set), we found that 6 chemicals (etoposide, paclitaxel, erlotinib, Bortezomib, and Crizotinib) that were in common among all three studies. Bafilomycin-A1 and DeBQ were also present in the intersection between POPs and Seashore-Ludlow et al. (2015). When comparing AUC’s calculated in Seashore-Ludlow et al. (2015) to AUC’s from this study, among NSCLC cell lines present in both studies, we found that 6/8 chemicals are significantly correlated across 66 cell lines (crizotinib, erlotinib, dasatinib, bortezomib, etoposide, paclitaxel). The Iorio et al. (2016) study employed an I$_{50}$ metric. When comparing our ED$_{50}$ dataset to I$_{50}$ values calculated in Iorio et al. (2016), we find that 1/8 chemicals (etoposide, Pearson p < 0.05) correlated across 60 cell lines. Dashed red line indicates p = 0.05.
Figure S2. (continued).
Figure S2. (continued).
Figure S2. (continued).
Figure S2. Detecting “Prodrugs” and Drug Efflux Substrates, Related to Figure 2

(A) Elastic net modeling associated high mRNA expression of one of 4 known drug metabolism enzymes with sensitivity to 4 chemicals. Cell lines are ranked by ED_{50} response indicated as a heatmap (top rows, unbolded) with log2 FPKM values for the same cell lines plotted as a heatmap underneath (bolded). A legend to interpret the color scale is plotted to the right.

(B) Unsupervised hierarchical clustering of ED_{50} responses of 12 chemicals which correlate with high expression of a drug metabolism enzyme.

(C and D) (C) Structures and (D) chemical sensitivity profiles (ED_{50} values) of prodrugs in 100 cell lines. Cell lines are ordered according to average ED_{50} value. Error bars are plotted as ±1 SD from the mean ED_{50} across (n = 6/cell line).

(E–J) Percent remaining (ln scale) of (E) SW115205 (F) SW098382 (G) SW153609 (H) SW167255 (I) SW134963 and (J) SW147739 is plotted as a function of time of treatment with each compound (sensitive = blue; resistant = orange). Values are normalized to control treatment. In all cases, concentration of the parent compound is followed over time. In some cases, selective sensitivity corresponds to selective loss of the parent compound, which may indicate conversion of the parent molecule into a toxic compound in sensitive cells (i.e., SW098382). In other cases, there is equivalent loss of the parent compound in both sensitive and resistant cell lines, which may indicate selective sensitivity to a commonly produced metabolite, or due to production of distinct metabolites with distinct toxicity profiles (i.e., SW153609).

(K and L) Clones were selected after gene editing of the CYP4F11 locus and analyzed by immunoblot. Parental protein expression of CYP4F11 (lane 1) was compared to that of the clones, and the clones resulting in the lowest protein expression were selected for further follow-up. (K) Clone 8 was selected for H647 and (L) Clone 2 was selected for H460.

(M) H647 cells with CYP4F11 edited out of the genome with CRISPR and H647 cells transfected with a control sgRNA were treated for 72 hr with the concentrations of SW157765 shown and cell viability relative to a DMSO control measured.

(N and O) Dose response curves of cell lines outside the training set (n = 26) predicted to be sensitive (blue) and resistant (orange) to (N) SW126788 and (O) SW001286.

(P) An empirical CDF plot comparing SW001286 sensitivity of cell lines mutant for LKB1 (red) compared to wild-type (gray) (p = 0.0055, Scanning-KS test)

(Q) Cells were co-treated with 2 mg/mL NAC and either 5 mM (HCC44) or 1 mM (A549) SW001286. Viability was measured 72 hr post-treatment and values were normalized to the DMSO condition.

(R) CYP4F11, as a member of the p450 cytochrome class of enzymes, likely catalyzes a variety of reactions in biological settings. Considering that p450 oxidation of internal olefins has been previously characterized, we considered that the alkene-containing small molecules SW115205, SW157765, and SW157692 might be susceptible to this type of enzymatic activity. As such, studies were undertaken with SW157765 to characterize the metabolic products.

(S) We reasoned that if the small molecule acted as a covalent modifier of its target protein, both transient and prolonged exposure would elicit similar cellular responses as covalent modification would result in sustained protein perturbation even after the treatment media was removed. Groups of sensitive (orange) and resistant (blue) cell lines were treated with 5 μM SW157765 or DMSO for either (bottom) a single 72-hour continuous dose or (top) a 24 hour continuous treatment followed by a 48-hour incubation with compound free media.

(T) Comparison of structures of SW157765 and inactive analog, 500-1.

(U) 500-1 retains no activity when retested alongside SW157765 in a panel of 12 cell lines.

(V–Y) H2122 cells were pre-treated with SW157765 or the inactive analog, 500-1, for 8 hours prior to analysis on a Sciex 4000QTRAP system. The fragmentation sites monitored in the experiment for SW157765 (V) and 500-1 (X) are indicated. In (W) and (Y) The larger total ion chromatogram (TIC) on the left depicts the summation of all new multiple reaction monitoring (MRM) peaks for SW157765 (W) or 500-1 (Y) identified by LightSight software (AB Sciex) in the time 0 samples (gray) versus 8hr (red). Arrows indicate peaks (transitions) with higher abundance in the 8hr samples. The M3 peak is contained within the shoulder of M2. The smaller MS/MS plots show the daughter ion fragmentation for each of the potential metabolites shown in the larger chromatogram. Hypothesized metabolite structures are indicated. All potential oxidation sites appear to lie on the right hand side of the molecule because the daughter ion masses are unchanged. Based on similar fragmentation patterns for SW157765 M2 and 500-1 M1, the oxidation site for these metabolites was hypothesized to be on the 5-member furan ring. The fragmentation pattern for SW157765 M1, however, is unique, and was therefore hypothesized to occur on the internal alkene.

(V–X) Predicted fragmentation sites and resulting daughter ions of (V) SW157765 and (X) 500-1.

(Z) Predicted metabolite of SW157765.
Figure S3. Glucocorticoid Sensitivity Is Predicted by NOTCH2 Mutations, Related to Figure 3

(A) Location and type of NOTCH2 mutations in NSCLC cell lines is shown by lollipop plot. Mutations are ordered based on annotated amino acid position along protein length. Height of symbols indicate frequency of non-synonymous mutations found in the UTSW cell line panel (blue = GC sensitive cell line). PFAM annotated domains in the NOTCH2 protein are shown below. Missense mutations are indicated as a black solid line and nonsense are indicated as a red solid line.

(B) Effect of individual siRNA oligonucleotides on protein expression of NR3C1 compared to H2073 cells treated with a non-targeting control (NC) 72 hr post-treatment with 5 μM hydrocortisone was measured by immunoblot.

(C) Changes in HES1 protein levels in nuclear (N) and cytosolic (C) fractions 72 hr after treatment with hydrocortisone was measure by immunoblot. β-actin and Lamin B1 serve as loading controls for the total lysate and nuclear fractions, respectively.

(D) Flow cytometric histograms for DNA content after 3-day exposure to hydrocortisone (3 μM) or DMSO measured in GC responsive and non-responsive cell lines. Nocodazole (100 ng/mL) was added 48 hr post-treatment to force accumulation of proliferating cells in G2/M over the course of the next 24 hr.

(legend continued on next page)
(E) Transient overexpression of HES1 or empty vector control cDNA in 3 GC responsive cell lines. Protein expression levels are shown 72 hr post-hydrocortisone treatment (5 μM).

(F) DNA content for the indicated populations in Figure 3J is measured by flow cytometry of cells stained with propidium iodide.

(G) Proposed mechanism for GC induced cell cycle arrest in lung cancer. Mutations in NOTCH2 are loss of function resulting in lower notch signaling and higher basal NR3C1 expression, priming cells to respond to the GC signal. Upon GC stimulation, sensitive cell lines will initiate a positive feedback loop to upregulate NR3C1 expression and downregulate HES1 expression. One consequence of higher GC signaling will be a selective reduction of CyclinD1, leading to G1 arrest.
Figure S4. Biological Diversity among Robust Chemical/Genetic Associations, Related to Figure 4

(A) A one-sided KS test was used to calculate p values assessing ability of the associated elastic net models to predict sensitivities in a test panel of cell lines (Figure 4A) and to assess preservation of chemical sensitivities in 3D spheroid models (4B).

(legend continued on next page)
(B) GSEA calculated p values for top gene sets predicted to be upregulated in TTC21B mutant, SW036310 sensitive cell lines compared to SW036310 resistant cell lines.

(C) Immunofluorescent staining of acetylated tubulin (green) marking cilia. DNA is stained with DAPI (blue). Cilia can be seen in the well characterized cilia forming cell line, 10T1/2, as well as two TTC21B mutant SW036310 sensitive cell lines (H647, H157) but not in two SW036310 resistant cell lines (H460, HCC1171). Scale bars, 10 µm.

(D and E) Dose response curves are shown for cell lines outside the training set that are predicted to be sensitive (blue) or resistant (orange) to (D) SW140154 and (E) SW151511.

(F) Dose response curves of cell lines grown in 3-D spheroid models in response to SW140154 (n = 8/experiment). Cell lines that were sensitive (blue) or resistant (orange) to SW140154 in standard 2-D cell-culture were evaluated in 3D (n = 8/dose).

(G) Fold change of 13 differentially regulated genes in response to treatment with 10 µM SW151511 for 24 hr in 2 sensitive and 2 resistant lines is shown. Values represent the log₂ fold change of the gene expression in SW151511 treated cells compared to cells treated with the DMSO vehicle.
Figure S5. Chemical Response Associations among KRAS Mutant NSCLC Lines, Related to Figure 5

(A) An empirical CDF plot compares the AUC values SW174769, SW087454, and SW056245 of cell lines with co-occurring mutations in KRAS and NUP214 (left), RAS_Class and PTPRT (center) and KRAS and TTC21B (right) (red) to wild-type cell lines (blue; p < 0.0002, Scanning KS test).

(B) Dose response curves of cell lines grown in 3D spheroid models in response to SW174769 (left), SW087454 (middle) and SW056245 (right). Cell lines that were sensitive (blue) or resistant (orange) to each chemical in standard 2-D cell-culture were evaluated in 3D (n = 8/dose).

(C) Deletion in KEAP1 as visualized by integrative genomics viewer (IGV). H1437 cancer cell lines (bottom panel) have a large deletion of exon2 (red box) as compared to DNA from a matched normal tissue sample (top panel).

(D) A lollipop plot compares NRF2 mutation statuses and locations in TCGA LUAD and LUSC tumor datasets and in the UTSW cell panel. Mutations are ordered based on annotated amino acid position along protein length. Panels indicate frequency of non-synonymous mutations found in TCGA LUAD’s (blue) and LUSC’s (red), in the UTSW cell line panel (blue = SW157765 sensitive cell line; orange = SW157765 resistant cell line). PFAM annotated domains in the NRF2 protein are diagrammed below.

(E) A CDF plot compares RNaseq mRNA expression (z-scores) of genes in the NRF2-regulated set of genes in cells in the training panel of 100 cell lines sensitive to SW157765 that were not found to contain variants in known NRF2 pathway related genes (n = 8; red) compared to SW157765 resistant cells (blue). (KS test p < 2.2 E-16).

(F) A CDF plot compares RNaseq mRNA expression (z-scores) of genes in the NRF2-regulated set of genes in a panel of 3 cell lines not included in the training panel of 100 cell lines. The indicated cell lines have significantly upregulated mRNA expression of NRF2-related genes and were tested for sensitivity to SW157765 (Figure 5H). P values are calculated based on a one-sided KS-test.

(G) mRNA expression values of NRF2 in response to siNC or siNRF2 48 hr post siRNA transfection are normalized to the non-targeting control. siNRF2 oligos were transfected as either 3 individual oligos or in a pooled format as indicated.
eventually converge on lactate, with LacM1 (one labeled carbon) being a reporter of PPP activity. Glucose carbons that are shunted toward glycolysis will result in labeling of LacM2 (2 labeled carbons). In the serine biosynthetic pathway (blue), the glycolytic precursor 3PG will be used converted in a series of steps to form serine (H), producing a SerM2 labeling pattern. In the citric acid cycle (red), the 3 carbon end product of glycolysis, pyruvate, will be shunted toward the TCA cycle. The first step of which involves an oxidative decarboxylation, resulting in two labeled carbons incorporated into citrate (CitM2).

Figure S6. SW157765-Sensitive Cell Lines Define a KRAS Mechanistic Subtype Addicted to GLUT8-Mediated Glucose Transport, Related to Figure 6

(A) Binder selection pipeline. ~14,000 purified proteins were arrayed in 96 well format, with one protein per well. Chemicals were added to each well and binding was determined based on retention times in LC/MS. Plotted are LC/MS binding signals (y axis) for all proteins assayed (x axis). As a control, FK-506 had the strongest signal for binding to well annotated binding partners, FK-506 binding protein family members (FKBP) 1A, 2, 3, 4, and 5. The family member with the strongest binding annotated in the literature, FKBP1A, had the highest LC/MS binding signal.

(B) (Left) GLUT8 model with docked poses of SW157765 (orange) and the SW157765-epoxide metabolite (cyan). Residues (in ball and stick) within 3.5 Å of SW157765 metabolite docked pose are shown. Both chemicals are predicted to make hydrogen bonds with Asn312. This interaction is consistent with GLUT1 co-crystal structure of glucose where Asn286 exhibits H-bond interaction with C4-OH of glucose. Further, this residue is also known to involve in conformation switching of GLUT3 from outward open to outward occluded form. SW157765 as well as the metabolite exhibited van der Waals interactions with equivalent residues in corresponding co-crystal structures, and these residues were experimentally confirmed to be crucial in inhibitor binding. The presence of a double bond next to furan seems to be a necessary constraint that helps restrain SW157765 or the epoxide metabolite in a specific conformation. Residues from TM4 are colored gray, TM5 purple, TM7 green, TM10 orange and TM11 pink (Right) Close-up view of SW157765 and SW157765-epoxide metabolite overlay highlighting deviation in docked poses.

(C) A thermal-stability shift assay comparing protein levels of GLUT8 in H647 cells treated with either DMSO, 5 μM 500-1, or 1 μM SW157765. Cells were heated to the indicated temperatures prior to lysis. Glut13 serves as a loading control.

(D) Cells were deprived of glucose from the culture medium for 5 days and viability was measured as relative DNA content at day 5 versus day 0. The boxplot compares viabilities in response to glucose withdrawal in SW157765 sensitive (blue) and resistant (orange) cells. GLUT1 mRNA was measured with qPCR 96 hr post-transfection with siGLUT1 or siNC. Values are normalized to the siNC.

(E and F) mRNA (E) and Protein expression of GLUT1 in H647 96 hr post-transfection with siGLUT1 or siNC oligonucleotides was measured by either qPCR or immunoblot.

(G) Viability of SW157765 sensitive and resistant cell lines transfected with siRNA’s targeting GLUT1 normalized to control (NC) was measured at 96 hr post-transfection.

(H) In the serine biosynthetic pathway, a glycolytic precursor, 3PG, is converted in a series of steps to serine (Ser), which is subsequently cleaved to produce glycine (Gly) and a one carbon intermediate that can then enter the folate cycle for production of purines and thymidines. PHGDH activity is rate limiting for the pathway. Steady-state flux through this pathway can be determined by measuring incorporation of [13C6] into all three carbons of serine (SerM3) and both carbons of glycine (GlyM2).

(I) Boxplots comparing RNaseq based log2 FPKM expression for the enzymes in the serine biosynthetic pathway in SW157765 sensitive (blue) and resistant (orange) cell lines.

(J) Relative viability of H647 cells at 24 hours post-SW157765 treatment

(K and L) Incorporation of [13C2] into (K) 1 carbon of lactate (LacM1) and (L) 2 carbons of citrate (CitM2), reporting activity in the pentose phosphate pathway and citric acid cycle, respectively, was measured 24 hr after treatment of H647 cells with SW157765 (5 μM) or DMSO.

(M) Heatmap comparing SW157765 sensitivities. Cell lines are ordered according to SW157765 AUC values (top panel). KEAP1, KRAS mutant cells are indicated in black (second panel). The following panels indicate expression levels (log2 RNaseq FPKM) of CYT4F11, PHGDH, and PSAT. Unanticipated non-responders (HCC4019, HCC44, DFCI024, and H2030) are highlighted in red.

(N) Protein expression of PHGDH in HCC44 parental cells and those stably expressing PHGDHV490M or PHGDHm3599D under the control of a tetracycline-inducible promoter was assessed by immunoblot. Doxycycline (1 μg/mL) was used to induce expression of the PHGDH construct, though the promoter is quite leaky.

(O) Fate of 1,2 labeled glucose upon entering the cell. Once labeled glucose enters the cell, it is phosphorylated to form glucose 6-phosphate (G6P). G6P can enter the PPP pathway (green), the first step of which involves an oxidative decarboxylation where one of the labeled carbons will be released as CO2. The pathway will eventually converge on lactate, with LacM1 (one labeled carbon) being a reporter of PPP activity. Glucose carbons that are shunted toward glycolysis will result in labeling of LacM2 (2 labeled carbons). In the serine biosynthetic pathway (blue), the glycolytic precursor 3PG will be used converted in a series of steps to form serine (H), producing a SerM2 labeling pattern. In the citric acid cycle (red), the 3 carbon end product of glycolysis, pyruvate, will be shunted toward the TCA cycle. The first step of which involves an oxidative decarboxylation, resulting in two labeled carbons incorporated into citrate (CitM2).