SUPPLEMENTARY MATERIALS AND METHODS

Cell lines and cell culture

Cells were cultured at 37°C with 5% CO2. HEK293T and GP2-293 were maintained in Dulbecco's modified Eagle's medium/Ham's F-12 (DMEM/F12, Corning) medium supplemented with 10% fetal bovine serum (FBS, Gemini Bio) and 100 units/ml penicillin and 100 μg/ml streptomycin (Gibco). MCF10A cells and MCF10A derivatives were cultured in DMEM/F12 medium supplemented with 5% horse serum (Thermo Fisher Scientific) and 100 units/ml penicillin and 100 μg/ml streptomycin; and 20 ng/ml EGF (Peprotech), 10μg/ml insulin (Sigma), 0.5 mg/ml hydrocortisone (Sigma), and 100 ng/ml cholera toxin (Sigma). To generate MCF10A derivatives (MCF10A-vector, MCF10A-PIK3CA, and MCF10A-MYC), 6 μg of plasmid DNAs pMSCVneo, pLP-LNCX-PIK3CA-H1047R (neo) and MSCV-neo-HA-Myc and 6 μg of pMD2.G were packaged into retroviruses using GP2-293 cells according to the manufacturer (ProFection® Mammalian Transfection System, Promega), and viruses were collected to infect MCF10A cells and selected for stable infection using G418.

Mice

Six-week-old female hsd:athymic nude-*Foxn1*^{nu} mice were obtained from Envigo. The mice were maintained in accordance with Institutional Animal Care and Use Committee procedures and guidelines under an approved protocol.

Construction of cells stably expressing Cas9

The EFS promoter in lentiCas9-Venus (Addgene #70267) was replaced with the Cbh promoter to increase the expression of Cas9-Venus in MCF10A cells (lentiCas9-Venus-Cbh is deposited with Addgene). To produce the lentivirus, 9 μg of lentiCas9-Venus-Cbh, 6 μg of psPAX2 (Addgene #12260) and 4 μg of pMD2.G (Addgene #12259) were transfected into HEK 293T cells (ProFection® Mammalian Transfection System, Promega). Then the lentiviral supernatant was harvested to infect cells. At 48 hours post-infection, to obtain cells with consistent expression and non-toxic levels of Cas9, cells expressing a narrow range (5%) of Venus closest to the mean expression were selected by flow cytometry and expanded for the following experiments.

Construction of the TSG dual sgRNA plasmid library

The TSG dual sgRNA plasmid library was constructed by following the methodology developed by Prashant Mali's laboratory (1) with some modifications. The first step involved synthesizing 35,940 oligonucleotides (146-mers; Twist Bioscience) that contained, starting from the 5'-end, left homology arm (30 nucleotides), followed by the targeting portion of the first sgRNA (20 nucleotides), two BsmI sites (40 nucleotides), the targeting sequence of the second sgRNA (20 nucleotides) and the right homology arm (36 nucleotides). The common and variable oligonucleotide sequences were as follows,
5'-
tatatatcttgtggaaaggacgaaacACCGNNNNNNNNNNNNNNNNNNNGTTTTgagacgTAGGGAT
AACAGGGTAATcgtctcGTTTGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAA
TAGCAAGTTAAAATAAGGCT-3'. To complete the first step of the plasmid library construction, the pooled synthesized oligos were amplified by PCR (98℃ for 3 min; 18 cycles of 98℃ for 10 s, 55℃ for 15 s, 72℃ for 15s; 72℃ for 5 min extension) with CloneAmp™ HiFi PCR Premix (Takara) and assembled into the linearized lentiGuide-puro vector (Addgene #52963) digested using a FastDigest Esp31 via In-Fusion® HD Cloning kit (Takara). In the second step, a fragment containing one sgRNA scaffold and one mouse U6 promoter (1) was amplified from the plasmid pUC57-mini-scaffold-mU6 (deposited on Addgene) and inserted into the Esp31-digested 1st-step plasmid library with T4 DNA ligase (New England BioLabs). In both steps, the in-fused or ligated products were transformed into Stellar competent cells (Takara) and 40 transformations were performed in each step to ensure 50X coverages of each dual sgRNA. Plasmid libraries were extracted via QIAGEN Plasmid Plus Midi Kit (Qiagen) and stored at -20 °C.

Construction of TSG dual sgRNA cell libraries

A total of 6 μg TSG dual sgRNA plasmid library DNA, 6 μg of psPAX2 (Addgene #12260) and 4 μg of pMD2.G (Addgene #12259) were transfected (ProFection® Mammalian Transfection System, Promega) into one 10 cm dishes containing freshly (overnight) plated HEK293T cells (~70% confluent). After 36-48 hours, viral supernatant was collected, and the titer was determined by infecting cells with serial dilutions and scoring the percentage of cell survival after 1.5 μg/ml puromycin selection (3-4 days). To construct the cell libraries, cell expressing Cas9-Venus (Cbh) were transduced with TSG dual sgRNA virus libraries at multiplicity of

infection (MOI) of 0.15 - 0.3. After 24 h of infection, the virus was removed, and cells were then selected in medium containing 1.5 μg/ml puromycin for 4 days or until all of the non-infected cells had died. To maintain 100 to 200-fold representations of every dual sgRNA, three million of cells were seeded into each of eight 15 cm dishes, so that a total of 24 million cells were transduced.

Dual sgRNA amplicon sequencing

Dual sgRNA amplicons were prepared for deep sequencing from plasmid, cell genomic DNA (gDNA) and tissue gDNA. Harvested cell pellets and homogenized tumor tissues were stored at −80 °C until gDNA extraction with an E.Z.N.A. Tissue DNA Kit (Omega Bio-tek) and a DNeasy Blood and Tissue Kit (Qiagen), respectively. Dual-sgRNA cassettes were amplified through two steps of PCR. Briefly, the first step was performed in 50 μl reactions with 10 ng of plasmid, or 500 ng of cell gDNA or 1μg of tumor gDNA with CloneAmp™ HiFi PCR Premix (Takara). The PCR primers were as follows:
F5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT[NNNNNNNN][i5-index] TATATATCTTGTGGAAAGGACGAAACACCG; R5'-CTCGGCATTCCTGCTGAACCGCTCTTCCGATCT[NNNNNNNN][i7-index] CCTTATTTTAACTTGCTATTTCTAGCTCTA. The thermocycling parameters for amplification were: 95 °C for 30 s; N cycles of 98 °C for 15 s, 55 °C for 15 s, and 72 °C for 45 s; and 72 °C for 5 min. N equals 3 cycles for plasmid, 15 cycles for cell gDNA and 22 cycles for tumor gDNA. For plasmid and cell gDNA, the PCR products were pooled and purified with a NucleoSpin Gel and PCR Clean-up kit (Takara). To maintain > 100 folds of coverages for dual sgRNAs, 60 PCR reactions were performed for plasmid and cell gDNA, and 30 reactions for tumor gDNA. A second PCR was performed with CloneAmp™ HiFi PCR Premix (Takara) in 50 μl PCR reactions. The PCR primers were as follows:
PE2-
5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
PE1-
5'CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT. The thermocycling parameters for amplification were: 95 °C for 30 s; N cycles of

98 °C for 15 s, 68 °C for 15 s, and 72 °C for 45 s; and 72 °C for 5 min. N equals 22 cycles for plasmid, 15 cycles for cell gDNA and 7-8 cycles for tumor gDNA. For PCR of plasmid and cell gDNA, 15 tubes of the 1st round PCR products were pooled and purified with one column and used as templates for one reaction of the 2nd round PCR. For PCR of tissue gDNA, 50 ng of 1st round PCR products were used for each of the 2nd round PCR. PCR products from the 2nd round PCR were purified using Agencourt AMPure Beads (Beckman Coulter). In total, 35 µl of prewarmed beads were added to 50 µl of PCR products and were mixed well. After incubation at room temperature for 5 min, the beads were separated from the solution using a magnetic rack and washed twice with 80% freshly made ethanol. The beads were incubated in ethanol for 30 s for each wash, while keeping the tubes on the magnetic rack. The ethanol was removed, the beads were air dried the beads for 1 min, and the DNA was eluted with 20 µl of elution buffer. The purified amplicons were confirmed by agarose gel electrophoresis, quantified by Qubit (Life Technologies, Q32854), then submitted for pair-end sequencing on an Illumina HiSeq 4000 system with 20% Phix DNA spike-in.

CRISPR mutation abundance profiling

To obtain the CRISPR mutational profiles, DNA fragments (~600 bp) around the NF2_2 sgRNA and TP53_4 sgRNA targeting sites were amplified from tumor gDNA. The primers for NF2 amplification were as follows: forward primers (30 bp upstream of NF2_2 sgRNA): 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT[sample index]$_6$ GATCTACTGCCCTCCTGAGG; reverse primers (522 bp downstream of NF2_2 sgRNA): 5'-CTCGGCATTCCTGCTGAACCGCTCTTCCGATCT[sample index]$_6$ CCACACACAAGGCAGCTTGT. The primers for TP53 amplification were as follows: forward primers (36 bp upstream of TP53_4 sgRNA): 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT[sample index]$_6$ CTGTGCAGCTGTGGGTTGATT; reverse primers (513 bp downstream of TP53_4 sgRNA): CTCGGCATTCCTGCTGAACCGCTCTTCCGATCT[sample index]$_6$ TCGATCTCCCAACCTCGTGAT. PE1 and PE2 primers were used to make the DNA libraries for the 2nd round of PCR. The CRISPR-induced INDEL mutations were profiled by deep sequencing. After mapping the sequencing reads to the human reference genome sequence

GRCh39, the Integrative Genomics Viewer (IGV) browser was used to visualize the indels around the target sites.

Generation of CROPseq-Guide-sgRNAs

A CROP-seq vector was modified from a regular CRISPR lentiviral vector expressing sgRNA cassette (U6-sgRNA-scaffold). Basically, in a CROP-seq vector, the sgRNA cassette is placed into the 3' LTR region, so that the sgRNA is expressed as part of the selective marker mRNA and is detectable by scRNAseq (Datlinger et al., 2017). CROPseq-Guide-Puro was purchased from Addgene (Addgene #86708) and CROPseq-Guide-Blasti was constructed by replacing a puromycin ORF sequence with a blasticidin ORF sequence (deposited at Addgene). Individual sgRNA oligos were cloned into the CROPseq-Guide-Puro and CROPseq-Guide-Blasti as previously described (https://media.addgene.org/cms/files/Zhang_lab_LentiCRISPR_library_protocol.pdf). Sanger sequencing was performed to confirm the inserted sgRNA sequences. CROP-seq viruses were packaged in HEK293T cells, and the titers were measured as described in methods of constructing lentiGuide-Puro library above.

Enrichment PCR for CROPseq screenings

To enrich the transcripts carrying the sgRNA sequences and the corresponding 10X cell barcode and unique molecular identifier (UMI) from the same cell, a two-step PCR was conducted. As templates for enrichment PCR, 10 ng of four cDNA libraries before fragmentation from scRNA-seq above were used. The 1st round of PCR was performed in a 50 µl reaction with CloneAmp™ HiFi PCR Premix (Takara). Four reactions were performed for each sample. The PCR primers for step 1 - PCR were as follows: Forward: Enrich_XYZ1 5'-CTACACGACGCTCTTCCGATCT (10 X Read 1 primer seq); Reverse: Enrich_XYZ2 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATATCTTGTGGAAAGGACGAAACAC. The thermocycling parameters for amplification were: 95 °C for 30 s; 20 cycles of 98 °C for 15 s, 60 °C for 15 s, and 72 °C for 45 s; and 72 °C for 5 min. PCR products were purified using NucleoSpin® Gel & PCR Clean-up Kit (Takara), and were eluted with 20 µl of elution buffer, of which 60 ng were used for the 2nd round of PCR. The primers for step 2 - PCR were as follows: Forward primers (SI-PCR primer): 5'-

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC; Reverse primers: Enrich.XYZ3 5' CAAGCAGAAGACGGCATACGAGAT [Chromium i7 sample index] GGTTTACTGTGACTGGAGTTCAGACGTGT. The i7 index sequences were obtained from PN-120262: Chromium i7 Multiplex Kit. The thermocycling parameters for amplification were: 95 °C for 30 s; 8 cycles of 98 °C for 15 s, 54 °C for 15 s, and 72 °C for 45 s; and 72 °C for 5 min. The expected band was around 450 bp. Double side size-selection with SPRIselect beads (Beckman Coulter) was conducted to purify step-2 PCR products. 100 µl of pooled PCR products from two PCR tubes were mixed with 60 µl of (0.6 X) SPRIselect Reagent for 15 times. After 5 min incubation at RT, beads were separated with Magnet H (10X Chromium). Next, 150 µl of the supernatant were transferred to a new tube strip and 20 µl of (0.8X) SPRIselect Reagent was added. The mixture was mixed for 15 times and incubate at RT for 5 min. The beads were separated with Magnet L and 165 µl of supernatant was removed. The beads were washed with 200 µl of 80% ethanol twice, then centrifuged briefly, with the ethanol removed as much as possible while keeping the tubes on the magnet. The beads were air-dried, and DNA was eluted from the beads in 20 µl of EB buffer. The DNA was stored at 4°C ≤ 72 h or at −20°C for long-term storage. The sgRNA enrichment libraries were separately indexed and sequenced as spike-ins alongside the scRNA-seq libraries with a molar ratio of 1:25 using the following configuration: R1: 26, i7 index: 8, i5 index: 0, R2: 98.

Crystal violet staining

DKO, SKO1, SKO2 and control cells were separately seeded onto 6-well plates at density of 10^5 cells per well with full medium. After the cells attached overnight, the medium was replaced with minimal assay medium. When the DKO cells grew confluent, cells were ready for crystal violet staining. Cells were washed twice with ice-cold PBS after aspirating the medium and were then fixed with ice-cold methanol for 10 minutes. Fixed cells were stained with 0.5% crystal violet solution at RT for 10 minutes. Lastly, plates were carefully rinsed with water until free crystal violet was no longer visible. Images of cells were collected using an EVOS Imaging system (Life technologies) after the plates were completely dry.

Cell cycle determination

Cells were harvested, resuspended in ice old DPBS, and fixed by slowly adding drop wise cold 70% ethanol. Fixed cells were stored at 4℃ to -20℃ for at least 2 hours. Fixed cell pellets were washed with ice old DPBS twice, and then stained with 500 µl of 100 µg/ml propidium Iodide (PI) for 30 min at RT. Cell cycle phases were determined by quantitation of DNA content on flow cytometry.

Computational analysis

Quantification of abundance of dual sgRNAs

Sequencing data containing dual sgRNAs were processed with custom written software in Python as previously described (2), with modifications. Briefly, sequences were parsed to isolate the two sgRNA regions (30 or 32 bp each) and two UMI regions (9 bp each), sorted by their sample indices (6 bp each), and removed if they failed to pass any of three quality filters: (1) the average Illumina quality score for both sgRNA regions must be greater than 30, (2) the first sgRNA(forward reads) must match the regular expression: '\D*?(.CCG|A.CG|AC.G|ACC.)\D{19,21}?(.TTT|G.TT|GT.T|GTT.)\D*', (3) the second gRNA (reverse reads) must match the regular expression: '\D*?(.AAC|A.AC|AA.C|AAA.)\D{19,21}?(.AAA|C.AA|CA.A|CAA.)\D*'. Each sgRNA was independently clustered with Bartender (-d 2 -z -1 -l 5 -s 1) (3), which rendered cluster centroid and UMI for each parsed sgRNA. The sgRNAs and their cluster centroids and UMIs in the forward and reverse reads of the same line were combined. The number of unique UMIs belonging to the same combined cluster centroids was considered as the count of each cluster centroids combination. Then the combined cluster centroids were mapped to the pre-known combined sgRNA sequences without allowing any mismatch to obtain the ultimate counts for each known dual sgRNA.

Fitness measurement based on time-series data *in vitro*

For *in vitro* screenings, assuming each cell lineage C expressing a specific dual sgRNAs constructs grows exponentially, the cell number of each cell lineage $N_c(t)$ will increase over time t as:

$$Nc(t) = Nc(t_0) \times 2^{(f_c + f_0)t}$$

where $N_c(t_0)$ is the cell number of lineage C at time point 0; f_c is the fitness of cell lineage C; f_0 is the fitness of cells expressing double non-targeting sgRNAs.

According to (1) , the solution to estimate f_c is:

$$f_c = \frac{Cov(X_c, t)}{Var(t)} + \delta$$

where X_c is the experimentally measured log2 relative frequency of each cell lineage.

As we know, in a simple linear regression model,

$$Y = X\beta + \delta$$

Here,

$$\hat{\beta} = \frac{Cov(Y,X)}{Var(X)}$$

Therefore, we calculated the slope of linear regression of log2 relative frequency over time to estimate the fitness of each cell lineage C:

$$f_c = slope\ of\ lm(X_c \sim time)$$

where lm is a R package to fit a simple linear regression model.

The sgRNA-level fitness f_c of each dual sgRNA was measured as above, and fitness at gene level f was averaged from all corresponding dual-sRNA-level fitnesses. To assess the reproducibility of biological replicates, gene-level $\boldsymbol{f}$ was separately calculated for each replicate in the three conditions. Standard Pearson correlation was applied to compare f from the two replicates.

Genetic interaction scoring *in vitro*

An sgRNA-level genetic interaction score of each DKO dual sgRNA pair was computed based on a "Max model" as follows (Figure S4):

$$Max\ model: \pi_c = f_{DKO} - \max(f_{SKO1}, f_{SKO2})$$

where π_c is the sgRNA-level genetic interaction score; f_{DKO} is f_c of a DKO dual sgRNA; f_{SKO1} and f_{SKO2} are f_c of its corresponding SKO dual sgRNAs, respectively. Gene-level interaction scores π of 1325 TSG pairs were then calculated by averaging all the corresponding sRNA-level interaction scores. The statistical significance of a genetic interaction score π of a given TSG pair was assessed by comparing a set of f_{DKO} values of average 25 DKO dual sgRNAs with a set of the greater f_{SKO} values for the "Max model" using a one-sided Student's t-test. The Benjamini-Hochberg procedure was then utilized to adjust p values for multiple comparisons.

Genetic interaction score π and adjusted p values (p.adj) were determined individually for each replicate in the three conditions by following the procedure above. Standard pearson correlation was applied to compare π from the two replicates. Lastly, genetic interaction scores of each TSG pair calculated from two replicates were aggregated by using inverse-variance weighted average method. Two adjusted p values were combined by Fisher's method using the function of "sumlog" in the R package "metap". Growth-promoting GI pairs were defined as synergistic interactions between TSG pairs ($Weighted\ average\ \pi > 0,\ p.adj < 0.05$), where their dual deletion significantly promote cell growth ($f_{DKO} > 0, p.adj < 0.05$). Cytoscape software (4) was utilized to visualize the growth-promoting interactions of synergistic TSG combinations.

$$Inverse\ variance\ weighted\ average\ \pi:$$

$$\left(\frac{\pi_{replicate1}}{Variance_{replicate1}} + \frac{\pi_{replicate2}}{Variance_{replicate2}}\right) / \left(\frac{1}{Variance_{replicate1}} + \frac{1}{Variance_{replicate2}}\right)$$

Fitness measurement and enrichment characterization *in vivo*

For *in vivo* screenings, fitness (or effect size) was measured based on the relative abundance of dual sgRNAs in cell libraries and tumors, instead of time-series data. *In vivo*, gene-level effect size f was defined as $\log_2$ fold change (LFC). First, the number of reads in each sample were normalized by converting raw counts to reads per million (RPM). Second, gene-level RPM

abundance of each SKO, DKO or control was calculated by taking the 70th percentile of RPMs of all its corresponding dual sgRNAs on a sample-by-sample basis. Third, mean gene-level RPMs were aggregated from all the cell and tumor samples, respectively. Lastly, LFC of mean RPMs of tumors versus cells was calculated as f.

$$f = LFC = log_2(FC) = log_2(\frac{mean\ RPM_{tumors}}{mean\ RPM_{cells}})$$

Statistical significance and 95% confidence interval (CI) were determined using bootstrapping. P values were adjusted by the Benjamini-Hochberg procedure.

In addition, we fit a linear regression model for the $\log_2$ average RPM abundance in tumors and cells. Significant outliers were characterized using the "outlierTest" function in the "car" R package, which computed the studentized residuals of the linear regression and calculated the corresponding P values. P values were adjusted by the Benjamini-Hochberg procedure. Standard Pearson correlation was applied to compare residuals derived from the linear fit of gene-level abundance and LFC. Results derived from the linear fit of sgRNA-level abundance were used to tabulate the number of enriched oncogenic dual sgRNAs for a given SKO or DKO.

Genetic interaction scoring *in vivo*

Genetic interaction score π of a given TSG pair was computed directly based on the gene-level f using a "Max model". The statistical significance of genetic interaction scores was determined by comparing RPMs in all the tumor samples with cell samples using the bootstrapping permutation method. Multiple comparison correction for P values the Benjamini-Hochberg procedure. Tumor-promoting genetic interactions were defined as synergistic interactions between TSGs ($\pi > 0, p.adj < 0.05$), where dual deletion of the gene pairs significantly promoted tumorigenesis ($f_{DKO} > 0,\ p.adj < 0.05$). Considering larger sampling noise due to a small sample size in PIK3CA tumors, a higher threshold was set for π ($\pi > 1.7$). Cytoscape software was utilized to visualize the tumor-promoting genetic interaction networks.

Genetic interaction profiling

Genetic profiles similarities were measured for all the 1326 TSG pairs by computing the Peasrson correlation coefficients (PCCs) from the genetic interaction matrix.

Generating gene expression matrices and assigning cell identity

The FASTQ files of the 10X scRNA expression library and its corresponding enrichment PCR library of each sample were aligned to GRCh38-3.0.0 human genome assembly and CRISPR feature reference using the count pipeline of Cell Ranger software (version 3.0.1, 10X Genomics by following the manual of Feature Barcoding Analysis of 10X Genomics. It outputs a unified feature-barcode matrix containing gene expression counts alongside a CSV file containing sgRNA identity information for the individual cells. Four CROP-seq expression matrices (S1_full_D0, S2_full_D6, S3_tgf_D6 and S4_min_D6) were aligned individually and single cell transcriptomes of 5,000 to 9,000 cells were profiled. Expression matrices were created into four individual Seurat objects using the Seurat R package (5), and the sgRNA identity information was aligned and loaded into the corresponding metadata of each Seurat object. On average, ~85% of the profiled cells expressed one or two sgRNAs (Figure S5B), which are of our interest, consisting of 55 of different DKO cells with two sgRNAs targeting two distinct TSGs, 11 of different SKO cells with one TSG sgRNA and control cells expressing two control sgRNAs. In the rest of cells, none of the sgRNAs was detected in ~10% of them, and ~5% of them expressed three or more sgRNAs (Figure S5B). Cells without detectable sgRNAs were likely to be debris, since they had much lower RNA expression and much higher percentage of mitochondria genes; cells expressing three or more sgRNAs were likely to be multiplets, since they carried much more RNA contents than other cells. Cells were excluded from further analyses based on the following criteria: (1) fewer than 800 expressed genes; (2) more than 5000 expressed genes; (3) mitochondrial genes expression accounted for $\geq$ 10% of the total RNA content. Genes expressed in fewer than three cells were excluded.

HDBSCAN clustering and UMAP projection on mean expression of perturbations

Clustering and UMAP projection here were performed using the Python codes described in (6). Briefly, normalized mean expression profiles of all perturbations (11 SKOs, 55 DKOs and 1 control) were computed by averaging genes across all the cells with the same perturbation. Only genes with mean expression greater than 0.25 were retained. The expression of each gene was standardized by dividing by the standard deviation across all mean expression profiles. Unsupervised learning was performed on the mean expression profiles of all the 67 different

perturbation identities to find stable clusters using the HDBSCAN library in Python. Lastly, UMAP dimension reduction was conducted on the mean expression profiles to visualize the projection, and points were colored according to cluster identity determined by HDBSCAN clustering. All the parameters used in HDBSCAN and UMAP were the same as those in Norman et al., 2019.

Gene Set Enrichment Analysis (GSEA)

Raw counts of cells bearing different perturbations (e.g., SKO cells vs. control cells) or grown under different conditions (e.g., controls in minimal medium vs. controls in full medium) were extracted using the "subset" function in Seurat 3.1.5, and were saved in a .txt file. A .cls file containing the phenotype labels (e.g., perturbation identity) of all the single cells was generated. The .txt format expression data and cls phenotype data were loaded into GSEA software 4.1.0 as input datasets. GSEA was performed as described previously (7)). Hallmark gene sets collection (h.all.v7.2.symbols.gmt) was utilized (8). Other parameters were as follows: Number of permutations:1,000; Collapse/Remap to gene symbols: No_collapse; Permutation type: phenotype for cells at different conditions, gene_set for comparison of SKOs/DKOs with control cells. Each single cell was treated as an independent replicate sample.

Differential expression analysis

DEGs induced by SKO or DKO perturbations were obtained by comparing single cell expression profiles of the perturbed cells with control cells. Briefly, we extracted raw counts of cells with a unique SKO or DKO perturbation; generated an expression matrix by combining it with raw counts of controls; and ran DESeq2 on genes with mean expression greater than 0.1 (5069 genes) using the DESeq2 package (9). Lastly, the "apeglm" algorithm was used to shrink log2 fold changes (LFCs) (10). The whole procedure was performed for all the SKO and DKO perturbations in parallel in Python. Each single cell was treated as an independent replicate sample. $P.adj < 0.05$ was used to identify significantly differentially expressed genes.

Measurement of transcriptional differences (TDs) and transcriptional interactions (TIs)

In procedures adapted from Norman et al., 2019, we used a linear regression model to decompose the effects of a DKO perturbation (δAB) in terms of the action of two of its corresponding SKO perturbations (δA and δB):

$$\delta AB = c_1 \delta A + c_2 \delta B + \varepsilon$$

where δAB denotes the differential expressions induced by each DKO perturbation (LFCs of 5069 genes from the comparison of DKO cells with control cells in the DESeq2 analysis). δA and δB denote the LFCs induced by the two respective SKO perturbations. c_1 and c_2 are regression coefficients that are fit to the data. ε is a vector of values measuring the deviations of the DKO effects from the model predicted effects - transcriptional differences. We defined the metric variance of ε as transcriptomic interaction score to scale the epistatic transcriptional differences.

Categorization of DEGs based on TDs

We classified all the differentially expressed genes (DEGs) in the DKOs into 3 different categories: synergistic, additive and buffering based on their transcriptional differences. DEGs with LFC > 0.03 were categorized into upregulated synergistic genes when TDs $\geq$ 0.05, upregulated buffering genes when TDs $\leq$ -0.05 and upregulated additive genes when -0.05 $<$ TDs $<$ 0.05. DEGs with LFC < -0.03 were categorized into downregulated synergistic genes when TDs $\leq$ -0.05, downregulated buffering genes when TDs $\geq$ 0.05 and downregulated additive genes when -0.05 $<$ TDs $<$ 0.05.

Gene ontology enrichment analysis

Genes in different epistatic transcriptional categories were submitted to http://geneontology.org/ powered by PANTHER (11) for term enrichment analysis in all the nine annotation data sets. A false discovery rate (FDR) < 0.05 was used to determine if a GO enrichment is statistically significant.

Enrichment scores of TFs for gene-specific synergistic DEGs

First, we determined how many genes out of the total 5069 genes were direct targets of p53, FOXO1 or YAP1 based on either the published lists or CHIP-Atlas databases. The target gene list of p53 as a activator was obtained from Fischer, 2017 (12) and the target list of DREAM

complex (TP53 repressor) was obtained from Fischer et al., 2016 (13). A gene was determined to be a direct target of FOXO1 when it was predicted to be bound by FOXO1 in SRX3230384.Hep_G2 dataset of CHIP-Atlas. It would be a target of FOXO1 as a transcriptional activator, if its expression was upregulated by PTEN SKO (LFC > 0.01) and a target of FOXO1 as a transcriptional repressor if its expression was downregulated by PTEN SKO (LFC < -0.01). Six datasets of CHIP-Atlas (SRX4213898.MCF_10A, SRX4213899.MCF_10A, SRX5287688.MCF.7, SRX5287705.MCF.7, SRX883576.MDA.MB.231 and SRX883577.MDA.MB.231) were combined to determine the binding target list of YAP. They include two experiments in MCF10A and four in two other breast cancer cell lines MCF7 and MDA-MB-231. We considered a gene as the target of YAP1 when the binding scores in MCF10A cells (positive score in either of the two) and breast cancer cell lines (positive score in either of the four) were both positive. It would be a target of YAP as a transcriptional activator, if its expression was upregulated by NF2 SKO (LFC > 0.01); a target of YAP as a transcriptional repressor if its expression was downregulated by NF2 SKO (LFC < -0.01). All the datasets were downloaded from CHIP-Atlas using hg38 as the human genome reference build and 1 kb as the cutoff for the distance from transcription start site (TSS) of the genes. Next, we counted up how many gene-specific synergistic DEGs were direct transcriptional targets of the individual major TFs. Lastly, the enrichment score of TFs were calculated as:

$$\frac{number\ of\ TF\ targets\ in\ gene-specific\ synergistic\ DEGs}{number\ of\ TF\ targets\ in\ all\ 5069\ reference\ genes}$$

Gene expressions in human breast cancer samples with dual or single alterations

The METABRIC dataset was downloaded from cBioPortal, where the genomic and RNA expression data of 1,904 breast cancer patients were used for study of the effects of dual alterations on gene expression. Alteration status of TP53 was determined only based on its mutational status (Mutated vs. Wild type). Alteration status of PTEN was determined based on its mutational and copy number status, where the tumors were categorized as being altered for PTEN when they harbor either PTEN point mutations or deep deletions. Alteration status of NF2 was determined based on the copy number variation of YAP1 and WWTR1, where the tumors were categorized as being altered for NF2 when they carry either amplified YAP1 or WWTR1. Gene expressions in patients with single and dual alterations were then compared.

Gene expression-based survival analysis for breast cancer

We used an online tool called GOBO (http://co.bmc.lu.se/gobo) (14) to predict the prognostic outcomes for different gene expression levels of sets of genes in a 1881-sample breast cancer dataset.

DESCRIPTIONS OF SUPPLEMENTARY TABLES

Supplementary Table 1

This table lists 52 TSG names, TSG sgRNA sequences, all the oncogenic interactions in vivo, full PIK3CA network, co-occurrence analysis and the epistatic expression of *CDK4*, *DNMT1* and *SPRK1* in breast cancer patients.

.

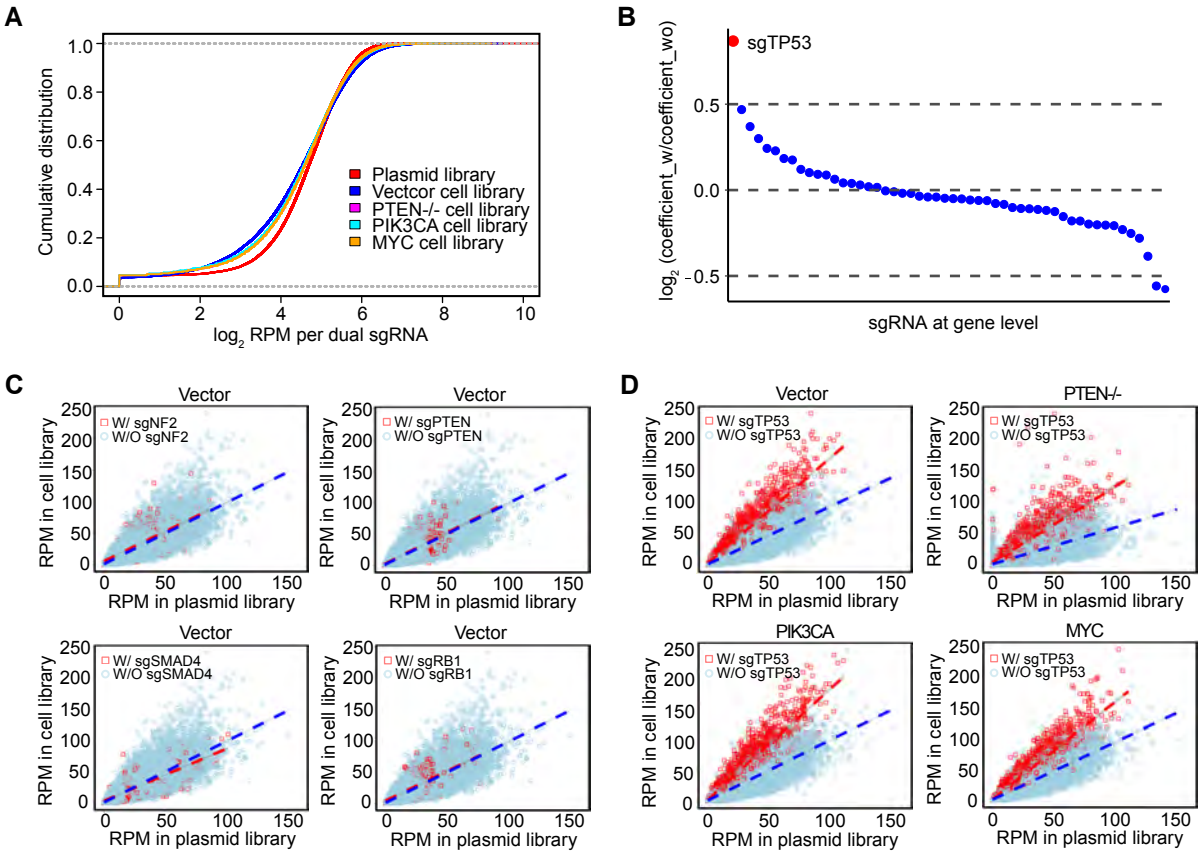
SUPPLEMENTARY FIGURES

Supplementary Figure S1

Related to Figure 1

Evaluation of the TSG plasmid and cell libraries for combinational CRISPR screenings. (A) Cumulative distributions of abundance of dual-sgRNAs in the plasmid library and four cell libraries. (B) log2 fold change (LFC) of linear regression coefficients for sgRNAs at gene level. LFC of each gene is the average of LFCs calculated from comparison of the four cell libraries versus plasmid library. (C) Scatter plots of the relative abundance of dual-sgRNAs with or without some specific sgRNA in the MCF10A vector cell library and plasmid library. Dual-sgRNAs with or without sgNF2 (top left), sgPTEN (top right), sgSMAD4 (bottom left) and sgCBFB (bottom right). (D) Scatter plots of the relative abundance of dual-sgRNAs with or without sgTP53 in four cell libraries (top left: MCF10A-vector cell library; top right: MCF10A-PTEN-/- cell library; bottom left: MCF10A-PIK3CA cell library; bottom right: MCF10A-MYC cell library) and plasmid library.

Figure S1


A
Cumulative distribution
log2 RPM per dual sgRNA
Plasmid library
Vectcor cell library
PTEN-/- cell library
PIK3CA cell library
MYC cell library
B
sgTP53
log2 (coefficient_w/coefficient_wo)
sgRNA at gene level
C
Vector
W/ sgNF2
W/O sgNF2
W/ sgPTEN
W/O sgPTEN
W/ sgSMAD4
W/O sgSMAD4
W/ sgRB1
W/O sgRB1
RPM in cell library
RPM in plasmid library
D
Vector
PTEN-/-
PIK3CA
MYC
W/ sgTP53
W/O sgTP53

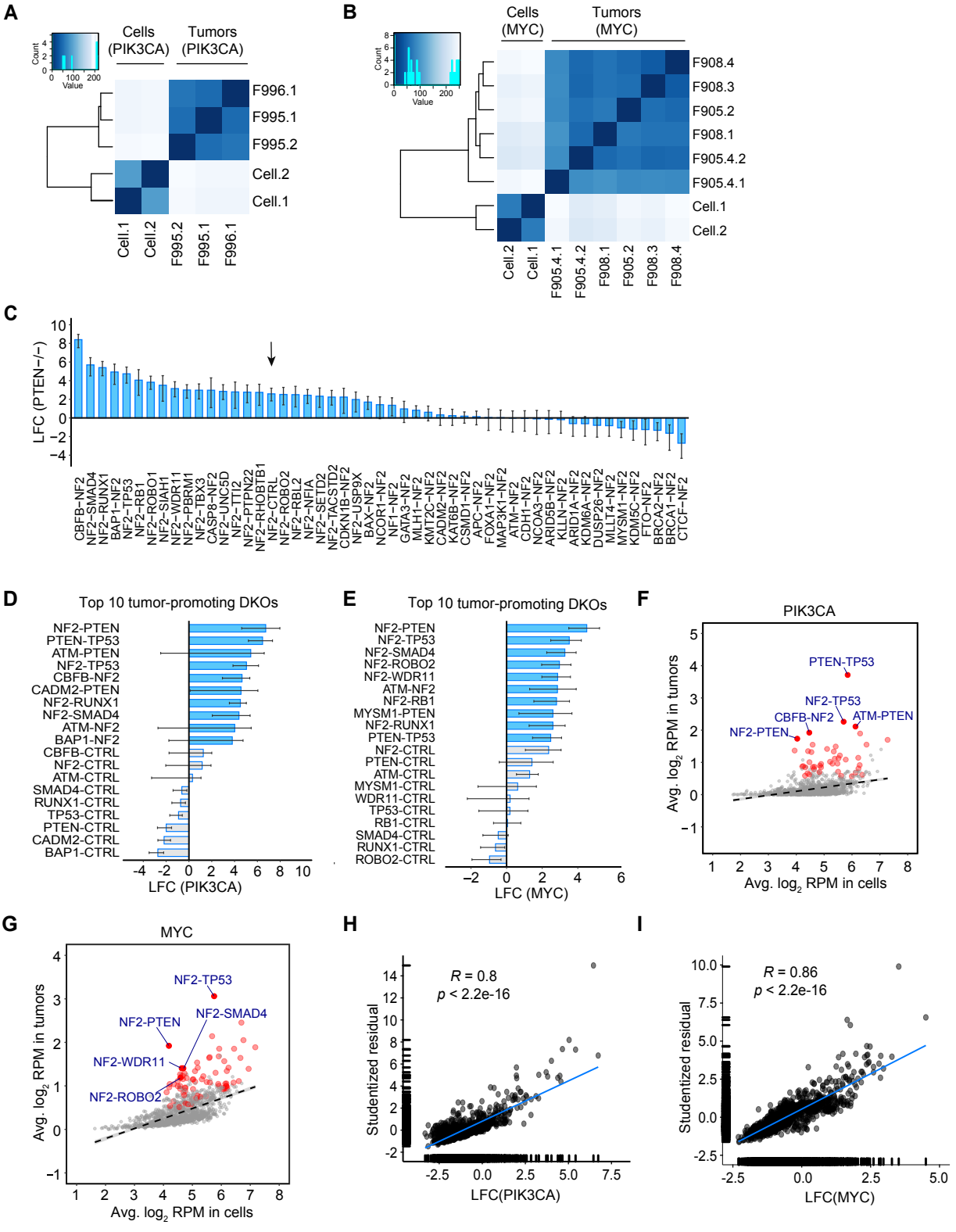
Supplementary Figure S2

Related to Figure 1

Tumor-promoting effects of TSG perturbations in the PTEN-/-, PIK3CA and MYC contexts.

(A) and (F) Heatmaps showing the Euclidean distances among all the cell and tumor samples; (A) for PIK3CA context and (F) for MYC context. (B) and (G) The top 10 tumor promoting DKOs and its corresponding SKOs in the PIK3CA and MYC contexts; (B) for PIK3CA and (G) for MYC. (C) The effect size at gene level (LFCs) of NF2 SKO and its DKOs in the PTEN-/- contexts. (D) and (H) Scatter plots of average $\log_2$ RPM in the tumors versus pre-injected cells; (D) for PIK3CA and (H) for MYC. All the oncogenic perturbations are highlighted in light and dark red and the top 5 ones are labeled. The linear regression line is shown (dashed). (E) and (I) Scatter plot and correlation analysis of the tumorigenic effects measured by quantile analysis versus average based regression for PIK3CA and MYC context; (E) for PIK3CA and (I) for MYC.

Figure S2


A
Cells (PIK3CA)
Tumors (PIK3CA)
F996.1
F995.1
F995.2
Cell.2
Cell.1
B
Cells (MYC)
Tumors (MYC)
F908.4
F908.3
F905.2
F908.1
F905.4.2
F905.4.1
Cell.1
Cell.2
C
LFC (PTEN−/−)
D
Top 10 tumor-promoting DKOs
NF2-PTEN
PTEN-TP53
ATM-PTEN
NF2-TP53
CBFB-NF2
CADM2-PTEN
NF2-RUNX1
NF2-SMAD4
ATM-NF2
BAP1-NF2
CBFB-CTRL
NF2-CTRL
ATM-CTRL
SMAD4-CTRL
RUNX1-CTRL
TP53-CTRL
PTEN-CTRL
CADM2-CTRL
BAP1-CTRL
LFC (PIK3CA)
E
Top 10 tumor-promoting DKOs
NF2-PTEN
NF2-TP53
NF2-SMAD4
NF2-ROBO2
NF2-WDR11
ATM-NF2
NF2-RB1
MYSM1-PTEN
NF2-RUNX1
PTEN-TP53
NF2-CTRL
PTEN-CTRL
ATM-CTRL
MYSM1-CTRL
WDR11-CTRL
TP53-CTRL
RB1-CTRL
SMAD4-CTRL
RUNX1-CTRL
ROBO2-CTRL
LFC (MYC)
F
PIK3CA
PTEN-TP53
NF2-TP53
CBFB-NF2
ATM-PTEN
NF2-PTEN
Avg. log2 RPM in tumors
Avg. log2 RPM in cells
G
MYC
NF2-TP53
NF2-SMAD4
NF2-PTEN
NF2-WDR11
NF2-ROBO2
Avg. log2 RPM in tumors
Avg. log2 RPM in cells
H
R = 0.8
p < 2.2e-16
Studentized residual
LFC(PIK3CA)
I
R = 0.86
p < 2.2e-16
Studentized residual
LFC(MYC)

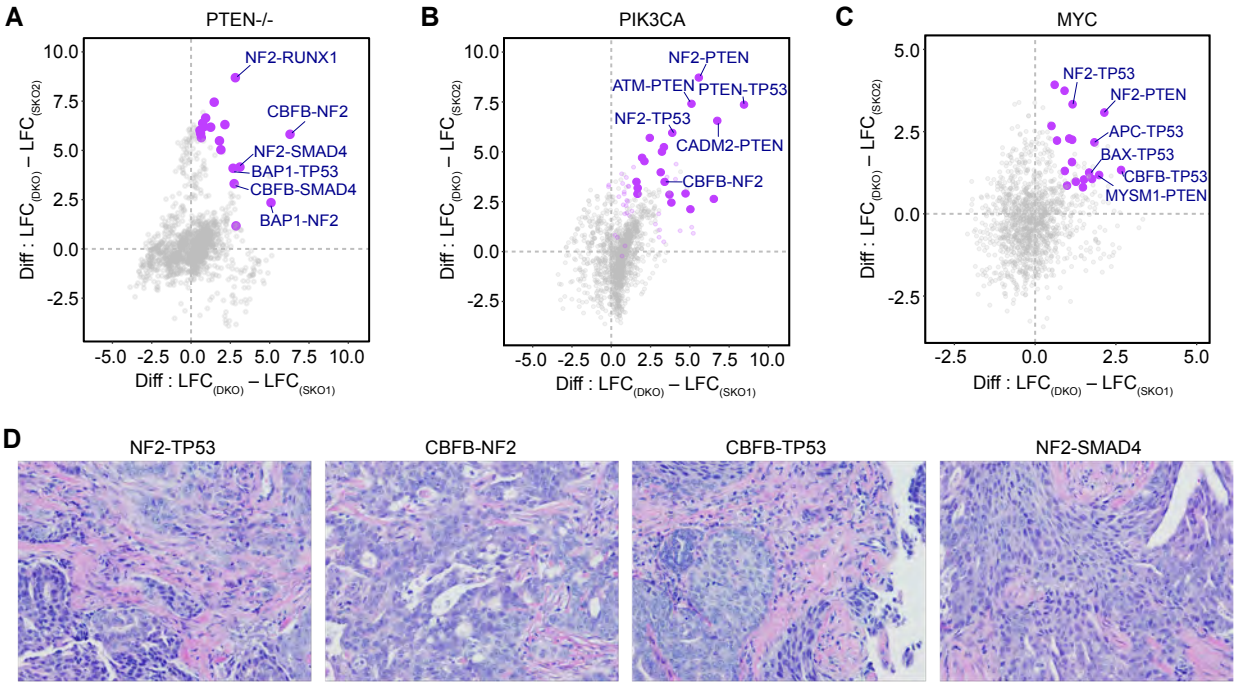
Supplementary Figure S3

Related to Figure 2

Histological analysis of the oncogenic GIs.

(A) (B) and (C) Scatter plot of differential LFCs of DKOs from the two respective SKOs in the PTEN-/-, PIK3CA and MYC contexts. Oncogenic GI pairs are highlighted in purple. (D) H&E staining of tumors induced by four oncogenic GIs validated in Figure 2F.

Figure S3


A
PTEN-/-
NF2-RUNX1
CBFB-NF2
NF2-SMAD4
BAP1-TP53
CBFB-SMAD4
BAP1-NF2
Diff : LFC(DKO) – LFC(SKO2)
Diff : LFC(DKO) – LFC(SKO1)
B
PIK3CA
NF2-PTEN
ATM-PTEN
PTEN-TP53
NF2-TP53
CADM2-PTEN
CBFB-NF2
C
MYC
NF2-TP53
NF2-PTEN
APC-TP53
BAX-TP53
CBFB-TP53
MYSM1-PTEN
D
NF2-TP53
CBFB-NF2
CBFB-TP53
NF2-SMAD4

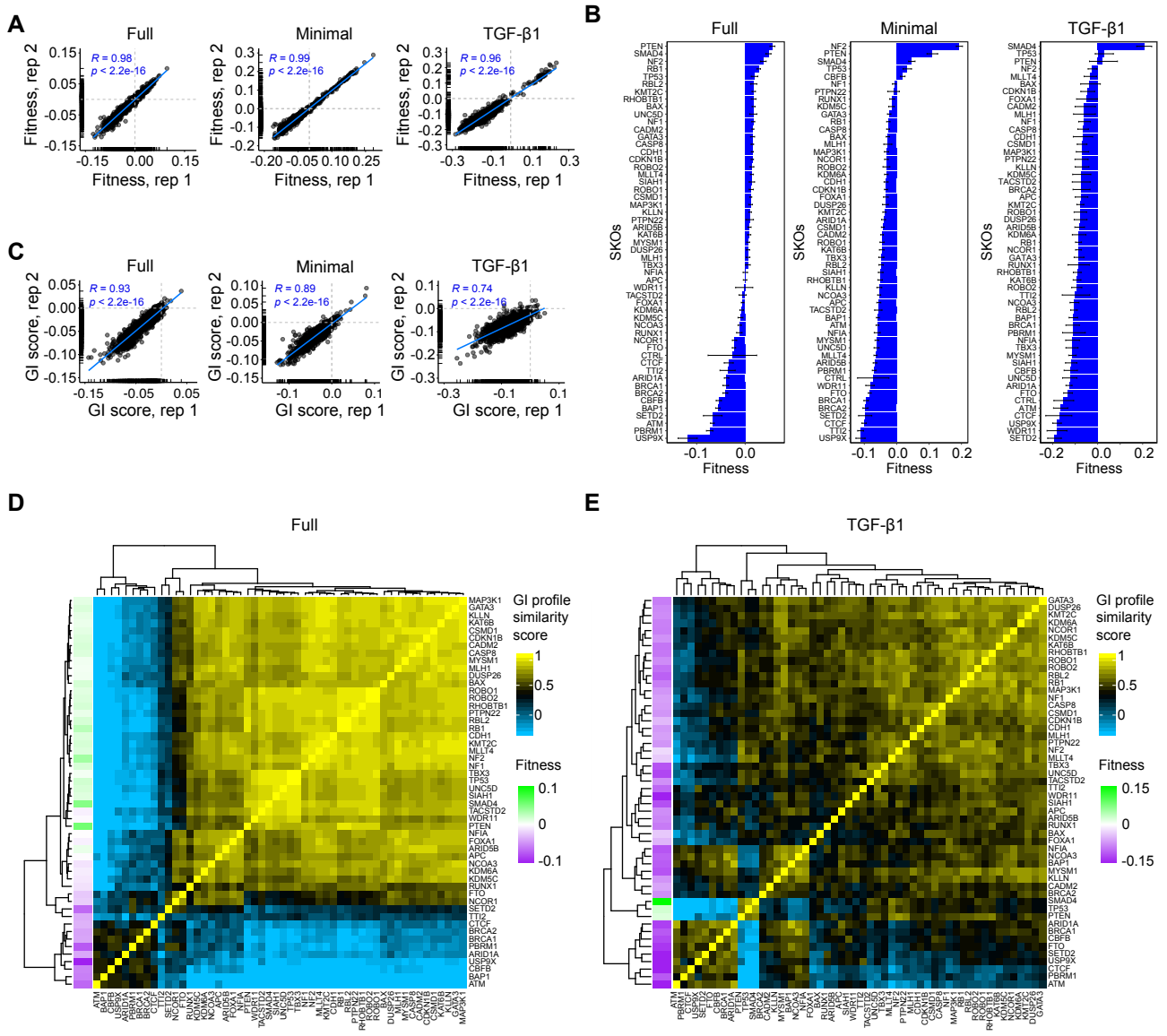
Supplementary Figure S4

Related to Figure 3 and 4

Biological reproducibility of fitness and GI scores *in vitro*.

(A) Scatterplots of fitnesses measured from the two biological replicates in full medium (left), minimal medium (middle) and medium supplemented with TGF-β1 (right). (B) Gene-level fitness of 52 SKOs in full (left), minimal (middle) and TGF-β1 (right) supplemented media. (C) Scatter plots of GI scores calculated from the two biological replicates in full medium (left), minimal medium (middle) and medium supplemented with TGF-β1 (right). (D) and (E) GI profiles of TSGs in full medium and TGF-β1 medium.

Figure S4


A
Full
Minimal
TGF-β1
R = 0.98
p < 2.2e-16
R = 0.99
p < 2.2e-16
R = 0.96
p < 2.2e-16
Fitness, rep 2
Fitness, rep 1
B
Full
Minimal
TGF-β1
SKOs
Fitness
C
Full
Minimal
TGF-β1
R = 0.93
p < 2.2e-16
R = 0.89
p < 2.2e-16
R = 0.74
p < 2.2e-16
GI score, rep 2
GI score, rep 1
D
Full
GI profile similarity score
Fitness
E
TGF-β1
GI profile similarity score
Fitness

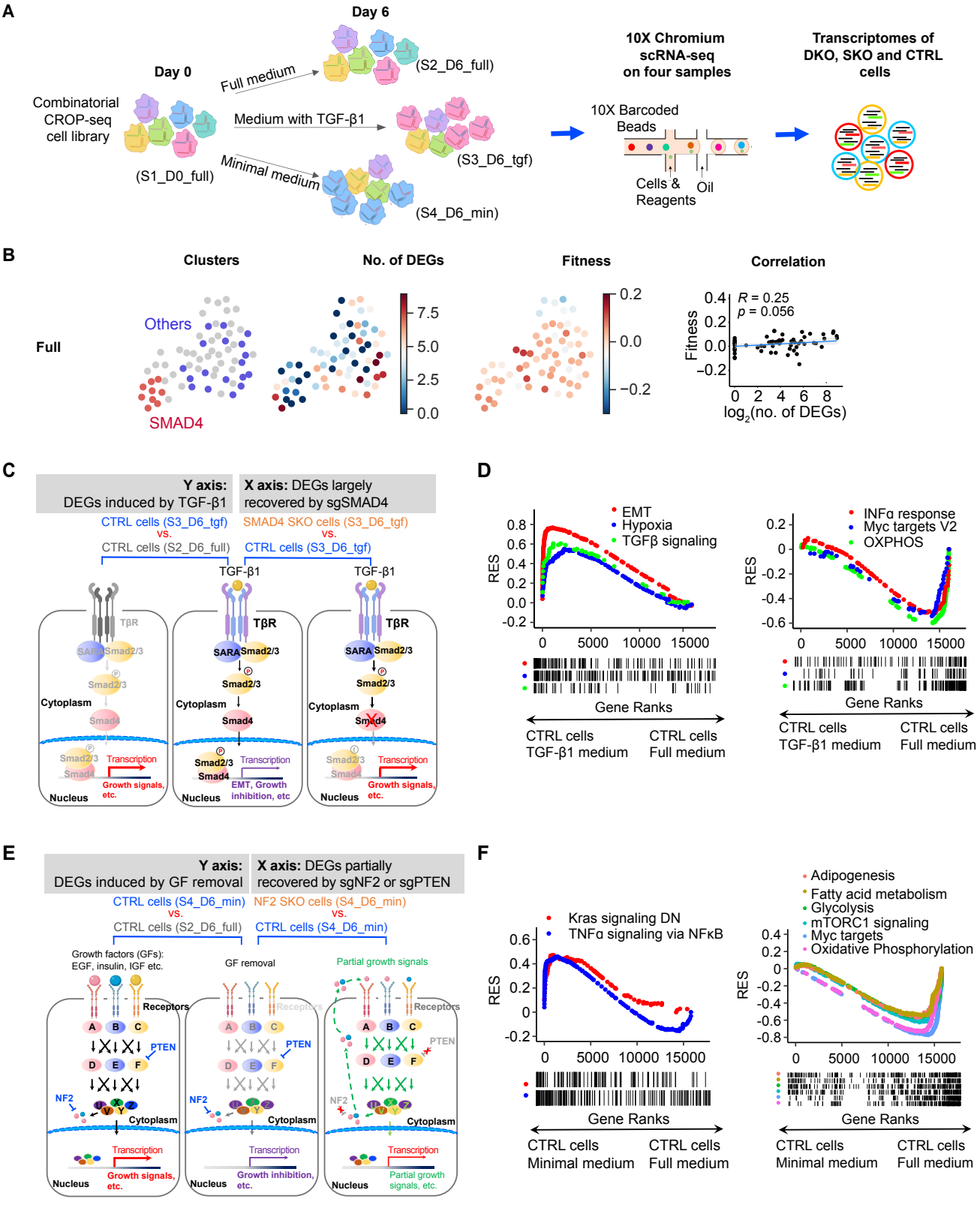
Supplementary Figure S5

Related to Figure 5

Transcriptional profiles of combinational CROP-seq under different conditions.

(A) Experimental design for combinatorial CROP-seq. (B) Relationship of transcriptional states and fitness in full medium. (C) Cartoon revealing a proposed mechanism of TGF-β1 induced differential expressions and sgSMAD4 induced recovery. (D) GSEA analysis showing that enriched and underrepresented hallmark gene sets induced by TGF-β1. (E) Cartoon revealing a proposed mechanism of growth-factor removal induced differential expressions and sgNF2 or sgPTEN induced recovery. (F) GSEA analysis showing that enriched and underrepresented hallmark gene sets by removal of growth factors.

Figure S5


A
Day 6
Day 0
Full medium
(S2_D6_full)
Combinatorial CROP-seq cell library
Medium with TGF-β1
(S3_D6_tgf)
(S1_D0_full)
Minimal medium
(S4_D6_min)
10X Chromium scRNA-seq on four samples
10X Barcoded Beads
Cells & Reagents
Oil
Transcriptomes of DKO, SKO and CTRL cells
B
Clusters
No. of DEGs
Fitness
Correlation
Full
Others
SMAD4
7.5
5.0
2.5
0.0
0.2
0.0
−0.2
R = 0.25
p = 0.056
Fitness
log2(no. of DEGs)
C
Y axis: DEGs induced by TGF-β1
X axis: DEGs largely recovered by sgSMAD4
CTRL cells (S3_D6_tgf) vs. CTRL cells (S2_D6_full)
SMAD4 SKO cells (S3_D6_tgf) vs. CTRL cells (S3_D6_tgf)
TGF-β1
TβR
SARA
Smad2/3
Cytoplasm
Smad4
Transcription
Growth signals, etc.
Nucleus
EMT, Growth inhibition, etc
D
EMT
Hypoxia
TGFβ signaling
INFα response
Myc targets V2
OXPHOS
RES
Gene Ranks
CTRL cells TGF-β1 medium
CTRL cells Full medium
E
Y axis: DEGs induced by GF removal
X axis: DEGs partially recovered by sgNF2 or sgPTEN
CTRL cells (S4_D6_min) vs. CTRL cells (S2_D6_full)
NF2 SKO cells (S4_D6_min) vs. CTRL cells (S4_D6_min)
Growth factors (GFs): EGF, insulin, IGF etc.
GF removal
Partial growth signals
Receptors
PTEN
NF2
Growth inhibition, etc.
Partial growth signals, etc.
F
Kras signaling DN
TNFα signaling via NFκB
Adipogenesis
Fatty acid metabolism
Glycolysis
mTORC1 signaling
Myc targets
Oxidative Phosphorylation
CTRL cells Minimal medium
CTRL cells Full medium

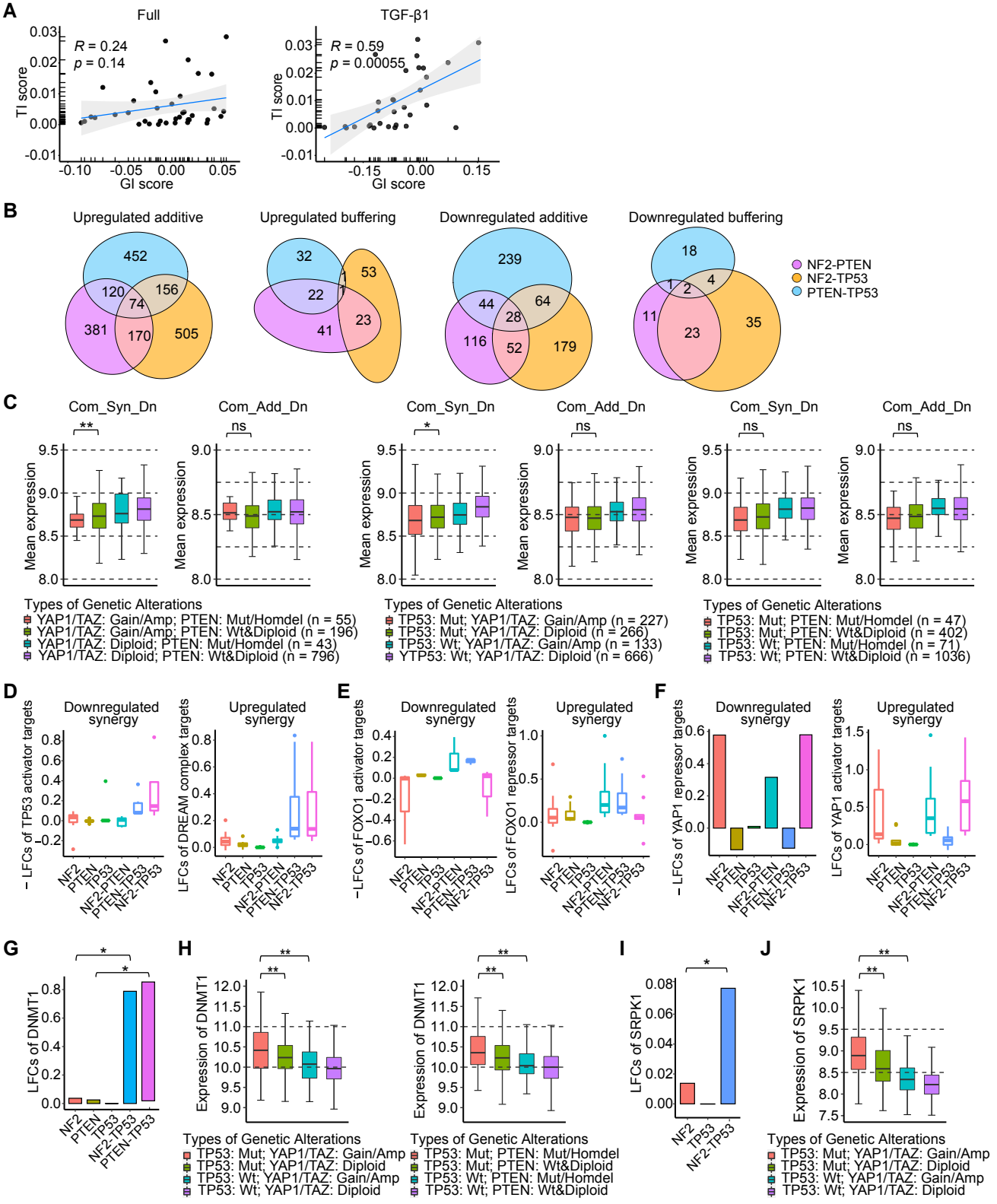
Supplementary Figure S6

Related to Figure 6 and 7

Transcriptional epistasis in DKOs and breast cancer patients.

(A) Correlation of TI scores and fitness GI scores, displayed as scatter plots, in full (left) and TGF-β1 supplemented media (right). (B) Number of additive and buffering DEGs in the three DKOs. (C) Average expression of the 9 common downregulated synergistic DEGs and 28 common downregulated additive DEGs in the tumors with dual and single alterations. **p < 0.05, *p < 0.1; one-sided Student's t-test. ns: not significant. (D) Synergistic expressions of p53 and DREAM complex targets in TP53-specific DKOs. (E) Synergistic expression of targets of FOXO1 activator and repressor in PTEN-specific DKOs. (F) Synergistic expression of targets of YAP1 repressors and activators in NF2-specific DKOs. (G) and (H) Synergistic expression of *DNMT1* in *TP53*-deleted DKO cells and *TP53*-mutated double mutant breast cancer patients. **p < 0.05; *p < 0.1; one-sided Student's t-test. (I) and (J) Synergistic expression of *SRPK1* in *NF2-PTEN* DKO cells and *NF2-PTEN* double mutant breast cancer patients. **p < 0.05; *p < 0.1; one-sided Student's t-test.

Figure S6


A
Full
TGF-β1
R = 0.24
p = 0.14
R = 0.59
p = 0.00055
TI score
GI score
B
Upregulated additive
Upregulated buffering
Downregulated additive
Downregulated buffering
NF2-PTEN
NF2-TP53
PTEN-TP53
C
Com_Syn_Dn
Com_Add_Dn
Mean expression
Types of Genetic Alterations
YAP1/TAZ: Gain/Amp; PTEN: Mut/Homdel (n = 55)
YAP1/TAZ: Gain/Amp; PTEN: Wt&Diploid (n = 196)
YAP1/TAZ: Diploid; PTEN: Mut/Homdel (n = 43)
YAP1/TAZ: Diploid; PTEN: Wt&Diploid (n = 796)
TP53: Mut; YAP1/TAZ: Gain/Amp (n = 227)
TP53: Mut; YAP1/TAZ: Diploid (n = 266)
TP53: Wt; YAP1/TAZ: Gain/Amp (n = 133)
YTP53: Wt; YAP1/TAZ: Diploid (n = 666)
TP53: Mut; PTEN: Mut/Homdel (n = 47)
TP53: Mut; PTEN: Wt&Diploid (n = 402)
TP53: Wt; PTEN: Mut/Homdel (n = 71)
TP53: Wt; PTEN: Wt&Diploid (n = 1036)
D
Downregulated synergy
Upregulated synergy
– LFCs of TP53 activator targets
LFCs of DREAM complex targets
E
– LFCs of FOXO1 activator targets
LFCs of FOXO1 repressor targets
F
– LFCs of YAP1 repressor targets
LFCs of YAP1 activator targets
NF2
PTEN
TP53
NF2-PTEN
PTEN-TP53
NF2-TP53
G
LFCs of DNMT1
H
Expression of DNMT1
I
LFCs of SRPK1
J
Expression of SRPK1

SUPPLEMENTARY REFERENCES

1. Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, et al. Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. Nat Methods 2017;14:573–6.

2. Schlecht U, Liu Z, Blundell JR, St.Onge RP, Levy SF. A scalable double-barcode sequencing platform for characterization of dynamic protein-protein interactions. Nat Commun 2017;8:1–9.

3. Zhao L, Liu Z, Levy SF, Wu S. Bartender: a fast and accurate clustering algorithm to count barcode reads. Bioinformatics 2018;34:739–47.

4. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 2003;13:2498–504.

5. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888-1902.e21.

6. Norman TM, Horlbeck MA, Replogle JM, Ge AY, Xu A, Jost M, et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. Science 2019;365:786–93.

7. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci 2005;102:15545–50.

8. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst 2015;1:417–25.

9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

10. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinforma Oxf Engl. 2019;35:2084–92.

11. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res 2019;47:D419–26.

12. Fischer M. Census and evaluation of p53 target genes. Oncogene 2017;36:3943–56.

13. Fischer M, Grossmann P, Padi M, DeCaprio JA. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. Nucleic Acids Res 2016;44:6070–86.

14. Ringnér M, Fredlund E, Häkkinen J, Borg Å, Staaf J. GOBO: Gene Expression-Based Outcome for Breast Cancer Online. PLOS ONE 2011;6:e17911.