

# CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities

<https://doi.org/10.1038/s41586-020-2099-x>

Received: 25 April 2019

Accepted: 10 January 2020

Published online: 11 March 2020

 Check for updates

Kyuho Han<sup>1</sup>✉, Sarah E. Pierce<sup>2</sup>, Amy Li<sup>1</sup>, Kaitlyn Spees<sup>1</sup>, Gray R. Anderson<sup>1</sup>, Jose A. Seoane<sup>3,4</sup>, Yuan-Hung Lo<sup>3</sup>, Michael Dubreuil<sup>1,2</sup>, Micah Olivas<sup>1</sup>, Roarke A. Kamber<sup>1</sup>, Michael Wainberg<sup>5</sup>, Kaja Kostyrko<sup>6</sup>, Marcus R. Kelly<sup>2</sup>, Maryam Yousefi<sup>1</sup>, Scott W. Simpkins<sup>1</sup>, David Yao<sup>1</sup>, Keonil Lee<sup>1</sup>, Calvin J. Kuo<sup>2,3</sup>, Peter K. Jackson<sup>2,7</sup>, Alejandro Sweet-Cordero<sup>6</sup>, Anshul Kundaje<sup>1,5</sup>, Andrew J. Gentles<sup>8</sup>, Christina Curtis<sup>1,2,3,4</sup>, Monte M. Winslow<sup>1,2,9</sup> & Michael C. Bassik<sup>1,2,10</sup>✉

Cancer genomics studies have identified thousands of putative cancer driver genes<sup>1</sup>. Development of high-throughput and accurate models to define the functions of these genes is a major challenge. Here we devised a scalable cancer-spheroid model and performed genome-wide CRISPR screens in 2D monolayers and 3D lung-cancer spheroids. CRISPR phenotypes in 3D more accurately recapitulated those of *in vivo* tumours, and genes with differential sensitivities between 2D and 3D conditions were highly enriched for genes that are mutated in lung cancers. These analyses also revealed drivers that are essential for cancer growth in 3D and *in vivo*, but not in 2D. Notably, we found that carboxypeptidase D is responsible for removal of a C-terminal RKRR motif<sup>2</sup> from the α-chain of the insulin-like growth factor 1 receptor that is critical for receptor activity. Carboxypeptidase D expression correlates with patient outcomes in patients with lung cancer, and loss of carboxypeptidase D reduced tumour growth. Our results reveal key differences between 2D and 3D cancer models, and establish a generalizable strategy for performing CRISPR screens in spheroids to reveal cancer vulnerabilities.

Despite the large increase in the catalogue of mutations observed across diverse cancer types (the ‘long tail’)<sup>1</sup>, it is frequently unclear which mutations are functional cancer drivers. Therefore, it is a central challenge to scalably investigate these genes in relevant cancer models to assign causality and identify cancer-specific vulnerabilities.

Existing *in vitro* and *in vivo* models are useful for defining the biological properties of cancer<sup>3–7</sup>, but each has limitations. Genetically engineered mouse models recapitulate tumour growth and microenvironment, but are limited by scalability, time and cost<sup>8</sup>. Xenograft-based models are limited in scale, and can be difficult to manipulate *in vitro*. Genome-scale investigation of cancer growth and drug sensitivity has largely relied on *in vitro* 2D cell culture<sup>9–12</sup>, which lacks many features of disease, such as hypoxia<sup>13</sup>, altered cell–cell contacts<sup>14</sup> and rewired metabolism<sup>15</sup>. *In vitro* organoid models alleviate some of these concerns<sup>3,16</sup>, but are much less scalable.

CRISPR has enabled substantially improved genetic screening in *in vitro* and *in vivo* cancer models<sup>9,11,12,17–19</sup>. Efforts such as DepMap have characterized cancer dependencies using genome-scale CRISPR screens in hundreds of cell lines, revealing previously undiscovered cancer drivers<sup>10,20–22</sup>. Nonetheless, it has been difficult to evaluate how differences in culture systems affect the ability to accurately uncover cancer drivers *in vivo*.

Here we devised a scalable method to propagate lung adenocarcinoma spheroids, and performed genome-wide CRISPR screens in both

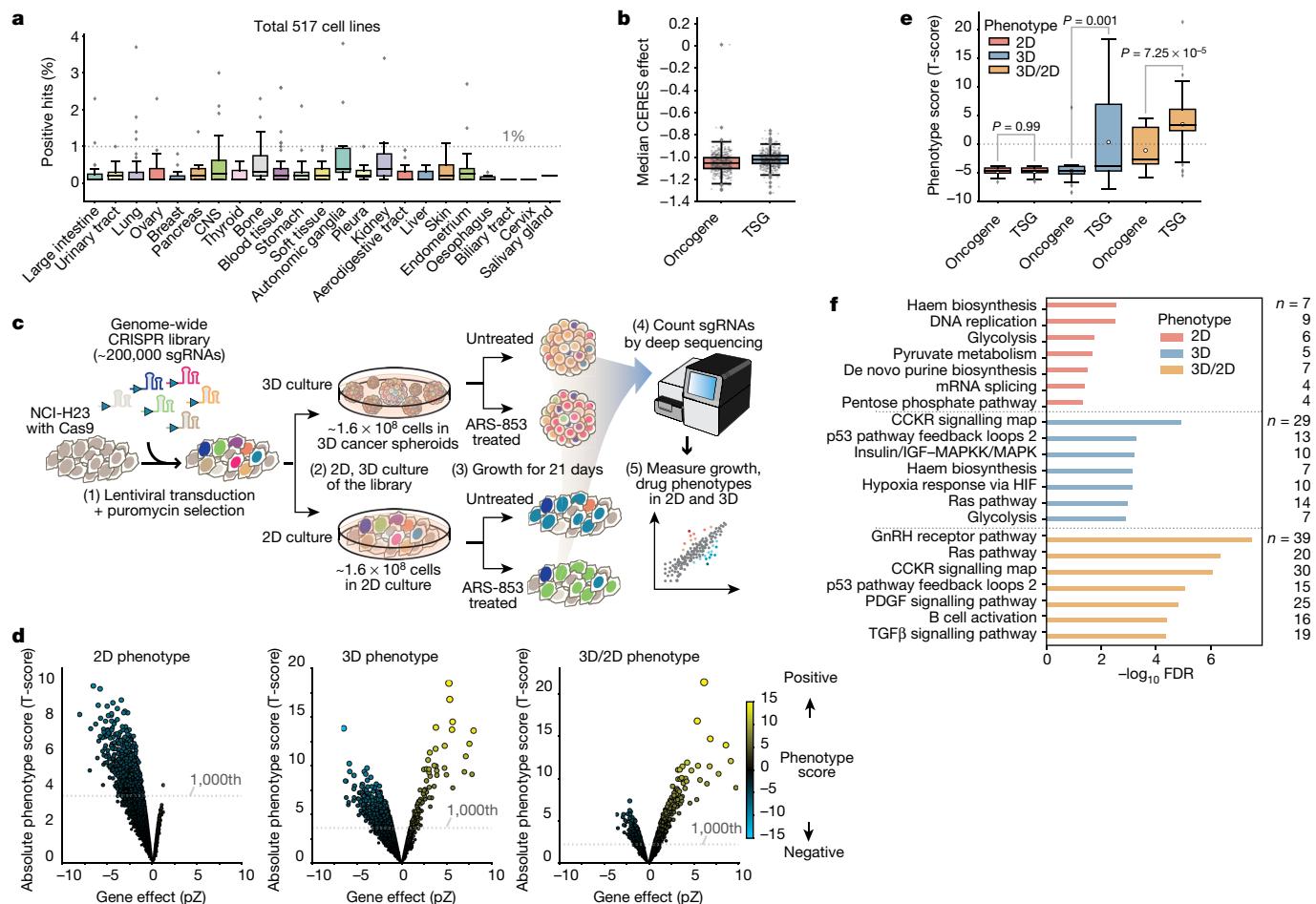
2D monolayers and 3D spheroids. Growth phenotypes in 3D more accurately resembled those observed in tumours. Furthermore, genes with differentially stronger effects in 3D were enriched for those significantly mutated in human lung cancers. Among these genes, we identified carboxypeptidase D (CPD), a poorly characterized carboxypeptidase, as an important enzyme for maturation of insulin-like growth factor 1 receptor (IGF1R). Together, these results demonstrate a strategy for genome-scale CRISPR screens in 3D spheroids to identify actionable cancer vulnerabilities.

## Scalable 3D spheroid system for CRISPR screens

Although CRISPR screens performed in 2D monolayers have produced a wealth of information<sup>9–12,23</sup>, they often fail to replicate key aspects of tumour biology<sup>24</sup>. This is illustrated by phenotypes measured across more than 500 screens from the DepMap project. Although this resource has uncovered many valuable biological findings<sup>10,20–22</sup>, less than 1% of the top 1,000 hits show a positive growth effect (Fig. 1a). Indeed, inactivation of known tumour-suppressor genes often results in negative phenotypes in 2D culture (Fig. 1b).

We sought to develop a scalable 3D spheroid system to enable high-throughput screens that more closely approximate *in vivo* cancers. We optimized seeding density and methylcellulose concentrations

<sup>1</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. <sup>2</sup>Program in Cancer Biology, Stanford University School of Medicine, Stanford, CA, USA. <sup>3</sup>Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. <sup>4</sup>Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. <sup>5</sup>Department of Computer Science, Stanford University, Stanford, CA, USA. <sup>6</sup>Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA. <sup>7</sup>Baxter Laboratory, Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. <sup>8</sup>Departments of Medicine and Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA. <sup>9</sup>Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. <sup>10</sup>Program in Chemistry, Engineering and Medicine for Human Health (ChEM-H), Stanford University, Stanford, CA, USA. <sup>✉</sup>e-mail: kyuhohan@stanford.edu; bassik@stanford.edu



**Fig. 1 | Genome-wide screens performed on 3D cultures improve detection of cancer genes and pathways compared with those performed in 2D.**

**a**, Percentage of positive hits among the top 1,000 hits in the DepMap dataset<sup>10</sup>. Each point represents one cell line. **b**, Median CERES<sup>10</sup> effects of oncogenes and tumour-suppressor genes (TSGs) (annotated in COSMIC<sup>30</sup>) among top 1,000 hits of 517 DepMap cell lines; each data point represents one cell line.

**c**, Schematic for CRISPR screens in H23 cells. **d**, Distributions of phenotypes. Values on the x axis show the effect size of each gene and the y axis shows absolute phenotype score (T-score) (see Methods). Dot size represents

absolute T-score. **e**, Phenotype scores for oncogenes and TSGs in top 1,000 hits in each condition. P-values calculated using two-sided t-test. **f**, Enriched pathways among the top 1,000 hits from each condition analysed using PANTHER overrepresentation test (see Methods). Significance of enriched pathways were measured with Fisher's exact test and the Benjamini–Hochberg false-discovery rates (FDR) were subsequently computed (xaxis). The number of genes in enriched pathways is shown on the right. In all box plots, box limits mark upper and lower quartiles, whiskers represent 1.5× the interquartile range and points show outliers.

(Extended Data Fig. 1a, Supplementary Video 1; see Methods) to enable propagation of approximately 200 million cells in 3D spheroids in low-attachment plates. This enabled us to perform genome-wide CRISPR screens in H23 lung adenocarcinoma cells grown in either 2D monolayers or 3D spheroids (Fig. 1c) using our custom single guide RNA (sgRNA) library<sup>25</sup>. Since H23 cells contain a KRAS<sup>G12C</sup> mutation, we also screened with ARS-853<sup>26–28</sup>, a cysteine-reactive KRAS inhibitor that frequently has stronger effects in 3D<sup>29</sup>.

### 3D phenotypes reflect cancer dependencies

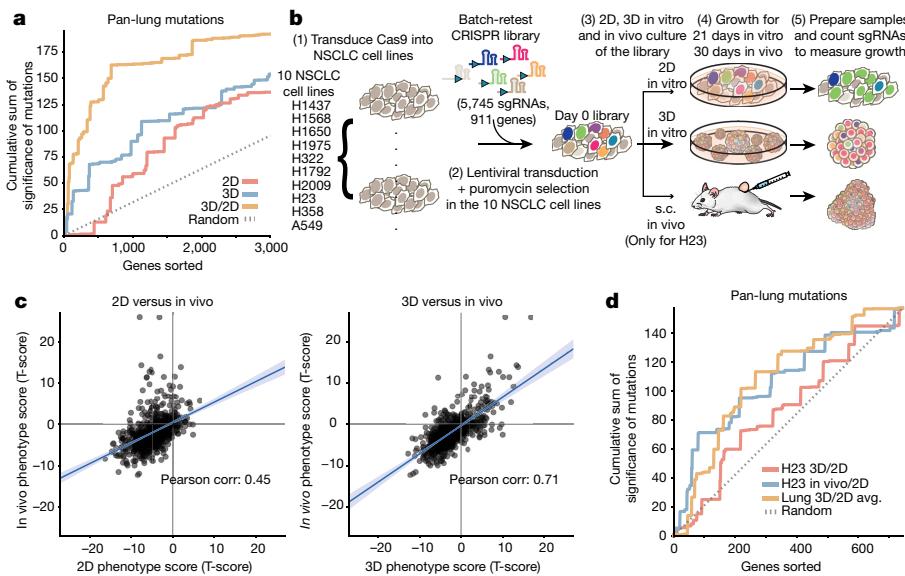
Reproducibility and quality of 3D screening data were equivalent to those of the 2D data (Extended Data Fig. 1b–d, Supplementary Table 1), and it was immediately apparent that CRISPR screens in 3D uncovered many more positive growth phenotypes, whereas most hits from the 2D screens had negative phenotypes (Fig. 1d). This became more apparent when we examined genes with differential effects in 3D by normalizing 3D phenotypes against the corresponding 2D phenotypes (3D/2D) (see Methods). We next analysed phenotypes for oncogenes and tumour suppressor genes (TSGs) annotated in the COSMIC database<sup>30</sup> within the top 1,000 hits in 2D or 3D conditions. Both groups were similar in 2D, showing negative

median-growth phenotypes (Fig. 1e). In 3D spheroids, however, oncogenes and TSGs exhibited markedly different behaviours, with knockout of TSGs showing more positive-growth phenotypes; this was clearer when the 3D/2D phenotype was considered (Fig. 1e, Extended Data Fig. 1e).

Pathway-enrichment analysis revealed that a distinct set of cancer-specific pathways—including p53 and Ras pathways (known drivers in H23 cells)—was enriched among hits in 3D and 3D/2D phenotypes, whereas 2D hits were generally related to common essential cellular functions such as DNA replication (Fig. 1f). Together, these data suggest that screens in 3D more accurately capture features of cancer genes and pathways (Extended Data Fig. 1f).

### 3D hits are frequently mutated in cancer

We further investigated the phenotypes for genes frequently mutated in lung adenocarcinoma and squamous cell carcinoma<sup>31</sup> (hereafter, ‘pan-lung’). When genes were sorted by the absolute value of their phenotypic strength, inactivation of the ten most-frequently mutated genes in the Pan-lung cancer cohort<sup>31</sup> showed weaker and more widely distributed effects in 2D (Extended Data Fig. 1g, Supplementary Table 2). By contrast, many of these frequently mutated genes showed stronger



**Fig. 2 | Genes with differential 3D/2D phenotypes are enriched for significantly mutated lung cancer genes.** **a**, Cumulative sum of the significance of 11,249 pan-lung cancer genes from 1,144 patients with lung cancer<sup>31,51</sup> measured by MutSig2CV<sup>31</sup>. The x-axis shows phenotypes sorted by strength in 2D, 3D and 3D/2D. The top 3,000 genes are shown. **b**, Schematic for batch-retest CRISPR screens. s.c., subcutaneous injection. **c**, Comparisons between in vitro and in vivo phenotypes in the H23 batch-retest screens. Data

are fit by linear regression (blue line); shaded bands indicate 95% confidence intervals. Pearson corr., Pearson correlation coefficient. **d**, Significance of 744 pan-lung cancer genes measured by MutSig2CV<sup>31</sup>, displayed as cumulative sum plots against genes sorted by absolute values of 3D/2D phenotypes in H23 cells, average (avg.) 3D/2D phenotypes across 10 lung cancer lines, and H23 in vivo/2D phenotypes in batch-retest screens.

phenotypes in 3D spheroids. Notably, the 3D/2D phenotypes showed a further improved ability to detect strong phenotypes for genes that are frequently mutated in lung cancer. This is consistent with the pathway enrichment analysis described above, and suggests that analysis of genes with differentially strong effects in 3D may increase the power to identify cancer drivers.

To systematically confirm this, we compared absolute CRISPR phenotypes (sorted by phenotypic strength) with the cumulative sum of significance of pan-lung cancer mutations<sup>31</sup> (Fig. 2a, Supplementary Table 3). Again, genes with stronger phenotypes in 3D and, to a greater extent, those with stronger phenotypes in 3D/2D, were enriched for significant lung-cancer mutations. We reasoned that two factors probably contribute to this enrichment. First, normalizing 3D with 2D phenotypes may unmask cancer-specific genes by minimizing the otherwise dominating effects of core essential genes (for example, ribosomal genes) that are critical for both 2D and 3D growth (Extended Data Fig. 1g). Second, as previously suggested<sup>32</sup>, 3D spheroids are more likely to mimic in vivo tumours.

Additional genome-wide screens in H1975 and H2009 lung cancer lines confirmed key advantages of 3D spheroids, including improved detection of cancer pathways and identification of the known drivers for each of these lines (EGFR–PI3K and p53–KRAS, respectively, Extended Data Figs. 2, 3, Supplementary Discussion).

### 3D spheroids better match tumour xenografts

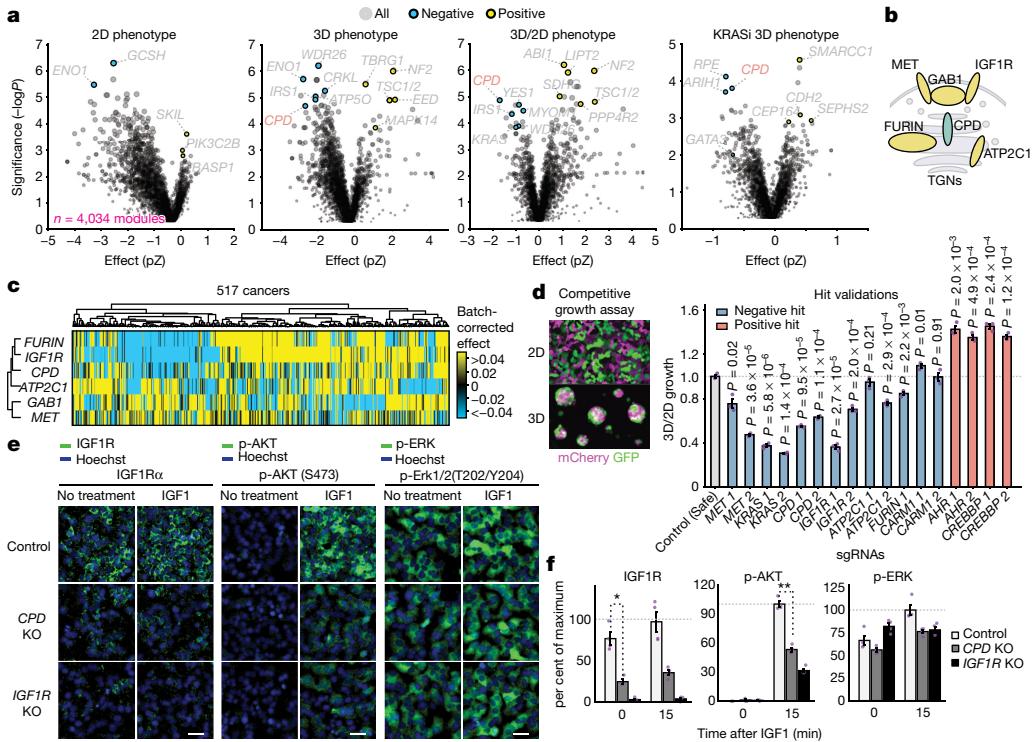
To systematically compare CRISPR screens in 2D monolayers, 3D spheroids and tumour xenografts, we generated a small batch-retest sgRNA library targeting 911 top hits with differential 3D growth effects in our genome-wide screens (Fig. 2b, Supplementary Table 4). We transduced this library into H23 cells and compared growth in subcutaneous xenograft tumours with growth in 2D and 3D cultures. We optimized a protocol (see Methods) for in vivo CRISPR screening, and obtained highly reproducible data from tumour xenografts (Extended Data Fig. 4a, b, Supplementary Table 5). Notably, phenotypes of genes from 3D screens were much more closely correlated with those in mouse xenograft than those from 2D screens (Fig. 2c, Supplementary Discussion).

To search for common 3D-selective vulnerabilities in lung adenocarcinoma, we used the same batch-retest library to perform 2D and 3D screens across multiple cancer lines. We again observed marked differences between 2D and 3D screens in all lines (Supplementary Table 5). Averaging 3D/2D phenotypes across ten cell lines further increased detection of significant mutations observed in patients with lung cancer compared with phenotypes from the H23 cell line alone (Fig. 2d). Of note, comparison of in vivo phenotypes with 2D phenotypes (in vivo/2D) in H23 cells also increased detection of significant mutations compared with the in vitro 3D/2D phenotypes. Notably, top sensitizing hits from the averaged 3D/2D phenotypes include several known regulators of RAS–MAPK pathways<sup>33</sup> such as GRB2, SHOC2, PTPN11 (also known as SHP2), GAB1 and MAPK1.

### CPD module shows selective 3D growth effects

Given that genes with strong 3D/2D phenotypes are enriched for lung cancer mutations, we reasoned that these might include novel therapeutic targets. To identify such targets, we defined functional gene modules on the basis of their correlated phenotypes in DepMap<sup>22</sup> and examined their phenotypes. Simultaneous depletion of multiple genes from the same functional group should help define vulnerabilities within pathways/complexes; indeed, we identified a number of differentially enriched modules from expected genes, including KRAS, mTOR and Hippo pathways (Supplementary Discussion).

Notably, a module composed of genes correlated with CPD was the most strongly depleted in the 3D/2D phenotype (Fig. 3a, Extended Data Fig. 4c) and showed strong synthetic lethality with the KRAS(G12C) inhibitor specifically in 3D. This suggested that CPD and its functional interactors could be promising therapeutic targets. CPD is a poorly characterized member of the metallocarboxypeptidase family that cleaves C-terminal arginines and lysines from polypeptides<sup>34</sup>; it is localized in the *trans*-Golgi network<sup>35</sup>. CPD is correlated with *FURIN*, *ATP2C1*, *IGF1R*, *MET* and *GAB1* in a functional module (Fig. 3b, c, Extended Data Fig. 4d–f), but not with a control olfactory receptor gene. Given that *FURIN* and *ATP2C1* are critical for processing of *IGF1R* and *MET* in the *trans*-Golgi<sup>36–38</sup>, we hypothesized that CPD might have a related role.



**Fig. 3 | The CPD module is critical for 3D spheroid growth and IGF1R function.** **a**, Top negative modules (blue). Top positive modules (yellow). The y axis shows significance of enrichment for co-essential modules ( $P$  values, two-sided Mann–Whitney  $U$ -test); the x axis shows average gene effects of co-essential modules (see Methods). KRASi, KRAS inhibitor. **b**, Proteins encoded by genes in the CPD co-essential module and their localization along the TGN–plasma membrane. **c**, Cluster map showing batch-corrected CERES gene effects for the CPD module. **d**, CPD module and selected top 3D/2D hits were validated with individual sgRNAs in competitive growth assays. Left

micrographs show example images of growth assays for control sgRNA in 2D and 3D. Data are mean  $\pm$  s.e.m.,  $n = 3$ ;  $P$  values calculated by two-sided  $t$ -test between the control and gene-targeting sgRNAs. Control (safe) sgRNAs are described in the Methods. **e**, Control, CPD-knockout (KO) and *IGF1R*-knockout H23 cells grown in 2D were stimulated with IGF1 (100 ng ml $^{-1}$ ) for 15 min and levels of IGF1R and activities of downstream effectors were measured by immunofluorescence. Scale bars, 20  $\mu$ m. **f**, Quantification of immunofluorescence in **e**. \* $P = 6.4 \times 10^{-4}$ , \*\* $P = 1.24 \times 10^{-5}$  (mean  $\pm$  s.e.m.,  $n = 4$ ; two-sided  $t$ -test).

To interrogate interactions within the *CPD* module in H23 cells, we measured all pairwise genetic interactions of 145 selected genes with strong 3D/2D phenotypes using CRISPR double-knockout screening<sup>39</sup> (Extended Data Fig. 5, Supplementary Tables 6, 7). Similar to their behaviour in DepMap, genetic interaction patterns of *FURIN* and *IGF1R* showed strong correlation with those of *CPD*.

Given the strong 3D/2D phenotypes of genes within the *CPD* module, we validated individual genes within the module and other strong hits using competitive growth assays and small-molecule inhibitors (Fig. 3d, Extended Data Fig. 6). We also observed that inducible knockdown of *CPD* in vitro in established H23 3D spheroids using tetracycline-inducible dCas9-KRAB<sup>17</sup> markedly reduced growth of spheroids (Extended Data Fig. 7), suggesting that targeting *CPD* can have an effect on further growth of established spheroids.

### IGF1R signalling is inhibited by *CPD* deletion

Since our data suggested *CPD* functionally interacts with *IGF1R*, we examined how *CPD* deletion affected IGF1R signalling pathways. We first measured protein levels of IGF1R and phosphorylation of its downstream effectors AKT and ERK1/2 following treatment with IGF1 (Fig. 3e, f) in H23 cells grown in 2D. We observed significant reduction of IGF1R protein levels and AKT phosphorylation in *CPD*-deficient H23 cells compared with control cells. By contrast, phospho-ERK1/2 levels were high and unchanged, probably owing to constitutively active KRAS in H23 cells. Levels of IGF1R were also significantly reduced in *CPD*-deficient H23 spheroids (Fig. 3g, h). In addition, *CPD* deletion reduced levels of IGF1R and phospho-AKT upon IGF1 addition in H322, A549 and H358 cells

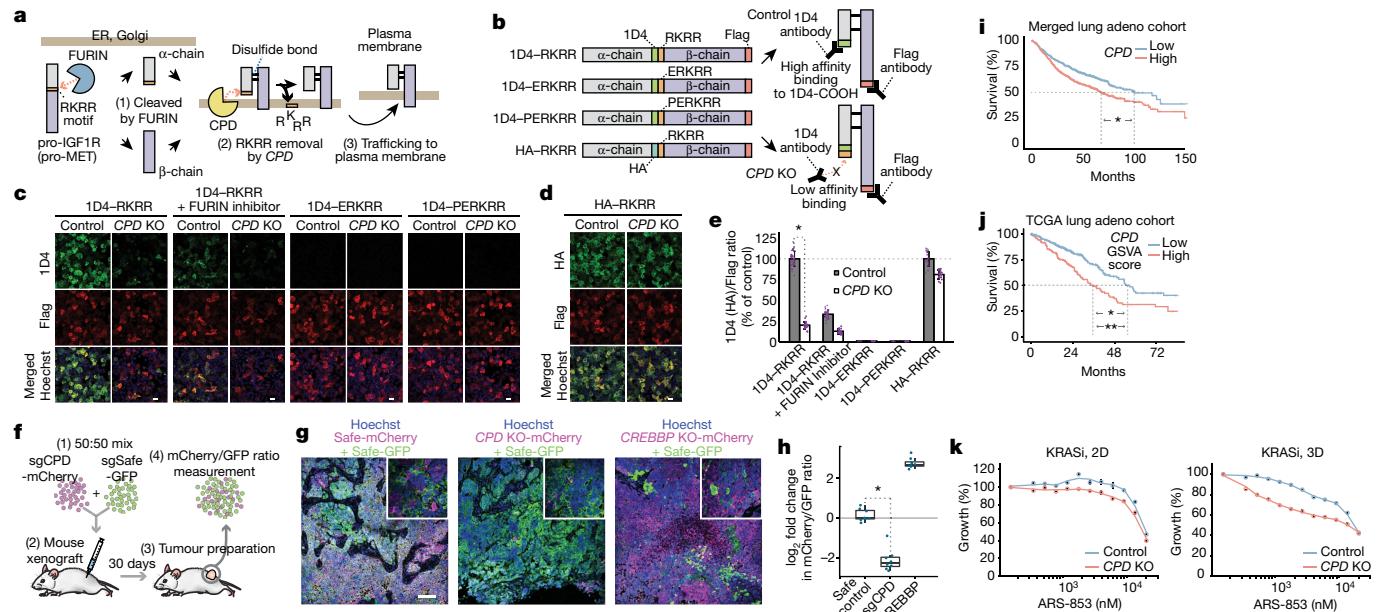
(Extended Data Fig. 8). Lastly, we found that the effect of *CPD* deletion can be rescued by treating H23 cells with excess IGF1, but not by treatment with epidermal growth factor (EGF) or hepatocyte growth factor (HGF) (Extended Data Fig. 9a, b), suggesting that much of the 3D-selective *CPD*-knockout phenotype can be attributed to its regulation of *IGF1R*.

### CPD removes the IGF1R $\alpha$ C-terminal RKRR motif

Since *CPD* is a carboxypeptidase, we considered whether IGF1R might be a substrate. IGF1R is translated as a single polypeptide (pro-IGF1R), which is cleaved by *FURIN* into  $\alpha$ - and  $\beta$ -chains<sup>2</sup> (Fig. 4a). pro-IGF1R does not end in lysine or arginine, and thus should not be a substrate for *CPD*; however, *FURIN* cleaves pro-IGF1R immediately after a central RKRR motif<sup>40</sup>, leaving these four positively charged amino acids at the C terminus of the  $\alpha$ -chain, creating a potential *CPD* substrate.

To test whether the RKRR motif is removed by *CPD*, we developed an assay to measure appearance of the 1D4 epitope<sup>41</sup>. Using the Rho1D4 antibody, which requires a free carboxylate group for binding, we could detect the presence of the 1D4 epitope specifically at the C terminus of a protein. We thus created an IGF1R reporter with a 1D4 epitope inserted immediately upstream of the RKRR motif (Fig. 4b). A Flag epitope on the  $\beta$ -chain was used to measure total protein levels.

When we transduced control H23 cells with the reporter, we observed strong 1D4 and Flag signals, suggesting that RKRR is removed and the 1D4 epitope is exposed at the C terminus of  $\alpha$ -chain (Fig. 4c, e). Deletion of *CPD* markedly reduced 1D4 staining, whereas total Flag–IGF1 remained unchanged, suggesting that *CPD* removes the RKRR motif. Consistent with these results, a *FURIN* inhibitor reduced the 1D4 signal



**Fig. 4 | CPD is a carboxypeptidase for the IGF1R  $\alpha$ -chain and loss of CPD inhibits in vivo tumour growth.** **a**, Proposed model of the CPD–IGF1R interaction. **b**, Schematic of 1D4 reporters used to test the model in **a**. **c**, Flag and 1D4 immunofluorescence from the 1D4 reporters, measured in control or CPD-knockout H23 cells grown in 2D culture, untreated or treated with FURIN inhibitor. Scale bars, 40  $\mu$ m. **d**, Immunofluorescence of HA-RKRR reporter in control or CPD-knockout H23 cells. Scale bar, 40  $\mu$ m. **e**, Ratios of 1D4 (or HA) to Flag immunofluorescence relative to the control 1D4–RKRR or to the control HA–RKRR for conditions in **c** and **d**. \* $P=1.38 \times 10^{-39}$ , two-sided  $t$ -test ( $n=19, 30, 18, 12, 20, 21, 18, 18, 18$  and 18 from left to right, mean  $\pm$  s.d.). **f**, Schematic for the competitive tumour-growth assay. **g**, Immunofluorescence of mCherry and GFP signal in day 30 tumour sections. Scale bar, 100  $\mu$ m. Original magnifications: 10 $\times$  (main images), 20 $\times$  (insets). Immunofluorescence experiments were repeated on two tumours per condition. **h**, Changes in mCherry/GFP ratios between day 0 and day 30 following tumour

transplantation. \* $P=4.3 \times 10^{-39}$ , two-sided  $t$ -test;  $n=8$  tumours per group). In the box plot, the centre line shows the median, box limits mark upper and lower quartiles, whiskers represent 1.5 $\times$  the interquartile range and points show outliers. **i**, Kaplan–Meier plot for patients with lung adenocarcinoma with high or low CPD expression. A median split was used and curve separation was assessed by two-sided log-rank test.  $n=1,106$ , \* $P<0.0001$ . **j**, Variation of the set of genes downregulated by CPD deletion in H23 spheroids were scored by gene-set variation analysis (CPD GSVA score; see Methods). Kaplan–Meier plot for survival of 479 patients with lung adenocarcinoma, divided into two groups with high or low CPD GSVA scores. Curve separation assessed by two-sided log-rank test (\* $P=9 \times 10^{-5}$ ) and Cox proportional-hazard test (\*\* $P=7.68 \times 10^{-4}$ ). **k**, CPD deletion sensitizes H23 cells to ARS-853, an inhibitor of KRAS(G12C). H23 cells in 2D or 3D culture treated with control or CPD sgRNA and indicated doses of ARS-853 for 72 h. Live cells were quantified using alamar blue ( $n=4$ , mean  $\pm$  s.e.m.).

in both control cells and cells in which *CPD* was deleted. FURIN inhibition would be expected to prevent cleavage of pro-IGF1R and exposure of the RKRR motif. Insertion of even a single amino acid between the 1D4 and the RKRR motif diminished the 1D4 signal, demonstrating the precise requirement for the removal of RKRR. An IGF1R reporter with a control haemagglutinin (HA) epitope upstream of RKRR showed a strong HA signal in both control and *CPD*-depleted cells (Fig. 4d, e). Similarly, CPD-mediated removal of the RKRR motif was observed in H322 and A549 cells (Extended Data Fig. 9c, d). Together, these data demonstrate that CPD is a carboxypeptidase that is required for IGF1R maturation. Notably, pro-MET is also cleaved by FURIN after a KRKKR motif. Although we observed toxicity upon expression of a MET 1D4 reporter in H23 cells, we were able to express the reporter in H322 cells—deletion of *CPD* prevented removal of the KRKKR motif in these cells (Extended Data Fig. 9e). Therefore, MET is also a probable substrate of CPD.

### CPD as a therapeutic target for lung cancers

Given the known role of IGF1R signalling in cancers<sup>42</sup>, we further assessed whether *CPD* deletion affects in vivo tumour growth. We performed competitive growth assays by subcutaneous injection of a mixed pool of H23 cells that expressed either a sgRNA targeting *CPD* (labelled with mCherry) or a control sgRNA (labelled with GFP) into mice (Fig. 4f). Immunofluorescence images of tumour sections showed that the tumours were dominated by GFP-expressing cells, indicating

that cells deficient in *CPD* did not readily form tumours (Fig. 4g). By contrast, deletion of *CREBBP*, a strongly positive hit in the 3D spheroids, promoted tumour growth, as reflected by dominant mCherry signal in the tumours. Flow cytometry measurement of mCherry:GFP ratios confirmed these results (Fig. 4h).

We next investigated whether expression levels of CPD were prognostic for patient survival. In a meta-analysis of expression signatures from around 18,000 human tumours with survival outcomes using PRECOG<sup>43</sup>, high expression of *CPD* is a strong indicator for poor prognosis of patients with lung adenocarcinoma (Extended Data Fig. 10a, b). A Kaplan–Meier plot generated from the merged data confirmed this result (Fig. 4i). We also showed that high expression of genes downregulated in a *CPD* knockout (identified by RNA-seq) is an indicator of poor prognosis in patients (see Methods, Fig. 4j, Extended Data Fig. 10c, Supplementary Table 8).

KRAS mutations occur in about 17% of lung cancers<sup>44</sup>, and inhibitors have been developed<sup>26–29,45</sup> for the KRAS(G12C) mutant, the most common KRAS mutant in lung adenocarcinomas<sup>31,45</sup>. Since inhibition of IGF1R can inhibit growth of KRAS-mutant lung cancer<sup>46</sup> and *CPD* was a top synthetic lethal hit with ARS-853 in our screens (Fig. 3a), we examined how deletion of *CPD* affects the response of H23 cells to ARS-853. *CPD* deletion greatly sensitized H23 cells to the drug, particularly in 3D culture (Fig. 4k). Consistent with this, expression of genes downregulated in *CPD*-knockout spheroids more strongly predict the survival of patients with lung adenocarcinoma with *KRAS* mutations than with wild-type *KRAS* (Extended Data Fig. 10d, e).

We further investigated potential synergy between ARS-853 and loss of CPD in additional *KRAS*<sup>G12C</sup>-mutant lung cancer cell lines (Extended Data Fig. 10f,g). We observed even greater synergy in H358 cells, whereas no synergy was detected in H1792 cells. Of note, H1792 cells do not have a phenotype for loss of CPD (Supplementary Table 5), and show negligible IGF1R expression (Extended Data Fig. 10h). This suggests that IGF1R expression and/or dependency and *KRAS* mutation may serve as biomarkers for combinatorial therapies targeting CPD and KRAS(G12C) in lung cancers.

## Conclusions

Here we have demonstrated a robust strategy to conduct genome-scale CRISPR screens in 3D spheroids. Phenotypes in 3D more closely match expectations for oncogenes and TSGs, and are better aligned with those in tumour xenografts. Accurate *in vitro* modelling of loss-of-function phenotypes in tumours is likely to become important for personalization of therapeutic strategies (Supplementary Discussion). For example: CREBBP inhibitors have been used to treat various cancers<sup>47</sup>; however, in certain lung cancer lines tested here, *CREBBP* knockout had a negative effect on 2D growth, but a profoundly positive effect on growth in 3D spheroids and mouse xenografts (Fig. 3d, Fig. 4g, h, Supplementary Table 5), arguing against the use of CREBBP inhibitors in these cases.

Of note, genes with differentially strong effects in 3D culture versus 2D culture are enriched for frequently occurring lung cancer mutations. This could be because these genes govern the transition to more aggressive 3D growth, a hallmark of tumorigenesis<sup>24</sup>. This probably includes genes involved in cell adhesion or genes that enable responses to ‘tumour-like’ stresses in the spheroids, such as hypoxia or cell crowding.

Ongoing efforts to investigate the roles of matrix composition<sup>48</sup>, nutrient conditions<sup>49</sup>, cancer-associated fibroblasts<sup>50</sup> and tumour-infiltrating immune cells<sup>16</sup> have enabled substantial improvements in *in vitro* and patient-derived organoid models of cancer. The ability to systematically and scalably determine which genes are required for growth and survival in response to such distinct environmental cues should facilitate both improved models for drug-target identification and a better understanding of cancer growth.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2099-x>.

- Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17–37 (2013).
- Khatib, A.-M. et al. Inhibition of proprotein convertases is associated with loss of growth and tumorigenicity of HT-29 human colon carcinoma cells: importance of insulin-like growth factor-1 (IGF-1) receptor processing in IGF-1-mediated functions. *J. Biol. Chem.* **276**, 30686–30693 (2001).
- Unger, C. et al. Modeling human carcinomas: physiologically relevant 3D models to improve anti-cancer drug development. *Adv. Drug Deliv. Rev.* **79–80**, 50–67 (2014).
- Siolas, D. & Hannon, G. J. Patient-derived tumor xenografts: transforming clinical samples into mouse models. *Cancer Res.* **73**, 5315–5319 (2013).
- Sharpless, N. E. & Depinho, R. A. The mighty mouse: genetically engineered mouse models in cancer drug development. *Nat. Rev. Drug Discov.* **5**, 741–754 (2006).
- Boj, S. F. et al. Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015).
- Debnath, J. & Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nat. Rev. Cancer* **5**, 675–688 (2005).
- Winters, I. P., Murray, C. W. & Winslow, M. M. Towards quantitative and multiplexed *in vivo* functional cancer genomics. *Nat. Rev. Genet.* **19**, 741–755 (2018).
- Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
- Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
- Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343**, 80–84 (2014).
- Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **568**, 511–516 (2019).
- Kim, J. W., Ho, W. J. & Wu, B. M. The role of the 3D environment in hypoxia-induced drug and apoptosis resistance. *Anticancer Res.* **31**, 3237–3245 (2011).
- Loessner, D. et al. Bioengineered 3D platform to explore cell–ECM interactions and drug resistance of epithelial ovarian cancer cells. *Biomaterials* **31**, 8494–8506 (2010).
- Dang, C. V. & Semenza, G. L. Oncogenic alterations of metabolism. *Trends Biochem. Sci.* **24**, 68–72 (1999).
- Neal, J. T. et al. Organoid modeling of the tumor immune microenvironment. *Cell* **175**, 1972–1988 (2018).
- Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
- Wang, T. et al. Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras. *Cell* **168**, 890–903 (2017).
- Chen, S. et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
- Pan, J. et al. Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens. *Cell Syst.* **6**, 555–568 (2018).
- Boyle, E. A., Pritchard, J. K. & Greenleaf, W. J. High-resolution mapping of cancer cell networks using co-functional interactions. *Mol. Syst. Biol.* **14**, e8594 (2018).
- Wainberg, M. et al. A genome-wide almanac of co-essential modules assigns function to uncharacterized genes. Preprint at *bioRxiv* <https://doi.org/10.1101/827071> (2019).
- Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576 (2017).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Morgens, D. W. et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).
- Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–551 (2013).
- Patricelli, M. P. et al. Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state. *Cancer Discov.* **6**, 316–329 (2016).
- Lito, P., Solomon, M., Li, L.-S., Hansen, R. & Rosen, N. Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism. *Science* **351**, 604–608 (2016).
- Janes, M. R. et al. Targeting KRAS mutant cancers with a covalent G12C-specific inhibitor. *Cell* **172**, 578–589 (2018).
- Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45** (D1), D777–D783 (2017).
- Campbell, J. D. et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).
- Pampaloni, F., Reynaud, E. G. & Stelzer, E. H. K. The third dimension bridges the gap between cell culture and live tissue. *Nat. Rev. Mol. Cell Biol.* **8**, 839–845 (2007).
- Tidymar, W. E. & Rauen, K. A. The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation. *Curr. Opin. Genet. Dev.* **19**, 230–236 (2009).
- Song, L. & Fricker, L. D. Purification and characterization of carboxypeptidase D, a novel carboxypeptidase E-like enzyme, from bovine pituitary. *J. Biol. Chem.* **270**, 25007–25013 (1995).
- Varlamov, O. & Fricker, L. D. Intracellular trafficking of metallocarboxypeptidase D in AtT-20 cells: localization to the trans-Golgi network and recycling from the cell surface. *J. Cell Sci.* **111**, 877–885 (1998).
- Alarcón, C. et al. A Kex2-related endopeptidase activity present in rat liver specifically processes the insulin proreceptor. *Biochem. J.* **301**, 257–265 (1994).
- Komada, M. et al. Proteolytic processing of the hepatocyte growth factor/scatter factor receptor by furin. *FEBS Lett.* **328**, 25–29 (1993).
- Grice, D. M. et al. Golgi calcium pump secretory pathway calcium ATPase 1 (SPCA1) is a key regulator of insulin-like growth factor receptor (IGF1R) processing in the basal-like breast cancer cell line MDA-MB-231. *J. Biol. Chem.* **285**, 37458–37466 (2010).
- Han, K. et al. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* **35**, 463–474 (2017).
- Bassi, D. E., Fu, J., Lopez de Cicco, R. & Klein-Szanto, A. J. P. Proprotein convertases: “master switches” in the regulation of tumor growth and progression. *Mol. Carcinog.* **44**, 151–161 (2005).
- Hodges, R. S., Heaton, R. J., Parker, J. M., Molday, L. & Molday, R. S. Antigen–antibody interaction. Synthetic peptides define linear antigenic determinants recognized by monoclonal antibodies directed to the cytoplasmic carboxyl terminus of rhodopsin. *J. Biol. Chem.* **263**, 11768–11775 (1988).
- King, H., Aleksic, T., Haluska, P. & Macaulay, V. M. Can we unlock the potential of IGF-1R inhibition in cancer therapy? *Cancer Treat. Rev.* **40**, 1096–1105 (2014).
- Gentles, A. J. et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.* **21**, 938–945 (2015).
- Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nat. Rev. Cancer* **11**, 761–774 (2011).
- Lindsay, C. R. & Blackhall, F. H. Direct Ras G12C inhibitors: crossing the rubicon. *Br. J. Cancer* **121**, 197–198 (2019).
- Molina-Arcas, M., Hancock, D. C., Sheridan, C., Kumar, M. S. & Downward, J. Coordinate direct input of both KRAS and IGF1 receptor to activation of PI3 kinase in KRAS-mutant lung cancer. *Cancer Discov.* **3**, 548–563 (2013).
- Lasko, L. M. et al. Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature* **550**, 128–132 (2017).
- Gjorevski, N. et al. Designer matrices for intestinal stem cell and organoid culture. *Nature* **539**, 560–564 (2016).
- Huch, M. & Koo, B.-K. Modeling mouse and human development using organoid cultures. *Development* **142**, 3113–3125 (2015).
- Thoma, C. R., Zimmermann, M., Agarkova, I., Kelm, J. M. & Krek, W. 3D cell culture systems modeling tumor growth determinants in cancer target discovery. *Adv. Drug Deliv. Rev.* **69–70**, 29–41 (2014).
- Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

### Cell lines

Ten non-small-cell lung carcinoma cell lines: NCI-H1437, NCI-H1568, NCI-1650, NCI-1975, NCI-H322, NCI-H1792, NCI-H2009, NCI-H23, NCI-H358 and A549, were purchased from the American Type Culture Collection. All cell lines were authenticated using the Human 9-Marker STR Profile test provided by IDEXX BioResearch and tested for mycoplasma contamination. Cells were cultured in RPMI1640 (Gibco) supplemented with 10% FBS (HyClone), penicillin–streptomycin (Genesee), and GlutaMAX (Gibco). These 10 cell lines were transduced with a spCas9 lentiviral vector with a blasticidin selection marker (Addgene no. 52962), and selected with blasticidin ( $10 \mu\text{g ml}^{-1}$ ). Single-cell clones of these selected cell lines were individually tested for their Cas9-cutting efficiency by lentiviral infection with pMBC306<sup>39</sup>, a self-GFP-cutting reporter that has both GFP and an sgRNA against GFP on the same backbone. Single clones with high Cas9-cutting efficiency were established and used in the CRISPR screens and other biological assays.

### Large-scale 3D spheroid cultures

To culture lung cancer cells as 3D spheroids at genome scale, we used either pre-treated ultra-low attachment plates (Corning, no. 3261) or polyhema (Sigma, no. P3932) coated tissue culture plates. Methylcellulose (0.75%; Fisher, no. M-352) was added in RPMI1640 growth medium to prevent excessive aggregation of cells in spheroid culture and to maintain even spheroid size. To determine an appropriate cell density for CRISPR screens, we tested multiple seeding densities of H23 cells ranging from 20,000 cells per  $\text{cm}^2$  to 150,000 cells per  $\text{cm}^2$ , with 500  $\mu\text{l}$  of growth medium per  $\text{cm}^2$ . H23 cells were seeded at multiple densities and their growth and death rates were monitored in an automated fluorescent microscope optimized for live-cell imaging (IncuCyte S3 or IncuCyte ZOOM, Essen Bioscience). Cell growth rates were monitored by mCherry expressed in the cell line and death rates were monitored by Sytox Green signal, which was added at 100 nM final concentration at the beginning of the experiment. Here, the number of live cells in spheroids was estimated by dividing total integrated mCherry intensities of spheroids by the average integrated mCherry intensity of single live cells measured at the initial cell-seeding phase. The number of dead cells was estimated similarly by dividing total integrated Sytox Green intensities of spheroids by the average integrated Sytox Green intensity of a single dead cell. We chose a cell density (50,000 cells per  $\text{cm}^2$ ) that showed about 30% peak cell death rate within 24 h after initial seeding. For all subsequent experiments, cells were initially seeded at 50,000 cells per  $\text{cm}^2$  density in 500  $\mu\text{l}$  of RPMI 1640 medium containing 0.75% methylcellulose. Spheroids were then split every 3–4 days. To passage cells, cancer spheroids were collected in methylcellulose media and diluted with PBS (~3 medium volumes) to reduce viscosity of the medium before centrifugation. Spheroids were then centrifuged at 800g for 15 min and medium and PBS was removed from the spheroid pellets. Accutase (Innovative Cell Technologies, no. AT104) was added to the pellets to dissociate the spheroids into single cells. We used 10 ml of accutase per 100 million cells in spheroids and incubated them for about 30 min until spheroids were fully dissociated into single cells. The single cells were then reseeded at the starting density (50,000 cells per  $\text{cm}^2$ , 500  $\mu\text{l}$  growth medium per  $\text{cm}^2$ ).

### Genome-wide and batch-retest CRISPR screens

The genome-wide CRISPR library and the batch-retest library were synthesized by Agilent and cloned as previously described<sup>25</sup>. The genome-wide CRISPR library was designed to have ~210,000 sgRNAs targeting 21,000 coding genes (10 sgRNAs per gene), with 13,500 negative control sgRNAs that are either scrambled, non-targeting sgRNAs or safe-sgRNAs targeting nonfunctional regions of human genomes. To design the batch-retest library, genes with 3D/2D phenotypes with

T-score cut-off (lower than -2.5 or higher than 3) were first selected from the H23 genome-wide screens. We also included hits obtained in the 2D and 3D screens in the presence of the KRAS inhibitor, with phenotypes having a T-score cutoff (lower than -2.5 or higher than 2.5). In addition, we included genes with known clinical drugs or druggable genes (such as kinase, phosphatase and other enzymes) and manually curated RAS-pathway genes that were hits in both 2D and 3D. The batch-retest library had 5,466 sgRNAs targeting these 911 hit genes (6 sgRNAs per gene) and 273 safe-sgRNAs. In brief, oligo pools for the libraries were synthesized (Agilent), PCR-amplified, digested with BstXI and BpuI restriction enzymes, and ligated into pMBC320 vector containing an mU6 promoter to drive sgRNA expression and a EF1a promoter to drive expression of mCherry fused to puromycin with a T2A linker. The plasmid libraries were then transfected into HEK239T cells to produce lentiviral pools, which were subsequently transduced into H23 cells and other indicated lung cancer cell lines. Cells were infected with the libraries at a multiplicity of infection of 0.3–0.5, and after 48 h were selected with puromycin ( $2 \mu\text{g ml}^{-1}$ ) for 3–5 days until the library-infected cell population was at least 90% mCherry positive (indicating presence of lentivirus). Cells were expanded for another 2–3 days and aliquots were saved as T0 stocks in liquid nitrogen. At the same time, the remaining cells were plated as 2D monolayer cultures or as 3D spheroids using the protocol described above. To maintain library complexity, the screens were performed at  $\sim 1,000 \times$  cell number coverage per sgRNA for the genome-wide screens (~200 million cells) and at  $\sim 2,000 \times$  cell number coverage for the batch-retest screens (~10 million cells). All screens were performed in biological replicates. In the genome-wide screens, we included an arm in which H23 cells were treated with ARS-853 at 5  $\mu\text{M}$  throughout the screens. Both 2D and 3D cultures were split every 3–4 days to keep cells in log growth phase throughout the screens. At day 21, cells were collected and stored in multiple cryovials (no. of cells in each cryovial for at least  $\sim 1,000 \times$  library coverage) in liquid nitrogen for further processing. Genomic DNA was extracted from the samples with Qiagen Blood Maxi Kit (Qiagen, no. 51194). sgRNA cassettes were PCR-amplified from genomic DNA using Herculase II Fusion polymerase (Agilent, no. 600679) and deep-sequencing adapters and sample barcodes were added during the PCR<sup>25</sup>. Finally, sgRNA compositions in the samples were measured with deep-sequencing on NextSeq 550 system (Illumina). Enrichments or disenrichments of sgRNAs either between T0 and end time point samples or between drug untreated and treated samples were then used to calculate growth or drug resistance phenotypes.

### Construction of CDKO library and CDKO screen

The  $145 \times 145$  CRISPR double-knockout (CDKO) library was constructed as previously described<sup>39</sup>. In brief, 145 genes that have most negative 3D/2D phenotypes were chosen for the CDKO library. The three sgRNAs that showed the strongest effects in the genome-wide screens were chosen for each gene. A total of 463 sgRNAs (435 gene-targeting sgRNAs and 28 safe sgRNAs) were PCR-amplified from pooled oligo chips (Agilent) and cloned into pMBC320 and pKHH030, which are lentiviral vectors with mU6 or hU6 promoters, respectively. hU6-sgRNA-tracrRNA cassettes were then digested from the single-knockout CRISPR library based on pKHH030 and ligated into the single-knockout CRISPR library based on pMBC320 downstream of the mU6-sgRNA-tracrRNA cassettes. This generated the  $145 \times 145$  CDKO library, which had 214,368 double-sgRNAs corresponding to 10,440 gene pairs. The CDKO screen was performed as other CRISPR screens at  $\sim 1,000 \times$  cell number coverage per sgRNA pair for 21 days in 2D monolayer H23 cells (~200 million cells). The screens were carried out in two experimental replicates starting from the same T0 population. Genomic DNA from both T0 and Day 21 samples were isolated and frequencies of double-sgRNAs were quantified by deep sequencing using a modified paired-end, single index protocol on NextSeq 550 as previously described<sup>39</sup>.

### Calculation of growth and drug resistance phenotypes

Effect sizes for sgRNAs were calculated as previously described<sup>17,39</sup>. In brief, log<sub>2</sub> fold enrichments of sgRNAs were first measured between two samples: T0 and day 21 samples for 2D and 3D phenotypes, T0 and day 30 samples for *in vivo* phenotypes, 2D day 21 and 3D day 21 samples for 3D/2D phenotypes, 2D day 21 and ARS-853 treated 2D day 21 samples for KRASi 2D phenotypes, and finally 3D day 21 and ARS-853 treated 3D day 21 samples for KRASi 3D phenotypes. The 3D/2D phenotypes were obtained by calculating enrichment of sgRNAs (read counts of sgRNAs) by comparing 2D day 21 samples with 3D day 21 samples directly. For any given phenotype, a median log<sub>2</sub> fold enrichment of all negative control sgRNAs (non-targeting and safe sgRNAs) was measured and this median value was subtracted from log<sub>2</sub> fold enrichments of all sgRNAs to account for systematic bias in screens. Lastly, log<sub>2</sub> fold enrichments of all sgRNAs were divided by the standard deviation of negative control sgRNAs to yield phenotype Z scores (pZ) of sgRNAs which we used as effect size of sgRNAs. Effect size of a gene is the median value of all sgRNAs that target the gene. We used modified t-value scores as our phenotype scores for genes, which account for both consistency and strength of all sgRNA effects for given genes.

Our phenotype scores based on t-value scores were computed as: phenotype score (T-score) =  $(U_{\text{gene}} - U_{\text{ctrl}})/\sqrt{(S_{\text{var}}/N_{\text{exp}} + S_{\text{var}}/N_{\text{ctrl}})}$ , where  $U_{\text{gene}}$  is the median effect of all sgRNAs (pZ) for a given gene,  $U_{\text{ctrl}}$  is the median effect of all negative control sgRNAs (pZ), and  $S_{\text{var}}$  is  $\text{Var}_{\text{gene}} \times (N_{\text{exp}} - 1) + \text{Var}_{\text{ctrl}} \times (N_{\text{ctrl}} - 1)$ , where  $\text{Var}_{\text{gene}}$  is the variance of sgRNA effects (pZ) for a given gene,  $N_{\text{exp}}$  is the number of sgRNAs for a given gene and  $N_{\text{ctrl}}$  is the average number of sgRNAs per gene in a given screen.

To combine data from two experimental replicates, normalized pZ scores of sgRNAs from two replicates were pooled together and gene effects and phenotype scores were calculated from the pooled sgRNAs as described above.

### Calculation of genetic interaction scores

Genetic interactions of gene pairs in the CDKO library were computed as previously described<sup>39</sup>. In brief, the single-knockout phenotype of an sgRNA was calculated from phenotype Z scores of all double sgRNAs that have that sgRNA paired with control (safe) sgRNAs. Safe control sgRNAs target regions of the genome predicted to be non-functional<sup>25</sup>. The expected double-knockout phenotype of a double-sgRNA pair was computed by summing single-knockout phenotypes of two sgRNAs in the pair. The difference between the expected double-knockout phenotype and the observed double-knockout phenotype of a given double sgRNA was then defined as the raw genetic interaction (GI) score of the double sgRNA. The raw GI of the double sgRNA was then normalized by the standard deviation of 200 double sgRNAs that have the most similar expected double-knockout phenotypes to account for systematic bias of genetic interactions along increasing phenotype strength of double-sgRNAs. These normalized genetic interactions (norm GIs) of double sgRNAs were then used to calculate genetic interactions at the level of gene pairs. Three sgRNAs were assigned for each gene in the library, which gave a total of 9 combinations (3 × 3) for the gene pair in one orientation. Since there are two possible orientations for a gene pair (for example, A–B and B–A), there are at most 18 double sgRNAs that target a gene pair. The norm GI of a gene pair is simply the median value of all double-sgRNAs against the gene pair. We used GI<sub>T</sub> score and GI<sub>M</sub> score as statistical scores to measure genetic interactions of gene pairs<sup>39</sup> in the CDKO library. In brief, the GI<sub>T</sub> score for a given gene pair was calculated on the basis of the modified t-value score and GI<sub>M</sub> score is signed log<sub>10</sub> P value measured by Mann–Whitney *U*-test. Both scores take into account the strength and consistency of norm GIs of double sgRNAs, adjusted by observed noise levels reflected in non-interacting double-sgRNA controls that have at least one safe sgRNA in each pair. Mann–Whitney *U*-test *P* values were multiple-test corrected to compute

adjusted FDRs using Benjamini–Hochberg procedure. In the 145 × 145 matrix of GI<sub>T</sub> scores, genes were hierarchically clustered with correlation distance calculated by Pearson correlation coefficients to generate the GI map. These correlation distances were also used to rank genes by their similarities to CPD in terms of their GI patterns. To combine data from two experimental replicates, norm GIs of double sgRNAs from two replicates were pooled together and norm GIs of genes and GI scores were then computed as described above.

### Annotation of cancer genes, TSGs and oncogenes

The Catalogue of Somatic Mutations in Cancer (COSMIC<sup>30</sup> v.86) was used to annotate genes as tumour suppressors or oncogenes. COSMIC is an expert-curated database of 719 somatic mutations for which roles in cancer are manually annotated by experts in the field. There are seven defined roles of the mutations in the database: oncogene, oncogene fusion, TSG, TSG fusion, fusion, oncogene–TSG and oncogene–TSG-fusion. For analysis of gene phenotypes and comparison to roles in cancer, we pooled genes in oncogene and oncogene-fusion categories and defined them as oncogenes. Genes in TSG and TSG-fusion categories were defined as TSGs.

### Analysis of lung cancer mutations

Comparisons between CRISPR phenotypes of genes and their significance as lung cancer mutations were performed using previously published data for lung cancers<sup>31</sup>. In the dataset, exome sequences and copy number profiles of 660 lung adenocarcinoma and 484 lung squamous cell carcinoma tumour–normal pairs were analysed. This generated a list of 11,249 genes that were reported to be mutated at least once in the lung cancer samples. Their mutational significances were computed with MutSig2CV<sup>51</sup> and also provided in the dataset. Sign-flipped log<sub>10</sub> MutSig2CV *q* values were then summed and displayed as cumulative sum plots along genes sorted by different screening phenotypes.

### Analysis of DepMap CRISPR datasets

The Avana dataset (v.18Q4) with CERES effects of ~18,000 genes across 517 cell lines was downloaded from the DepMap website (<https://depmap.org/portal/download/>). To measure the percentage of positive hits in the CERES cell lines, absolute CERES effects were used to sort genes in descending order in each cell line. The first 1,000 genes were selected and the percentage of genes with positive CERES effects was measured in the 1,000 genes for each cell line. Cell lines were then grouped by their tissues of origin and the percentage of positive hits in each cancer were plotted as box plots (Fig. 1a). To define 50 core essential genes, we averaged CERES effects across the 517 cell lines. Genes were then sorted by average CERES effect in ascending order and the 50 genes with the most negative or toxic average CERES effects were defined as ‘core’ essential genes. To measure correlation of genes in terms of their cancer dependencies, CERES effects were first subject to a PCA-based correction method for genome-wide screening data<sup>21</sup>. This bias correction was shown to bolster the sensitivity and specificity of detecting true co-essentiality of gene pairs. Pearson correlation coefficients of genes were measured in the matrix of batch-corrected CERES effects.

### Identification of enriched co-essential functional modules

We used generalized least squares (GLS) to map co-essential interactions across all pairs of genes in the Avana dataset (v.18Q3) while automatically accounting for relatedness between cell lines<sup>22</sup>; unlike conventional approaches to co-essentiality mapping based on Pearson correlation, this approach yields non-inflated *P* values. We applied GLS to the matrix of CERES effects corrected with the PCA-based correction method described above<sup>21</sup>. We then applied the ClusterONE clustering algorithm<sup>52</sup>, originally developed to discover protein complexes *de novo* from protein–protein interaction data, to cluster genes into ‘co-essential modules’ in an unbiased fashion, based on their co-essentiality profiles across all other genes. Specifically, we ran ClusterONE on

# Article

the gene-by-gene matrix of GLS *P* values after row-wise FDR correction, with edge weights set to one minus the FDR *q* value<sup>53</sup>. To determine which co-essential modules were enriched in the different screening phenotypes, the probability that the distribution of members in a given module in terms of their phenotypes scores was significantly different from that of all genes was measured using Mann–Whitney *U*-test. Sign-flipped log<sub>10</sub> Mann–Whitney *U*-test *P* values and median effects of members in co-essential modules were plotted in volcano plots as *y* axis and *x* axis, respectively (Fig. 3a, Extended Data Fig. 4c). The most enriched co-essential modules from different screen phenotypes were then analysed. While we used GLS to define co-essential modules, we used batch-corrected CERES effects for visualizing co-essentiality of gene pairs in all scatter plots and cluster maps (Fig. 3c, Extended Data Fig. 4e, f).

## PANTHER pathway-enrichment analysis

To determine which pathways were enriched among the top hits from the different screen phenotypes, we uploaded the top 1,000 hits from each screen phenotype into the gene ontology knowledgebase website (<http://geneontology.org/>). We then performed the PANTHER overrepresentation test with PANTHER pathways<sup>54</sup> as the annotation dataset. Significance of enriched pathways was measured with Fisher's exact test and pathways that passed 5% FDR cutoff were displayed as significantly enriched pathways for each phenotype with the indicated Log10 FDR.

## Subcutaneous transplantation and analysis of subcutaneous tumours

Ten- to twelve-week-old female NSG mice<sup>55</sup> of similar weights were used for cell transplantation experiments. To determine the number of H23-derived cell lines to inject, several dilutions of cells ( $2 \times 10^5$ ,  $1 \times 10^6$ ,  $2 \times 10^6$  and  $4 \times 10^6$ ) were injected into both flanks and both shoulders of one NSG recipient mouse per dilution ( $n = 4$  mice; 16 tumours total). After ten days, 4 out of 4 palpable tumours formed from the  $4 \times 10^6$  cell injections, compared to 0 out of 4 for  $2 \times 10^5$  cell injections, 1 out of 4 for the  $1 \times 10^6$  cell injections and 1 out of 4 for the  $2 \times 10^6$  cell injections; therefore  $4 \times 10^6$  or more cells were used for all subsequent injections. For the batch re-test CRISPR screens, H23 cells were transduced with the library as described above. After selecting the cells with puromycin,  $8 \times 10^6$  library-transduced cells in 100  $\mu$ l PBS were injected into both flanks of NSG recipient mice. ( $n = 10$  mice; 20 tumours total). Ideally, this would represent ~13,000× cell number coverage for the library, although the actual cell number coverage per sgRNA was likely much lower since a large portion of injected cells would not contribute to tumour development after subcutaneous transplantation. Four weeks after transplantation, tumours were removed and homogenized using a tissue blender (Omni International, no. TH115-PCR), which was cleaned between each sample. Ten tumours from left flanks were pooled together as one experimental replicate and the other 10 tumours from right flanks were pooled together as the second experimental replicate. Genomic DNA was then extracted from these two pools using QIAamp DNA Blood Maxi Kit (Qiagen, no. 51194) with the manufacturer's protocol. To PCR-amplify sgRNA cassettes from genomic DNA for deep sequencing, we used ~15× more genomic DNA than what we would use for samples from in vitro CRISPR screens<sup>25,39</sup>. In brief, we scaled a reaction based on ~10  $\mu$ g of genomic DNA in 100  $\mu$ l of PCR reaction for each ~300 sgRNAs in the library. This was to account for genomic DNA that came from tumour infiltrating mouse cells. Amplified PCR samples were sequenced on a NextSeq 550 as described above. For the competitive growth assays in tumours, total  $4 \times 10^6$  H23-derived cells with roughly equal numbers of mCherry (gene-targeting sgRNAs) and GFP (safe sgRNAs) expressing cells in 100  $\mu$ l PBS were injected into both flanks of four NSG recipient mice per genotype ( $n = 12$  mice total across three groups; 24 tumours total). Thirty days after transplantation, subcutaneous tumours were individually dissected, roughly chopped using dissecting scissors and further dissociated into a single-cell suspension using collagenase IV,

dispase and trypsin at 37 °C for 30 min with rotation. After digestion, samples were passed through a 40- $\mu$ m filter and maintained in PBS with 2% FBS, 2mM EDTA, and 1U ml<sup>-1</sup> DNase before analysis by fluorescence-activated cell sorting (FACS). For FACS analysis, mCherry/GFP ratio was determined at day 0 before subcutaneous injection and at day 30 from dissociated tumours. Log fold change of mCherry/GFP ratio between these two time points was calculated and normalized to the control mix (safe mCherry/safe GFP) (Fig. 3h). The Stanford Institute of Medicine Animal Care and Use Committee approved all animal studies and procedures.

## Histologic preparation and immunohistochemistry

Tumours from the in vivo competition assay were fixed with 4% formalin in PBS overnight and transferred to 70% ethanol before paraffin embedding. Paraffin-embedded tumours were sectioned into 4- $\mu$ m-thick slices, deparaffinized with xylene and ethanol and antigen-retrieved in citrate buffer. Immunohistochemical staining for GFP (Abcam, ab13970, 1:250) and mCherry (Abcam, ab167453, 1:250) was performed on these 4- $\mu$ m-thick sections. Alexa Fluor 488 secondary antibody (ThermoFisher Scientific, A-11039) and Alexa Fluor 594 secondary antibody (ThermoFisher Scientific, A-11012) were added with Hoechst to visualize GFP, mCherry and nuclei in the subsequent immunofluorescence imaging. Images were taken on an inverted epifluorescence microscope (Eclipse Ti, Nikon) using 10× and 20× objectives.

## The 1D4 reporter system

A 1D4 epitope<sup>41</sup> was placed just upstream of the RKRR motif in the IGF1R  $\alpha$ -chain and a Flag epitope was placed at the C terminus of the IGF1R  $\beta$ -chain (1D4–RKRR) (Fig. 4b). One or two additional amino acids were inserted between the 1D4 epitope and the RKRR motif in the control reporters (1D4–ERKRR, 1D4–PERKRR). An additional control reporter had an HA epitope instead of 1D4 (HA–RKRR reporter).

## Immunofluorescence imaging

For immunofluorescence imaging, cells were either fixed with 4% paraformaldehyde in PBS for 15 min at room temperature, or fixed with ice cold methanol at 4 °C for 15 min; for the CPD antibody (A305-514A-M, ThermoFisher), we used methanol fixation and used paraformaldehyde fixation for all other antibodies. Cells were washed twice with PBS and subsequently permeabilized with 0.2% Triton X-100 in PBS for 15 min at 4 °C for paraformaldehyde-fixed samples. Cells were blocked with 3% BSA in PBS for 1 h at room temperature. Primary and secondary antibodies were diluted in PBS containing 3% BSA. Cells were first incubated with the primary antibodies overnight at 4 °C. Cells were then washed three times with PBS and incubated with the secondary antibodies and Hoechst for 2 h before a triple wash in PBS. To quantify IGF1R-signalling activities in 2D monolayer cells, cells were processed in a 96-well multi-well plate and imaged either on inverted epifluorescence microscope (ImageXpress Micro, Molecular Devices) using a 10× objective or on a spinning-disk confocal microscope (Eclipse Ti, Nikon, CSU-W1, Yokogawa) using a 20× objective. More than four sites were acquired from each well and fluorescence signals were quantified across multiple image sites per condition. For the 1D4 assays, CPD staining and IGF1R staining in 3D spheroids, cells were processed in glass-bottom 24-well plates and imaged using the spinning-disk confocal microscope with a 10× or 20× objective. Primary antibodies were obtained from the following sources: IGF1R  $\alpha$ -chain and CPD antibodies from ThermoFisher (AHR0321, A305-514A-M); antibodies to MET, phospho-AKT (Ser437), phospho-ERK1/2 (Thr202/Tyr204) and Flag from Cell Signaling Technology (no. 8198, 4060, 4370 and 14793); Rho1D4 antibody from Millipore (MAB5356).

## Individual sgRNA validations using automated microscopy

H23 cell lines expressing the indicated sgRNAs were seeded either in tissue-culture treated (2D monolayers) or ultra-low-attachment (3D

spheroids) 24-well plates and loaded into an inverted epifluorescence microscope (IncuCyte S3 or IncuCyte ZOOM, Essenbioscience) compatible with live-cell imaging. For the competition assays, ~50,000 cells expressing gene-targeting sgRNA (mCherry) were mixed with ~50,000 cells expressing safe sgRNA (GFP) and seeded into a well in 24-well plates. Images were taken every 4 h for the next 72 h. mCherry/GFP ratios were then compared between 0 h and 72 h time points to track fold changes in the ratio. Fold changes in the ratios of samples were then normalized by the fold change in the ratio of safe mCherry and safe GFP mix to estimate relative 2D and 3D growth phenotypes of sgRNAs to the control. In addition, the normalized 3D fold changes were divided by the normalized 2D fold changes to estimate 3D/2D growth phenotypes of sgRNAs. For imaging colony size from H23 knockout cell lines, ~100,000 cells expressing gene-targeting sgRNAs (mCherry) were seeded into ultra-low attachment 24-well plates in the presence of 100 nM Sytox Green. Size and cell death of 3D spheroids from each knockout line was then monitored for the next 72 h. All experiments were performed in triplicate and sequences of sgRNAs used for the validation are listed in Supplementary Table 10.

### Rescue experiment with growth factors

The competitive growth assays between CPD null H23 cells and control H23 cells were performed in presence of 50 ng ml<sup>-1</sup> of IGF1 (PHG0071, ThermoFisher), EGF (E9644, Sigma-Aldrich) or HGF (294-HG-005, R&D Systems). The competitive growth assay was performed as described in the sgRNA validation experiments, but in this case, the indicated growth factor was added at the beginning of the experiment to measure its ability to rescue gene loss phenotypes.

### Drug-titration experiments

For the drug-titration experiments, ~16,000 cells were seeded into tissue-culture treated 96-well plates in RPMI 1640 growth medium (2D monolayers) or ultra-low attachment 96-well plates in RPMI 1640 growth medium with 0.75% methylcellulose. Cells were then grown for the next 72 h in presence of titrated inhibitors. At the 72 h point, 1/10th volume of alamarBlue reagent (ThermoFisher, DAL1100) was added to cells and incubated ~2 h for 2D monolayer cells and ~10 h for 3D spheroids at 37 °C. Fluorescence signals were then measured in a fluorescence plate reader (TECAN, no. 30016056; excitation at 560 nm, emission at 590 nm) to estimate relative number of live cells at different dosages of the inhibitors. Wild-type H23 cells were used in the experiments where efficacies of small molecule inhibitors were compared between 2D and 3D. To test whether CPD deletion sensitizes cells against ARS-853, H23 cells with safe sgRNA and with *CPD* sgRNA (no fluorescent marker) were used. Small inhibitors were obtained from the following sources: savolitinib from Selleckchem (no. S7674), linsitinib from VWR (no. 10189-468), FURIN inhibitor I from Sigma Aldrich (no. 344930) and ARS-853 from Cayman Chemical (no. 1629268-00-3).

### Immunoblotting

Cells were lysed in RIPA buffer containing phosphatase and protease inhibitor cocktails (Roche, no. 11697498001). Lysates were then incubated on ice for 15 min, then clarified at 16,000g, 4 °C, for 10 min. Protein was quantified using the Bradford method and lysates were made with NuPage Sample Buffer (4×). Membranes were then probed with KRAS and GAPDH antibodies (1:1,000 dilution) from ThermoFisher (no. 415700, AM4300). The following secondary antibodies were used at a 1:5,000 dilution: anti-rabbit or anti-mouse IRDye-conjugated secondary antibodies from Fisher Scientific (no. NC9401841, NC9401842, NC0110517 and NC9030091). Finally, membranes probed with the IRDye-conjugated antibodies were imaged on an infrared imaging system (Li-Cor, Odyssey CLx). Uncropped western blots are shown in Supplementary Fig. 1.

### Knocking down genes in established spheroids

To knockdown genes in established spheroids, we transduced rtTA and inducible KRAB-dCas9-T2A-mCherry<sup>17</sup> under control of a tet-on promoter into H23 cells. These cells were treated with doxycycline for two days and were sorted for mCherry signal by FACS to select cells that can reliably induce dCas9 expression upon doxycycline treatment. Doxycycline was withdrawn from the sorted cells and cells were sorted again for loss of mCherry signal to establish an inducible CRISPRi cell line that can turn off dCas9 upon doxycycline withdrawal. This cell line was transduced with CRISPRi sgRNAs against CPD and KRAS. These cells were then seeded to form spheroids for 48 h, after which doxycycline was added at 0.2 µg ml<sup>-1</sup> concentration to induce knockdown target genes in the established spheroids. Growth of spheroids was then monitored for the next 5 days in an automated microscope (IncuCyte S3, Essen Bioscience).

### PRECOG analysis

PRECOG analysis was performed as previously described<sup>43</sup>. In brief, lung adenocarcinoma datasets were merged by normalizing CPD expression within each cohort so that its mean and s.d. were 1 across stage 1 patients. The merged set of 1,321 patients was split into high versus low CPD on the basis of the median expression of CPD across the entire dataset. Kaplan–Meier analysis was used to assess association with overall survival, with *P* value calculated by log-rank test. PRECOG Meta-Z scores for genes in the CPD module across different cancer types were obtained from the PRECOG website (<https://precog.stanford.edu/>).

### RNA-seq experiment and analysis

H23 cells expressing control (safe) sgRNA or *CPD* sgRNA were cultured as 2D monolayers or 3D spheroids in 100-mm tissue culture plates. RNA was extracted with TRIzol (ThermoFisher, 15596026) and processed with a RNA-seq library preparation kit (Illumina, RS-122-2101) to produce libraries for deep sequencing on NextSeq 550. Library preparation and sequencing were performed according to the manufacturer's protocol. Sequencing reads were mapped to the combined indices of cDNAs and non-coding RNA transcripts from GRCh38 genome reference using Kallisto<sup>56</sup>. Differentially regulated genes between the two different conditions were analysed using Sleuth<sup>57</sup>. Here, Sleuth computed FDRs for differential regulation of transcripts. If a gene has multiple transcripts, the best FDR value from all the transcripts was chosen to represent the FDR for differential regulation of the gene. We then defined a set of differentially regulated genes using 5% FDR cut-off. Genes significantly downregulated in *CPD*-deleted 3D spheroids compared to control 3D spheroids were further analysed for their predictive power for survival rates of patients with lung cancer.

### TCGA outcome analysis in downregulated genes upon CPD deletion

TCGA lung adenocarcinoma gene expression data (FPKM-UQ) and outcome and clinical data were downloaded from gdc.cancer.gov. We used GSVA<sup>58</sup> to study the association with outcome of the genes associated with the *CPD*-deleted phenotype. RNA-seq counts were normalized using Limma voom<sup>59</sup>. Outcome data was censored to seven years. Kaplan–Meier plots were generated using the survminer package from Bioconductor. High versus low *CPD* GSVA score was defined using the 1/3 upper versus 1/3 lower quantiles. Log-rank test *P* values are reported. Additionally, we built a Cox proportional-hazard model to account for key clinical covariates including age, stage, gender and TP53 and KRAS status. We also studied the interaction between *CPD* GSVA score and KRAS mutation status using a Cox proportional-hazard model with the same covariates.

# Article

## Statistical analysis

The statistical significance used to compare the averages of two different experimental groups in all box plots and bar graphs in this study was computed using unpaired, two-tailed Student's *t*-test. No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Sequencing data from all CRISPR screens and RNA-seq experiments are available under BioProject accession number PRJNA535417. All other data are available from the corresponding author upon reasonable request.

## Code availability

All screening data were analysed with custom Python scripts (v.2.7) that are available at [https://github.com/biohank/CRISPR\\_screen\\_analysis](https://github.com/biohank/CRISPR_screen_analysis). Custom Matlab scripts (v.2015b) were used to quantify signals from all immunofluorescence images and to analyse FACS data: these scripts can be requested from K.H.

52. Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* **9**, 471–472 (2012).
53. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
54. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* **41**, D377–D386 (2013).
55. Shultz, L. D. et al. Human lymphoid and myeloid cell development in NOD/LtSz-scid /IL2R $\gamma^{null}$  mice engrafted with mobilized human hemopoietic stem cells. *J. Immunol.* **174**, 6477–6489 (2005).
56. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
57. Pimentel, H. J., Bray, N., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **14**, 687–690 (2016).
58. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
59. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

**Acknowledgements** We thank J. Sage and members of the Bassik laboratory for discussions and critical reading of the manuscript. This work was supported by the NIH Director's New Innovator Award Program (1DP2HD084069), NIH/NCI 1U01CA199261 to M.C.B., P.J. and A.S.-C., and NIH/NCI 1U01CA217851 to C.J.K., C.C. and M.C.B. This work was also partly supported by a Stanford SPARK Translational Research Grant. We also thank D. Mochly-Rosen and K. Grimes at Stanford University for their support of this work. K.H. is supported by the Walter V. and Idun Berry award. S.E.P. is supported by National Science Foundation.

**Author contributions** K.H. and M.C.B. conceived and designed the study. K.H. designed the scalable 3D culture system for genome-wide CRISPR screens. K.H., A.L. and K.S. performed the CRISPR screens. K.H. and S.E.P. performed *in vivo* mouse experiments including the *in vivo* CRISPR screens and the competition assay. K.H., S.E.P. and M.Y. performed immunohistochemistry on tumour sections from the competition assay. G.R.A., M.D., M.O. and R.A.K. performed the western blots for IGF1R in various cancer cell lines. K.H. designed the 1D4 reporter and performed the immunocytochemistry. A.L. performed the western blot for the 1D4 assay. K.H. performed and analysed the RNA-seq. J.A.S. analysed the TCGA data for patient survival outcomes associated with genes downstream of CPD. M.W. defined co-essential modules from the DepMap CRISPR dataset. K.H. and K.K. generated Cas9-expressing non-small-cell lung carcinoma lines. A.J.G. performed PRECOG analysis. K.H., K.S. and K.L. validated the hits from the CRISPR screens. K.H. performed the drug validations. K.H. wrote the Python and Matlab scripts to analyse the screening data and to quantify the immunofluorescence signals in microscope images. K.H. and M.C.B. wrote the manuscript. All authors discussed the results and the manuscript.

**Competing interests** The authors, through the Office of Technology Licensing at Stanford University, have filed patent applications on methods for inhibiting tumour growth by inhibiting CPD as well as systems and methods for identifying CPD inhibitors and other tumour suppressors and/or oncogenes.

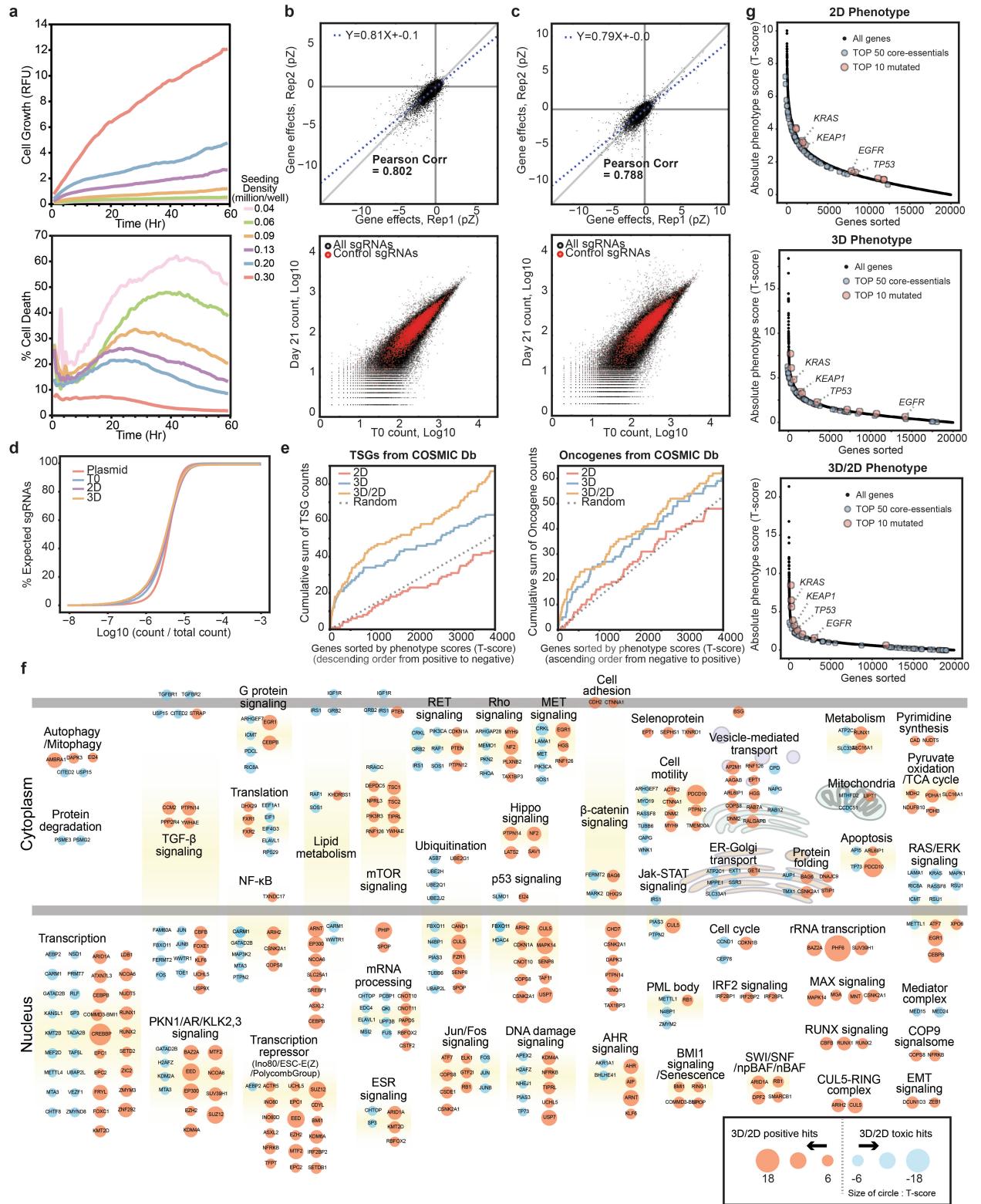
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2099-x>.

**Correspondence and requests for materials** should be addressed to K.H. or M.C.B.

**Peer review information** *Nature* thanks Charles M. Rudin, Nicola Valeri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

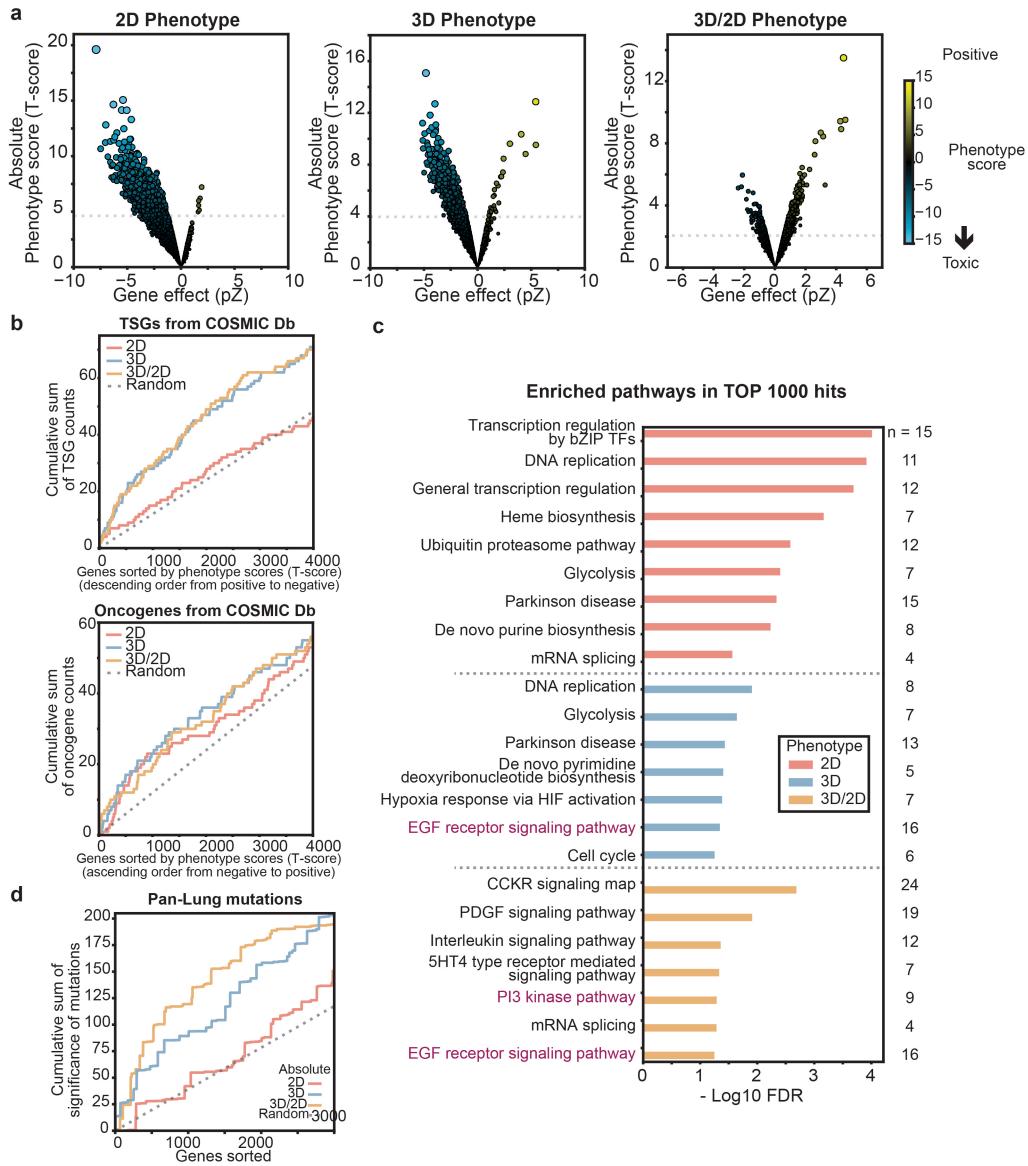


**Extended Data Fig. 1** | See next page for caption.

# Article

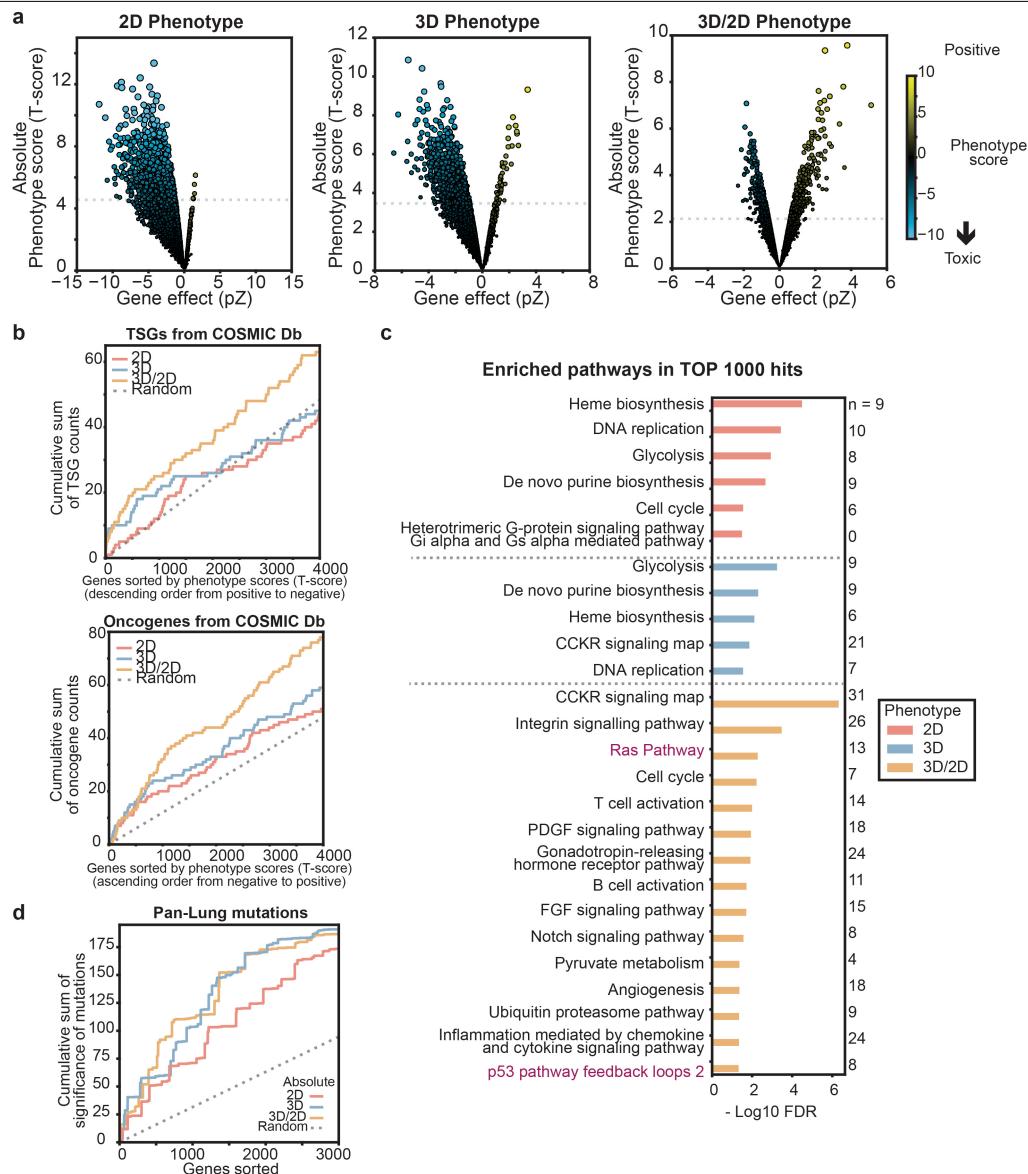
**Extended Data Fig. 1 | High quality and reproducibility of 2D and 3D genome-wide CRISPR screens and hits with differential effects in the two conditions.** **a**, H23 cells expressing mCherry were seeded at different densities in ultra-low attachment plates in the presence of 0.75% methylcellulose. Sytox Green was added at 100 nM concentration. Average mCherry signal and Sytox Green signal measured across single cells were used to estimate the total numbers of live cells and dead cells at each seeding density. Cell growth and death rates were then monitored simultaneously on a live-cell microscope for 60 h. We aimed for a cell death rate of about 30% during the initial growth phase of spheroids, and  $10^5$  cells per well ( $1.9 \text{ cm}^2$ ) was the chosen cell seeding density for our genome-wide screens in 3D spheroids. **b**, Two-dimensional growth phenotypes of 20,463 genes were highly reproducible between experimental replicates (top). Sequencing counts of 208,687 sgRNAs in a T0 sample and a day 21 sample from the 2D genome-wide screens (bottom) show that most negative-control sgRNAs (red dots) are not enriched or disenriched between T0 and day 21 (black dots). This indicates the complexity of the genome-wide library was maintained throughout the 2D screen. In the top plot, the data are fit by a linear regression line (blue dotted line). The grey line marks a 1:1 diagonal. **c**, The quality and reproducibility of the 3D screens were comparable to those of the 2D screens, suggesting that the scalable 3D spheroid culture system is on a par with traditional 2D culture methods for its performance in genome-scale CRISPR screens.  $n = 20,463$  genes (top);  $n = 208,687$  sgRNAs

(bottom). In the top plot, the data are fit by a linear regression line (blue dotted line). The grey line marks a 1:1 diagonal. **d**, Cumulative distribution of sequencing reads for sgRNAs in the genome-wide CRISPR library. Read counts were normalized by total reads for each sample and the cumulative sums of sgRNAs were plotted as relative percentages of the number of expected sgRNAs. **e**, Cumulative sums of TSG counts (left) or oncogene counts (right) are plotted against genes sorted by their 2D, 3D or 3D/2D phenotypes (T-score) from the genome-wide screens in H23 cells. TSGs are expected to have positive growth phenotypes when deleted. Therefore, genes are sorted in descending order from the most positive to the most negative phenotypes in the left plot. Oncogenes are expected to have negative or toxic growth phenotypes and genes are sorted in ascending order in the right plot. Black dotted line, randomly sorted genes. The first 4,000 genes are displayed. **f**, Summary of hits with differential 3D/2D phenotypes. Top positive (red-filled circles) and negative (blue-filled circles) hits from the differential 3D/2D phenotypes reveal many cancer-relevant genes associated with transcriptional regulation, cell motility, cell adhesion and energy metabolism. Cancer-signalling pathways such as Ras–MAPK, TGF $\beta$ , MET, Rho,  $\beta$ -catenin and Hippo signalling are highly represented. Sizes of circles are proportional to 3D/2D phenotype scores. **g**, The 10 most significant pan-lung cancer genes<sup>31</sup> and 50 top core essential genes are marked. Genes sorted by absolute phenotype (T-score) in 2D, 3D and 3D/2D (see Methods).



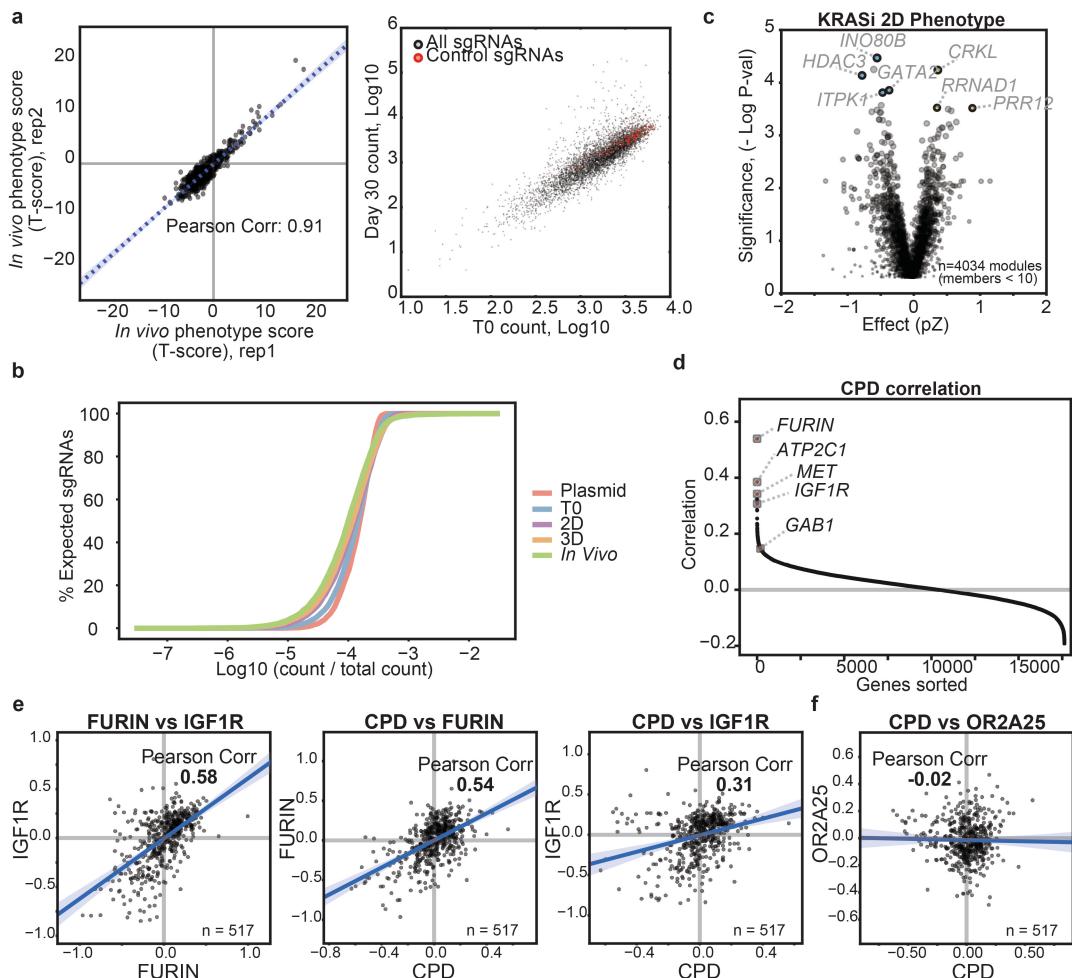
**Extended Data Fig. 2 | Genome-wide 2D and 3D CRISPR screens in H1975, a lung adenocarcinoma line with *EGFR<sup>LSSR</sup>* mutation.** **a**, Distributions of 2D and 3D phenotypes are shown as volcano plots. They axis represents absolute T-score for each gene, and the x axis represents effect size of each gene. Size of dots represents absolute T-score of genes. **b**, Prediction of TSGs or oncogenes with 2D, 3D, 3D/2D phenotypes in H1975 cells. Cumulative sums of TSGs counts (top panel) or oncogenes counts (bottom panel) are plotted against genes sorted by their 2D, 3D, or 3D/2D phenotypes (T-score) from the genome-wide screens in H1975 cells. These data indicate 3D or differential 3D/2D phenotypes show marked improvement for prediction of TSGs when compared to the 2D phenotypes, with marginal improvement for predicting oncogenes. In the box plots, centre lines mark median, box limits mark upper and lower quartiles,

whiskers show  $1.5 \times$  interquartile range and points indicate outliers. **c**, Enriched pathways among the top 1,000 hits from each culture condition were analysed using PANTHER overrepresentation test. Significance of enriched pathways was measured with Fisher's exact test and the Benjamini–Hochberg FDR was subsequently computed (x axis). The EGFR signalling pathway, a known driver for H1975 cells, is enriched in only 3D or 3D/2D phenotypes. Number of genes for enriched pathways are marked to the right of bars. **d**, The cumulative sum of the significance of 11,249 pan-lung cancer mutations from 1,144 patients with lung cancer as measured by MutSig2CV is displayed on the y axis, whereas the x axis shows phenotypes for genes sorted by their strength in 2D (solid red line), 3D (solid blue line) or 3D/2D (solid yellow line). Black dotted line, randomly sorted genes. Top 3,000 genes are shown.



**Extended Data Fig. 3 | Genome-wide 2D and 3D CRISPR screens in H2009, a lung adenocarcinoma line with KRAS<sup>G12A</sup> mutation.** **a**, Distributions of 2D and 3D phenotypes are shown as volcano plots. The y axis represents absolute T-score for each gene, and the x axis represents effect size of each gene. Size of dots represents absolute T-score of genes. **b**, Prediction of TSGs or oncogenes with 2D, 3D and 3D/2D phenotypes in H2009 cells. Cumulative sums of TSG counts (top) or oncogene counts (bottom) are plotted against genes sorted by their 2D, 3D or 3D/2D phenotypes (T-score) from the genome-wide screens in H2009 cells. These data indicate that 3D phenotypes, and in particular the differential 3D/2D phenotypes show improved prediction of both TSGs and oncogenes when compared with 2D phenotypes. In the box plots, centre lines mark median, box limits mark upper and lower quartiles, whiskers show

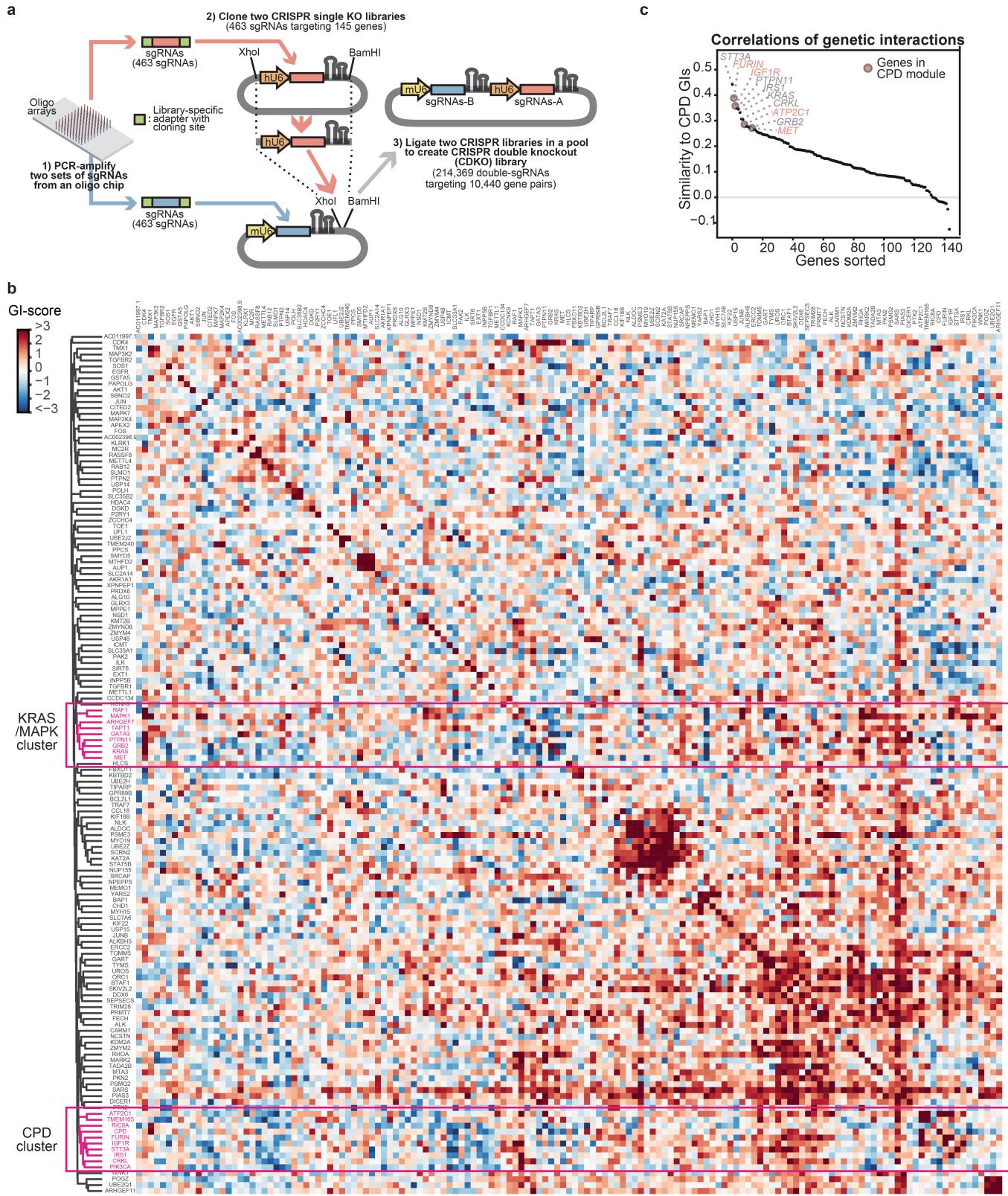
1.5× interquartile range and points indicate outliers. **c**, Enriched pathways among the top 1,000 hits from each culture condition were analysed using PANTHER overrepresentation test. Significance of enriched pathways was measured with Fisher's Exact test and the Benjamini–Hochberg FDR was subsequently computed (x axis). The Ras pathway, a known driver for H2009 cells, is enriched in 3D/2D phenotypes. Numbers of genes for enriched pathways are marked to the right of bars. **d**, The cumulative sum of the significance of 11,249 pan-lung cancer mutations from 1,144 patients with lung cancer as measured by MutSig2CV is displayed on the y axis, while the x axis shows phenotypes for genes sorted by their strength in 2D (solid red line), 3D (solid blue line) or 3D/2D (solid yellow line). Black dotted line, randomly sorted genes. Top 3,000 genes are shown.



**Extended Data Fig. 4 | High quality and reproducibility of optimized in vivo CRISPR screens and analysis of the CPD co-essential module.** **a**, A CRISPR sgRNA library targeting 911 hits with differential growth effects in 3D versus 2D (Supplementary Table 4) was introduced into H23 cells, and introduced by subcutaneous injection into NSG mice. After 30 days, tumours were collected and sgRNAs were amplified. In vivo growth phenotypes of 911 genes were highly reproducible between experimental replicates (left). Sequencing counts of T0 samples and day 30 samples from the in vivo batch-retest screens (right). In the left plot, the data are fit by a linear regression line (blue dotted line). **b**, Cumulative distribution of sequencing reads for sgRNAs in the batch-retest library in H23 cells. Read counts were normalized by total reads for each sample and the cumulative sums of sgRNAs were plotted as relative percentages of the number of expected sgRNAs. **c**, The 4,034 co-essential gene

modules based on the DepMap CRISPR dataset are plotted as volcano plots for KRASi 2D phenotype scores. The y axis shows significance of enrichments of co-essential modules as measured in  $\log P$  values from the two-sided Mann–Whitney  $U$ -test (see Methods); the x axis shows average gene effects of members in CERES modules. **d**, Genes in the CPD module are indicated among 17,634 genes sorted by their correlations to CPD. Pearson correlation coefficients between CPD and other genes are measured in batch-corrected CERES effects in the DepMap CRISPR dataset. **e**, CERES effects of CPD, FURIN and IGF1R are shown as correlation plots. CERES effects are batch-corrected before plotting<sup>21</sup>. Blue lines, regression lines. Blue shaded translucent bands, 95% confidence intervals. **f**, Lack of correlation between CPD and OR2A25, an olfactory receptor, in their CERES effects across 517 cancer lines.

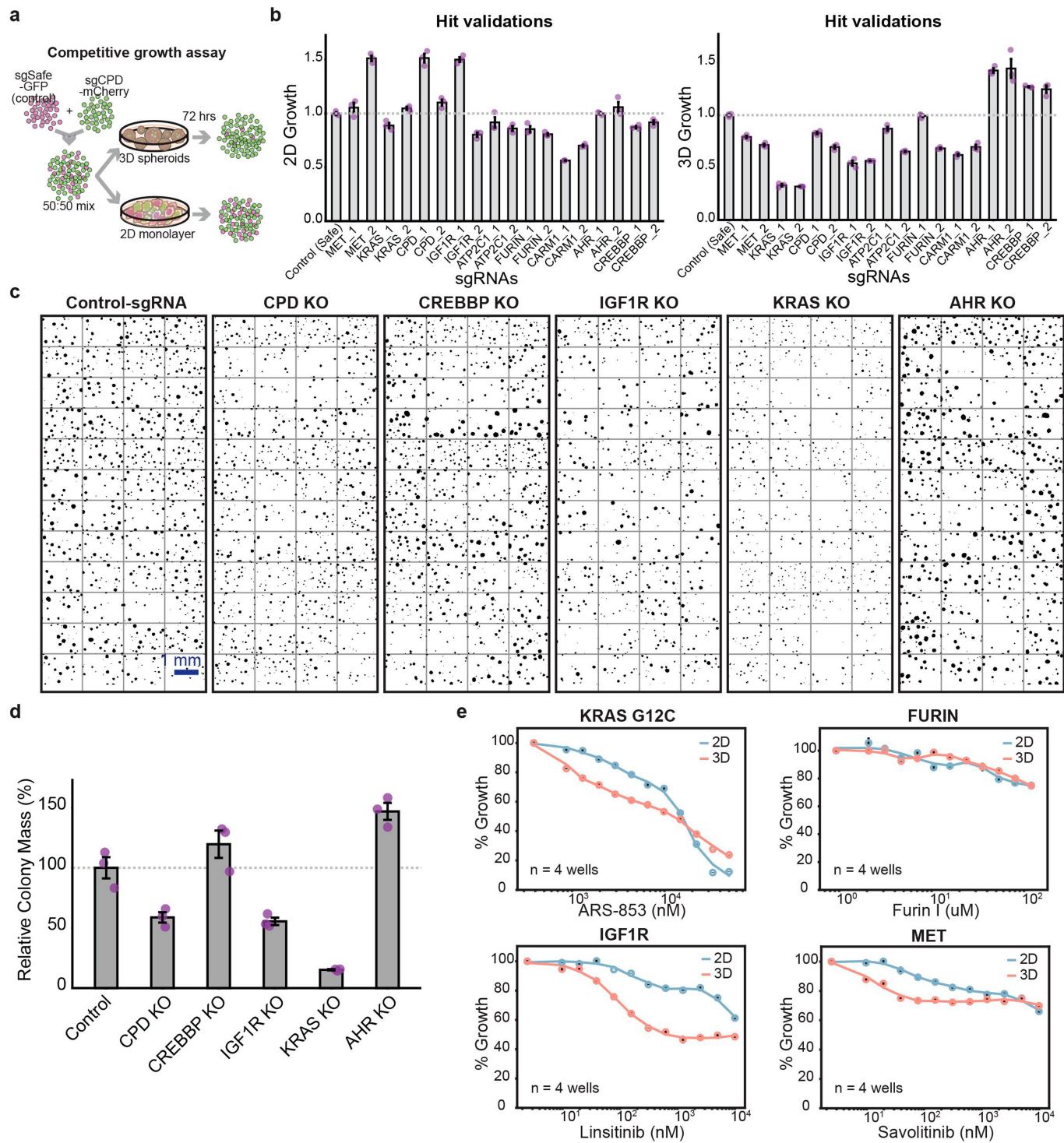
# Article



**Extended Data Fig. 5 | Analysis of *CPD* co-essential module with a  $145 \times 145$  gene genetic-interaction map.** **a**, Cloning of CDKO library. A total of 463 sgRNAs targeting 145 hits from the 3D/2D phenotypes were PCR-amplified from an oligonucleotide array. These 145 hits include members of the *CPD* co-essential module. sgRNAs were separately cloned into two lentiviral vectors with either a mU6 or a hU6 promoter to generate two CRISPR single-knockout libraries. hU6-sgRNA cassettes were then cut out from one library and ligated into the other library containing the mU6 promoter. This generated a CDKO library with all possible pairwise combinations of the 463 sgRNAs (214,369 double sgRNAs). This CDKO library was used to measure genetic interactions

(GIs) of 10,440 gene pairs ( $145 \times 145$  combinations). **b**, The  $145 \times 145$  genetic-interaction map; the  $145 \times 145$  matrix of genetic-interaction scores are shown as a heat map. The 145 genes are clustered by the similarities of their genetic interactions (Pearson correlation coefficients of genetic interactions) in the map. Members of the *CPD* co-essential module form a cluster (marked with red box) in this genetic-interaction map, consistent with their correlations in the DepMap CRISPR dataset. **c**, A genetic-interaction map validates the *CPD* co-essential module in H23. Correlations of genetic interactions are used to sort 145 genes on the basis of their similarities to genetic interactions of *CPD*. Genes in the *CPD* module are marked with red dots along the sorted genes.

# Article

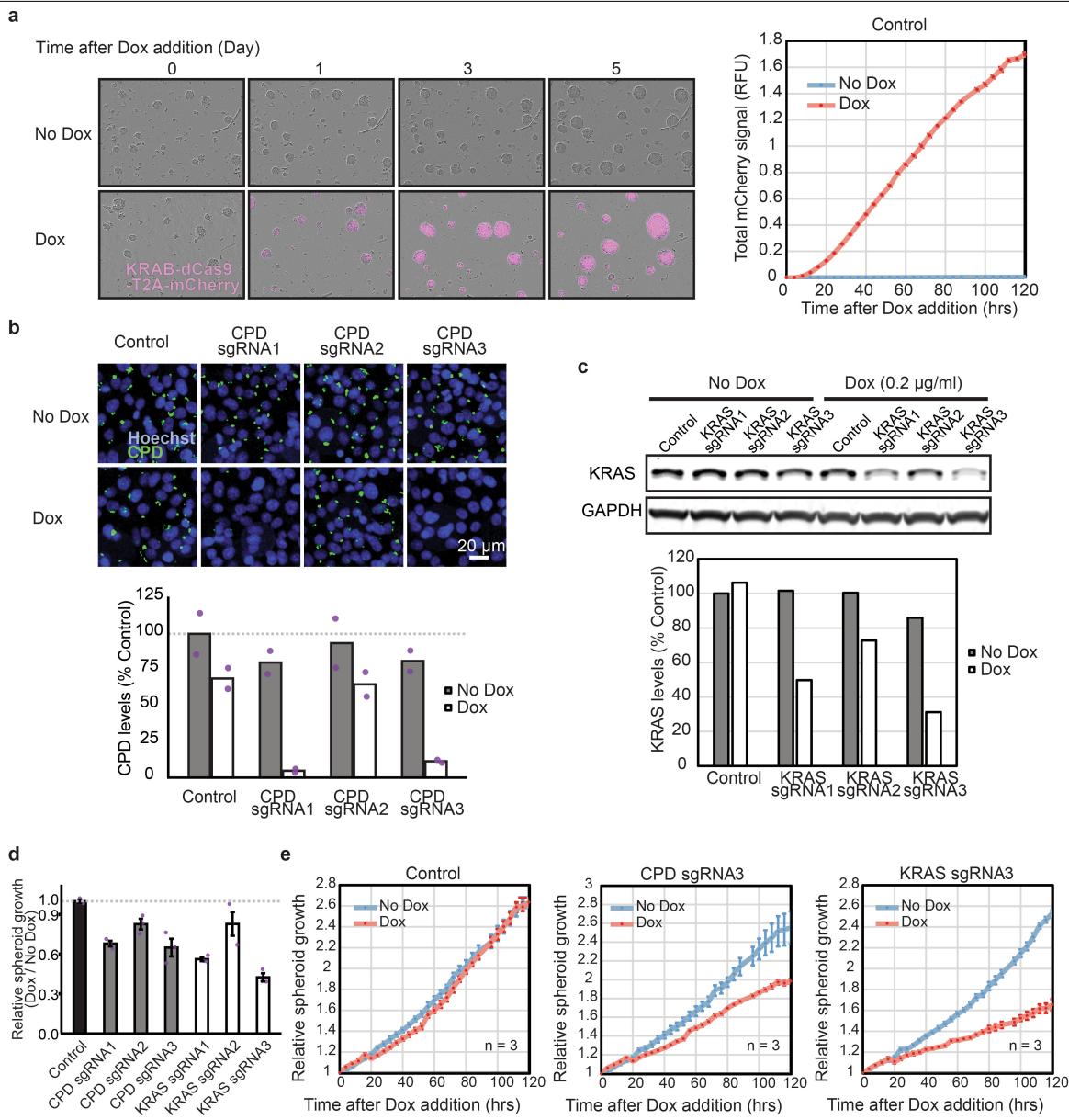


Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | Validation of individual sgRNAs targeting top hits with differential 3D/2D growth effects.** **a**, A schematic showing the competitive growth assay used to validate individual sgRNAs in 2D and 3D conditions. Cells expressing a gene-targeting sgRNA (mCherry) are mixed with cells expressing a control sgRNA (safe sgRNA, encoding GFP). Relative changes of mCherry to GFP ratios are monitored to compute growth phenotypes of gene-targeting sgRNAs. **b**, Genes within the *CPD* module and selected top hits with differential effects in 3D versus 2D growth were targeted with individual sgRNAs and subjected to competitive growth assays in both 2D and 3D culture. Relative 2D and 3D growth phenotypes of individual sgRNAs were measured by tracking changes in ratios of mCherry (gene-targeting sgRNAs) to GFP (control sgRNA) in the assays by automated fluorescence microscopy. ( $n = 3$  wells in a 24-well plate, mean  $\pm$  s.e.m.). **c**, Binary masks of H23 spheroids with the indicated gene knockouts. H23 knockout cell lines expressing sgRNAs against top hits from the 3D/2D phenotypes were seeded at equal density on ultra-low

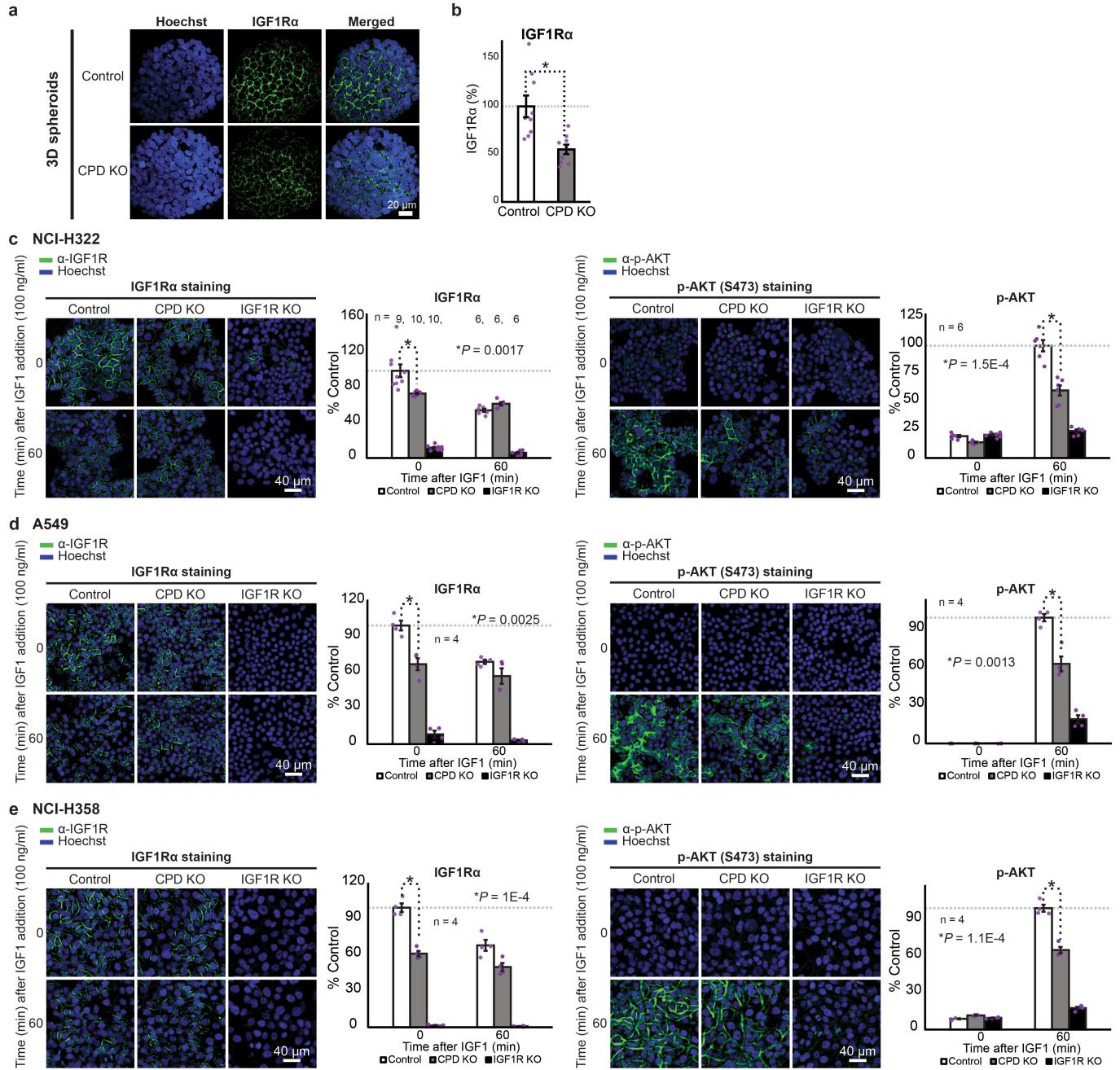
attachment plates. 3D spheroids generated from the knockout lines were imaged in a fluorescent microscope 72 h after seeding. For each knockout line, 48 images were taken from three wells in a 24-well plate using a  $10\times$  objective. Binary masks were then generated from mCherry signals of 3D spheroids. Forty-eight images were then stitched together to be shown as one large image for each knockout. **d**, Relative colony masses of H23 spheroids with gene knockouts are quantified and displayed in bar graphs ( $n = 3$  wells in a 24-well plate, mean  $\pm$  s.e.m.). **e**, Genes in the *CPD* module and *KRAS* were targeted with corresponding small-molecule inhibitors. Cells were seeded in 96-well plates in 2D (blue line) and 3D (red line) conditions, and grown in the presence of titrating doses of inhibitors for 72 h. Live cells were quantified with alamar blue assays. Relative growth of treated cells compared with the untreated samples are plotted in the drug titration curves.  $n = 3$  wells in a 96-well plate for linsitinib and  $n = 4$  for all other drugs; mean  $\pm$  s.e.m.

# Article



**Extended Data Fig. 7 | Induced CPD knockdown in established H23 spheroids slows growth.** **a**, Doxycycline (Dox; 0.2  $\mu$ g ml $^{-1}$ ) was added to established spheroids at 48 h after initial seeding. Spheroids were expressing both mCherry and KRAB-dCas9 separated by a T2A sequence under the same doxycycline-inducible promoter. Addition of doxycycline rapidly induced KRAB-dCas9-T2A-mCherry expression in spheroids. ( $n=3$  wells in a 24-well plate, mean  $\pm$  s.e.m.). **b**, Immunofluorescence staining of CPD (green) showed that CPD sgRNAs 1 and 3 robustly reduced CPD levels in H23 cells expressing the inducible KRAB-dCas9 upon doxycycline addition. CPD sgRNA 2 was less effective. Mean intensities of CPD immunofluorescence of two biological replicates were measured in the bottom bar plot. **c**, Immunostaining of KRAS (green) by western blot showed that KRAS sgRNAs 1 and 3 robustly reduced KRAS levels in H23 cells expressing the inducible KRAB-dCas9 upon doxycycline addition. KRAS sgRNA 2 was less effective. These experiments were repeated twice to confirm the result. **d**, Relative spheroid growth, five days after doxycycline addition, comparing doxycycline-treated and untreated

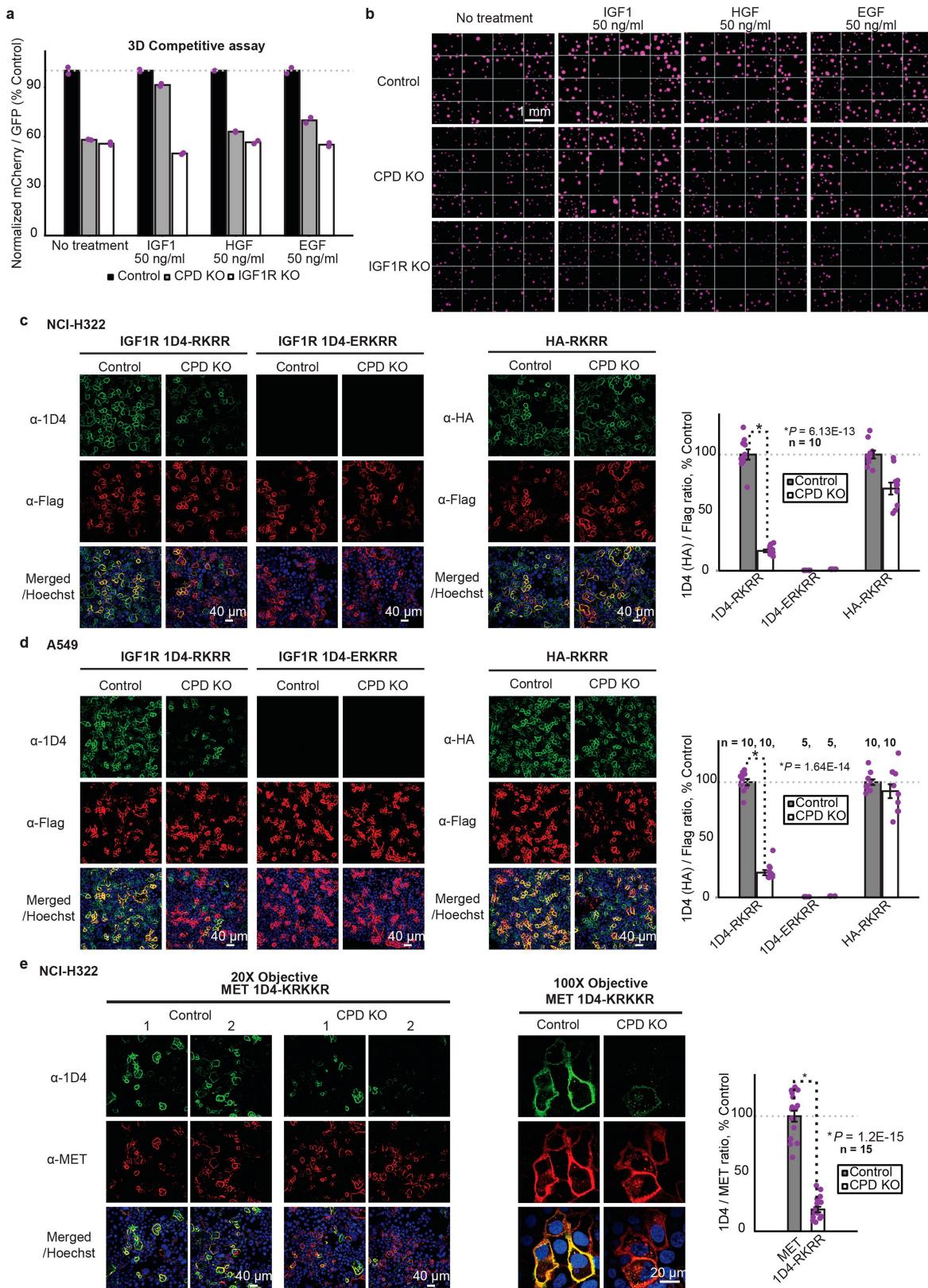
samples, measured in control cells and cells expressing CPD and KRAS sgRNA cells. H23 cells with inducible KRAB-dCas9-T2A-mCherry were first transduced with gene-targeting sgRNAs using a lentivirus that also expressed a GFP marker. Cells were seeded and allowed to form spheroids for 48 h. Doxycycline was then added and growth of spheroids in doxycycline-treated or untreated samples was monitored by GFP signal for another five days. Spheroids expressing CPD sgRNAs 1 or 3 and spheroids expressing KRAS sgRNAs 1 or 3 showed markedly reduced growth upon doxycycline addition, whereas spheroids expressing control sgRNA did not show any difference between doxycycline-treated and untreated samples ( $n=3$  wells in a 24-well plate, mean  $\pm$  s.e.m.). **e**, Growth of spheroids expressing control sgRNA, CPD sgRNA 3 or KRAS sgRNA 3 were monitored after doxycycline addition. Cells were seeded to form spheroids in the first 48 h and growth of spheroids was monitored by GFP fluorescence for the next 5 days ( $n=3$  wells in a 24-well plate, mean  $\pm$  s.e.m.).



**Extended Data Fig. 8 | CPD deletion inhibits the IGF1R pathway in H322, A549 and H358 cells.** **a**, Representative immunofluorescence images of IGF1R $\alpha$  chain (green) in control and CPD-knockout H23 spheroids. **b**, Quantification of immunofluorescence in **a**. IGF1R $\alpha$  chain intensities averaged across nine spheroids per condition.  $P=2.2\times 10^{-3}$  ( $n=9$ , two-sided  $t$ -test; mean  $\pm$  s.e.m.).

**c–e**, IGF1R and phosphorylated AKT levels were quantified from immunofluorescence images for H322 (**c**), A549 (**d**) and H358 (**e**) cells. The dotted grey line marks a 100% level ( $P$  values calculated using two-sided  $t$ -test; mean  $\pm$  s.e.m.).

# Article

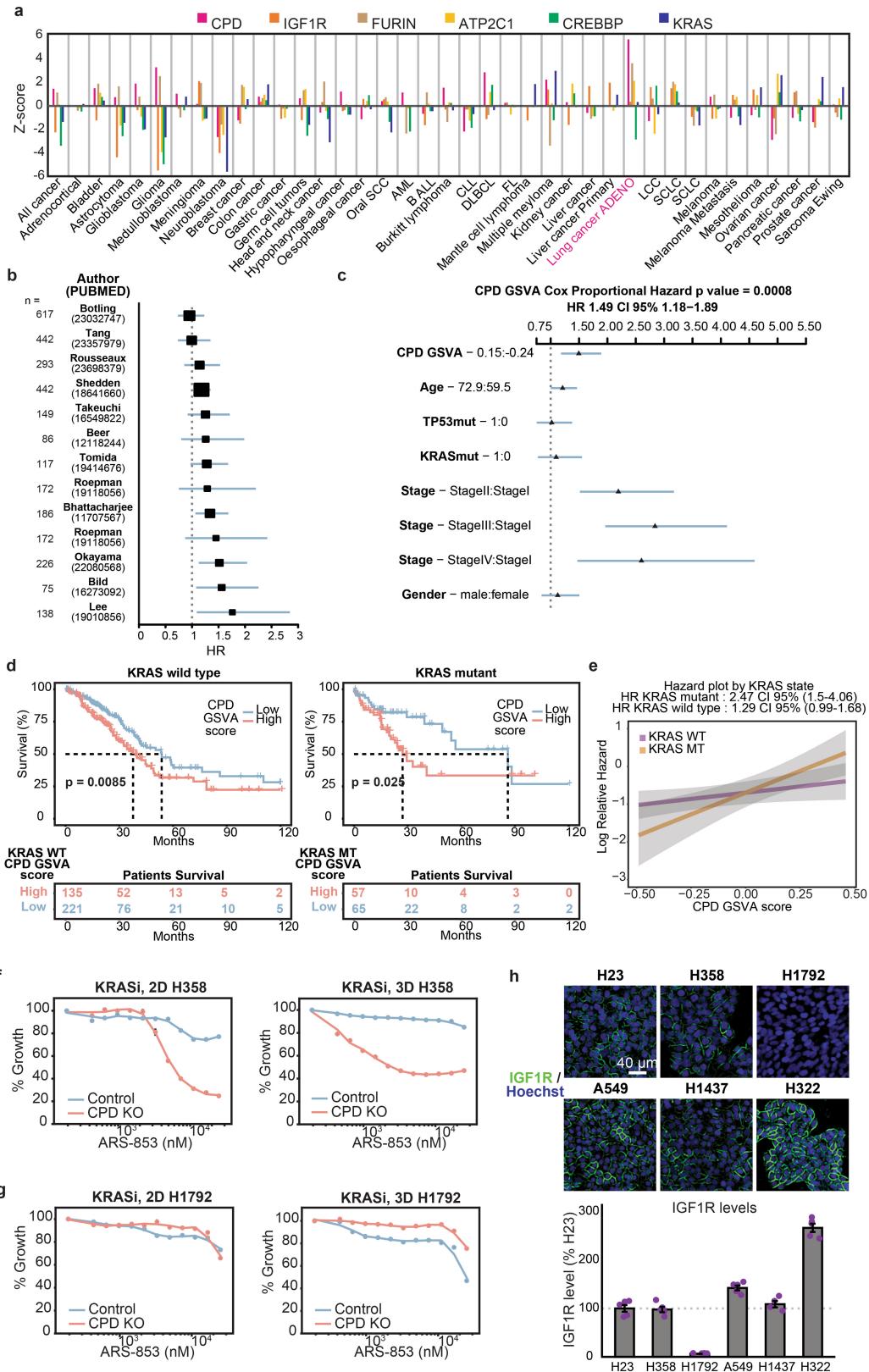


Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | CPD deletion acts through the IGF1R pathway to inhibit 3D growth in H23 cells and CPD removes the FURIN-recognition motif from the C terminus of IGF1R and MET  $\alpha$ -chain.** **a**, The growth phenotype observed upon *CPD* deletion in H23 cells is rescued by addition of excess IGF1 (50 ng ml<sup>-1</sup>) to the growth medium. A *CPD*- or *IGF1R*-targeting sgRNA with mCherry cDNA and a safe sgRNA with GFP cDNA were transduced into H23 cells separately, mixed in 50:50 ratio, and cultured in 3D spheroids for 72 h. Ratios of mCherry to GFP at 72 h, normalized to the ration at T0, were plotted in the bar graphs. Deletion of either *CPD* or *IGF1R* reduced 3D growth of spheroids, as reflected in the reduced mCherry-to-GFP ratios compared with control. Treating cells with excess IGF1 ligand (50 ng ml<sup>-1</sup>) rescued *CPD*-deletion phenotypes, whereas addition of EGF or HGF did not. This suggests that partial inhibition of the IGF1R pathway by *CPD* deletion can be compensated by over-activation of the pathway with the excess IGF1 ligand. IGF1 could not rescue the *IGF1R* deletion phenotype ( $n=2$  wells in a 24 well plate; mean  $\pm$  s.e.m.). **b**, Control, *CPD*-knockout and *IGF1R*-knockout spheroids

were treated with the indicated growth factors. Sixteen images of mCherry fluorescence in spheroids expressing a gene-targeting sgRNA vector with mCherry marker were stitched together to create the images shown. **c, d**, IGF1R-1D4 reporters (see Fig. 4b) showed that removal of the FURIN-recognition site RKRR from the C terminus of IGF1R  $\alpha$ -chain after FURIN cleavage is severely impaired by *CPD* deletion in H322 (**c**) and A549 (**d**) cells. *P*values calculated using two-sided *t*-test; mean  $\pm$  s.e.m. **e**, A MET-1D4-KRKRR reporter (with 1D4 epitope inserted upstream of the FURIN-recognition site KRKKR in MET, as with IGF1R in Fig. 4b) showed that removal of KRKKR from the C terminus of MET  $\alpha$ -chain is severely impaired by *CPD* deletion in H322 cells. Total *MET*-reporter levels were measured using an antibody against MET and ratios of 1D4 to MET signal were used to assess the degree of the KRKKR processing in control and *CPD*-null background. Error bars show s.e.m. of biological replicates in a 96-well plate. *P*values calculated using two-sided *t*-test; mean  $\pm$  s.e.m.

# Article



Extended Data Fig. 10 | See next page for caption.

**Extended Data Fig. 10 | Targeting *CPD* may have therapeutic effects in patients with lung cancer.** **a**, Meta-Z-scores of genes in the *CPD* module across different cancer types, from PRECOG analysis<sup>43</sup>. Positive Z-score predicts that high expression of a given gene is associated with poor prognosis of disease. Pink bars show that high *CPD* expression predicts poor prognosis of lung adenocarcinoma (ADENO) ( $Z\text{score} = 5.59$ , PRECOG meta-FDR =  $3.23 \times 10^{-6}$ ). **b**, Forest plot showing hazard ratios (HR) of *CPD* measured from different datasets (authors and PubMed IDs for the datasets are indicated on the y axis). The HR is the increase in risk of death for each unit increase in expression of *CPD* (see Methods). Blue error bars indicate 95% confidence intervals. Number of patient samples used for each study is listed to the left of the plot. **c**, Forest plot showing the hazard ratios from an adjusted two-sided Cox proportional-hazard model, using the *CPD* GSVA score as a continuous variable adjusted by age, *TP53*, *KRAS*, stage and gender. **d**, Kaplan–Meier plots for patients with lung cancer with wild-type (left) or mutant (right) *KRAS*. Variation of a gene set downregulated by *CPD* deletion in H23 spheroids was first scored by GSVA (*CPD* GSVA score) in patients with lung cancer. Differences in survival among patients with lung cancer with high versus low *CPD* GSVA score are illustrated in Kaplan–Meier plots. High *CPD* GSVA scores are significantly associated with

poor prognosis in both wild-type and mutant *KRAS* patient groups. However, the separation between high and low *CPD* GSVA groups is larger among *KRAS*-mutant patients than wild-type patients, suggesting an interaction between *CPD* and *KRAS* mutations in patients with lung cancer. *P* values calculated using a two-sided log-rank test. **e**, Hazard plots illustrating the two-sided Cox proportional log relative hazard by expression levels of *CPD* in *KRAS*-mutant versus *KRAS* wild-type samples. Grey shading corresponds to 95% confidence intervals. **f**, *CPD* deletion sensitizes H358 cells to ARS-853, a *KRAS* inhibitor. H358 cells with control safe sgRNA (blue line) or *CPD* sgRNA (red line) were treated with increasing doses of ARS-853 for 72 h in both 2D (top) and 3D (bottom). Live cells were then quantified using alamar blue assay. Relative growth of treated cells compared with the untreated cells is plotted against ARS-853 concentration.  $n = 4$  wells in a 96-well plate, mean  $\pm$  s.e.m. **g**, *CPD* deletion does not show synergy with ARS-853 in H1792 cells. Similar plots as in **f** were generated for H1792 cells ( $n = 4$  wells in a 96-well plate, mean  $\pm$  s.e.m.). **h**, IGF1R was quantified from immunofluorescence images of IGF1R staining across six lung cancer cell lines. H1792 cells show very low IGF1R expression compared with the other five cell lines.  $n = 4$  for H1437,  $n = 5$  for all other cell lines, mean  $\pm$  s.e.m.

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

## Software and code

Policy information about [availability of computer code](#)

### Data collection

NextSeq Control Software (v2.2.0, Illumina) was used for deep-sequencing of CRISPR screening (Fig. 1,2) and RNA-seq samples (Figure 4). Incucyte ZOOM/S3 control softwares (v2018a, ESSEN) were used to acquire microscopy images for the competition assay (Figure 3). MetaXpress (v1.7, Molecular Devices) was used to acquire microscopy images for the IF experiments on the IGF1R signaling pathways (Figure 3). NISElements (v4.4, Nikon) was used to acquire confocal IF microscopy images (Figure 4). BD CSampler controller (v227, BD Biosciences) was used to acquire FACS data for the in vivo competition assay (Figure 4).

### Data analysis

All screening data were analyzed with custom Python scripts (v2.7). They are available at [https://github.com/biohank/CRISPR\\_screen\\_analysis](https://github.com/biohank/CRISPR_screen_analysis). Custom Matlab scripts (v2015b) were used to quantify signals from all IF images and to analyze FACS data and these scripts will be available upon request to K.H.. RNA-seq data was mapped using Kallisto (Bray et al., 2016) and differentially regulated genes were analyzed using Sleuth (Pimentel et al., 2017). PRECOG analysis was performed with custom python scripts previously described in Gentles et al., 2015; request for the scripts can be made to A.J.G.. Custom R scripts (v3.5.3) were used to study association between expression scores of CPD-downstream targets and clinical outcome of TCGA lung adenocarcinoma and they will be available upon request to J.A.S.. Incucyte ZOOM or S3 software (v2018a, ESSEN) was used for automated microscopy analysis. Significance of lung cancer mutations were analyzed using MutSig2CV (Lawrence et al., 2013) available at <https://software.broadinstitute.org/cancer/cga/mutsig>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data from all CRISPR screens and RNA-seq experiments are available under BioProject accession number PRJNA535417.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](http://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

10- to 12-week old, female NSG mice of similar weights were used for cell transplantation experiments. To determine the number of cells to inject for H23-derived cell lines, several dilutions of cells ( $2 \times 10^5$ ,  $1 \times 10^6$ ,  $2 \times 10^6$ , and  $4 \times 10^6$ ) were injected into both flanks and both shoulders of one NSG recipient mouse per dilution (n=4 mice; 16 tumors total). After ten days, 4/4 palpable tumors formed from the  $4 \times 10^6$  cell injections, compared to 0/4 for  $2 \times 10^5$  cell injections, 1/4 for the  $1 \times 10^6$  cell injections, and 1/4 for the  $2 \times 10^6$  cell injections, and so  $4 \times 10^6$  or more cells were used for all subsequent injections.

For the gene-expression/patient survival analysis in Fig 4i, see "PRECOG analysis" in the Method. For the gene-expression/patient survival analysis in Fig 4j, see "RNA-seq experiment and analysis" and "TCGA outcome analysis in downregulated genes upon CPD deletion" in the Method.

For all CRISPR screens and RNA-seq experiments, experiments were performed in two replicates. We determined to perform these experiments in two replicates based on the current standard in the field, which has yielded enough power to detect meaningful/specific biological effects.

For in vitro validation assays, sample size is indicated in the figure legend for each experiment. The sample size was determined based on previous experience for each experiment to detect specific effects and it was not predetermined with any statistical methods.

Data exclusions

No data was excluded

Replication

Each in vivo presented in the paper was repeated in multiple mice (20 tumor flanks in 10 mice for in vivo CRISPR screening in Fig. 2c, d; 4 mice per genotype for the competitive growth assay in Fig. 4f-h)

All other in vitro validation assays were successfully replicated and noted in the figure legends. Detailed materials and methods for non-standard in vitro experiments are described in the Method section.

Randomization

For all the mice experiments, mice were randomly allocated to each experimental groups

Blinding

N/A

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

KRAS (ThermoFisher 415700, Lot # SJ260203, Dilution 1:1000 for Western Blot), IGF1R- $\alpha$  (ThermoFisher AHR0321, Lot # UB2717301, 1:200 for IF), CPD (ThermoFisher A305-514A-M, Lot # 1, 1:200 for IF), MET (Cell Signaling Technology 8198, Lot # 4, Dilution 1:200 for IF), phospho-AKT (Ser437) (Cell Signaling Technology 4060, Lot # 24, 1:200 for IF), phospho-ERK1/2 (Thr202/Yr204) (Cell Signaling Technology 4370, Lot # 17, Dilution 1:200 for IF), Flag (Cell Signaling Technology 14793, Lot # 5, Dilution 1:1000 for IF), GAPDH (ThermoFisher AM4300, Lot # 274128, Dilution 1:1000 for western blot), Rho1D4 (Millipore MAB5356, Lot # 3068439, Dilution 1:2000 for IF)

### Validation

CPD antibody was validated by IF (Ext. Data Fig. 7b) on CPD knockdown H23 cells.

IGF1R antibody was validated by IF (Fig. 3e,f) on IGF1R knockout H23 cells and by IF (Ext. Data Fig. 8) on IGF1R knockout NCI-H322, A549, and NCI-H358 cell lines.

Rho1D4 antibody was validated by IF on Rho1D4 reporter expressing H23 cells (Fig. 4c,e), NCI-H322 and A549 cells (Ext. Data Fig. 9c-e)

See the following "Manufacturers Statement" for the other antibodies :

K-Ras Antibody (415700) is specific for human K-Ras (K-Ras2, Ki-Ras, c-K-ras, GTPase KRas) protein (accession # NP\_004976.2, P01116), which is 100% homologous with mouse, 95% with rat, and 94% with bovine, respectively. Reactivity has been confirmed on western blots with human HeLa and WI-38 cell lysates as well as rat KNRK and mouse NIH 3T3 cell lysates, and identifies the target band at~21 kDa. The reactivity has been also confirmed with rat KNRK cells using immunoprecipitation and with HeLa cells by immunofluorescence. Based on amino acid sequence homology, reactivity with bovine is also expected. Product Citations: Kopp, F., Wagner, E. & Roidl, A. The proto-oncogene KRAS is targeted by miR-200c. *Oncotarget* 5, 185–195 (2014).

Met (D1C2) Rabbit mAb recognizes endogenous levels of total human Met protein. Monoclonal antibody is produced by immunizing animals with a synthetic peptide corresponding to residues near the carboxy terminus of human Met protein. Product Citations: Matsumoto, S. et al. GREB1 induced by Wnt signaling promotes development of hepatoblastoma by suppressing TGF $\beta$  signaling. *Nat. Commun.* 10, 3882 (2019); Willbold, R. et al. Excess hepsin proteolytic activity limits oncogenic signaling and induces ER stress and autophagy in prostate cancer cells. *Cell Death Dis.* 10, 601 (2019); Jiang, S. et al. WNT5B governs the phenotype of basal-like breast cancer by activating WNT signaling. *Cell Commun. Signal.* 17, 109 (2019).

Phospho-Akt (Ser473) (D9E) Rabbit mAb detects endogenous levels of Akt only when phosphorylated at Ser473. Species Reactivity: Human, Mouse, Rat, Hamster, Monkey, D. melanogaster, Zebrafish, Bovine. Monoclonal antibody is produced by immunizing animals with a synthetic phosphopeptide corresponding to residues around Ser473 of human Akt. Product Citations: Du, M. et al. Osthole inhibits proliferation and induces apoptosis in BV-2 microglia cells in kainic acid-induced epilepsy via modulating PI3K/Akt/mTOR signalling way. *Pharm. Biol.* 57, 238–244 (2019); Wang, Y., Li, B. & Zhang, X. *Scutellaria barbata* D. Don (SBD) protects oxygen glucose deprivation/reperfusion-induced injuries of PC12 cells by up-regulating Nrf2. *Artif. Cells Nanomed. Biotechnol.* 47, 1797–1807 (2019); Li, X., Zhang, Q. & Yang, Z. Silence of MEG3 intensifies lipopolysaccharide-stimulated damage of human lung cells through modulating miR-4262. *Artif. Cells Nanomed. Biotechnol.* 47, 2369–2378 (2019).

Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (D13.14.4E) Rabbit mAb detects endogenous levels of p44 and p42 MAP Kinase (Erk1 and Erk2) when dually phosphorylated at Thr202 and Tyr204 of Erk1 (Thr185 and Tyr187 of Erk2), and singly phosphorylated at Thr202. This antibody does not cross-react with the corresponding phosphorylated residues of either JNK/SAPK or p38 MAP kinases. Species Reactivity: Human, Mouse, Rat, Hamster, Monkey, Mink, D. melanogaster, Zebrafish, Bovine, Dog, Pig, S. cerevisiae. Monoclonal antibody is produced by immunizing animals with a synthetic phosphopeptide corresponding to residues surrounding Thr202/Tyr204 of human p44 MAP kinase. Product Citations: Li, X., Ma, A. & Liu, K. Geniposide alleviates lipopolysaccharide-caused apoptosis of murine kidney podocytes by activating Ras/Raf/MEK/ERK-mediated cell autophagy. *Artif. Cells Nanomed. Biotechnol.* 47, 1524–1532 (2019); Gao, Y. et al. Mechanical strain promotes skin fibrosis through LRG-1 induction mediated by ELK1 and ERK signalling. *Commun Biol* 2, 359 (2019); Wang, S. et al. Enhancement of Macrophage Function by the Antimicrobial Peptide Sublancin Protects Mice from Methicillin-Resistant *Staphylococcus aureus*. *J Immunol Res* 2019.

Flag DYKDDDDK Tag (D6W5B) Rabbit mAb detects exogenously expressed DYKDDDDK proteins in cells. The antibody recognizes the DYKDDDDK peptide, which is the same epitope recognized by Sigma's Anti-FLAG® antibodies, fused to either the amino-terminus or carboxy-terminus of the target protein. Monoclonal antibody is produced by immunizing animals with a synthetic DYKDDDDK peptide. Product Citations: Wang, D. et al. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat. Commun.* 10, 4284 (2019); Ji, L. et al. USP7 inhibits Wnt/ $\beta$ -catenin signaling through promoting stabilization of Axin. *Nat. Commun.* 10, 4184 (2019); Chuang, S. K., Vrla, G. D., Fröhlich, K. S. & Gitai, Z. Surface association sensitizes *Pseudomonas aeruginosa* to quorum sensing. *Nat. Commun.* 10, 4118 (2019).

## Eukaryotic cell lines

### Policy information about [cell lines](#)

#### Cell line source(s)

NCI-H1437, NCI-H1568, NCI-H1650, NCI-H1975, NCI-H322, NCI-H1792, NCI-H2009, NCI-H23, NCI-H358, and A549 are from ATCC

#### Authentication

All cell lines were authenticated by the vendor (ATCC). Authentication includes an assay to detect species specific variants of the

cytochrome C oxidase I gene (COI analysis) to rule out inter-species contamination and short tandem repeat (STR) profiling to distinguish between individual human cell lines and rule out intra-species contamination. After transducing Cas9 into the 10 cell lines, we authenticated them again by Human 9-Marker STR Profile test provided by IDEXX BioResearch.

Mycoplasma contamination	We tested the 10 cell lines for mycoplasma contamination with IDEXX BioResearch and all cell lines were negative for Mycoplasma sp.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	N/A

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	10- to 12-week old female NSG mice of similar weights
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected from the fields.
Ethics oversight	The Stanford Institute of Medicine Animal Care and Use Committee approved all animal studies and procedures

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation	Subcutaneous tumors formed from injections of human cancer cell lines were individually dissected, roughly chopped using dissecting scissors, and further dissociated into a single-cell suspension using collagenase IV, dispase, and trypsin at 37 degrees for 30 minutes with rotation. After digestion, samples were passed through a 40μM filter and maintained in PBS with 2% FBS, 2mM EDTA, and 1 U/mL DNase before FACS analysis.
Instrument	BD Accuri
Software	BD CSampler controller was used to perform FACS on dissociated tumor cells and custom matlab scripts were used to analyze and display FCS data generated in BD CSampler software.
Cell population abundance	Purify of samples varied a lot among different tumor samples, but the purity of samples were determined by either GFP or mCherry fluorescence signals in cells.
Gating strategy	The same cell lines cultured in vitro were used to set the gates for the subcutaneously injected tumor cells.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.