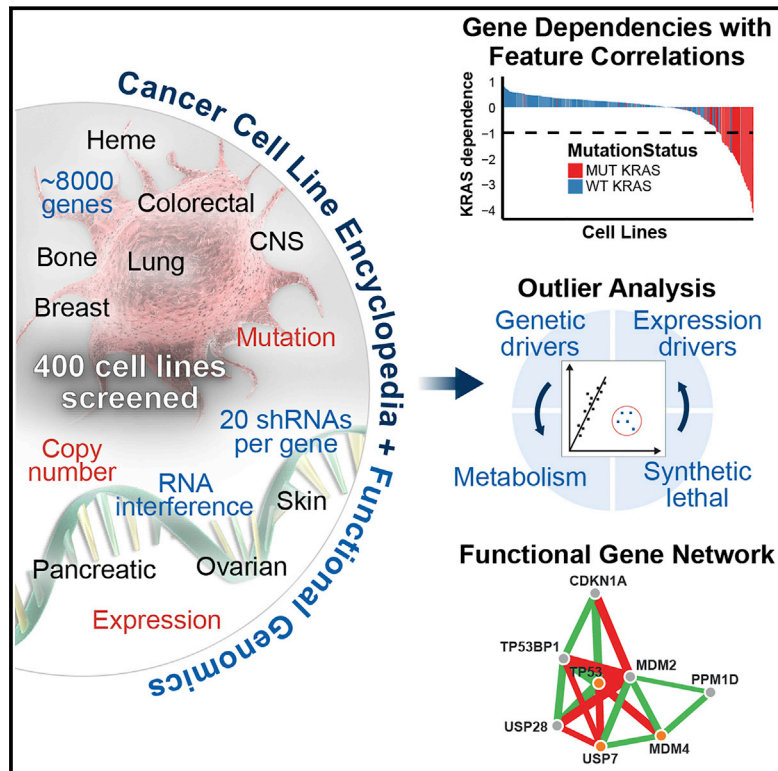


Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening

Graphical Abstract



Authors

E. Robert McDonald III, Antoine de Weck, Michael R. Schlabach, ..., Francesco Hofmann, Tobias Schmelzle, William R. Sellers

Correspondence

rob.mcdonald@novartis.com (E.R.M.), tobias.schmelzle@novartis.com (T.S.)

In Brief

A large-scale RNAi screen in 398 cancer cell lines reveals vulnerabilities of specific cancer subtypes.

Highlights

- Project DRIVE: deep RNAi interrogation of viability effects in cancer
- ~8,000 genes were targeted by a median of 20 shRNAs per gene in ~400 CCLE models
- Data robustly define cancer dependency genes falling into distinct outlier classes
- Enables genetic networks that reveal protein complexes and biological pathways



McDonald et al., 2017, Cell 170, 577–592
July 27, 2017 © 2017 Elsevier Inc.
<http://dx.doi.org/10.1016/j.cell.2017.07.005>

Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening

E. Robert McDonald III,^{1,2,*} Antoine de Weck,^{1,2} Michael R. Schlabach,^{1,2} Eric Billy,^{1,2} Konstantinos J. Mavrakis,^{1,2} Gregory R. Hoffman,^{1,2} Dhiren Belur,¹ Deborah Castelletti,¹ Elizabeth Frias,¹ Kalyani Gampa,¹ Javad Golji,¹ Iris Kao,¹ Li Li,¹ Philippe Megel,¹ Thomas A. Perkins,¹ Nadire Ramadan,¹ David A. Ruddy,¹ Serena J. Silver,¹ Sosathya Sovath,¹ Mark Stump,¹ Odile Weber,¹ Roland Widmer,¹ Jianjun Yu,¹ Kristine Yu,¹ Yingzi Yue,¹ Dorothee Abramowski,¹ Elizabeth Ackley,¹ Rosemary Barrett,¹ Joel Berger,¹ Julie L. Bernard,¹ Rebecca Billig,¹ Saskia M. Brachmann,¹ Frank Buxton,¹ Roger Caothien,¹ Justina X. Caushi,¹ Franklin S. Chung,¹ Marta Cortés-Cros,¹ Rosalie S. deBeaumont,¹ Clara Delaunay,¹ Aurore Desplat,¹ William Duong,¹ Donald A. Dvoske,¹ Richard S. Eldridge,¹ Ali Farsidjani,¹ Fei Feng,¹ JiaJia Feng,¹ Daisy Flemming,¹ William Forrester,¹ Giorgio G. Galli,¹ Zhenhai Gao,¹ François Gauter,¹ Veronica Gibaja,¹ Kristy Haas,¹ Marc Hattenberger,¹ Tami Hood,¹ Kristen E. Hurov,¹ Zainab Jagani,¹ Mathias Jenal,¹ Jennifer A. Johnson,¹ Michael D. Jones,¹ Avnish Kapoor,¹ Joshua Korn,¹ Jilin Liu,¹ Qiumei Liu,¹ Shumei Liu,¹ Yue Liu,¹ Alice T. Loo,¹ Kaitlin J. Macchi,¹ Typhaine Martin,¹ Gregory McAllister,¹ Amandine Meyer,¹ Sandra Mollé,¹ Raymond A. Pagliarini,¹ Tanushree Phadke,¹ Brian Repko,¹ Tanja Schouwey,¹ Frances Shanahan,¹ Qiong Shen,¹ Christelle Stamm,¹ Christine Stephan,¹ Volker M. Stucke,¹ Ralph Tiedt,¹ Malini Varadarajan,¹ Kavitha Venkatesan,¹ Alberto C. Vitari,¹ Marco Wallroth,¹ Jan Weiler,¹ Jing Zhang,¹ Craig Mickanin,¹ Vic E. Myer,¹ Jeffery A. Porter,¹ Albert Lai,¹ Hans Bitter,¹ Emma Lees,¹ Nicholas Keen,¹ Audrey Kauffmann,^{1,2} Frank Stegmeier,^{1,2} Francesco Hofmann,^{1,2} Tobias Schmelzle,^{1,2,3,*} and William R. Sellers^{1,2}

¹Novartis Institutes for Biomedical Research, Oncology Disease Area, Basel 4002, Switzerland; Cambridge, MA 02139, USA; and Emeryville, CA 94608, USA

²These authors contributed equally

³Lead Contact

*Correspondence: rob.mcdonald@novartis.com (E.R.M.), tobias.schmelzle@novartis.com (T.S.)

<http://dx.doi.org/10.1016/j.cell.2017.07.005>

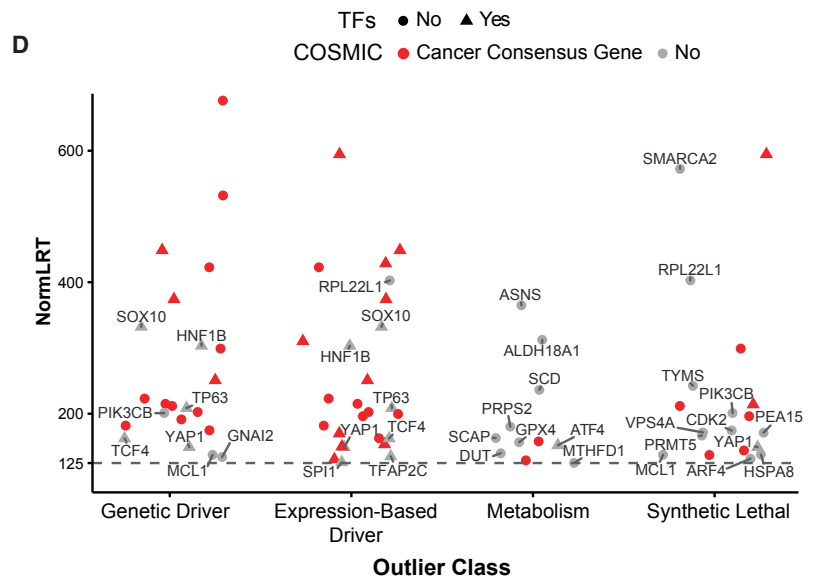
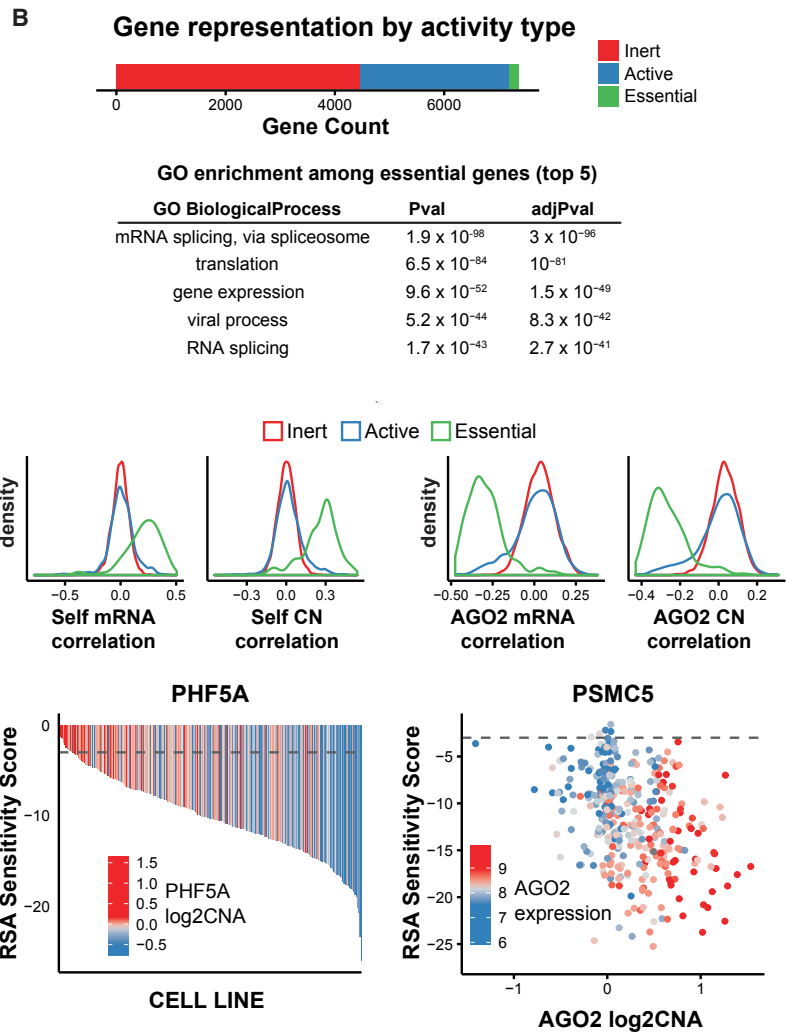
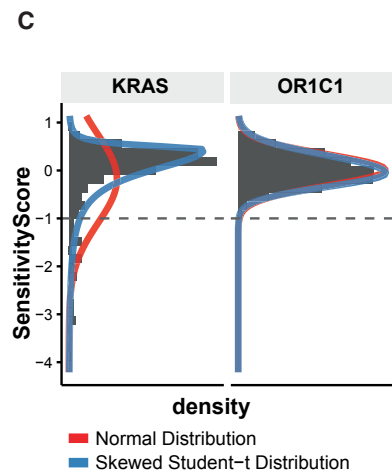
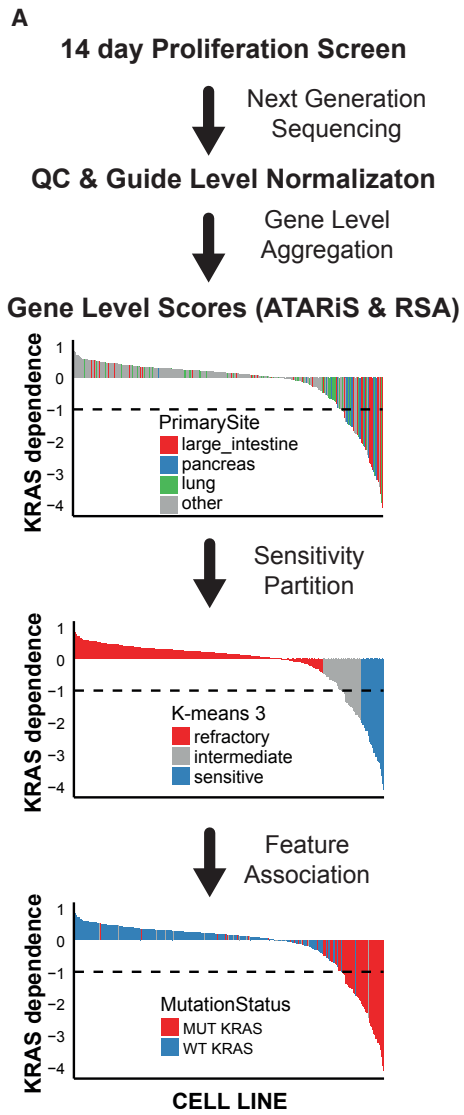
SUMMARY

Elucidation of the mutational landscape of human cancer has progressed rapidly and been accompanied by the development of therapeutics targeting mutant oncogenes. However, a comprehensive mapping of cancer dependencies has lagged behind and the discovery of therapeutic targets for counteracting tumor suppressor gene loss is needed. To identify vulnerabilities relevant to specific cancer subtypes, we conducted a large-scale RNAi screen in which viability effects of mRNA knockdown were assessed for 7,837 genes using an average of 20 shRNAs per gene in 398 cancer cell lines. We describe findings of this screen, outlining the classes of cancer dependency genes and their relationships to genetic, expression, and lineage features. In addition, we describe robust gene-interaction networks recapitulating both protein complexes and functional cooperation among complexes and pathways. This dataset along with a web portal is provided to the community to assist in the discovery and translation of new therapeutic approaches for cancer.

INTRODUCTION

Enabled by large-scale sequencing projects (Weinstein et al., 2013; Hudson et al., 2010), a detailed map of the nucleotide-

and chromosomal-level alterations comprising the disordered cancer genome has been provided. In concert, targeted therapeutics are now mainstays of therapy in melanoma, NSCLC, CML, and GIST while many more are in clinical testing (Pagliarini et al., 2015). Nonetheless, the discovery of new therapeutics for most cancers remains elusive. This has motivated the use of larger-scale gene knockdown experiments to discover functional gene requirements across diverse sets of cancer (Cheung et al., 2011; Marcotte et al., 2012). The power of RNA interference screens can, however, be hampered by the inability to distinguish between on- and off-target effects (low number of shRNAs/gene) or by the lack of statistical power to describe molecular correlates of knockdown effects (low number of models screened). To overcome these limitations, we recently performed a large-scale deep shRNA screen targeting 384 genes from the epigenome using ~17 shRNAs per gene across 58 cell lines and demonstrated robust detection of known and novel cancer vulnerabilities (Hoffman et al., 2014). Based on the success of these efforts, we launched a larger screen referred to as project DRIVE (deep RNAi interrogation of viability effects in cancer). In DRIVE, a lentiviral library was produced targeting 7,837 human genes with a median of 20 shRNAs per gene and used to screen 398 cancer cell lines in a pooled format to ascertain the effect on cell viability. Here we describe the known and novel genes that comprise distinct classes of cancer dependence including genetic, expression, metabolic, and synthetic lethal relationships. This effort brings us closer to a functional annotation of the cancer genome.



(legend on next page)

RESULTS

Functional Genomic Screening Reveals Four Categories of Dependency Outliers

In project DRIVE, we constructed deep coverage shRNA lentiviral libraries targeting 7,837 genes (Table S1). Gene content included known mutated/amplified cancer drivers, epigenetic genes, transcription factors (TFs), cell surface proteins, and druggable enzymes, as well as known pan-essential complexes. This library was screened across 398 cell lines to identify gene dependencies (Table S2). 2.7×10^8 cells were infected ($>1,000\times$ representation of the 151,504 member library) and passaged for 14 days after infection. Quantile normalized log fold change data were generated from next-generation sequencing (NGS) counts of individual shRNAs after screen versus the shRNA abundance in the library input.

shRNA activity was aggregated to gene-level activity by two complementary methods, ATARIS and RSA (Figure 1A). RSA uses all shRNA reagents against a given gene to calculate a statistical significance that knockdown of gene X leads to loss of viability (König et al., 2007). ATARIS gene level scores incorporate only shRNAs that have consistent activity across the entire dataset (Shao et al., 2013) and attempts to eliminate inert or potential off-target shRNAs. The ATARIS algorithm median centers shRNA level scores across all screened cell lines, resulting in a metric of relative effects so essential and inert gene profiles appear no different. For this reason, RSA was used to determine whether a gene was essential, active, or inert and all other analyses used ATARIS to focus on only shRNA reagents with consistent, on-target activity. As part of the Project DRIVE resource, a public portal has been created using the Shiny framework (Bassik et al., 2016) to allow visualization of gene profiles (Sensitivity Profile Viz tab at <https://oncologyinbr.shinyapps.io/drive/>).

Using RSA to categorize each gene as inert, active, or essential (Figure 1B), $>4,450$ of the genes showed no significant growth effects. Genes with an RSA value of ≤ -3 for $>50\%$ of cell lines were deemed essential. Assignment of GO terms for this class demonstrated enrichment of critical cellular processes such as translation, gene expression, and splicing. In addition, correlation analyses for the essential genes uncovered two principle features. First, gene dependence correlated with the expression and/or copy number of the target gene itself (Figure 1B). For example, cell lines with low expression or a copy number deletion of an essential gene were more sensitive to knockdown. This phenomenon is termed CYCLOPS for genes with heterozygous copy number deletion (Nijhawan et al.,

2012). Notably, *PHF5A* had the highest CN-dependence correlation and while previously reported to be required for GBM stem cells survival (Hubert et al., 2013), our data suggest that *PHF5A* is instead a broadly essential gene. The second principle feature of essential genes was that gene dependence correlated with high expression and/or copy number gain of *AGO2*, a component of the RNA-induced silencing complex (RISC) (Figure 1B). This strongly suggests that lines with increased *AGO2* have improved knockdown efficiency leading to enhanced shRNA activity against essential genes. For example, the profile of the proteasome component *PSMC5* is shown in relation to *AGO2* CN and expression (Figure 1B). These two confounders must be considered when genetic screens are conducted in small cell line sets. A complete list of putative CYCLOPS and pan-essential genes is provided (Tables S3 and S4).

To enable the discovery of features predictive of sensitivity, we developed a bioinformatics pipeline. K-means ($k = 3$) clustering of gene profiles was used to delineate sensitive and insensitive populations and these classifications were used to interrogate the feature sets of the CCLE (Figure 1A; Barretina et al., 2012). This workflow enabled the systematic identification of sensitivity predictors across the genes queried in Project DRIVE. We next wanted to identify selective cancer cell dependencies in an unbiased manner. To this end, an outlier analysis was performed to identify profiles that demonstrated robust dropout behavior in a subset of cancer cell lines (Figure 1C). A normality likelihood ratio test (NormLRT) was applied to the gene-level ATARIS scores to determine whether each gene profile had a distribution divergent from normal and to assign an outlier score based on the deviance between the normal distribution and the skewed t distribution. This approach is agnostic to the direction of the skew and therefore both growth suppressors and enhancers are detected. Importantly, it is also sensitive to a single cell line outlier. Lastly, NormLRT incorporates the magnitude of the phenotypic effect in order to focus on genes with robust dependency. The top outliers (NormLRT > 125) fall into distinct groups of dependence (Table S5). The majority (88%) can be categorized into four classes: genetic dependence, expression-based dependence, metabolic genes/enzymes, and synthetic lethals (Figure 1D), with some genes belonging to multiple classes. The first two classes encompass genes in which the feature correlations predictive of dependence are mutation, copy number amplification, or high-level mRNA expression of the gene itself. Many of these genes are found on the COSMIC gene consensus list (Forbes et al., 2017). In the synthetic lethality and metabolism classes, the features correlated with cancer cell line dependence appear to relate

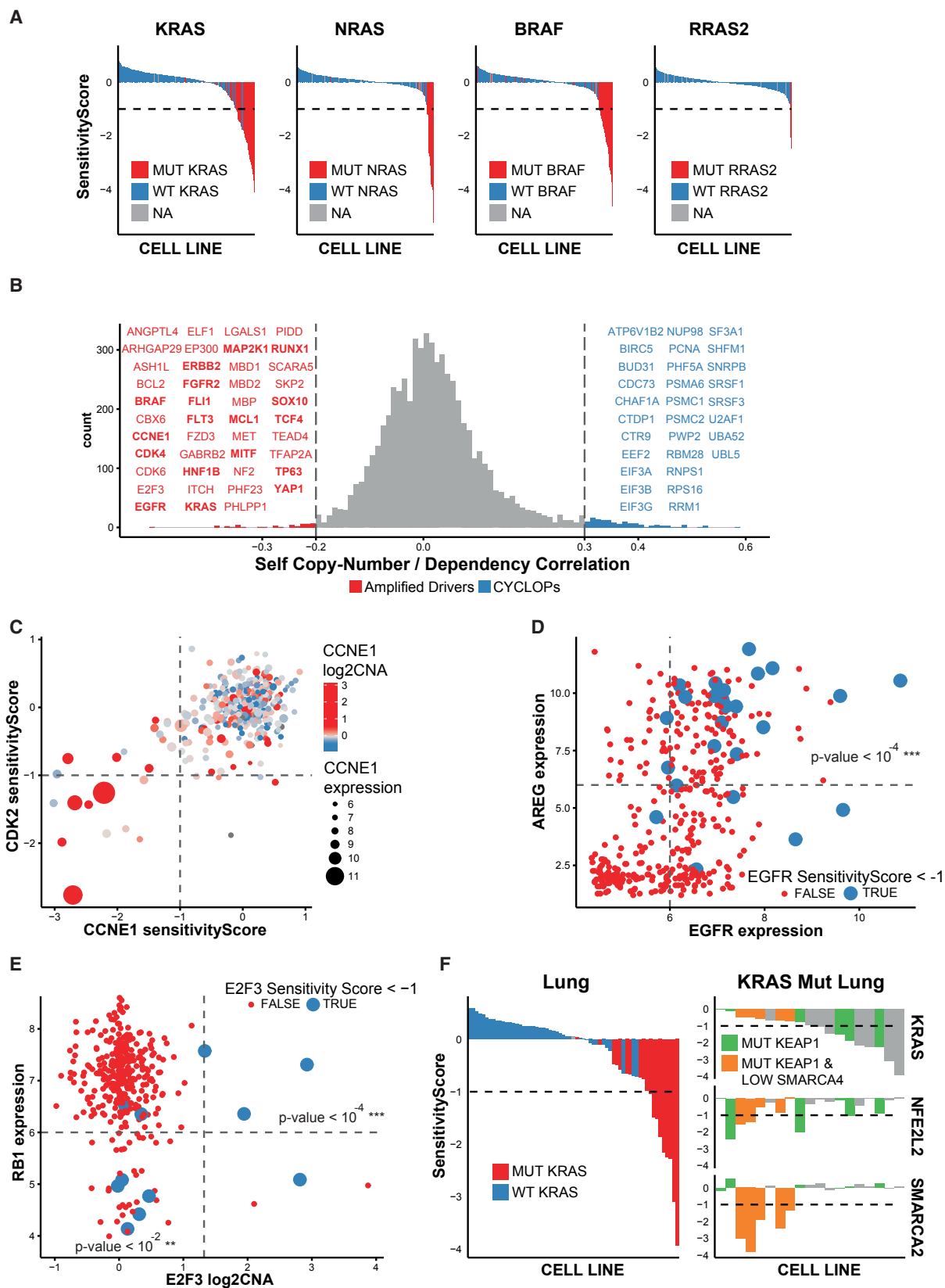
Figure 1. Project DRIVE Informatics and Outlier Analysis

(A) Project DRIVE screening and analytical workflow. 2-week pooled shRNA viability screens were followed by an NGS readout and shRNA gene-level aggregation by RSA and ATARIS. Feature correlation was performed using k-means clustering ($k = 3$) to identify sensitive and insensitive populations with resulting display of top sensitivity correlations. KRAS example is shown.

(B) DRIVE gene activity categorization into inert, active, or essential profiles. Essential gene dependency correlations include self-CN/expr and *AGO2* CN/expr correlations. RSA waterfall plot for *PHF5A*, colored by *PHF5A* copy number. RSA sensitivity cutoff shown at -3 indicates that this is an essential gene in most lines tested. *PSMC5* RSA sensitivity plotted against *AGO2* CN and colored by *AGO2* expression.

(C) Normality LRT compares the fit of a skewed Student's t distribution and a normal distribution. Profiles with better skewed Student's t distribution fit over normal have high NormLRT scores.

(D) Top DRIVE outliers by class colored by their presence in COSMIC. Non-COSMIC genes labeled and TFs shown as triangles.



(legend on next page)

more closely to an altered cell state or genetic alterations in other genes. Genes in these two classes have largely not been identified in cancer genomics efforts and are not contained on the cancer consensus list.

Genetic Drivers of Cancer Dependency and Modifiers of Oncogene Addiction

Mutated oncogenes (e.g., *NRAS*, *BRAF*, and *KRAS*) are among the most robust dependencies observed in DRIVE (Figure 2A). As expected, *BRAF* mutation and dependence are mainly observed in colon, thyroid, and melanoma lines, while *NRAS* mutation and dependence is most prevalent in melanoma. *KRAS* mutation and dependence occurs in colon, pancreatic, and lung lineages. Consistent with TCGA (Cancer Genome Atlas Network, 2012) and emerging clinical data (Mayer et al., 2017), *PIK3CA* shRNAs show activity across lineages with enrichment in *PIK3CA*^{mut} ER+ breast lines. In the DRIVE data, we observe a novel genetic dependence involving *RRAS2* (TC21) (Figure 2A). Both CAL51 and A2780 lines harbor *RRAS2*^{Q72L} mutations that are analogous to the *KRAS*^{Q61} position. Interestingly, NCI-H1048 harbors a G23S mutation that may mimic position G12 or G13 of other RAS proteins. While *RRAS2* mutation appears to be a relatively rare event in the primary tumors sequenced as part of TCGA, position Q72 is the most recurrent and our data suggest that this creates an oncogenic dependence.

Recurrent gene amplification of either mutated alleles or wild-type alleles are a common mechanism for oncogene activation. Forty genes demonstrate a strong correlation between dependence and gene amplification (Figure 2B, genes in red). For *KRAS* and *BRAF*, amplification of the mutant allele likely contributes to dependence. On the other hand, wild-type *ERBB2*, *MET*, and *FGFR2* undergo extreme CN gain (>20 copies) associated with cancer dependence. High-level amplification of *ITCH* in a single thyroid cell line supports a role for *ITCH* in anaplastic thyroid cancer (Ishihara et al., 2008). The 19q12 amplicon, seen in primary breast, ovarian, and endometrial cancers, harbors *C19orf12*, *PLEKHF1*, *URI1*, *POP4*, and *CCNE1* (Cyclin E). Multiple genes in the amplicon exhibit elevated expression leading to controversy as to whether a single gene is the driver. We found that 19q12 amplified lines are dependent on *CCNE1* and *CDK2* (Figure 2C) while the remaining genes showed no correlation between CN and dependence. In contrast to prior data (Theurillat et al., 2011), *URI1* appears to be an essential gene rather than a 19q12 driver gene based on its dependency correlations with its own low expression and high AGO2 (Figure S1). We also observed a set of genes in which

knockdown was strongly correlated with copy number, yet the magnitude of the effect on cancer growth appeared to be modest. These strong correlations appear to arise from the dual contribution of slight copy number gain in dependent lines and gene deletion in the least dependent lines. This pattern seems unlikely to be indicative of key cancer dependent genes. Genes in this category include *ASH1L*, *ELF1*, *MBD1*, *MBD2*, *PHLPP1*, *PLXNA4*, and *VAPB*.

In using deep shRNA libraries we hoped to discover not only the direct dependence on key drivers, but also genes that might act as modifiers. Indeed, modifiers of oncogene dependence can be observed in DRIVE. For example, *EGFR* dependence in lung and other solid tumors is associated not only with amplification and expression of *EGFR* but also with high levels of Amphiregulin (*AREG*) (Figure 2D). Dependence on the TF E2F3 is most correlated with *E2F3* gene amplification and/or with loss of expression of RB (Figure 2E). In lung cancer, we observed a substantial number of *KRAS* mutant cell lines in which *KRAS* is dispensable for growth. Instead, these *KRAS* mutant lung lines are susceptible to *NFE2L2* (NRF2) and *SMARCA2* (BRM) knockdown correlated with loss-of-function mutations in *KEAP1* and/or low *SMARCA4* (BRG1), respectively (Figure 2F). This has important implications for treating *KRAS* mutant cancers with MAPK pathway inhibitors, as co-occurring mutation is predicted to lead to de novo resistance.

Expression Correlation Analysis Identifies Oncogenes and Lineage-Specific Transcription Factors

The second class of cancer-dependent genes was comprised of those in which high-level expression of the target gene was a top correlated feature. To further explore this class, we directly correlated gene expression and gene dependence. 57 genes had a negative correlation (−0.3 or better) where high levels of target gene expression were associated with significant growth effects upon shRNA-induced target knockdown (Figure 3A, genes in red). Hits shown in bold were originally identified in the unbiased outlier analysis. Genes in this class include many previously described genetic drivers, as amplification and mutation often leads to increased expression. Additionally, a number of genes are unique to this class and show either a pattern of broad expression correlations across different lineages (e.g., *ZEB1*, *BCL2L1*) or a more lineage-restricted pattern (e.g., *MPL* and *CCND3*). Notably, many of the expression outliers are lineage-restricted TFs (Figure 1D) (see below).

The transcriptional repressor *ZEB1* shows a strong correlation between gene expression and gene dependence observed

Figure 2. Genetic Drivers and Modifiers of Oncogene Dependence

- (A) ATARIS waterfall plots for top mutation-driven dependencies, colored by their respective mutation status.
- (B) Self copy number/dependency Pearson correlation for each gene identifies amplified genetic drivers as well as CYCLOPS genes. Outliers shown in bold. Only CYCLOPS genes with a correlation >0.4 are listed.
- (C) ATARIS profiles of *CCNE1* (x axis) and *CDK2* (y axis) where each dot is a cell line colored by its *CCNE1* copy number and sized by its *CCNE1* expression. See also Figure S1.
- (D) *EGFR* expression versus ligand *AREG* expression, colored by *EGFR* dependence as measured by ATARIS sensitivity score. p value from Fisher exact test shown for upper right quadrant shows statistical significance for *EGFR* dependence with hi *AREG* and hi *EGFR*.
- (E) E2F3 dependency (size and color) plotted with its copy number (x axis) and RB1 expression (y axis). p value from Fisher exact test shown for upper right and lower left quadrant shows statistical significance for E2F3 dependence with high E2F3 CN or low RB1 expression.
- (F) *KRAS* dependency profile in lung lineage colored by mutation status. On the right, *KRAS* mutant lung lines only are co-plotted with *KRAS*, *NFE2L2*, and *SMARCA2* dependence, colored by mutation/expression status of *KEAP1* and *SMARCA4*.

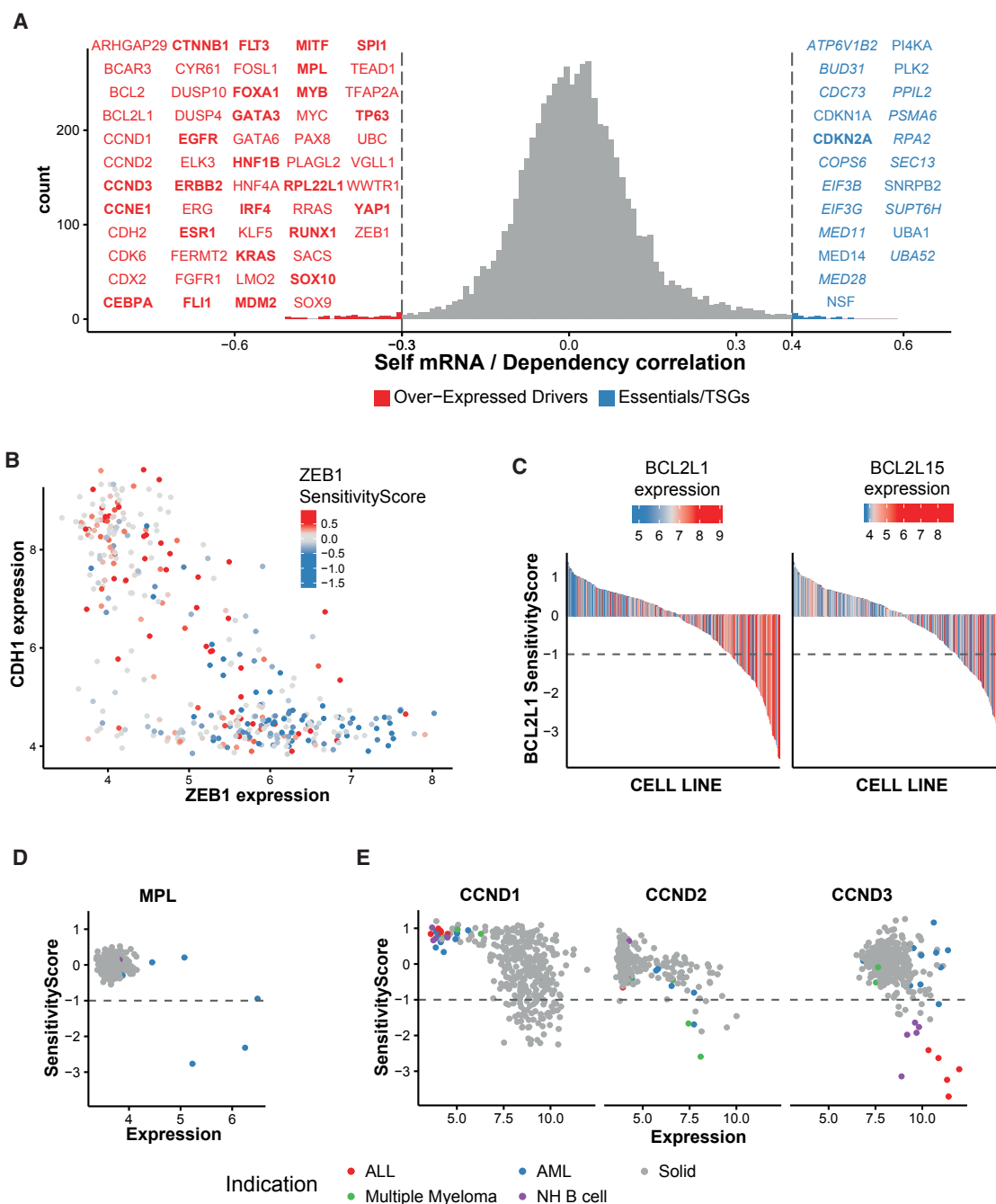


Figure 3. Expression-Based Driver Dependency Outlier Class

(A) Self mRNA expression/dependency Pearson correlations for all genes in DRIVE are plotted. Outliers shown in bold. CYCLOPS shown in italics.

(B) ZEB1 expression versus CDH1 expression, colored by ZEB1 dependence.

(C) High-level BCL2L1 expression (color) predicts its dependence (left waterfall). BCL2L1 dependence is also correlated with BCL2L15 expression (right waterfall).

(D) MPL dependence plotted versus expression, colored by hematopoietic subtypes or solid tumor cell line.

(E) Dependence on individual D-type cyclins plotted versus expression and colored by hematopoietic subtypes or solid tumor cell line.

across multiple lineages. ZEB1 dependence is linked to both self-expression and lack of expression of CDH1, a transcriptional repression target of ZEB1 (Figure 3B; Grootclaes and Frisch, 2000). *BCL2L1* (BCLXL) is another example where high

expression is predictive of dependence across lineages (Figure 3C). Interestingly, the lines most sensitive to BCLXL depletion also express high levels of BCL2L15, a poorly studied BH2/BH3-containing family member with weak pro-apoptotic

activity (Coultais et al., 2003). *MPL* is the receptor for thrombopoietin and sustains gain-of-function mutations in myeloproliferative neoplasms including essential thrombocythemia and primary myelofibrosis (Pardananani et al., 2006). In DRIVE, only four hematopoietic cell lines have appreciable *MPL* expression and three demonstrate *MPL* dependence (Figure 3D). The D-type cyclins interact with either CDK4 or CDK6 to govern control of the G1/S transition (Otto and Sicinski, 2017). Each cyclin shows a strong expression-dependence correlation (Figure 3E). *CCND3* shows hematopoietic-specific expression and dependence in collaboration with CDK6 in ALL and DLBCL while *CCND1* is active in solid tumor cell lines in collaboration with CDK4. *CCND2* shows activity in a smaller subset of lines in both hematopoietic and solid tumor lineages but maintains a strong expression correlation. These data suggest that within a given cell line, a single D-type cyclin in combination with CDK4 or 6 is required to mediate passage through the restriction point. Hence, greater therapeutic specificity might be achieved through the disruption of specific D-type cyclin-CDK4/6 interactions.

Within the expression outlier class is a large collection of lineage TFs (Figure 4A). Certain lineages demonstrate master TF regulators where tissue-restricted expression is highly correlated with sensitivity to gene depletion (*MYB*, *PAX8*, *CTNNB1*, *SOX10*, and *MITF*). *SOX10* is expressed in neural crest progenitors during development (Bondurand and Sham, 2013) and is expressed in neural crest-derived tumors including melanoma and glioma. The *SOX10* relationship is binary where any detectable expression equates to dependence (Figure 4B).

HNF1B functions developmentally to control distinct aspects of kidney, pancreatic, and liver tissue specification with adult tissue expression also seen in the gonads and lung. In DRIVE, cancer dependence for *HNF1B* is observed in kidney, pancreatic, liver, lung, and ovarian cell lines and shows a correlation between expression and dependence (Figure 4B). In normal development, *HNF1B* functions with *HNF1A*. However, in the cancer cell line datasets we failed to observe *HNF1A* and B co-dependence, suggesting that dependence might be driven by *HNF1B* homodimers.

Within the hematopoietic lineage, specific subtypes have unique TF dependencies (Figure 4C). Indeed, hierarchical clustering based on TF dependence largely segregates these models by disease subtype. For example, models of DLBCL are uniquely dependent on *MEF2B* where mutations have been previously described (reviewed in Pasqualucci and Dalla-Favera, 2015). Both *MYB* and *RUNX1* are required for proliferation across multiple leukemia subtypes (ALL and AML). *RUNX1* forms the core binding TF (CBF) in collaboration with *CBFB* to drive hematopoiesis (Tracey and Speck, 2000). *RUNX1* knockdown was most closely phenocopied by knockdown of its binding partner *CBFB* and, to a lesser degree, *MYB*. shRNAs against *FLI1* and *SPI1* demonstrated robust activity across many AML lines whereas knockdown of *CEBPA* showed specific activity in the M5 subtype of AMLs. These observations are consistent with the model that myeloid leukemias suffer from a block in terminal differentiation arising as a result of aberrant TF activity (Orkin and Zon, 2008; Rosenbauer and Tenen, 2007). Finally, *IRF4* expression is detectable across multiple lymphoma subtypes whereas

IRF4 dependency is uniquely detected in multiple myeloma as previously described (Shaffer et al., 2008).

While no specific TF requirements were identified for the maintenance of basal breast cancers, analysis of luminal breast cancer confirmed the network of TFs that coordinate the regulation of hormonal signaling including *ESR1*, *FOXA1*, *GATA3*, and *TFAP2C* (Figure 4D; Cyr et al., 2015; Lupien and Brown, 2009). In addition, we also observed models that are dependent on *TFAP2A*. Consistent with emerging clinical data, the luminal subtype also showed dependence on components of the *PIK3CA*/mTOR and *CDK4* pathways (Baselga et al., 2012; Hortobagyi et al., 2016). The difference between the TF network observed in the luminal subtype compared to the absence of specific TF requirements in the basal subtype raises the possibility that basal phenotypes are a default differentiation state resulting from the absence of super-imposed specification (Bernardo et al., 2013).

Synthetic Lethal Classes

In our usage, synthetic lethality (SL) included any altered cell state (genetic, metabolic, or otherwise) that was linked to a definable non-self genetic dependence. Within this class we observed distinct subsets including those where synthetic lethality was linked to a pathway, to loss of a paralog, or to collateral lethality. In the parallel pathways of de novo synthesis and salvage of thymidine, *TYMP* (thymidine phosphorylase) mediates the reduction of thymidine to thymine. This activity reduces cellular thymidine, which is normally produced by the activity of *TYMS* (thymidylate synthase) as part of the de novo pathway. A subset of cancer cell lines display an altered ratio of *TYMP* (high) to *TYMS* (low) expression leading to the apparent cellular dependence on residual *TYMS* (Figure S2A). Depletion of the anti-apoptotic protein *MCL1* showed a robust dependent phenotype and a complex set of correlations with multiple *BCL2* proteins from parallel pathways. Here, reduced expression of *BCL2L1* (*BCLXL*) accompanied by increased expression of the pro-apoptotic BH3-only member *BIM* was most predictive of sensitivity to *MCL1* downregulation. In contrast, cell lines with high *BCLXL* expression were refractory to *MCL1* knockdown (Figure 5A).

In addition to parallel pathway SL, we observed eight distinct vertical pathway SL outlier dependencies. These included dependence on β -catenin (*CTNNB1*) in the context of genetic loss of *APC* (Figure 5A). Amplification and overexpression of Cyclin E was correlated with sensitivity to knockdown of *CDK2* (Figure 2C). Similarly, *CBFB* dependence was strongly associated with high expression of its binding partner *RUNX1* in hematopoietic lines. Cancer cell lines that retained wild-type *p53* were sensitive to depletion of either *MDM2* or *MDM4* (Figure 5A), likely resulting from the inappropriate activation of *p53* after *MDM2*/*MDM4* depletion. We also observed the previously described SL pathway relationship between *PTEN* loss and *PIK3CB* (Wee et al., 2008).

There has been significant interest in identifying synthetic lethal interactions pertaining to the RAS pathway. In DRIVE, we do not detect robust synthetic lethal interactions for mutant *KRAS*, and previously published synthetic lethal interactions are not confirmed in this dataset (Figure S2B; reviewed in

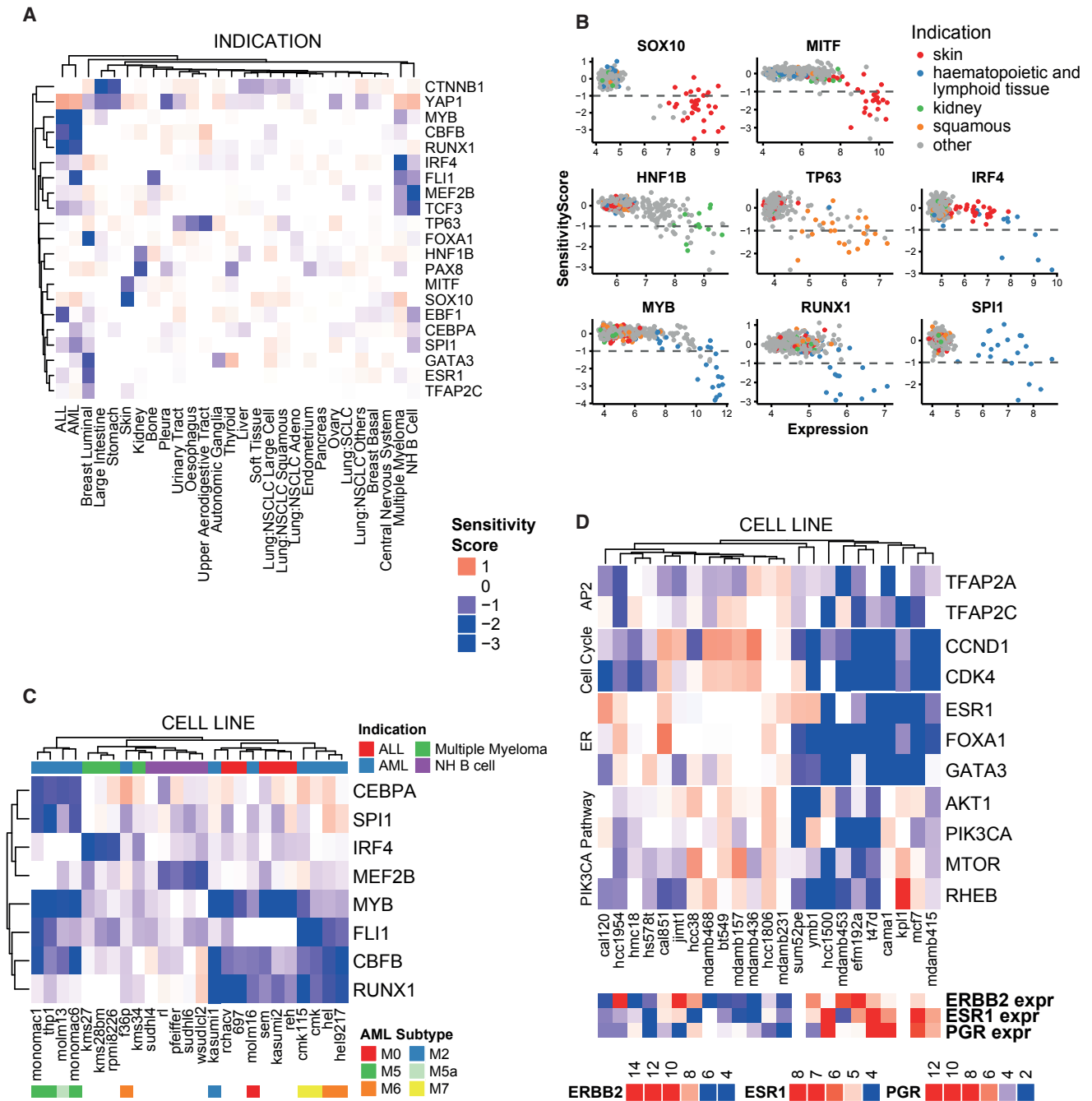


Figure 4. Lineage-Specific Transcription Factors

(A) ATARIS sensitivity score heatmap for each TF outlier per lineage. Shading is representative of average TF activity across all cell lines of a given lineage. Whenever dendrograms are shown, a hierarchical clustering was used with Euclidean distance and average linkage.

(B) TF expression (x axis) plotted versus its dependency score (y axis), colored by specific lineages or squamous histology.

(C) ATARIS sensitivity score heatmap for hematopoietic transcription factors. The hierarchical clustering results in subtype segregation (color side bar on top of dendrogram) based on functional activity of TFs.

(D) Breast lineage is plotted at the cell line level with sensitivity scores for TFs and other outliers. Genes grouped by functional class and receptor expression for each models is shown at the bottom.

Downward, 2015, and primary references therein; Kim et al., 2016; Zimmermann et al., 2013). Data supporting the notion that *RAF1* (CRAF) is required for Ras-dependent transformation

(Blasco et al., 2011; Karreth et al., 2011) can be observed but CRAF depletion does not phenocopy KRAS depletion (Figure S2B). This partial effect of CRAF depletion is also seen in

the *NRAS* mutant setting. The partial effect is unlikely to be the result of insufficient CRAF knockdown as depletion of a known positive regulator of CRAF activation, SHOC2, phenocopies CRAF depletion (Figure S2C). Mutant RAS family members may require suppression of multiple RAF family members or suppression of additional effector arms beyond the MAPK pathway to achieve efficacy. On the other hand, we do observe synthetic lethal interactions for *BRAF* in the melanoma setting. Targeting of a node downstream of mutant *BRAF* (*MEK1* or *ERK2*) was sufficient to induce vertical pathway lethality and in the case of MEK inhibition, this has been recapitulated clinically (Flaherty et al., 2012). In addition to positive regulators of MAPK signaling, we also found the surprising observation that loss of negative regulators downstream of activated *BRAF* in melanoma, such as *PEA15* and *DUSP4*, resulted in lethality, suggesting that either too little or too much flux through the MAPK pathway is detrimental in this setting.

We, and others, have described synthetic lethal interactions among paralogous genes where one paralog is genetically inactivated through mutation and/or deletion. Specifically, we found that the loss of the SWI/SNF complex member *BRG1* explains the sensitivity observed upon *BRM* depletion in NSCLC cell lines (Hoffman et al., 2014). Here, we observe a similar relationship for *ARID1A* and *ARID1B* (Helming et al., 2014) in which mutation of *ARID1A* renders cells dependent on *ARID1B* (Figure 5B) and for *RPL22* and *RPL22L1*. In this latter instance, microsatellite instable (MSI+) cell lines have frequent single-nucleotide deletions in a mononucleotide repeat in *RPL22*, a ribosome subunit (Novetsky et al., 2013). *RPL22* mutant lines were selectively dependent on the paralog *RPL22L1* (Figure 5B). In all three cases, inactivating somatic mutations result in a dependence on the remaining paralog.

We also observe paralog SL relationships where low or absent expression (rather than mutation) of a paralog is associated with dependence on the other. As previously discussed, *CDK4* and *CDK6* show largely non-overlapping dependence in the CCLE. In fact, the best predictor of *CDK4* dependence in DRIVE is low expression of *CDK6*. Similarly, the best predictor of *YAP1* sensitivity is low expression of the paralog *WWTR1* (*TAZ*). *ARF4* and 5 are the only two class II members of the ADP-ribosylation factor family (Jackson and Bouvet, 2014). We observed a novel synthetic lethal relationship where low expression of *ARF5* predicts for *ARF4* dependency (Figure 5B). Finally, *VPS4A* and *B* are homologs of the essential yeast *VPS4* gene (McCullough et al., 2013). *VPS4B* is located at 18q21.33 and is frequently lost in concert with homozygous deletions of *SMAD4* (18q21.1) and this co-deletion event is associated with dependence on the paralog *VPS4A*.

The *VPS4A* paralog dependence is also an example of collateral synthetic lethality where bystander deletion of a neighboring gene leads to cancer dependence (Figure 5C). This was first described for the *ENO1* locus on 1p36 resulting in dependence on the paralog *ENO2* (Muller et al., 2012). We were unable to verify this finding due to the absence of relevant models in the CCLE. We and others previously reported a second example, namely the dependence on *PRMT5* linked to the co-deletion of *MTAP* and the tumor suppressor *CDKN2A* (Kryukov et al., 2016; Marjon et al., 2016; Mavrikis et al., 2016). More recently,

a collateral lethal relationship was described for *SMAD4* and the nearby gene *ME2*, leading to dependence on the paralog, *ME3* (Dey et al., 2017). shRNAs for *ME3* were not included in the DRIVE library and hence we cannot detect this event. While not a top outlier, we find that ubiquitin B (*UBB*) undergoes CN loss with *p53* and is correlated with dependence on the paralog ubiquitin C (*UBC*) (Figure S3A). Finally, we observed that frequent heterozygous deletion of *p53* results in heterozygous loss of three essential genes, *POLR2A*, *MED11*, and *AURKB*, and is associated with increased sensitivity to knockdown of these genes compared to cell lines with normal CN (Figure S3B). In aggregate, we have detected synthetic lethal interactions for some of the most prevalent tumor suppressors found in cancer (Figure 5D).

DRIVE Sensitivity Network Uncovers Signaling Pathways, Protein Complexes, and Lineage Biology

The density of shRNAs used per gene, the extensive set of cell lines tested, and the robustness of the observed dependency correlations led us to ask whether the viability effects of each gene knockdown would be varied and robust enough to find correlated gene-gene interactions based only on the gene knockdown data. To test this, a similarity metric was calculated for each pair of gene profiles (21,493,846 pairwise calculations; see STAR Methods for details). This can be represented as a network where each gene represents a node and each edge represents the pairwise similarity between genes. The DRIVE Sensitivity Network (DSN) interactive tool (Network Viz tab at <https://oncology.nibr.shinyapps.io/drive/>) allows for a single- or multi-gene entry point to explore their local correlation neighborhoods. In Figure 6, a global genetic interaction map is represented using the tSNE method that, despite the limitations of 2D, enables a global view of the network and conserves many of its original features. Each point represents a gene dependence profile, and the local “neighborhood” proximity between two points represents the similarity in the original high-dimensional network between two gene profiles. Individual subnetwork examples created using the DSN tool are also shown (Figure 6).

Within the global network, we observe a dense and large neighborhood comprised of essential genes (large circles) as defined in Figure 1B. Despite the essential nature of this neighborhood, there still exists sufficient differential variation in growth effects across the cell lines to allow for the identification of large protein complexes including the proteasome (purple), ribosome, RNA polymerase complex, and the mediator complex (pink). The co-localization of a large number of non-essential genes within this cluster suggests that full inactivation of these genes (by CRISPR) might result in lethality. One pathway example in the essential neighborhood is the DNA replication checkpoint. This includes *ATR* as well as pathway components *RPA3* that senses ssDNA and *RAD17* and *TOPBP1* that help *ATR* transduce the signal to the downstream kinase *CHEK1*. In contrast to the essential gene neighborhood, the mTORC1 arm of the mTOR signaling complex shows minimal activity across the dataset but still shows robust correlations. Starting with the negative regulator *TSC2*, its direct downstream target, *RHEB*, as well as *mTOR* and *RAPTOR* can be seen as anti-correlations. Additional Regulator complex components are also seen in this mTOR

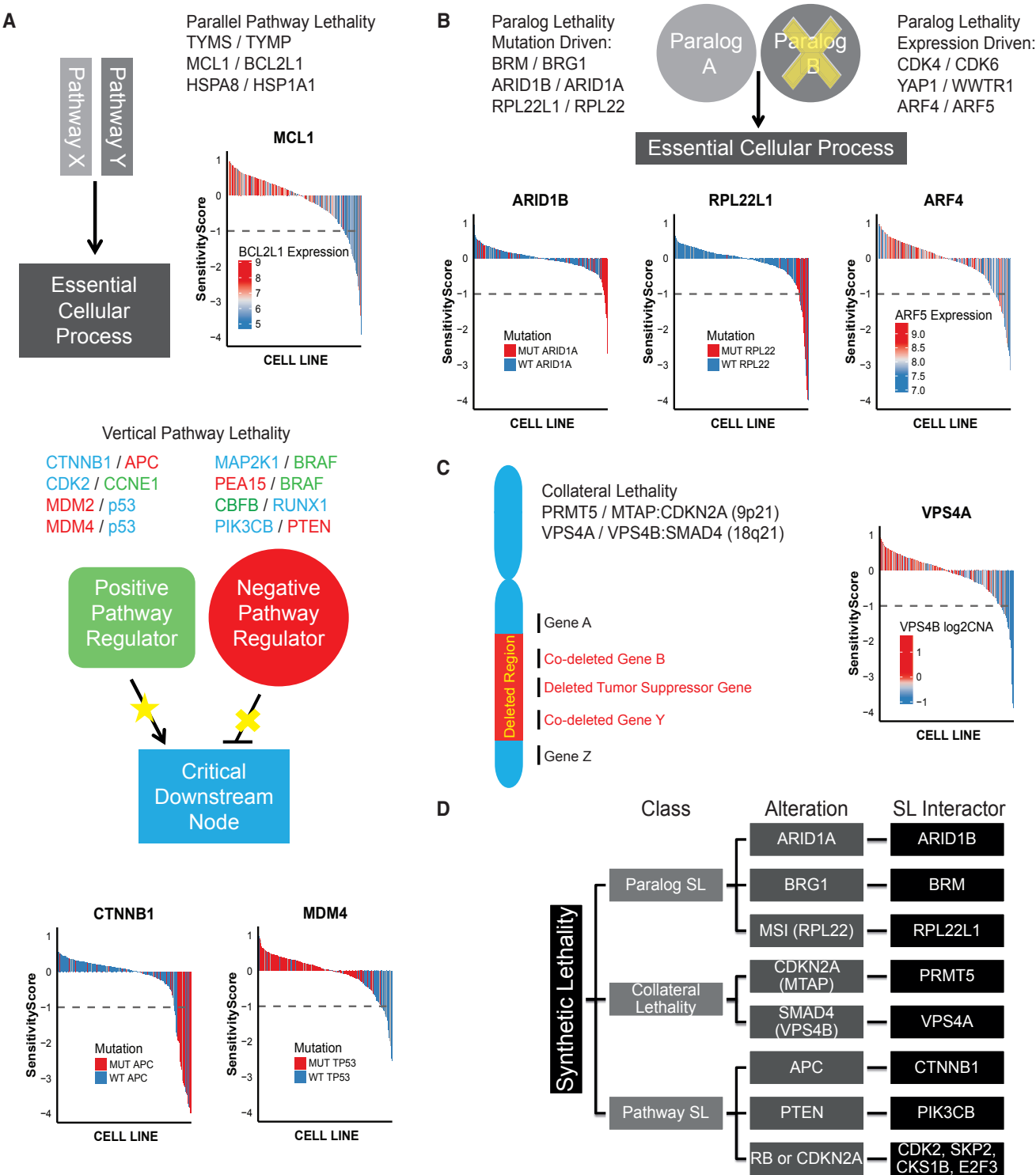


Figure 5. Synthetic Lethal Outlier Class
(A) Vertical and parallel pathway lethality outliers with DRIVE outliers listed first and pathway predictors listed second. For vertical pathway examples, relationship is indicated by color. Example waterfall plots are colored by respective predictive feature.
(B) Paralog outliers are either mutation driven or expression driven. Example waterfall plots are colored by respective predictive feature.

(legend continued on next page)

amino acid-sensing neighborhood including SEH1L, RRAGC, WDR24, and MIOS.

The recapitulation of the p53 pathway is robust in DRIVE. Known negative regulators (e.g., MDM2, MDM4, USP7, PPM1D) all show dependence effects in the WT p53 setting. Conversely, knockdown of p53 or positive regulators (e.g., TP53BP1, USP28, and CDKN1A) promote the growth of cells harboring WT p53. As an illustration of progressive cluster build using the DSN interactive tool, we show the p53 network when a range of 2–10 nearest neighbors (NNs) is used (Figure S4). Here, as the neighborhood is grown from 2 NNs to 10, the absolute correlation coefficients (ranging from 0.35 to 0.86) decrease as indicated by the thickness of the lines between genes. As such, all members of the 2, 3, 4, and 5 NN networks are known components of the p53 pathway whereas new associations seen in the 10 NN network are less robust (by definition) and are not known p53 pathway members.

One can also observe network relationships between genes that appear to have similar functional consequences when depleted from cells. For instance, the PRC2 complex members form a tight interaction network suggesting very similar functional consequences of EZH2, EED, and SUZ12 knockdown consistent with their function as a unit. In addition, the NN location of the MLL/menin (MEN1) and ASH1L methyltransferase complexes suggests a functional concordance between these proteins and the PRC2 complex. Interestingly, the histone acetyltransferases p300 and the MOZ complex, comprised in part by KAT6A (MYST3, MOZ) and BRPF1, is located nearby to all three methyltransferases suggesting an interplay of acetylation and methylation with likely similar effects on cancer dependence.

Two cell cycle neighborhoods are observed within the DSN. The CCNE1/CDK2 axis is active in cell lines with low RB expression or high CCNE1 expression. Here, E2F3 appears uniquely required for CDK2-mediated cell cycle progression along with SKP2 and CKS1B. These proteins collaborate to degrade p27 (CDKN2B) (Ganoth et al., 2001), hence their knockdown would likely stabilize p27 leading to CDK2 inhibition. Interestingly, an E2 enzyme, UBE2Q1, is found in the CDK2 node suggesting that it may take part in p27 degradation. Not surprisingly, anti-correlations with CDKN2A and RB1 are also observed. A second sub-network of cell-cycle control genes regulating G2/M is observed in the DSN. This includes Cyclin A2 and B1 along with FOXM1 that controls transcriptional events required for this transition including upregulation of CDC25B that controls activation of CDK1. Interference with the timely activation of the anaphase promoting complex (FBXO5) or the mitotic spindle checkpoint (BUB1B, MAD2L1) has apparently deleterious effects on a similar group of cell lines. BTG1, a G0/G1 cell cycle inhibitor, is anti-correlated within this network, consistent with its role in inhibiting early cell cycle progression.

The critical balance between BCL2 family members has been highlighted as both MCL1 and BCL2L1 (BCLXL) demonstrate complex dependency profiles with expression of both pro- and

anti-apoptotic family members in order to maintain cell survival. WSB2 (WD40 repeat and SOCS box protein) has an outlier profile that does not contain robust features predictive of dependence yet does demonstrate multiple connections to the apoptosis pathway in the DSN. The most robust association is anti-correlation of WSB2 and BAX. WSB2-positive correlations include MCL1 and C8orf33. WSB2 is thought to be a component of DDB1/CUL4 E3 ligase complexes (He et al., 2006). The anti-correlation between WSB2 and BAX raises the possibility that WSB2 acts as a negative regulator of BAX protein levels through ubiquitin-mediated degradation.

In addition to pathway and protein complex networks, genes with lineage-specific activity can be detected as DSN neighborhoods. In the melanoma sub-network, nodes are observed that highlight both TF dependencies (*MITF* and *SOX10*) and also oncogenic signaling pathways. The MAPK node is robust (*BRAF*, *MAPK1*) likely due to the prevalence of mutant *BRAF*. However, the importance of fine tuning this pathway is also highlighted as inhibition of the negative pathway regulators, *PEA15* and *DUSP4*, is similarly detrimental. The colon neighborhood, likely dominated by *APC* mutation, includes the obligate CTNNB1 partner, TCF7L2, as well as a collection of TFs including *GATA6*, *SMAD4*, and *YAP1*.

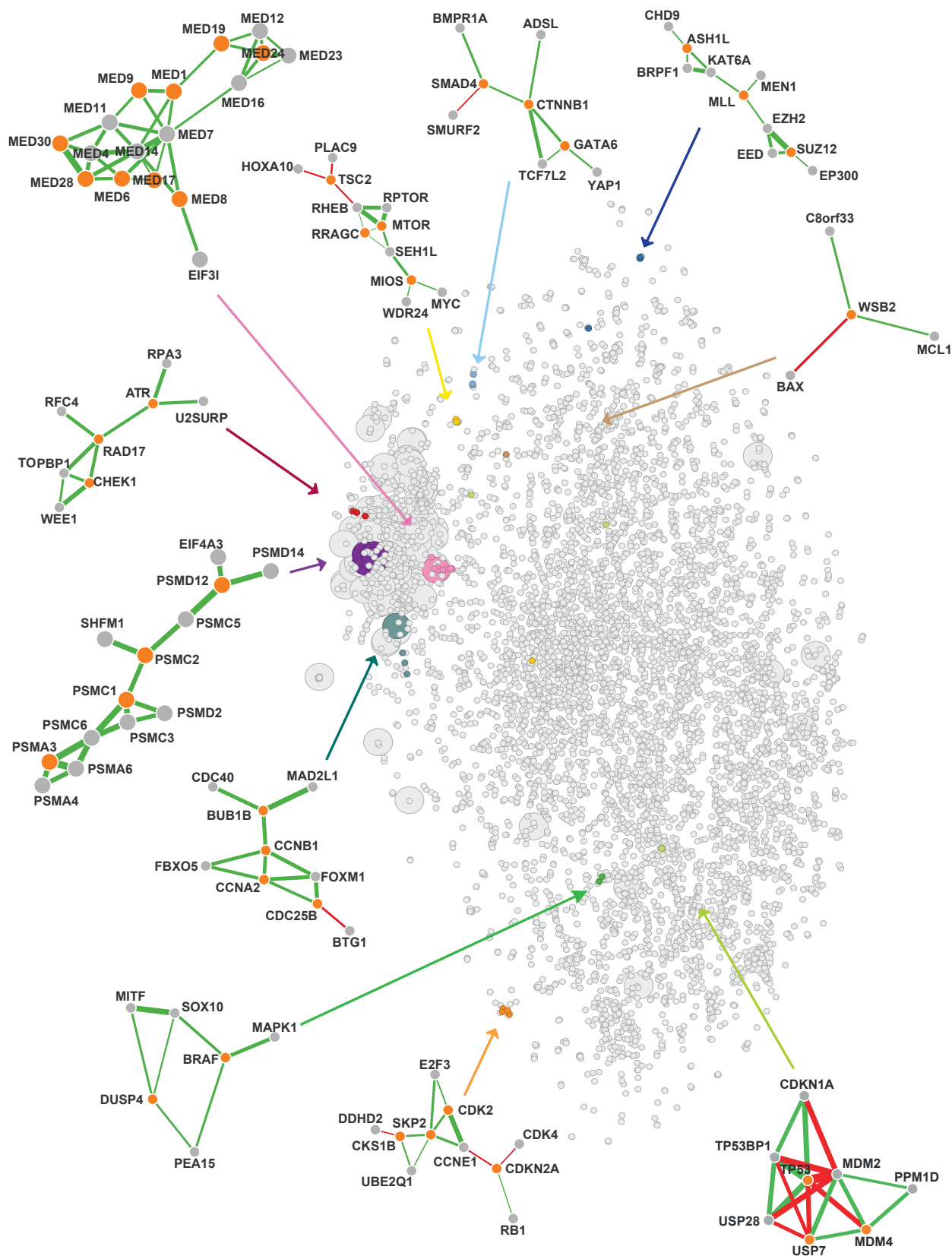
DISCUSSION

In Project DRIVE, a high-density shRNA library was used to assess the consequences of gene depletion for half of the expressed genome of the CCLE in 398 cancer cell lines. Here we provide the data and tools for interrogation as a resource for the further elucidation of therapeutic targets in cancer. The breadth of cell line coverage allows for a diversity of genomic backgrounds to be probed and for relationships with similar functional consequences to be identified. In addition, network relationships for essential genes can be assembled due to incomplete gene depletion by RNA interference. The robustness of the dependence data and in particular the nodal relationship of known protein complexes and pathways in the DRIVE Sensitivity Network suggests that the drawbacks of RNAi, namely off-target effects and inefficient target knockdown, were significantly addressed by the use of appropriate deep coverage libraries and computational methods. We retrospectively conducted a power analysis to compare our ability to detect outliers with a decreasing number of shRNAs/gene (Figure S5), allowing us to estimate the optimal number of shRNA/gene to recover the majority of the outliers for a dataset of this size. In this case, 15 shRNAs captures most of the observations seen with 20. This is consistent with a previous estimate made from one experimental system (Bassik et al., 2013). While CRISPR can more accurately model the null phenotype, there are acknowledged shortcomings such as the need to express a nuclease (e.g., Cas9) and off-target lethality stemming from excessive cuts in amplified genomic loci (Aguirre et al., 2016; Munoz et al.,

(C) Collateral lethal outliers are listed with depiction of concept. Tumor suppressor (TS) deletion leads to collateral loss of nearby genes (B and Y) creating a dependency due to loss of B or Y. VPS4A ATARIS waterfall plot is shown, colored by VPS4B copy number.

(D) Synthetic lethal outlier relationships identified in DRIVE associated with tumor suppressor alterations.

See also Figures S2 and S3.



(legend on next page)

2016). Moreover, RNAi-mediated discovery of synthetic lethal relationships with partial reduction of essential genes, such as in the case of PRMT5, demonstrate that hypomorphs are needed in some circumstances. As such, large-scale RNA interference and CRISPR datasets are likely to be complementary.

To further consider the robustness of this shRNA screen, we have compared small datasets generated by CRISPR studies to the results seen in DRIVE (Munoz et al., 2016). Here, we find a consistent and significant overlap of selective dependencies when comparing shRNA and CRISPR screens across five models with libraries of comparable content and depth. While CRISPR did uncover more essential genes, every essential gene scored by shRNA was also found by CRISPR, resulting in zero shRNA false positives. As further orthogonal validation of the DRIVE dataset, we have also intersected selected DRIVE profiles with sensitivity calls from pharmacological treatment with a number of small molecule inhibitors with clinically validated modes of action, namely EGFR (erlotinib), ERBB2 (lapatinib), BRAF (PLX4720), and PIK3CA (BYL719, alpelisib). We observed a very significant association of models sensitive to genetic target modulation (DRIVE) and sensitive to small molecule inhibitor treatment across the respective pairs (Figure S6). Taken together, the technical performance of deep shRNA screening has been validated by several orthogonal means and provides clear evidence for the robustness of DRIVE.

The molecular characterization of the CCLE allows for the exploration of correlations between gene dependence and genetic, expression, or other features of the cell line set. Two out of our three defined outlier classes (genetic drivers, expression-based drivers) are self-associations, meaning they are driven by characteristics of the target of the shRNA itself. The synthetic lethal class represents relationships in which a non-self association is most highly correlated with the shRNA knock-down phenotype. These can be further broken down by the specific relationships uncovered by DRIVE: pathway, paralog, or collateral lethality. The metabolism class also likely represents synthetic lethal relationships but the molecular features to explain the sensitivity are not yet characterized or are unclear. Two metabolic neighborhoods found in the network analysis (SCAP/SCD/SREBF1 and ASNS/ATF4/EIF2AK4) represent known biological pathways yet no currently available molecular feature explains sensitivity to these nodes. Similarly, dependence on WSB2 did not correlate with definable genetic or expression features yet the DRIVE network correlations with BCL2 family members form the basis of a testable hypothesis. Since our outlier approach is agnostic to the direction of the population skew, tumor suppressors CDKN2A and p53 are both detected. Knockdown of these tumor suppressors results in growth enhancement in cell lines that are wild-type for their respective genes. Some of the most robust DRIVE/DRIVE corre-

lations and anti-correlations are seen with p53 and its pathway components (Figure 6). Hence, these interactions can help robustly define the biology of the pathway rather than a specific therapeutic target.

Ultimately the goal of this project is to define drug targets for treating cancers within defined patient populations. The combination of the ATARIS gene summary and sensitivity feature prediction pipeline coupled with the outlier approach identifies robust and selective cancer dependencies and outlines a biological hypothesis for the phenotype. It allows for relationships to be discovered beyond those that could be predicted a priori, such as by hotspot mutational analysis. Not surprisingly, in Project DRIVE, we observe both known and novel mutation dependencies that begin to assign functional significance to the cancer mutational landscape. This dataset will also help to ascribe function (or not) to mutated genes unearthed by cancer genome-sequencing efforts that have lacked functional validation. Additionally, DRIVE can further refine what we know about some of the most prevalent lesions in cancer. The search for KRAS synthetic lethal interactions has been stymied by false discovery and the data here raise the likelihood that no single synthetic lethal gene will be found across all KRAS mutant tumors. While we affirm that the majority of KRAS mutant cancer cell lines are indeed dependent on continued expression of KRAS, we observed that some of the most commonly used KRAS mutant models are not KRAS dependent, when interrogated as monolayer cell cultures. Hence, studies of therapeutic approaches aimed at ablating KRAS dependence will need to carefully consider these findings.

The DRIVE dataset adds significantly to our understanding of cancer dependencies and synthetic lethal relationships, including those for the most prevalent tumor suppressors. The homozygous and heterozygous deletion of p53, homozygous deletion of CDKN2A, and homozygous deletion of SMAD4 all result in specific collateral synthetic lethality involving a diverse set of genes (UBC, POLR2A, AURKB, MED11, PRMT5, and VPS4A). For ARID1A/ARID1B, RPL22/RPL22L1, and SMARCA4/SMARCA2, inactivating mutations in the first paralog results in dependence on the second. For CDK4, YAP, and ARF4, lack of expression of their paralogs is predictive of dependence. Given the sub-genome nature of our library and our focus on the top outliers coupled with the expansion of the mammalian genome, this class of paralog relationships is only likely to grow. The identification of synthetic lethal relationships for most major tumor suppressors (Figure 5D) now creates the opportunity for novel therapeutics for these defined patient populations.

For project DRIVE, we have created a public portal of gene dependence profiles (<https://oncology.nibr.shinyapps.io/drive/>) that affords the cancer community a resource to validate or devalidate experimental findings across nearly 400 cell lines

Figure 6. DRIVE Sensitivity Network

DRIVE-DRIVE correlations are plotted using tSNE algorithm. Enlarged circles represent essential genes highlighting an essential gene subnetwork. Specific DSN examples include complexes (proteasome, mediator, epigenetic neighborhood), pathways (ATR, p53, G1/S, G2/M, mTOR, WSB2), and lineages (skin, colon). Each subnetwork is displayed around the global tSNE plot with genes used to seed the subnetwork shown in orange. Top three neighborhood correlations for each seed gene are shown with positive correlations in green and negative correlations in red (the top four correlations are used when building the p53 network). The thickness of the lines between two genes represents the strength of the correlation. Subnetwork seed genes are mapped back onto the global tSNE and colored by pathway to illustrate correlations. See also Figure S4.

rapidly by interactive query (Sensitivity Profile Viz tab). Furthermore, the DRIVE network associations can be used in combination with emerging large-scale datasets (e.g., proteomic, transcriptional) to confirm the discovery of novel complex and pathway components (Network Viz tab). The functional genomics now provided by project DRIVE complement the molecular characterization by TCGA and ICGC to bring us closer to an understanding of the molecular drivers of cancer.

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**
- **METHOD DETAILS**
 - Library Construction and Virus Production
 - Screening Approach
 - Purification of Genomic DNA & PCR for Library Production and Next Generation Sequencing
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - From Raw Counts to shRNA Level Scores to Gene Level Scores
 - Quality Control
 - Identifying Outlier Sensitivity Profiles
 - Feature Association
 - GO Term Enrichment
 - DRIVE Sensitivity Network
 - Coloring Scheme
 - shRNA Power Analysis
 - Compound Sensitivity Calls and Comparison to DRIVE Profiles
- **DATA AND SOFTWARE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2017.07.005>.

AUTHOR CONTRIBUTIONS

Conceptualization, F.H., A.L., M.R.S., T. Schmelzle, W.R.S., and F. Stegmeier. Methodology, D.B., E.B., A.d.W., E.F., J.G., G.R.H., F.H., I.K., A. Kauffmann, A.L., K.J. Mavrakis, E.R.M., N.R., D.A.R., M.R.S., T. Schmelzle, W.R.S., F. Stegmeier, M.S., K.V., O.W., K.Y., and J.Y. Software, A.d.W., D.A.D., R.S.E., D.F., J.G., M.D.J., G.M., T.A.P., B.R., D.A.R., K.V., and J.Y. Formal Analysis, E.B., S.M.B., D.C., M.C.-C., A.d.W., W.F., G.G.G., J.G., F.H., K.E.H., Z.J., M.J., A. Kauffmann, A. Kapoor, J.K., K.J. Mavrakis, E.R.M., R.A.P., M.R.S., T. Schmelzle, S.J.S., F. Stegmeier, V.M.S., R.T., and K.V. Investigation, D.A., E.A., R. Billig, D.B., J.B., J.L.B., R. Barrett, E.B., F.B., R.C., J.X.C., F.S.C., R.S.D., C.D., A.D., W.D., A.F., F.F., J.F., E.F., K.G., Z.G., F.G., V.G., K.H., M.H., G.R.H., T.H., J.A.J., I.K., L.L., Y.L., S.L., Q.L., J.L., A.T.L., K.J. Macchi, T.M., K.J. Mavrakis, P.M., A.M., S.M., T.P., N.R., D.A.R., M.R.S., T. Schouwey, F. Shanahan, Q.S., S.J.S., S.S., C. Stamm, C. Stephan, M.S., M.V., A.C.V., M.W., O.W., J.W., R.W., K.Y., Y.Y., and J.Z. Data Curation, E.B., A.d.W., G.R.H., K.J. Mavrakis, E.R.M., T.A.P., D.A.R., M.R.S., K.V., and J.Y. Writing - Original Draft, A.d.W., E.R.M., and W.R.S. Writing - Reviewing & Editing, A.d.W., F.H., A. Kauffmann, E.R.M., R.A.P., T. Schmelzle, and W.R.S. Visualization, A.d.W., A. Kauffmann, and E.R.M.

Supervision, E.B., G.R.H., A. Kauffmann, K.J. Mavrakis, E.R.M., T.A.P., D.A.R., M.R.S., S.J.S., and M.S. Project Management, H.B., F.H., N.K., A.L., E.L., C.M., V.E.M., J.A.P., T. Schmelzle, W.R.S., and F. Stegmeier.

ACKNOWLEDGMENTS

We would like to thank Tom Etter, Heather Keane, Giuseppe Manco, and Jen Taylor for operational assistance; Martine Marchant for FACS assistance; Linda Hinh, Jane Keiran, Dachuan Lei, Wei Ma, and Bing Zheng for technical assistance, and Mélusine Bleu, Delphine Gorses, Lucy Kirchhofer-Allan, Fanny Marque, Emilie Niemezyk, and Estelle Pfister for cell line characterization. We would like to thank Alan Abrams for Creative Art and Jeff Engelman for critical reading of the manuscript. The following investigators and/or institutes kindly provided cell lines: Kevin Jones (University of Utah), Celeste Simon (UPenn), Shinya Tanaka (Hokkaido University Graduate School of Medicine), M.J. Jager (Leiden University Medical Center), and Bruce Ksander (Schepens Eye Research Institute), Karolinska Institutet, and RIKEN BRC. During their involvement related to this reported work, all authors were employees and shareholders of Novartis.

Received: March 24, 2017

Revised: June 2, 2017

Accepted: July 6, 2017

Published: July 27, 2017

REFERENCES

- Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* 6, 914–929.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Baselga, J., Campone, M., Piccart, M., Burris, H.A., 3rd, Rugo, H.S., Sahmoud, T., Noguchi, S., Gnant, M., Pritchard, K.I., Lebrun, F., et al. (2012). Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N. Engl. J. Med.* 366, 520–529.
- Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J., Horlbeck, M.A., Chen, S., Mann, M., et al. (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* 152, 909–922.
- Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J., Horlbeck, M.A., Chang, W., Cheng, J., et al. (2016). shiny: Web Application Framework for R. R package version 0.14.2. <http://CRAN.R-project.org/package=shiny>
- Bernardo, G.M., Bebek, G., Ginther, C.L., Sizemore, S.T., Lozada, K.L., Miedler, J.D., Anderson, L.A., Godwin, A.K., Abdul-Karim, F.W., Slamon, D.J., and Keri, R.A. (2013). FOXA1 represses the molecular phenotype of basal breast cancer cells. *Oncogene* 32, 554–563.
- Blasco, R.B., Franco, S., Santamaría, D., Cañamero, M., Dubus, P., Charron, J., Baccarini, M., and Barbacid, M. (2011). c-Raf, but not B-Raf, is essential for development of K-Ras oncogene-driven non-small cell lung carcinoma. *Cancer Cell* 19, 652–663.
- Bondurand, N., and Sham, M.H. (2013). The role of SOX10 during enteric nervous system development. *Dev. Biol.* 382, 330–343.
- Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.
- Cheung, H.W., Cowley, G.S., Weir, B.A., Boehm, J.S., Rusin, S., Scott, J.A., East, A., Ali, L.D., Lizotte, P.H., Wong, T.C., et al. (2011). Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc. Natl. Acad. Sci. USA* 108, 12372–12377.
- Coultas, L., Pellegrini, M., Visvader, J.E., Lindeman, G.J., Chen, L., Adams, J.M., Huang, D.C., and Strasser, A. (2003). Bf1: a novel weakly proapoptotic

member of the Bcl-2 protein family with a BH3 and a BH2 region. *Cell Death Differ.* 10, 185–192.

Cyr, A.R., Kulak, M.V., Park, J.M., Bogachek, M.V., Spanheimer, P.M., Woodfield, G.W., White-Baer, L.S., O'Malley, Y.Q., Sugg, S.L., Olivier, A.K., et al. (2015). TFAP2C governs the luminal epithelial phenotype in mammary development and carcinogenesis. *Oncogene* 34, 436–444.

Dey, P., Baddour, J., Muller, F., Wu, C.C., Wang, H., Liao, W.T., Lan, Z., Chen, A., Gutschner, T., Kang, Y., et al. (2017). Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. *Nature* 542, 119–123.

Downward, J. (2015). RAS synthetic lethal screens revisited: Still seeking the elusive prize? *Clin. Cancer Res.* 21, 1802–1809.

Flaherty, K.T., Robert, C., Hersey, P., Nathan, P., Garbe, C., Milhem, M., Demidov, L.V., Hassel, J.C., Rutkowski, P., Mohr, P., et al.; METRIC Study Group (2012). Improved survival with MEK inhibition in BRAF-mutated melanoma. *N. Engl. J. Med.* 367, 107–114.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 45 (D1), D777–D783.

Fritsch, C., Huang, A., Chatenay-Rivauday, C., Schnell, C., Reddy, A., Liu, M., Kauffmann, A., Guthy, D., Erdmann, D., De Pover, A., et al. (2014). Characterization of the novel and specific PI3K α inhibitor NVP-BYL719 and development of the patient stratification strategy for clinical trials. *Mol. Cancer Ther.* 13, 1117–1129.

Ganoth, D., Bornstein, G., Ko, T.K., Larsen, B., Tyers, M., Pagano, M., and Herskho, A. (2001). The cell-cycle regulatory protein Cks1 is required for SCF(Skp2)-mediated ubiquitinylation of p27. *Nat. Cell Biol.* 3, 321–324.

Grooteclaes, M.L., and Frisch, S.M. (2000). Evidence for a function of CtBP in epithelial gene regulation and anoikis. *Oncogene* 19, 3823–3828.

He, Y.J., McCall, C.M., Hu, J., Zeng, Y., and Xiong, Y. (2006). DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. *Genes Dev.* 20, 2949–2954.

Helming, K.C., Wang, X., Wilson, B.G., Vazquez, F., Haswell, J.R., Manchester, H.E., Kim, Y., Kryukov, G.V., Ghandi, M., Aguirre, A.J., et al. (2014). ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat. Med.* 20, 251–254.

Hoffman, G.R., Rahal, R., Buxton, F., Xiang, K., McAllister, G., Frias, E., Bagdasarian, L., Huber, J., Lindeman, A., Chen, D., et al. (2014). Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl. Acad. Sci. USA* 111, 3128–3133.

Hortobagyi, G.N., Stemmer, S.M., Burris, H.A., Yap, Y.S., Sonke, G.S., Paluch-Shimon, S., Campone, M., Blackwell, K.L., André, F., Winer, E.P., et al. (2016). Ribociclib as first-line therapy for HR-positive, advanced breast cancer. *N. Engl. J. Med.* 375, 1738–1748.

Hubert, C.G., Bradley, R.K., Ding, Y., Toledo, C.M., Herman, J., Skutt-Kakaria, K., Girard, E.J., Davison, J., Berndt, J., Corrin, P., et al. (2013). Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev.* 27, 1032–1045.

Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabé, R.R., Bhan, M.K., Calvo, F., Eerola, I., Gerhard, D.S., et al.; International Cancer Genome Consortium (2010). International network of cancer genome projects. *Nature* 464, 993–998.

Ishihara, T., Tsuda, H., Hotta, A., Kozaki, K., Yoshida, A., Noh, J.Y., Ito, K., Imoto, I., and Inazawa, J. (2008). ITCH is a putative target for a novel 20q11.22 amplification detected in anaplastic thyroid carcinoma cells by array-based comparative genomic hybridization. *Cancer Sci.* 99, 1940–1949.

Jackson, C.L., and Bouvet, S. (2014). Arfs at a glance. *J. Cell Sci.* 127, 4103–4109.

Karreth, F.A., Frese, K.K., DeNicola, G.M., Baccarini, M., and Tuveson, D.A. (2011). C-Raf is required for the initiation of lung cancer by K-Ras(G12D). *Cancer Discov.* 1, 128–136.

Kim, J., McMillan, E., Kim, H.S., Venkateswaran, N., Makkar, G., Rodriguez-Canales, J., Villalobos, P., Neggers, J.E., Mendiratta, S., Wei, S., et al. (2016). XPO1-dependent nuclear export is a druggable vulnerability in KRAS-mutant lung cancer. *Nature* 538, 114–117.

König, R., Chiang, C.Y., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A., Bergauer, T., Orth, A., Krueger, U., Zhou, Y., and Chanda, S.K. (2007). A probability-based approach for the analysis of large-scale RNAi screens. *Nat. Methods* 4, 847–849.

Kryukov, G.V., Wilson, F.H., Ruth, J.R., Paulk, J., Tsherniak, A., Marlow, S.E., Vazquez, F., Weir, B.A., Fitzgerald, M.E., Tanaka, M., et al. (2016). MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science* 351, 1214–1218.

Lupien, M., and Brown, M. (2009). Cistromics of hormone-dependent cancer. *Endocr. Relat. Cancer* 16, 381–389.

Marcotte, R., Brown, K.R., Suarez, F., Sayad, A., Karamboulas, K., Krzyzanowski, P.M., Sircoulomb, F., Medrano, M., Fedyshyn, Y., Koh, J.L., et al. (2012). Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov.* 2, 172–189.

Marjon, K., Cameron, M.J., Quang, P., Clasquin, M.F., Mandley, E., Kunii, K., McVay, M., Choe, S., Kernysky, A., Gross, S., et al. (2016). MTAP deletions in cancer create vulnerability to targeting of the MAT2A/PRMT5/RIK1 axis. *Cell Rep.* 15, 574–587.

Mavrakis, K.J., McDonald, E.R., 3rd, Schlabach, M.R., Billy, E., Hoffman, G.R., deWeck, A., Ruddy, D.A., Venkatesan, K., Yu, J., McAllister, G., et al. (2016). Disordered methionine metabolism in MTAP/CDKN2A-deleted cancers leads to dependence on PRMT5. *Science* 351, 1208–1213.

Mayer, I.A., Abramson, V.G., Formisano, L., Balko, J.M., Estrada, M.V., Sanders, M.E., Juric, D., Solit, D., Berger, M.F., Won, H.H., et al. (2017). A phase Ib study of alpelisib (BYL719), a PI3K α -specific inhibitor, with letrozole in ER+/HER2- metastatic breast cancer. *Clin. Cancer Res.* 23, 26–34.

McCullough, J., Colf, L.A., and Sundquist, W.I. (2013). Membrane fission reactions of the mammalian ESCRT pathway. *Annu. Rev. Biochem.* 82, 663–692.

Muller, F.L., Colla, S., Aquilanti, E., Manzo, V.E., Genovese, G., Lee, J., Eisen-son, D., Narurkar, R., Deng, P., Nezi, L., et al. (2012). Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature* 488, 337–342.

Munoz, D.M., Cassiani, P.J., Li, L., Billy, E., Korn, J.M., Jones, M.D., Golji, J., Ruddy, D.A., Yu, K., McAllister, G., et al. (2016). CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* 6, 900–913.

Nijhawan, D., Zack, T.I., Ren, Y., Strickland, M.R., Lamothe, R., Schumacher, S.E., Tsherniak, A., Besche, H.C., Rosenbluh, J., Shehata, S., et al. (2012). Cancer vulnerabilities unveiled by genomic loss. *Cell* 150, 842–854.

Novetsky, A.P., Zigelboim, I., Thompson, D.M., Jr., Powell, M.A., Mutch, D.G., and Goodfellow, P.J. (2013). Frequent mutations in the RPL22 gene and its clinical and functional implications. *Gynecol. Oncol.* 128, 470–474.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 132, 631–644.

Otto, T., and Sicinski, P. (2017). Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* 17, 93–115.

Pagliarini, R., Shao, W., and Sellers, W.R. (2015). Oncogene addiction: pathways of therapeutic response, resistance, and road maps toward a cure. *EMBO Rep.* 16, 280–296.

Pardanani, A.D., Levine, R.L., Lasho, T., Pikman, Y., Mesa, R.A., Wadleigh, M., Steensma, D.P., Elliott, M.A., Wolansky, A.P., Hogan, W.J., et al. (2006). MPLS15 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood* 108, 3472–3476.

Pasqualucci, L., and Dalla-Favera, R. (2015). The genetic landscape of diffuse large B-cell lymphoma. *Semin. Hematol.* 52, 67–76.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

Rosenbauer, F., and Tenen, D.G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. *Nat. Rev. Immunol.* 7, 105–117.

- Shaffer, A.L., Emre, N.C., Lamy, L., Ngo, V.N., Wright, G., Xiao, W., Powell, J., Dave, S., Yu, X., Zhao, H., et al. (2008). IRF4 addiction in multiple myeloma. *Nature* 454, 226–231.
- Shao, D.D., Tsherniak, A., Gopal, S., Weir, B.A., Tamayo, P., Stransky, N., Schumacher, S.E., Zack, T.I., Beroukhi, R., Garraway, L.A., et al. (2013). ATARIS: computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Res.* 23, 665–678.
- Theurillat, J.P., Metzler, S.C., Henzi, N., Djouder, N., Helbling, M., Zimmermann, A.K., Jacob, F., Soltermann, A., Caduff, R., Heinzelmann-Schwarz, V., et al. (2011). URF is an oncogene amplified in ovarian cancer cells and is required for their survival. *Cancer Cell* 19, 317–332.
- Tracey, W.D., and Speck, N.A. (2000). Potential roles for RUNX1 and its orthologs in determining hematopoietic cell fate. *Semin. Cell Dev. Biol.* 11, 337–342.
- van der Maaten, L.J.P., and Hinton, G.E. (2008). Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Wee, S., Wiederschain, D., Maira, S.M., Loo, A., Miller, C., deBeaumont, R., Stegmeier, F., Yao, Y.M., and Lengauer, C. (2008). PTEN-deficient cancers depend on PIK3CB. *Proc. Natl. Acad. Sci. USA* 105, 13057–13062.
- Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M.; Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120.
- Zimmermann, G., Papke, B., Ismail, S., Vartak, N., Chandra, A., Hoffmann, M., Hahn, S.A., Triola, G., Wittinghofer, A., Bastiaens, P.I., and Waldmann, H. (2013). Small molecule inhibition of the KRAS-PDE δ interaction impairs oncogenic KRAS signalling. *Nature* 497, 638–642.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
TransIT-293 Transfection Reagent	Mirus	Cat # MIR2700
Puromycin	ThermoFisher Scientific	Cat # A1113802
Polybrene	Millipore	Cat # TR-1003-G
Titanium Taq	Takara	Cat # 639242
Critical Commercial Assays		
gDNA extraction: QIAamp DNA Blood Maxi Spin	QIAGEN	Cat # 51192
PCR purification: Agencourt AMPure XP PCR Cleanup	Beckman Coulter	Cat # A63882
Deposited Data		
Raw NGS data	This paper; Mendeley data portal	https://data.mendeley.com/datasets/y3ds55n88r/1
Web based sensitivity profile and network tool (processed data)	This paper	https://oncologynibr.shinyapps.io/drive/
Experimental Models: Cell Lines		
HEK293T/17 cells (for virus production)	ATCC	Cat # CRL-11268
383 cancer cell lines from original CCLE collection (Table S2)	Various commercial suppliers; see original CCLE publication for individual sources (Barretina et al., 2012)	https://portals.broadinstitute.org/ccle/home
15 non-CCLE cell lines (Table S2)	Various investigators and institutes	N/A
Oligonucleotides		
Primers for PCR indexing, see Table S6 (25 oligos)	This paper	N/A
Recombinant DNA		
55K_PoolA	Cellecta	Cat # 55K_PoolA-MS-NOVA
55K_PoolB	Cellecta	Cat # 55K_PoolB-MS-NOVA
27K_BGP2	Cellecta	Cat # 27KBGP2-MS-NOVA
13K-hTF	Cellecta	Cat # 13K-hTF-GH-NOVA
13K-hYAP	Cellecta	Cat # 13K-hYAP-GH-NOVA
13K-hEPI2	Cellecta	Cat # 13K-hEPI2-GHNOVA
pRS116cb-U6-sh-13kCB18-HTS6-UbiC-TagRFP-2A-Puro vector	Cellecta	Cat # SVSHU616-L
Cellecta packaging plasmids (psPAX2 and pMD2)	Cellecta	Cat # CPCP-K2A
Software and Algorithms		
EdgeR	Bioconductor	http://bioconductor.org/packages/edgeR
Limma	Bioconductor	http://bioconductor.org/packages/limma
R shiny	CRAN	https://CRAN.R-project.org/package=shiny
ATARIS algorithm	PubMed	23269662
RSA algorithm	PubMed	17828270
Rtsne	CRAN	https://CRAN.R-project.org/package=Rtsne
ggplot2	CRAN	https://CRAN.R-project.org/package=ggplot2
Other		
CellSTACK (5 layer)	Corning	Cat# 22250-152

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, Tobias Schmelzle (tobias.schmelzle@novartis.com). The DRIVE library is available from Collecta as outlined below. Cell lines can be attained directly from the original supplier.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

383 of the cell lines are part of the original Cancer Cell Line Encyclopedia (CCLE) and were handled accordingly while the remaining 15 lines were kindly provided by investigators (Table S2). All models are regularly tested for being free of mycoplasma and their identity verified, both for the banked stocks (i.e., starting material for screen) as well as upon completion of screen (extracted gDNA for NGS). A 79 SNP identification panel has been uploaded to the CCLE website by the CCLE team to allow investigators to readily identify CCLE cell lines.

METHOD DETAILS

Library Construction and Virus Production

The DRIVE library was synthesized by Collecta as independent pools and are available for order from Collecta using the following library ID numbers: 55K_PoolA-MS-NOVA; 55K_PoolB-MS-NOVA; 27KBGP2-MS-NOVA; 13K-hTF-GH-NOVA; 13K-hYAP-GH-NOVA; 13K-hEPI2-GHNOVA. Pool A, Pool B and a combination of the other libraries (BGP2, TF, YAP, EPI2 combination is referred to as BGPD pool) were cloned in the pRS16cb-U6-sh-13kCB18-HTS6-UbiC-TagRFP-2A-Puro vector. The DRIVE viral packaging was scaled up to 5-layer Cellstack (Corning, Cat No, 22250-152). 2.1x10⁸ 293T cells were plated on one 5-layer Cellstack 24hrs prior to transfection. Cells were transfected according to the manufacturer's recommended protocol (Collecta). For each 5-layer Cellstack, cells were transfected using 510.3 μ L of TransIT reagent diluted into 18.4 mL of OPTI-MEM that was combined with 75.6 μ g of the plasmid pool and 94.5 μ g of the Collecta packaging mix (containing the psPAX2 and pMD2 plasmids that encode Gag/Pol and VSV-G respectively). Virus was harvested at 72hrs post transfection, aliquoted, and frozen at -80°C for later use. Viral titers were measured and benchmarked against a reference virus that was used to assess infectability of each screened cell line.

Screening Approach

For each cell line the optimal puromycin dose required to achieve > 95% cell killing in 72 hr was determined by measuring cell viability with a Cell Titer Glo or a Methylene blue staining assay for a 6-point dose response ranging from 0 to 8 μ g of puromycin. The volume of virus required to give an MOI of 0.5 (for delivery of only one shRNA per cell) was determined using a 10 point dose response ranging from 0 to 400 μ L of viral supernatant in the presence of 8 μ g/ml polybrene. Infectivity was determined using titrated virus from the library backbone vector expressing RFP and measuring the % RFP-positive cells by FACS 4 days post-infection after three days of puromycin selection. Cell lines that could not pass these screening characterization criteria (puromycin sensitivity, reasonable infectivity and RFP positivity) did not enter the final screening queue.

The three DRIVE libraries were screened as independent pools across 398 cell lines. In short, cells were infected to maintain 1000x library representation during the duration of the 2 week viability screen. For large-scale infections, 90 million cells (per library) were plated 24 hr prior to infection in 5-layer CellSTACK culture chambers (Corning) in 500 mL of medium (one CellSTACK per library). On the day of infection, the culture media was replaced with 500 mL of fresh media containing 8 μ g/ml polybrene and required volume of virus for MOI of 0.5 was added. 24 hr after infection the culture media was replaced with 500 mL fresh media containing puromycin at the cell line-specific concentration. 72 hr following puromycin addition, cells were trypsinized, and 70-90 million cells were re-plated in 5-layer CellSTACK culture chambers. An aliquot of cells was used to measure transduction efficiency determined by measuring the % RFP positive cells and was typically > 90%–95%. Cells were maintained in culture and passaged as needed to ensure they did not exceed 90% confluence during the course of the screen. At each split, 70-90 million cells were passaged into new flasks, ensuring a representation of > 1000 cells/shRNA in the library and the % RFP positive cells was measured to ensure stability of the transduced population over time. 14 days after infection, cells were trypsinized and samples of 70-90 million cells were harvested by centrifugation and stored at -80°C prior to gDNA extraction.

Purification of Genomic DNA & PCR for Library Production and Next Generation Sequencing

Cell pellets were processed according to the QIAamp® DNA Blood Maxi Spin Protocol (Cat # 51192), and the resulting genomic DNA resuspended in 2mls QIAGEN buffer AE. Genomic DNA concentrations were measured using a Picogreen dye-binding assay giving a typical yield of 1 μ g gDNA per million cells. For Next Generation Sequencing (NGS) library generation, the barcodes are PCR amplified in 24 independent 100 μ L PCR reactions using 4 μ g of input gDNA per reaction with Titanium Taq, a single forward primer and one of 24 indexing oligos (as listed in Table S6) for 30 cycles. Library input DNA was also sequenced and referred to as plasmid counts in the provided raw data. 24 independent PCR reactions were pooled and purified using the Agencourt AMPure XP PCR cleanup kit (Beckman Coulter). The resulting products were analyzed by agarose e-gel to confirm the expected ~190bp product and the amount of purified product quantified using the Advanced Analytical Fragment Analyzer. Barcode representation was measured on the

Illumina 2500 platform. For good representation of each shRNA in the NGS data, 40–60 million raw Illumina sequence reads were required per sample averaging approximately 1000 reads per shRNA. Note that the individual plasmid pools for each shRNA library were spiked into each NGS flowcell at 15% of the total loading volume as normalization controls.

QUANTIFICATION AND STATISTICAL ANALYSIS

From Raw Counts to shRNA Level Scores to Gene Level Scores

The drop-out value for each shRNA was calculated using the Bioconductor R package EdgeR (Robinson et al., 2010). The plasmid and sample raw counts per shRNAs were normalized in pairs using the Timed Mean of M-values (TMM) normalization. In the rare event that the plasmid spike-in failed to generate sufficient counts (< 20 million total reads per plasmid) to be used as a normalization control then a virtual library was used in place of plasmid counts. This virtual library is obtained by equalizing the library size of the available plasmid libraries counts using the `equalizeLibSizes` function in edgeR and rounding the mean pseudo count of the resulting output to the nearest integer. The edgeR negative binomial model is fitted to obtain the log fold change (logFC) of counts between the sample and plasmid (the parameters used are common dispersion = 0.2 and prior count = 12). This is performed for each sample-plasmid pair to obtain a logFC per shRNA per cell line. Additionally this procedure is performed on each pool individually. The logFC are then normalized per sample using a quantile normalization to obtain a shRNA level sensitivity score. The shRNA level scores are further aggregated to gene level sensitivity scores using either the ATARIS algorithm (Shao et al., 2013) or the RSA algorithm (König et al., 2007). RSA uses all the shRNAs per gene to give a measure of the statistical significance of the drop-out of those 20 shRNAs compared to the background of the rest of the shRNAs logFC. The method is directional and only captures shRNA drop-out, not growth enhancing shRNAs. Additionally every gene incorporated in the screen is aggregated into a RSA score. ATARIS only uses the shRNAs which display consistent profiles across the 398 samples, thus mitigating the potential issue of off-targets. Additionally, inert genes will not generate ATARIS scores since the shRNAs are unlikely to correlate. The ATARIS algorithm can serially produce multiple solutions depending on the number of clusters of shRNA showing consistent phenotype across the cell lines. In those cases only the first solution is reported which generally represents the solution with the most shRNA (i.e., ~84% of cases). Finally, the procedure from raw counts to gene level scores is performed on each pool independently from each other.

Quality Control

Several QC metrics were considered when running the analysis pipeline. Those metrics were recorded and analyzed in aggregate to identify potential outlier samples. In particular at the read count level the total number of read counts in each sample, the frequency of the most abundant shRNA, and the third quartile of the read counts after library size scaling were considered and recorded. After logFC estimation of the individual shRNA the proportion of pan-lethal genes represented in the lowest quartile within each sample was also considered. The proportion of shRNAs with logFC below the quantile with $p = 0.05$ across samples was kept in order to identify potential hyper-sensitive lines. Finally some shRNAs are present in all three pools and the correlation of those shRNA logFC levels across the pools was assessed. Those metrics were then considered in aggregate to identify outliers which were removed.

Identifying Outlier Sensitivity Profiles

The distribution of the sensitivity scores can be used to identify potential genes of interest. In particular one hopes to prioritize genes which are neither inert nor essential, but display differential sensitivity across samples, i.e., a subset of samples is strongly sensitive to the knockdown, while the rest is unaffected. To identify these profiles a “Likelihood Ratio Test”-based method was used (i.e., Normality LRT) which identifies profiles whose distribution diverge the most from the normal distribution. For each gene both a normal distribution and a skew Student t distribution are fitted to the distribution of sensitivity scores across all cell lines. The Normality LRT score is twice the log of the likelihood ratio of the fitted skewed distribution over the likelihood of the fitted normal distribution, i.e., the difference between the deviance of the two models.

Feature Association

To find potential hypotheses explaining the observed sensitivity profiles an automated separation of the samples population into sensitive and insensitive lines was performed for each gene using a univariate k -means clustering method with $k = 3$. The two extreme populations were designated as sensitive and insensitive, while the intermediate cluster was not used further. A differential representation of the known CCLE features across the two remaining populations was systematically undertaken to identify potential explanatory features (mutation, CNA, gene expression). The type of test performed was dependent on the feature type: Fisher Exact test for mutation, Wilcoxon test for CNA and Bioconductor Limma (Ritchie et al., 2015) for expression. The resulting p values were adjusted for multiple testing within each feature type using the Benjamini Hochberg correction.

GO Term Enrichment

The enrichment of GO terms within the essential genes of the screen compared to the remaining genes of the screen was calculated. For this purpose “essential” genes were defined as genes with 50% or more cell lines displaying an RSA score below -3 upon knock-down. For this analysis only the GO terms with 50 or more genes were considered (not restricted to the genes in the screen). For each remaining biological process GO term a contingency table was built using the number of essential genes within the term, the number

of essential genes outside the term, the number of non-essential genes within the term and the number of non-essential genes outside the term in each of the entry of the 2x2 matrix. A Chi-square test was then performed on the resulting contingency table and the top 5 most enriched Biological Process GO terms among the essential genes were reported. The Benjamini Hochberg procedure was used to correct for multiple testing.

DRIVE Sensitivity Network

The DRIVE data was also used to build a co-sensitivity network (similarly to co-expression networks in expression datasets). In this case genes were used as vertices and the absolute value of the correlation between the sensitivity profiles of two genes defined the similarity of the gene pair (the correlation is performed across all samples using the Pearson correlation). In the DSN interactive tool the strength of the similarity is represented by the weight of the edges linking the genes. To visualize the topology of the whole network in two dimensions a t-distributed Stochastic Neighbor Embedding (tSNE) (van der Maaten and Hinton, 2008) is used whereby the distance between two genes is defined as the square of one minus their similarity. So that the distance matrix δ is defined as

$$\delta = (1 - \sigma)^2 = (1 - |\rho|)^2$$

with σ the similarity matrix and ρ the Pearson correlation matrix for the different gene profiles. t-SNE is a non-linear dimension reduction technique which models nearby points in high dimension by nearby points in the lower dimensional space. This however is generally not possible without information loss. This can be seen in Figure 6 whereby genes which are in each other's direct neighborhood in the high dimensional network (presented in the interactive tool) are not represented close by in the 2D tSNE representation (e.g., TP53, USP7 and MDM4). However some global aspects of the topology of the network are still conserved in particular the distribution of essential genes in the network.

Coloring Scheme

Throughout the various figures a consistent coloring scheme was used whenever we represented CN or expression values. Namely the color is saturated to blue or red among the lowest and highest 10% of samples respectively. The median value is colored light gray and a linear scale is used between the 10th and 50th percentile (blue to gray) and 50th to 90th percentile (gray to red).

shRNA Power Analysis

To perform the shRNA power analysis we used a subset of the BGPD pool, focusing on 1381 genes with 20 and only 20 shRNAs per gene. The ATARiS solutions were built for each of the 1381 genes and the outlierness by Normality LRT was calculated. The top 100 outlier genes were recorded and used for further analysis. The ATARiS calculation was then performed on those 100 genes with 5, 10 and 15 shRNAs selected at random. This process was repeated 10 times with random resampling of the selected shRNAs. In each iteration, we recorded the number of top outlier genes for which an ATARiS solution was found.

Compound Sensitivity Calls and Comparison to DRIVE Profiles

The compound sensitivity calls are based on the results available in the CCLE (Barretina et al., 2012) and in the BYL719 report (Fritsch et al., 2014) whereby the A_{\max} (maximum observed response) and EC_{50} (inflection point) using an 8-point dose-response matrix are combined to assess cell line sensitivity to compound treatment. Starting from the vector of responses A_{\max} or EC_{50} , we considered the distribution of response values (for A_{\max} , log-transformed EC_{50}) in order to assign cell lines into sensitive, and non-sensitive classes using a combination of EC_{50} and A_{\max} cutoffs in a compound specific manner. In particular the cell lines below both of the following cutoffs were deemed sensitive: $A_{\max} = -40\%$ and $EC_{50} = 1.95 \mu\text{M}$ for Erlotinib, $A_{\max} = -40\%$ and $EC_{50} = 1.95 \mu\text{M}$ for PLX4720, $A_{\max} = -40\%$ and $EC_{50} = 1.28 \mu\text{M}$ for Lapatinib and $A_{\max} = -30\%$ and $EC_{50} = 3.04 \mu\text{M}$ for BYL719. For DRIVE the cell lines with phenotypic values below -1 were deemed sensitive. The consistency of compound versus shRNA response was assessed using Fisher's exact tests on the contingency of their respective sensitivity call (i.e., sensitive versus non-sensitive).

DATA AND SOFTWARE AVAILABILITY

The raw data for project DRIVE is archived and publicly available in Mendeley (<https://data.mendeley.com/datasets/y3ds55n88r/1>). The cell line feature data including DNA copy number, mRNA expression and mutation calls is available on the CCLE portal (<https://portals.broadinstitute.org/ccle/home>). Algorithms used for the analyses are described and referenced in the individual "quantification and statistical analysis" subsections. The custom code will be provided upon request to the Lead Contact.

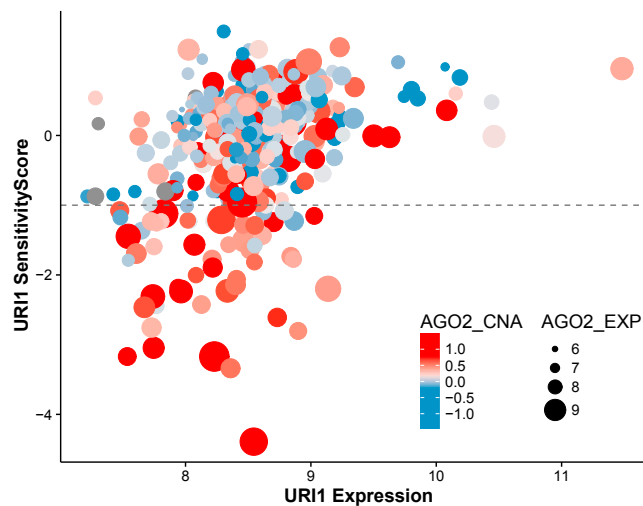


Figure S1. URI1 DRIVE Profile, Related to Figure 2C

URI1 Expression (x axis) is plotted compared to its Sensitivity Score (y axis). Each dot represents a cell line colored by its AGO2 CN and sized by AGO2 expression.

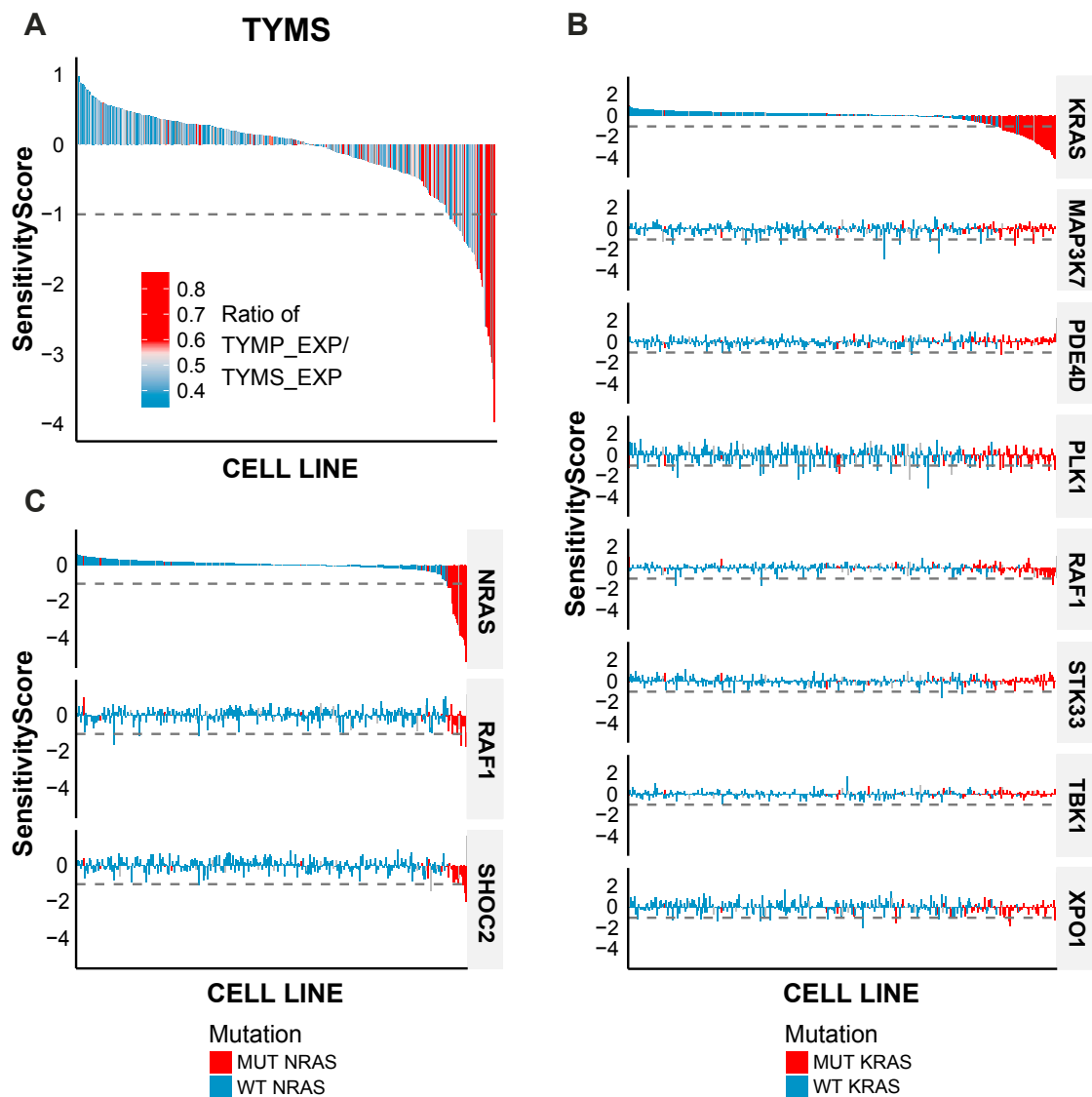


Figure S2. Pathway Synthetic Lethality Outliers, Related to Figure 5

(A) TYMS DRIVE profile. TYMS sensitivity score waterfall plot is colored by the ratio of TYMP/TYMS expression (i.e., high TYMP expression coupled with low TYMS expression indicative of sensitivity). (B) KRAS DRIVE profile with other reported synthetic lethal gene profiles. Cell lines are ordered by KRAS Sensitivity Score and colored by KRAS mutation status. Knockdown of a known MAPK pathway component, RAF1 (cRAF), is shown as a comparator for a gene that partially phenocopies KRAS knockdown. (C) NRAS, RAF1 and SHOC2 DRIVE profiles. Cell lines are ordered by NRAS sensitivity Score and colored by NRAS mutation status. The highest correlating DRIVE profiles to the SHOC2 profile are RAF1 and NRAS.

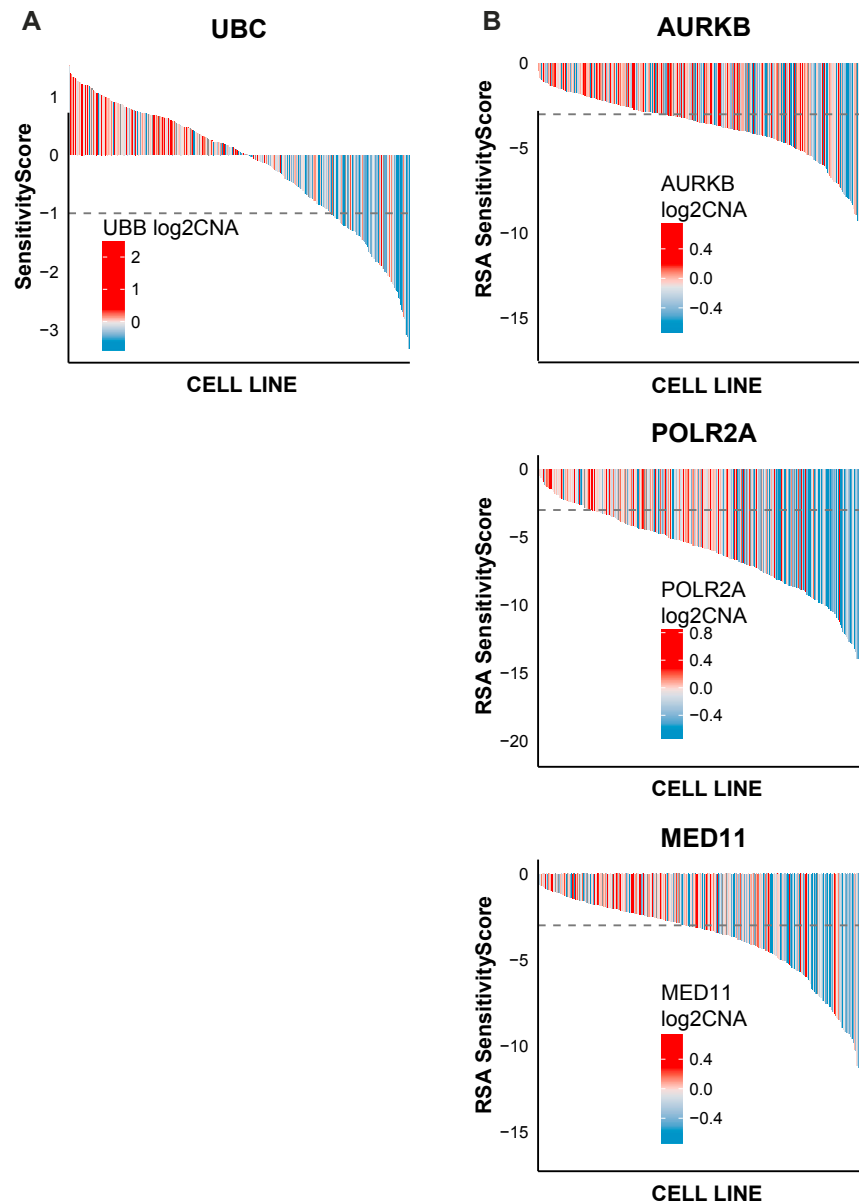


Figure S3. p53 Collateral Synthetic Lethality Examples, Related to Figure 5

(A) UBC DRIVE profile. UBC sensitivity is colored by the CN value of its paralogs, *UBB*. (B) RSA sensitivity plots for POLR2A, MED11 and AURKB. Cell lines are colored by their CN value for each respective gene.

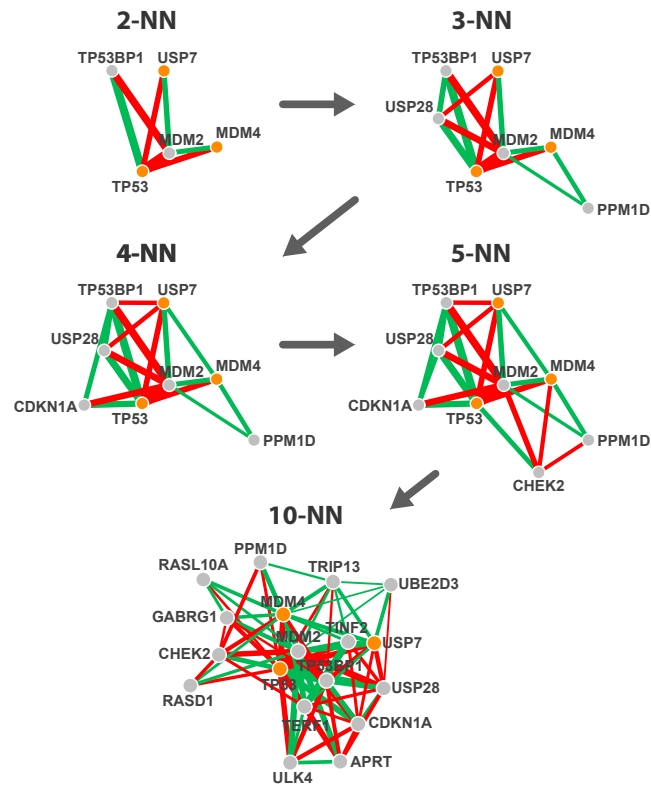


Figure S4. p53 DSN Nearest Neighbor (NN) Analysis, Related to Figure 6

Seed genes (p53, MDM4 and USP7; shown in orange) are used to nucleate the p53 neighborhood using the DSN interactive tool with a variable number of edges per gene (2, 3, 4, 5, and 10). All the edges from non-seed genes are also shown whenever they are part of their respective top N neighbors (2, 3, 4, 5, and 10 respectively). Positive correlations are shown in green and negative correlations in red while the thickness of the lines between two genes represents the strength of the correlation.

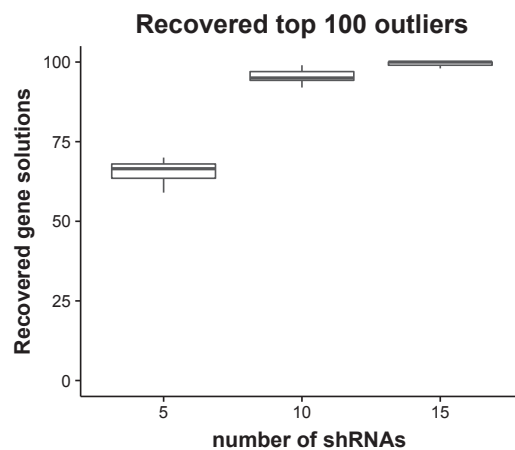


Figure S5. shRNA Power Analysis, Related to STAR Methods "shRNA Power Analysis"

The top 100 Outliers (by Normality LRT) were determined using the ATARIS solutions of 1381 genes containing 20 shRNAs per gene from the BGPD pool. 5, 10, or 15 shRNAs per gene were then randomly sampled for these top 100 outlier genes to generate ATARIS solutions. This process was repeated 10 times and the ability to generate an ATARIS solution using only 5, 10 or 15 shRNA reagents per gene is shown on the y axis.

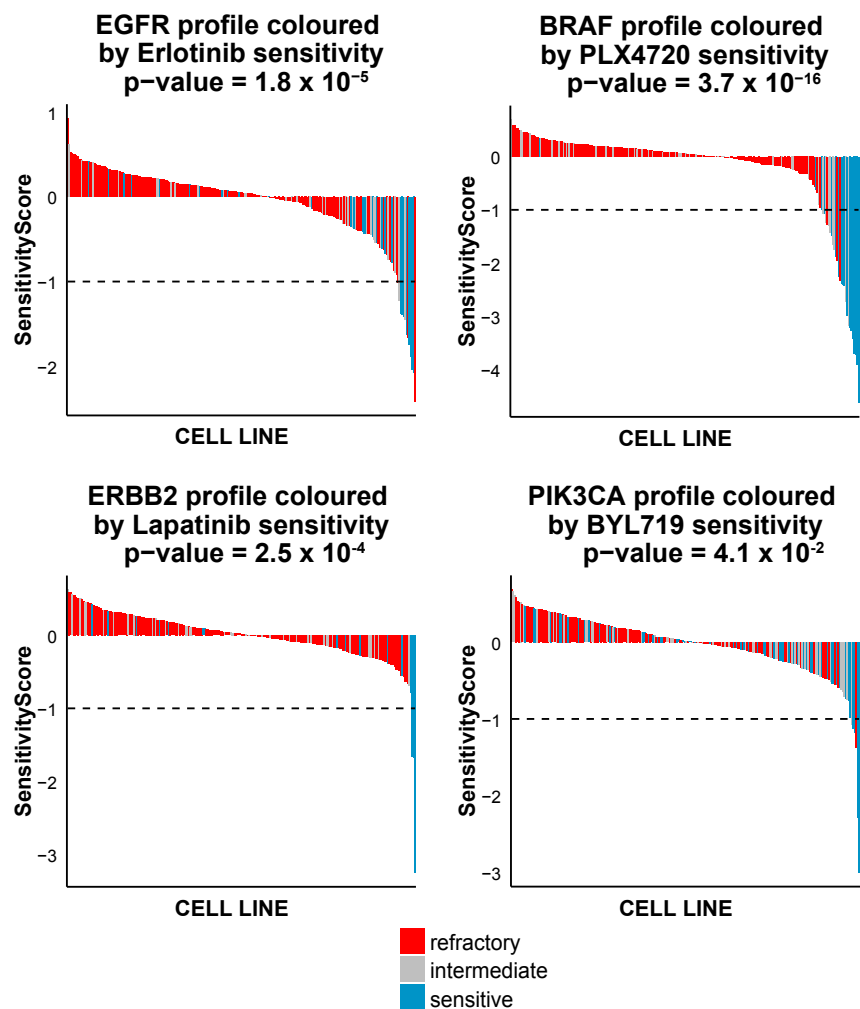


Figure S6. Compound Sensitivity Coloring of DRIVE Profiles, Related to STAR Methods "Compound Sensitivity Calls"
A Fisher's exact test was performed comparing the compound sensitivity call and the DRIVE sensitivity call based on a value < -1 .