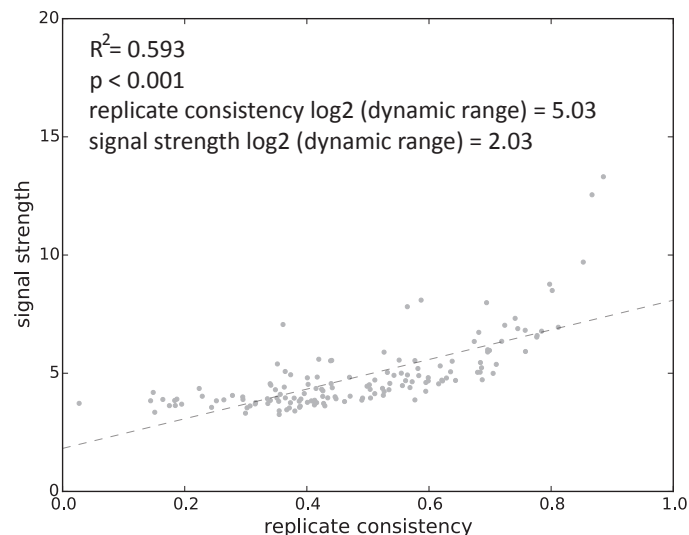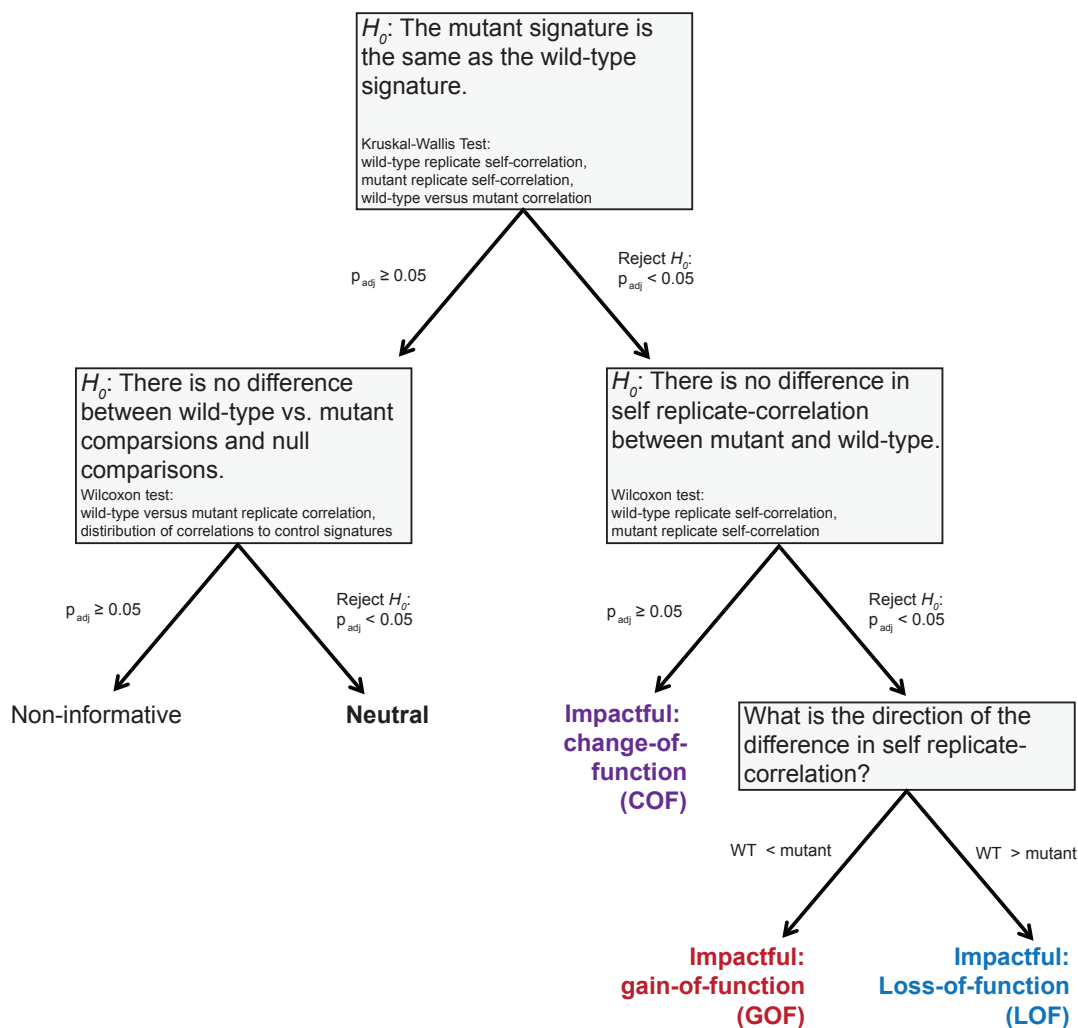# SUPPLEMENTAL DATA

A

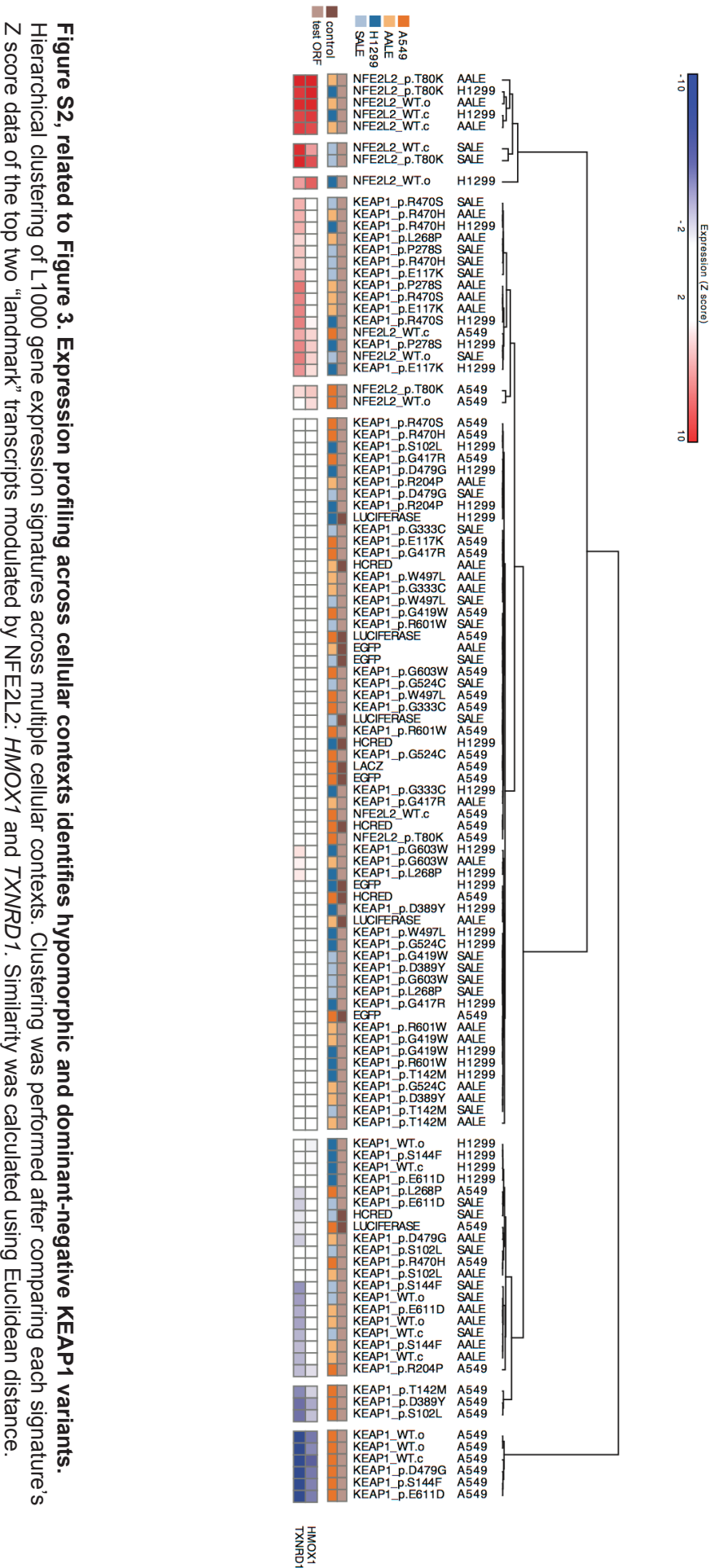**B**

C

Figure S1, related to Figure 1.

(A) Comparison of functional impact calls between SIFT, PolyPhen2 HumDiv, and MutationAssessor for mutations assessed by eVIP in this study.

(B) Relationship between replicate consistency and signal strength in 8-replicate L1000 profiling data. Replicate consistency for each ORF was determined by collapsing the weighted connectivity score (wtcs) of each pair-wise self-replicate comparison. Signal strength is calculated by taking the difference in the median Z-scores of the top 50 most up-regulated and bottom 50 most down-regulated genes in each signature.

(C) Expression-based variant impact phenotyping decision tree. The functional impact of a mutation is determined by comparing gene expression signatures induced by wild-type and mutant ORFs. The decision at the root of the tree gives the overall impact prediction p value for a given mutant variant. To determine if there is a functional impact, the self-correlation between replicate introductions of wild-type constructs are compared with replicates of mutant constructs and also wild-type versus mutant correlation. A Kruskal-Wallis test is performed on these three distributions of correlation to test the null hypothesis that there is no difference in expression signatures between wild-type and mutant.

Table S1, related to Figure 1, supplied as an Excel file.
File S1, related to Figure 1, supplied as an Excel file.

Table S2, related to Figure 2, supplied as an Excel file.
Table S3, related to Figure 2, supplied as an Excel file.
Table S4, related to Figure 2, supplied as an Excel file.
File S2, related to Figure 2, supplied as an Excel file.

Figure S2, related to Figure 3. Expression profiling across cellular contexts identifies hypomorphic and dominant-negative KEAP1 variants. Hierarchical clustering of L1000 gene expression signatures across multiple cellular contexts. Clustering was performed after comparing each signature's Z score data of the top two "landmark" transcripts modulated by NFE2L2: *HMOX1* and *TXNRD1*. Similarity was calculated using Euclidean distance.

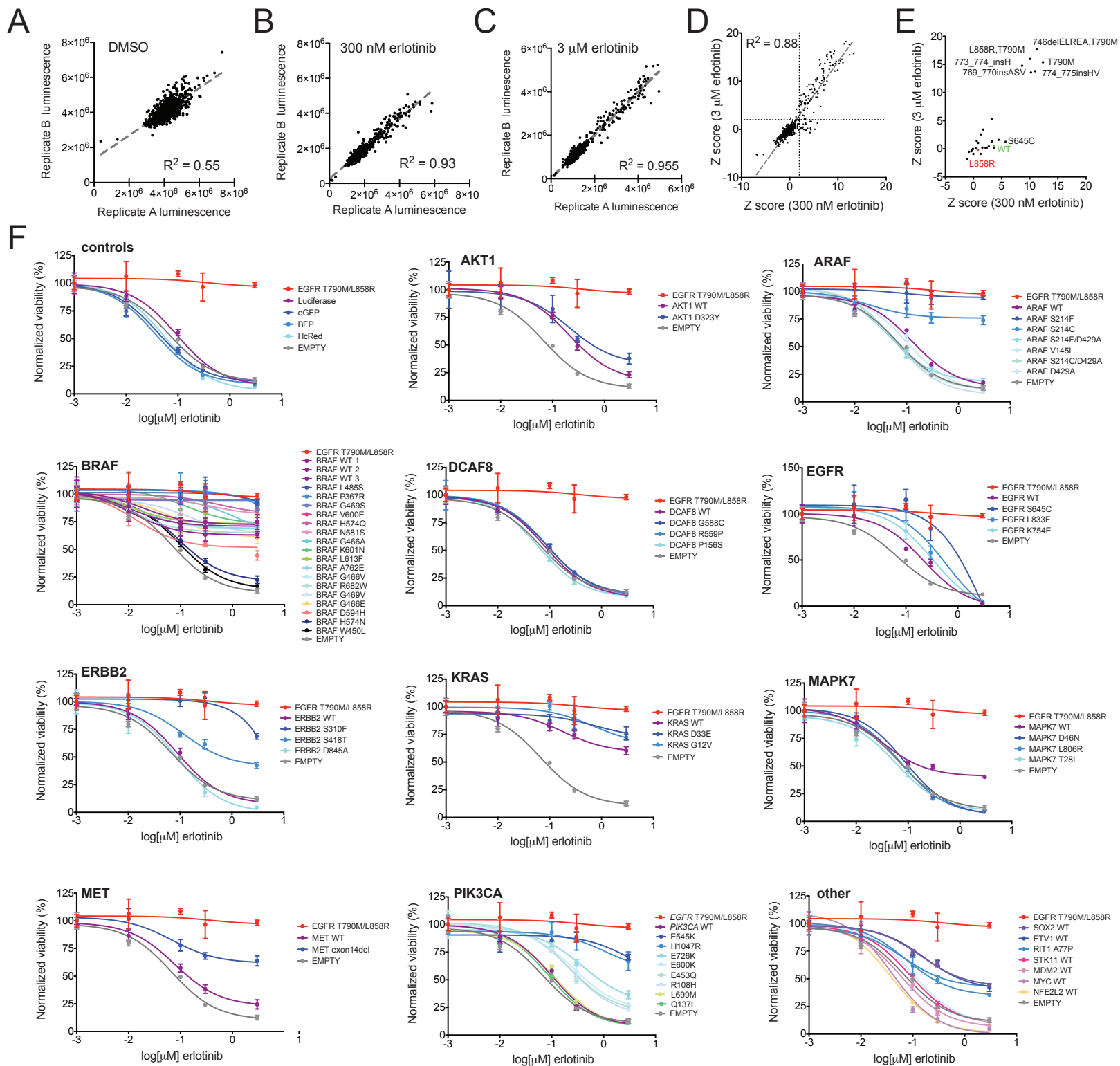File S3, related to Figure 3, supplied as an Excel file.

Figure S3, related to Figure 4. PC9 EGFR epistasis screen and validation identifies gain-of-function mutations epistatic to EGFR.

(A) Replicate data showing luminescence after CellTiterGlo treatment of PC9 cells infected with the mutated ORF library and treated with DMSO for 72 hours. A dashed line indicates the linear regression. Each point represents an individual well/data point.

(B) Replicate data showing luminescence after CellTiterGlo treatment of PC9 cells infected with the mutated ORF library and treated with 300 nM erlotinib for 72 hours. A dashed line indicates the linear regression. Each point represents an individual well/data point.

(C) Replicate data showing luminescence after CellTiterGlo treatment of PC9 cells infected with the mutated ORF library and treated with 3 µM erlotinib for 72 hours. A dashed line indicates the linear regression. Each point represents an individual well/data point.

(D) Relationship between cell viability (CellTiterGlo luminescence) of PC9 cells at different doses of erlotinib and after transduction with the mutated ORF library. Each point represents the average robust Z score of two replicate data points after calculating the robust Z score of each well's luminescence value with respect to the rest of the plate and then collapsing the two replicate Z scores by taking the mean. A dashed line indicates the linear regression.

(E) Data shown is the same as panel (D), limited to variants of EGFR.

(F) Validation data from retesting of wild-type and mutated ORFs in PC9 cells under conditions identical to the primary screen except at additional erlotinib doses and with a 96 hour DMSO/erlotinib incubation. Viability values (luminescence) are normalized to the lowest dose shown. Data shown is mean +/- standard deviation of n=4 replicates. Replicates were independent lentiviral transductions performed on the same day. "Empty" is non-transduced cells, which did not significantly differ in behavior from cells transduced with control ORFs such as GFP (see p values in Table S5). The "empty" and EGFR T790M/L858R data is the same data shown for reference on each plot.

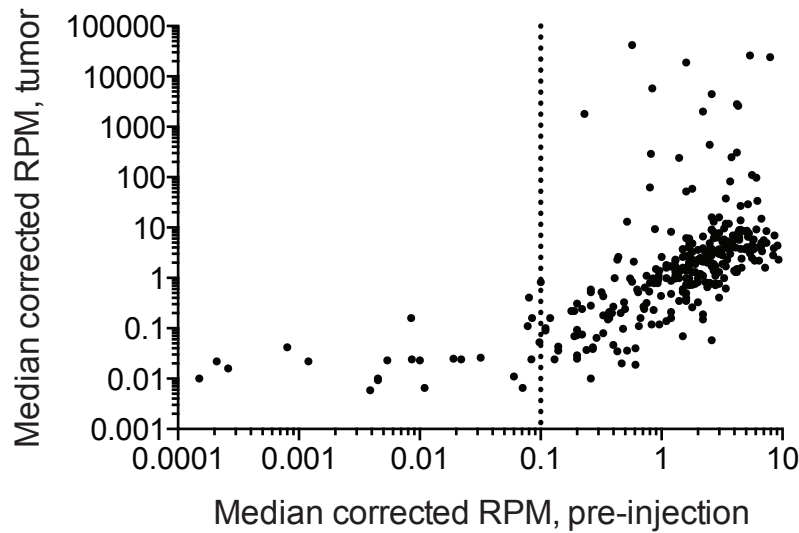Table S5, related to Figure 4, supplied as an Excel file.
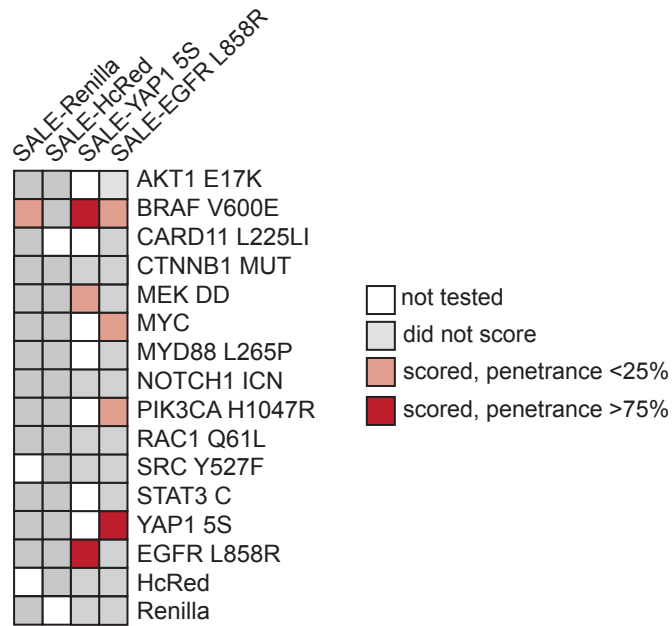
Figure S4, related to Figure 5. Pooled xenograft screening for identification of oncogenic mutations in vivo.
(A) Data from combination screening of SALE immortalized human lung epithelial cells. Stable cell lines expressing activated YAP1 (YAP1 5S) or EGFR (EGFR L858R) were generated and then transduced with the barcoded ORFs or controls (HcRed, Renilla). Cells were pooled and injected subcutaneously in immunocompromised mice. Sequencing data of barcodes in tumors was compared for enrichment with the pre-injection sample. ORF combinations enriched by 8 fold or higher are colored (pink, red). Among these combinations, those that scored in 75% or more of tumors are shown in red, while those that scored in under 25% are shown in pink. There were no combinations that scored in 25-75% of tumors. Each combination was tested in 10 or more tumors. White, not tested. Gray, did not score in any tumors.
(B) DNA barcode sequencing data adjusted to reads per million (RPM) and corrected for differing distributions of reads in each sample (Experimental Procedures). Each point represents the median RPM across multiple pre-injection samples (median n=3) or tumor samples (median n=6) for each ORF. A dashed line indicates the cutoff used to exclude ORFs with poor pre-injection representation from further analysis.

Table S6, related to Figure 5, supplied as an Excel file.
Table S7, related to Figure 5, supplied as an Excel file.
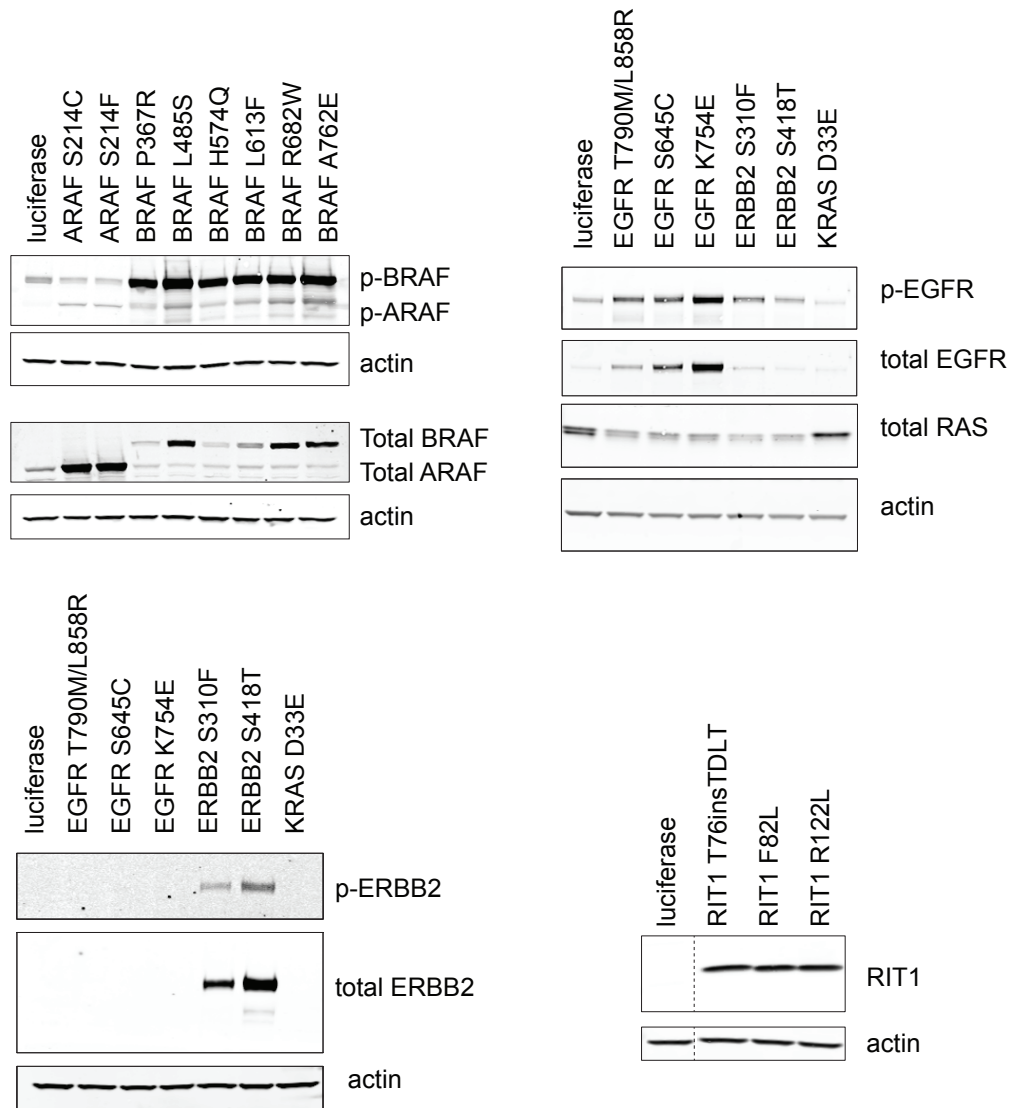
Figure S5, related to Figure 7. Western blot analysis of stable isogenic PC9 cell lines.
PC9 lung adenocarcinoma cells were transduced with pLX317 lentivirus encoding the variants indicated and pools of cells were selected with puromycin. Luciferase was used as an inert negative-control gene. Lysates were prepared and resolved on denaturing polyacrylamide gels before probing with the indicated antibodies, followed by fluorescent secondary antibodies and detection by a Li-Cor Odyssey.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Mutated cDNA library construction, lentivirus production, and transduction

pLX317 is a lentiviral expression vector that encodes a puromycin resistance cassette and an ORF-expression cassette under control of the EF1-alpha promoter. pLX317 contains a C-terminal V5 epitope tag, but all constructs in this study included a stop codon, precluding expression of the tag, unless otherwise indicated. 352 constructs (listed in **Table S1**) were included in the final analysis: 272 alleles of the 53 selected lung adenocarcinoma genes, 4 negative controls (BFP, RFP, HcRed, and Luciferase), and 76 overexpression control ORFs (constructs known to be effectively expressed and detected by L1000 gene expression profiling). Complete details of standard virus production pipelines can be found at the Broad Institute Genetic Perturbation Platform portal (www.broadinstitute.org/rnai/public/). Virus was produced in arrayed 96 well plates via triple transfection of HEK293T cells with each packaging vector (100 ng), envelope plasmid (10 ng), and each respective pLX317 plasmid (100 ng). Lentiviral-containing supernatants were harvested at 32-56 hours post-transfection and stored in polypropylene plates at -80 °C until use. For the eVIP experiment in A549 cells, infections were carried out in ten replicate copies, eight of which were used for gene expression profiling, and the remaining two used for assessing infection efficiency (+/- 1-2 ug/ml puromycin selection). For traditional expression analysis (AALE, SALE, H1299), infections were carried out in five replicate copies, three of which were used for gene expression profiling and the remaining two used for assessment of infection efficiency.

L1000 profiling

Luminex bead-based high-throughput gene expression profiling was performed as described (Peck et al., 2006). Briefly, mRNA in TCL lysates was captured by oligo-dT turbo capture plates (QIAGEN). Transcripts were amplified using gene-specific probes that also contain a bead id barcode sequence and biotin. Amplification products were hybridized to Luminex xMAP beads each coupled to complementary probe oligos corresponding to a specific mRNA. A total of 978 "landmark" genes were represented in the entire bead mixture. Note that probes can be directed at either coding sequence or untranslated regions (UTRs) of the target mRNA. For this reason, it is not expected that every ORF overexpression of a landmark gene would be detected by L1000. Beads were washed, and overall hybridization levels were detected by incubation with SAPE (Invitrogen) followed by quantification of the intensity of each bead color on a Luminex FLEXMAP 3D® machine. More detailed L1000 protocols can be found on the lincscloud website: http://support.lincscloud.org/

Expression-based variant impact phenotyping (eVIP)

Variant impact phenotyping based on gene expression changes uses two features of expression signatures induced by replicate introduction of a given ORF. The first feature is the strength of the signature. Replicate consistency was used as a quantitative measure of signal strength. All pairwise replicate correlations for a given allele's signature were computed using a previously established metric, the "weighted connectivity score" (Lamb et al., 2006) (www.lincscloud.org). This metric uses a weighted Kolmogorov-Smirnov statistic, similar to the computation used in gene set enrichment analysis (GSEA; (Subramanian et al., 2005), to score the enrichment of the most upregulated and downregulated genes in one signature (top 50 and bottom 50 Z-scores was used in this study) to the ranked Z-scores of another signature. Using this approach, ORFs that induce a strong gene expression change result in replicate profiles with a high signal-to-noise that are internally consistent. In contrast, in signatures that are similar to baseline/control, the top-ranked (by Z score) differentially expressed transcripts will have low fold changes and be inconsistent from replicate to replicate. For a given ORF, all pairwise weighted connectivity scores (wtcs) between replicates are calculated. To increase robustness against outlier replicates, the row medians of pairwise comparisons are used for the final distribution of self-correlation.

The second feature used for variant impact phenotyping is the signature identity. Mutant ORFs that function similarly to wild-type ORFs will induce similar gene expression changes and, therefore, have highly correlated gene expression profiles. Here, the correlation (wtcs) of each replicate signature of a mutant ORF with the replicate signatures of its cognate wild-type ORF is calculated and, for robustness from outlier replicates, the row and column medians of these pairwise comparisons are used for the final distribution of signature identity.

Predictions of variant impact are determined using the features of replicate consistency and signature identity in a decision tree approach we call expression-based variant impact phenotyping (eVIP) (**Figure 1C**). The root decision node of the tree tests for the null hypothesis that the mutant ORF signature is the same as (i.e., indistinguishable from) the wild-type ORF signature. This null hypothesis is rejected if there is a difference in the replicate consistency and identify of the mutant and wild-type ORF signatures. A Kruskal-Wallis test between the distribution of self-correlation of wild-type and mutant ORF signatures and the distribution of correlation between wild-type

versus mutant is performed to test for the null hypothesis. If the null hypothesis is rejected (Benjamini-Hochberg False Discovery Rate (B-H FDR < 5%), the variant is considered to have an impact on the wild-type function of the gene. The variant impact prediction at this root decision node is the primary prediction used in benchmark evaluation and comparison with in silico computational methods of variant impact predictions.

If the null hypothesis is rejected at the root decision node, then the variant is considered to impact the gene function and an additional decision is made to predict the directionality of the functional impact. If the replicate consistency of the mutant is greater than wild-type, the variant is predicted to have a gain-of-function impact (GOF). Conversely, if the replicate consistency of the mutant is less than wild-type, the variant is predicted to be loss-of-function (LOF). If no difference in replicate consistency can be determined, then the variant is classified as change-of-function. To determine if there is a difference in replicate consistency, an additional decision node tests for the null hypothesis that there is no difference in the distribution of self-correlation between wild-type and mutant ORFs using a Wilcoxon test. A difference in replicate consistency is identified when the null hypothesis is rejected (B-H FDR < 5%).

Finally, if the null hypothesis of the root decision node is not rejected, this may occur due to the variant having no impact on gene function (i.e., neutral) or that the wild-type and/or mutant signatures are noisy and, therefore, non-informative. To distinguish between these two possibilities, an additional decision is made by comparing the wild-type versus mutant replicate correlation distribution to a null distribution. The null distribution is determined by randomly selecting a test ORF and comparing it to a randomly selected control ORF and iterating this comparison 1,000 times. Thus, a variant is predicted to be neutral if the mutant ORF signature is indistinguishable from wild-type and if the mutant ORF signature is more similar to wild-type than random signature comparisons.

All eVIP code is deposited at www.github.com and the graphical output of eVIP is provided in **File S4**.

In silico prediction of mutation function
We compared functional impact calls from the tools SIFT, PolyPhen2 HumDiv, and MutationAssessor to functional impact predictions from eVIP. Variants may be called non-informative by eVIP or may have not-determined calls by the computational methods; therefore, only variants that were called by all four methods were included in the comparison resulting in 103 variant calls for comparison. For SIFT, variants with a "deleterious" call were considered impactful. For PolyPhen2 HumDiv "possibly damaging" and "probably damaging" calls were considered impactful. Finally, for MutationAssessor "high" and "medium" calls were considered impactful. To determine the expected agreement between the three computational methods, we used the proportion of impactful predictions each individual caller made on the 103 variants as the probability of each caller making an impactful prediction.

KEAP1/NRF2 analysis

For *KEAP1* LOH analysis, purity, ploidy, and absolute copy number of individual chromosomal segments were estimated for 660 lung adenocarcinomas using ABSOLUTE (https://www.broadinstitute.org/cancer/cga/absolute) (Campbell et al., *Nature Genetics* in press). LOH status was determined for *KEAP1* in 555 of these tumors. If multiple segments overlapped with *KEAP1*, the segment with the absolute copy number most different from the overall ploidy was used.

EGFR epistasis assay and validation
PC9 lung adenocarcinoma cells were plated at a density of 400 cells per well on 384 well plates in RPMI+10% FBS (R10). The next day, cells were transduced for 2-3 hours in arrayed format with 5 ul pLX317 lentivirus in the presence of 4 ug/ml polybrene. After transduction, media was removed and replaced with fresh R10. 72 hours post-infection media was changed to R10 containing 300 nM erlotinib, 3 uM erlotinib, or DMSO. Erlotinib was purchased from Selleckchem. 72 hours post-treatment, cell viability was determined using CellTiterGlo reagent (Promega) and luminescence quantified on an Envision MultiLabel Plate Reader (PerkinElmer). Infection efficiency of each ORF was determined by including an additional replicate plate which was treated with 1 ug/ml puromycin at 24 hours post-infection and comparing the CellTiterGlo luminescence of the puromycin-treated cells to the DMSO-treated cells for each respective well/ORF. Plate-level robust Z scores were calculated by plate for each ORF using the following equation:

$$Z\ score_x = \frac{Lum_x - Median\ (Lum_{all})}{1.4826 \times MAD(Lum_{all})}$$

Replicate-level Z scores were then averaged to a single Z score across the duplicate wells for each ORF at each drug concentration.

For the validation experiment, cells were plated and treated as above with erlotinib in quadruplicate per dose at the following concentrations: DMSO only, 1 nM, 10 nM, 100 nM, 300 nM, 3 uM. 96 hours after erlotinib addition, cell viability was assessed with CellTiterGlo. Each data-point was normalized to the cell viability in 1 nM erlotinib before dose-response curves were generated in GraphPad Prism using a three-parameter log(inhibitor) vs. response algorithm. Statistical differences between curves were determined using an extra sum-of-squares F test with a $P < 0.05$ considered to be significant.

For testing with additional inhibitors (**Figure 7**), stable PC9 isogenic cell lines were generated by lentiviral transduction in the presence of 4 μg/ml polybrene followed by selection with 1.5 μg/ml puromycin for 48-72 hours. Stable lines were expanded and ORF expression confirmed by Western blotting. For viability assays, PC9 cells were plated in 384 well plates at a density of 400 cells per well in 50 ul total volume. One day later, a 36-point dose curve of each inhibitor was performed using a D300e dispenser (Tecan). 96 hours post-inhibitor addition, cell viability was determined using CellTiterGlo according to manufacturer's instructions. Fractional viability was determined by normalizing to quadruplicate wells of no drug, and curve fitting was performed using GraphPad Prism four parameter inhibitor response with variable slope. AUC values were calculated by GraphPad Prism.

In vivo pooled tumor formation screen

Cell line: Human small airway epithelial cells immortalized with hTERT and SV40 early region (SALE) were a gift of the lab of William C. Hahn, Dana-Farber Cancer Institute. An activated form of YAP1 (YAP1[5S]) from Rosenbluh et al. 2012 (Addgene: 42562) was cloned into pLX311 and stably expressed in SALE cells (SALE-Y). YAP1[5S] corresponds to YAP1 isoform NM_001130145.2 with multiple serines mutated to alanines resulting in inhibition of phosphorylation and subsequent degradation of the protein. The following serines are mutagenized: S61, S109, S127, S128, S131, S163, S164, S397.

Infection and pooling: SALE-Y cells were plated in 96 well plates on Day 1 at 1000 cells per well in 100 ul SAGM (Lonza). On day 2, 5 ul of arrayed virus was added to the target cells. Plates were then spun at 2250 rpm for 30 minutes at room temperature. Media was changed 24 hrs post infection. 48 hrs post infection, another media change was performed to add media containing 1 ug/ml puromycin. Antibiotic selection was carried out for 72 hrs and puromycin-resistant cells were trypsinized and pooled at 7 days post infection. Adherent cells and cells in suspension were passaged. Pools were propagated until 19 days post infection to achieve required number of cells for injection.

Injection, monitoring and tumor genomic DNA isolation: On the day of injection, pools were trypsinized, quenched and washed with PBS. Two million cells in 200 ul PBS were injected per site. Three sites – inter-scapular area, right and left flanks – were injected in each mouse and 2 mice were injected per pool. Mice were monitored for tumor formation at injection sites. The longest dimension of each tumor was measured using digital calipers. Animals were sacrificed when any one tumor reached around 2 cm. Tumors were dissected and tumor samples were processed for genomic DNA (gDNA) extraction using the DNAeasy Blood and Tissue kit (Qiagen). Cell pellets of each pool on the day of injection and during passaging were also collected and processed for gDNA extraction.

PCR and sequencing: PCR of gDNA was performed to attach sequencing adaptors and barcode samples, which were divided into 50 ul reactions (total volume) containing a maximum of 10 μg gDNA. Per 96 well plate, a master mix consisted of 75 μL ExTaq DNA Polymerase (Clontech), 1000 μL of 10x Ex Taq buffer, 800 μL of dNTP provided with the enzyme, 50 μL of P5 stagger primer mix (stock at 100 μM concentration), and 2075 μL water. Each well consisted of 50 μL gDNA plus water, 40 μL PCR master mix, and 10 μL of a uniquely barcoded P7 primer (stock at 5 μM concentration). PCR cycling conditions: an initial 1 minute at 95˚C; followed by 30 seconds at 94˚C, 30 seconds at 52.5˚C, 30 seconds at 72˚C, for 28 cycles; and a final 10 minute extension at 72˚C. P5/P7 primers were synthesized at Integrated DNA Technologies (IDT). Samples were purified with Agencourt AMPure XP SPRI beads according to manufacturer's instructions (Beckman Coulter, A63880). Samples were sequenced on a HiSeq2000 (Illumina).

Data analysis: Reads were counted by first searching for the GACGA sequence in the primary read file that appears in the vector 5' to all ORF barcodes. The next 24 nts are the ORF barcode insert, which was then mapped to a reference file of all possible barcodes present in the library. The read was then assigned to a condition (e.g. a well on the PCR plate) on the basis of the 8nt barcode included in the P7 primer. The resulting matrix of read counts was first normalized to a reads per million within each condition by the following formula: read per barcode / total reads

per condition x 10^6. Reads per million was then $\log_2$-transformed after adding one to all values, which is necessary in order to take the log of barcode with zero reads.

Log_{10} transformed reads per million (LRPM) data were corrected for per-pool batch effects and systematic differences in pre vs. post-injection measurements using a linear model. Specifically, LRPM data points (y_i) were modeled as a function of pool id ($p_i \in \{0, \ldots, 11\}$) and injection status ($s_i \in \{0, 1\}$) by fitting

$$y_i = \alpha_0 + \alpha_1 \mathbf{1}_1(s_i) + \sum_{k=1}^{11} \beta_k \mathbf{1}_k(p_i) + \gamma_k \mathbf{1}_k(s_i p_i)$$

across the dataset of 11,265 LPRM values using maximum likelihood (ML). ($\mathbf{1}_k(x)$ is the indicator function that is 1 when $k = x$ and 0 otherwise). The 24 fitted parameters ($\alpha_0, \alpha_1, \beta_k, \gamma_k, k \in \{1, \ldots, 11\}$) have the following interpretation: α_0 is the model intercept, α_1 is the effect of post vs pre-injection, β_k is the effect of pool id k relative to pool 0, and γ_k the interaction effect between pool k and post-injection status for pools $k \in \{1, \ldots, 11\}$.

Residuals from the ML fit $\hat{y}_i$ were used to compute "corrected LRPM" values $z_i = y_i - \hat{y}_i$. Significant increases in corrected LRPM values between pre- and post-injection were assessed for each variant using a one-tailed t-test. False discovery rates were controlled using Benjamini and Hochberg method. All Tumorplex variant ranking analyses were implemented in R using stats and data.table packages (https://www.r-project.org/).

Antibodies and Western blotting

For Western blot analysis of BRAF expression, lysates were taken from stable A549 cells were generated via lentiviral transduction and selection in puromycin. Primary antibodies were all from Cell Signaling: B-Raf (55C6) Rabbit mAb #9433, Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) Antibody #9101, and p44/42 MAPK (Erk1/2) (3A7) Mouse mAb #9107. A mouse monoclonal antibody against Vinculin was from Sigma V9264. LiCOR secondary antibodies were: 926-68020 IRDye® 680LT Goat anti-Mouse IgG (H + L), and 925-32211 IRDye® 800CW Goat anti-Rabbit IgG (H + L). For analysis of isogenic PC9 cell lines, antibodies used were using protein-specific antibodies. Antibodies used were Cell Signaling Technology #4432 (ARAF), #4431 (phospho-ARAF), #2239 (EGFR), #3777 (phospho-EGFR), #2248 (ERBB2), #2247 (phospho-ERBB2), #3965 (RAS), and Abcam #53720 (RIT1).

SUPPLEMENTAL REFERENCES

Engel, B.E., Schabath, M.B., Thompson, Z.J., Eschrich, S.A., Brantley, S.G., Belock, A.R., Berglund, A., Gray, J.E., Beg, A.A., Haura, E.B., Cress, W.D. Characterization of three recurring STK11/LKB1 mutants in lung adenocarcinoma. [abstract]. In: Proceedings of the 105th Annual Meeting of the American Association for Cancer Research; 2014 Apr 5-9; San Diego, CA. Philadelphia (PA): AACR; Cancer Res 2014;74(19 Suppl):Abstract nr 4159. doi:10.1158/1538-7445.AM2014-4159

Olschwang, S., Boisson, C., and Thomas, G. (2001). Peutz-Jeghers families unlinked to STK11/LKB1 gene mutations are highly predisposed to primitive biliary adenocarcinoma. J Med Genet *38*, 356-360.

Palacios, J., and Gamallo, C. (1998). Mutations in the beta-catenin gene (CTNNB1) in endometrioid ovarian carcinomas. Cancer Res *58*, 1344-1347.

Prior, I.A., Lewis, P.D., and Mattos, C. (2012). A comprehensive survey of Ras mutations in cancer. Cancer Res *72*, 2457-2467.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.