SUPPLEMENTARY INFORMATION

SUPPLEMENTARY METHODS

Analysis of lung cancer gene expression and survival data

Lung cancer gene expression and survival data was examined using the Lung Cancer Portal tool, at the following link: https://qbrc.swmed.edu/projects/lungcancer/. A total of thirteen datasets were analyzed in Figure 1 and Supplementary Table S1 (1-13).

Data pre-processing and normalization - To reduce the variation from different data processing approaches used by individual laboratories, we downloaded raw data and processed the data ourselves when possible. For datasets that were derived using Affymetrix platforms and provided raw data, we processed each data set using the Robust Multi-array Average (RMA) approach with log2 transformation and quantile normalization (14). For datasets without raw data, the processed data was downloaded from the original publications. Global normalization by standardizing gene expression for each sample to zero mean and unit variance was implemented to normalize data across different studies (15). We summarized the gene level expression by averaging the probeset expression for each gene. For example, using the Affymetrix platform, PCDH7 expression was averaged across the following probes: 205534_at, 205535_s_at, and 210273_at. Affymetrix probe IDs analyzed for each PCDH family member are provided in Supplementary Table S2. TCGA lung adenocarcinoma RNA-

seq data (13) was downloaded from cBioPortal (RNASeq V2 data) and the cut-off for mRNA up-regulation was the z-score threshold of 1.5 (shown in Figure 1C).

Survival analysis - Survival analysis was estimated using the Kaplan-Meier product-limit method with the log-rank test (survival, R package). When evaluating the association between gene expression and patient survival, we selected the mean cut-off of gene expression to define patient risk groups. The following sample sizes were assessed in each of the following survival studies (Figure 1B-C): Shedden *et al. Nat Medicine,* 2008 = 442 (lung adenocarcinoma); Bild *et al. Nature,* 2006 = 58 (lung adenocarcinoma only); Bhattacharjee *et al. PNAS,* 2001 = 139 (lung adenocarcinoma only); TCGA Research Network, *Nature*, 2014 = 230 (lung adenocarcinoma). For the TCGA lung adenocarcinoma RNA-seq data, the cut-off for *PCDH7* mRNA up-regulation was the z-score threshold of 1.5 (Figure 1C).

Comparative analysis - Comparative analysis was implemented to measure the associations between the selected genes with tumor versus normal samples. Gene expression levels of the selected genes between the selected groups are then shown in boxplots and the differences were reported as p-values using ANOVA analysis.

Tissue microarray and statistical analysis

A tissue microarray (TMA) with 218 patients with Non Small Cell Lung Carcinoma (NSCLC) and 10 whole sections of normal lung tissues were used for the

study. Samples were prospectively collected from patients treated at the MD Anderson Cancer Center. All tissue specimens were collected under institutional review board approval. Complete clinico-pathological information of the cohort of NSCLC patients whose tissue blocks were used for the generation of the tissue microarray is provided in Supplementary Table S3. The TMA were constructed using a manual tissue arrayer (Veridiam model ATA100, Oceanside CA). From each patient, three cores of 1 mm diameter were obtained from tumor areas in the donor blocks selected by a pathologist (3 cores per patient) and transferred to a TMA block. Paraffin sections of the TMA blocks and resection samples were cut for IHC (4 μm-thick). IHC was performed using a Leica Bond Max automated stainer (Leica Biosystems). PCDH7 expression was detected using a mouse monoclonal antibody (clone 2G6, Abcam, 1:150). Leica Bond Polymer Refine detection kit was employed for visualization of the IHC reaction. All slides were scanned in an Aperio AT2 scanner (Leica Biosystems) for pathology evaluation. Cell membrane PCDH7 expression on the tumor and normal cells were evaluated by 2 pathologists and the expression was quantified using the H-score, calculated as the staining intensity (0, no staining; 1+, weak staining; 2+, moderate staining; 3+, strong staining) times the extension of the expression for each sample (reported as percentage of positive nuclei, 0% to 100%). The final H-score (0–300) was obtained by averaging the 3 cores of each patient.

Survival analysis was performed on patients with lung adenocarcinoma. Eight patients treated with chemotherapy were excluded from the analysis. Membrane expression of PCDH7 was categorized into three groups: Low (PCDH7$\leq$ 5), Middle

(PCDH7 between 5-80) and High (PCDH7≥ 80). These cut-off numbers, 5 and 80, were selected based on the rationale that high PCDH7 expression exhibits oncogenic activity and statistical considerations that the low, intermediate, and high groups have an adequate number of patients for comparisons. Comparisons were made between all three groups. Kaplan-Meier analysis and univariate and multivariate Cox regression of recurrence-free survival of PCDH7 membrane expression and other variables were performed (Supplementary Table S4). Univariate analysis was performed first. In multivariate Cox regression analysis, a backwards model selection method was used and all variables with p-values less than 0.2 from the univariate regression were entered in the initial model, which was based on the following parameters: Tumor size, Final T stage, Final mountain stage, and PCDH7 membrane expression. Then a recursive multivariate analysis was run and the final model was based on the following parameters: tumor size, final mountain stage, and PCDH7 membrane expression. All reported p-values are two-sided. All statistical calculations were performed in SAS 9.4 for Windows (SAS Institute Inc., Cary, NC).

RNA-Seq and GSEA analysis and Quantitative Real Time PCR

Total RNA was purified from HBEC-shp53-*PCDH7* or control HBEC-shp53-GFP cells using RNeasy Mini Kit (Qiagen). RNA-seq was performed by McDermott Center Next Generation Sequencing Core at UT Southwestern Medical Center, and the results were analyzed using Gene Set Enrichment Analysis. Validation of the upregulation of transcripts was performed using quantitative real-time PCR (qRT-PCR) with iTaq™

Universal Supermixes (BIORAD). Primer sequences are listed in Supplementary Table S6.

Analysis of PCDH7 isoforms in a panel of human lung cancer cell lines

Expression of PCDH7 isoforms (A, B, C, and D) was analyzed in 138 lung cancer cell lines using publicly available data from the Cancer Cell Line Encyclopedia (CCLE) (16). We downloaded paired-end RNA-Seq samples in the Binary Alignment/Map (BAM) format via Cancer Genomics Hub (https://cghub.ucsc.edu). Cufflinks was used to quantify isoform-level relative abundance in Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) (17). We then analyzed log2 values and normalized FPKM using upper-quartile normalization (18). This normalization is performed by first removing isoforms with 0 read value in all samples, then ensuring that the top quartile values in each sample have the same median. The following NCBI Reference Sequence IDs were used for each isoform: NM_002589 (A), NM_032456 (B), NM_032457 (C), and NM_001173523 (D).

Soft agar assays

For HBEC soft agar assays, I.0 ml agar (bottom layer) was prepared in 12-well plates with 0.5 % agarose (Lonza, Catalogue no: 50002) dissolved in K25 medium (KSFM supplemented with 50 µg/mL of bovine pituitary extract, 5 ng/mL of EGF, 25% FBS and 1% Antibiotic-Antimycotic (Invitrogen). 0.5-1 x 10^4 cells were re-suspended in 1.0 ml top agar (0.37% agarose dissolved in K25 medium), and immediately seeded on top of

bottom agar in triplicate. Once the top layer solidified, 1.0 ml K25 medium was added to the top, and the medium was changed every 3 days. For H1944 and PC9 soft agar assays, agar was prepared with 0.5% agarose dissolved in R25 (RPMI-1640 supplemented 25% FBS and 1% Antibiotic-Antimycotic). 0.5-1 X 10^4 cells were re-suspended in 1.0ml top agar (0.37% agar dissolved in R25 medium), and seeded on top of bottom agar in triplicate. R25 medium was used as feeding medium and changed every three days. After 4-6 weeks, colonies were fixed in 1ml 10% MeOH/10% acetic acid for 10 min, and stained with 500ul crystal violet (0.005%) for 1-2 hours. After staining, the crystal violet was removed and the plate was washed with PBS for 4 hours, and colonies counted.

***In vivo* xenograft assays**

In vivo tumorigenicity was evaluated by injection of cells into 5- to 6-week-old female NSG mice. 5×10^6 HBECs, or 1×10^6 lung cancer cells (H1944, HCC827, or PC9) were injected subcutaneously into one or both flanks of each animal in a volume of 0.2ml PBS. Mice were monitored every 3 to 5 days for tumor formation for up to 4 months, the experimental endpoint when tissues were harvested (for HBEC cells). Tumor volume was calculated using the formula (length x width^2)/2. A total of 5 mice were analyzed for each of the following groups: HBEC-shp53 cells with or without PCDH7 (n = 10 total mice). A total of 9 mice were analyzed for each of the following groups: HBEC-shp53-$KRAS^{G12V}$ with or without PCDH7 (n = 18 mice total). For PC9 xenografts, 3 mice were analyzed for each group (PC9-Control and PC9-PCDH7, 2 tumors per animal x n= 6

mice/12 tumors total). For H1944 xenografts, 10 mice were analyzed for each group (Control sgRNA and three independent PCDH7 sgRNAs (1,3, and 4), 2 tumors per animal, n = 80 tumors total).

Cell proliferation and drug sensitivity assays

500 HBEC cells were seeded into 96 well plates (6 replicates per sample) and cultured for 5 days. The CellTiter-Glo Luminescent Cell Viability Assay (Promega) was used to measured cell proliferation according to manufacturer's instructions.

To measure toxicity in response to Trametinib or GDC-0944, 1 x 10^3 lung cancer cells were plated into 96 well plates in triplicate. Cells were treated with the MEK inhibitor Trametinib (0.5 μM) or the ERK inhibitor GDC-0944 (5 μM) for 96 h. To measure toxicity in response to FTY720, 1 x 10^3 cells were plated into 96 well plates in triplicate and treated with 0.3125, 0.625, 1.25, 2.5, 5.0, 10, 20 μM FTY720. The CellTiter-Glo Luminescent Cell Viability Assay was performed 48-96 hours after drug treatment.

PP2A assays

2 x 10^7 cells were collected and cell lysates prepared in lysis buffer provided in the PP2A DuoSet (R&D systems, DYC3309-2) and protein concentrations were measured by BCA assay (Thermo, 23228 and 23224). For each sample, 150 μg protein lysate was incubated for 3 hours at 4° C in each well of a 96-well plate coated with capture antibody. After washing, a Serine/Threonine phosphatase substrate was added into

each well and the plate was incubated for 30 minutes at 37° C. Malachite green reagent A was added to each sample and incubated for 10 minutes at room temperature. Malachite green reagent B was then added to each well, mixed thoroughly and incubated for 20 minutes at room temperature. The absorbance of 620nm was measured on a microplate reader. Experiments were performed in triplicate three independent times.

Phospho-antibody arrays

Cell lysates were harvested from 1 x 10^7 cells in Lysis Buffer (R&D systems, ARY002B). The membrane was incubated with an array buffer to block for one hour on a rocking platform shaker, then probed, and washed according to manufacturer's instructions (R&D Systems). Signals were detected with Chemi Reagent Mix provided in the kit, and the membrane was exposed to X-ray film.

SUPPLEMENTARY REFERENCES

1. Fujiwara T, Hiramatsu M, Isagawa T, Ninomiya H, Inamura K, Ishikawa S, et al. ASCL1-coexpression profiling but not single gene expression profiling defines lung adenocarcinomas of neuroendocrine nature with poor prognosis. Lung cancer 2012;75:119-25.
2. Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PloS one 2010;5:e10312.
3. Jones MH, Virtanen C, Honjoh D, Miyoshi T, Satoh Y, Okumura S, et al. Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. Lancet 2004;363:775-81.
4. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PloS one 2008;3:e1651.
5. Lu TP, Tsai MH, Lee JM, Hsu CP, Chen PC, Lin CW, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2010;19:2590-7.
6. Sanchez-Palencia A, Gomez-Morales M, Gomez-Capilla JA, Pedraza V, Boyero L, Rosell R, et al. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. International journal of cancer Journal international du cancer 2011;129:355-64.
7. Stearman RS, Dwyer-Nield L, Zerbe L, Blaine SA, Chan Z, Bunn PA, Jr., et al. Analysis of orthologous gene expression between human pulmonary adenocarcinoma and a carcinogen-induced murine model. The American journal of pathology 2005;167:1763-75.
8. Su LJ, Chang CW, Wu YC, Chen KC, Lin CJ, Liang SC, et al. Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. BMC genomics 2007;8:140.
9. Tang H, Xiao G, Behrens C, Schiller J, Allen J, Chow CW, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. Clin Cancer Res 2013;19:1577-86.
10. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proceedings of the National Academy of Sciences of the United States of America 2001;98:13790-5.
11. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 2006;439:353-7.

12. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nature medicine 2008;14:822-7.

13. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511:543-50.

14. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249-64.

15. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002;30:e15.

16. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603-7.

17. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511-5.

18. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 2010;11:94.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1. High magnification images of PCDH7 IHC staining in normal human lung and lung adenocarcinomas. 40x unprocessed images are shown. In normal lung, alveolar cells are mainly negative for PCDH7 staining, while normal bronchial epithelium show low to moderate levels of expression. Examples of low, medium, and high PCDH7 staining are shown in lung adenocarcinomas (images are the same tissue cores as shown in Figure 1D).

Supplementary Figure S2. NSCLC tissue microarray analysis of PCDH7 immunoexpression. A, Analysis of PCDH7 immunoexpression in normal human lung (n = 10) and in NSCLC cases (n = 174). Values represent H-index numbers. Membrane expression of PCDH7 was categorized into three groups: Low (PCDH7$\leq$ 5), Medium (PCDH7 between 5-80) and High (PCDH7$\geq$ 80). **B,** Kaplan-Meier analysis demonstrating recurrence-free survival of lung adenocarcinoma patients with high, intermediate, and low expression of PCDH7. Log rank p-values are shown for the high vs. low comparison (p=0.02) and for all groups (p=0.067). Recurrence free survival is the time to the first recurrence after surgery. A patient is classified as having a recurrence based on imaging and/or biopsy of the recurring site.

Supplementary Figure S3. PCDH7 induces cellular transformation of HBEC cells and induces a gene expression program that promotes cell proliferation. A, Quantification of soft agar assays of HBEC-shp53 cells overexpressing GFP (as control), PCDH7 or PCDH10. **B,** Quantification of PCDH7 isoform expression using RNAseq

data from The Cancer Cell Line Encyclopedia (CCLE) (n=138 lung cancer cell lines). **C,** Western blot analysis of PCDH7 expression in HBEC-shp53 cells and in human NSCLC cells categorized by mutation status of *KRAS* and *EGFR*. **D-E,** RNA-seq analysis of HBEC-shp53 cells vs. HBEC-shp53-*PCDH7* cells. Gene Set Enrichment Analysis (GSEA) demonstrating that expression of gene sets is significantly increased in HBEC-shp53-*PCDH7* cells. **F,** Quantitative real time PCR (qRT-PCR) validation of PCDH7-induced gene expression changes. **G,** PCDH7 enhances cell proliferation of HBEC-shp53 cells with or without EGF (**p<0.01 vs. control).

Supplementary Figure S4. All PCDH7 isoforms induce MAPK signaling in HBEC-shp53-*KRAS*^{G12V} cells. Western blot demonstrating increased levels of phospho-ERK1/2 in HBEC-shp53-*KRAS*^{G12V} cells with enforced expression of PCDH7 isoforms A, B, C, or D.

Supplementary Figure S5. PCDH7 synergizes with mutant EGFR to promote tumor formation. Quantification of tumor volume in NSG mice injected with control HCC827 cells or HCC827 cells with enforced expression of PCDH7. N= 6 tumors per group (p<0.01, Student's t-test)

Supplementary Figure S6. CRISPR/Cas9-mediated inhibition of PCDH7 sensitizes NSCLC cells to MEK and ERK inhibitors. Cell viability assay demonstrating that PCDH7 ablation in *KRAS* mutant cells enhanced sensitivity to the MEK inhibitor

Trametinib and the ERK inhibitor GDC-0944 (*p<0.05, **p<0.01 vs. control, one-way ANOVA test).

Supplementary Figure S7. PCDH7 isoforms A, C, and D bind SET in *EGFR*-mutant cells. Western blot analysis of FLAG-tagged SET immunoprecipitates in PC9 cells.

Supplementary Figure S8. SET localization studies in HBECs and NSCLC cells and requirement of SET for PCDH7-induced MAPK pathway activation. A, Cellular fractionation and western blot analysis of SET, PP2A-A, and PCDH7 in HBEC-shp53-*KRAS*^{G12V} and HBEC-shp53-*KRAS*^{G12V}-*PCDH7* cells. Cytoplasmic and nuclear SET was detected. MYC and Histone H3 serve as controls that are expressed in the nucleus. **B,** Western blot showing cellular localization of SET in *KRAS* mutant and *EGFR* mutant lung cancer cell lines. **C,** Analysis of single cell *SET* knockout clones generated by CRISPR/Cas9-mediated genome editing. Western blot showing reduced phospho-ERK1/2 in HBEC-shp53-*KRAS*^{G12V}-*PCDH7* cells with SET sgRNA but not control sgRNA.

Supplementary Figure S9. Treatment of NSCLC cells with a SET inhibitor FTY720 induces cell death. Quantification of cell viability assays demonstrating that FTY720 sensitizes NSCLC cells to death after 48h treatment.

SUPPLEMENTARY TABLES

Supplementary Table S1. Expression and survival analysis of PCDH family members in lung adenocarcinoma.

Supplementary Table S2. Affymetrix probe IDs analyzed for each PCDH family member.

Supplementary Table S3. Summary of clinical and pathologic information of the cohort of NSCLC patient tissue blocks used for TMA generation.

Supplementary Table S4. Analysis of tissue microarray correlations.

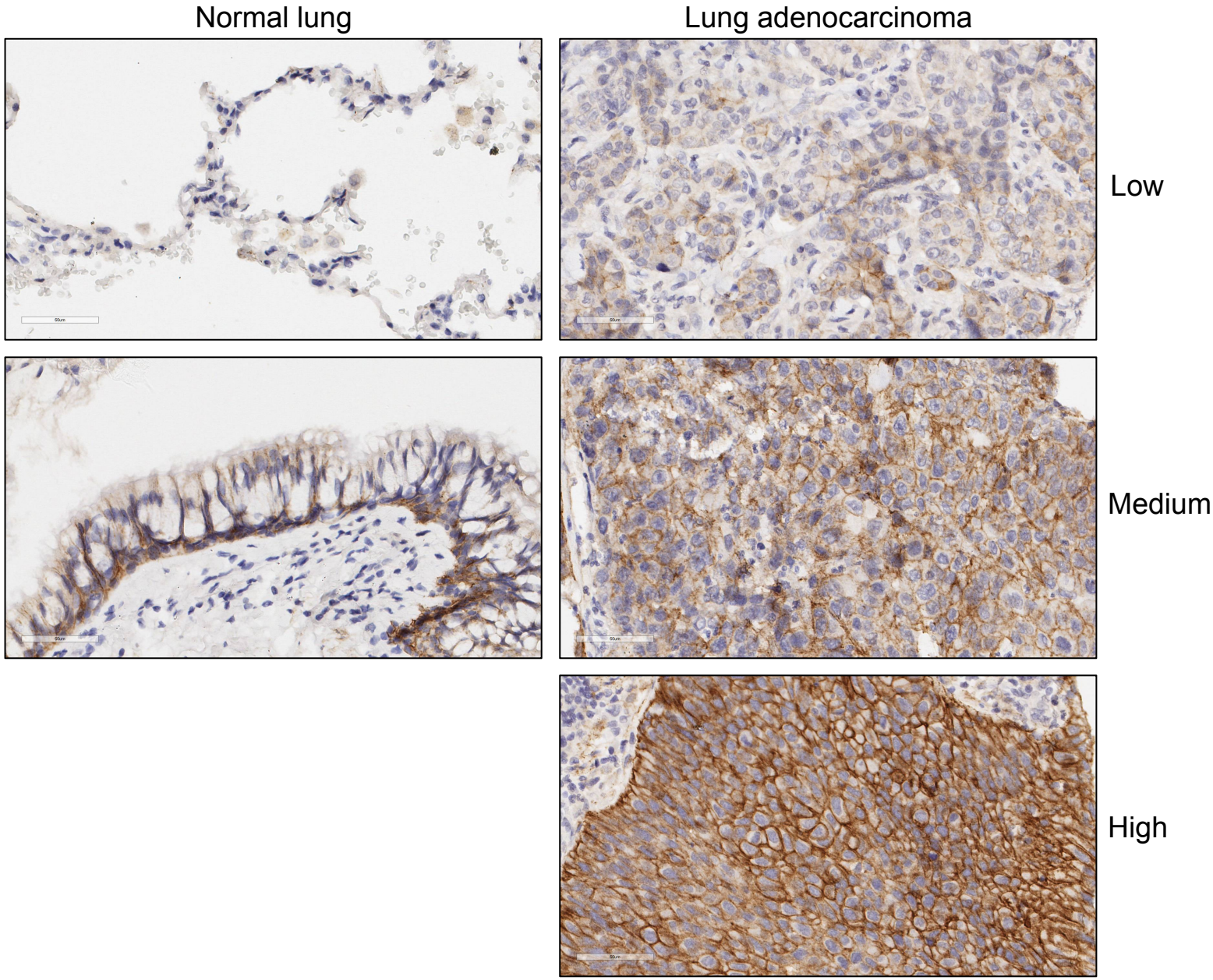
Supplementary Table S5. Correlation between *PCDH7* upregulation and alterations in *KRAS* or *EGFR*.

Supplementary Table S6. RNA-seq results from HBEC-shp53 and HBEC-shp53-*PCDH7* cells.

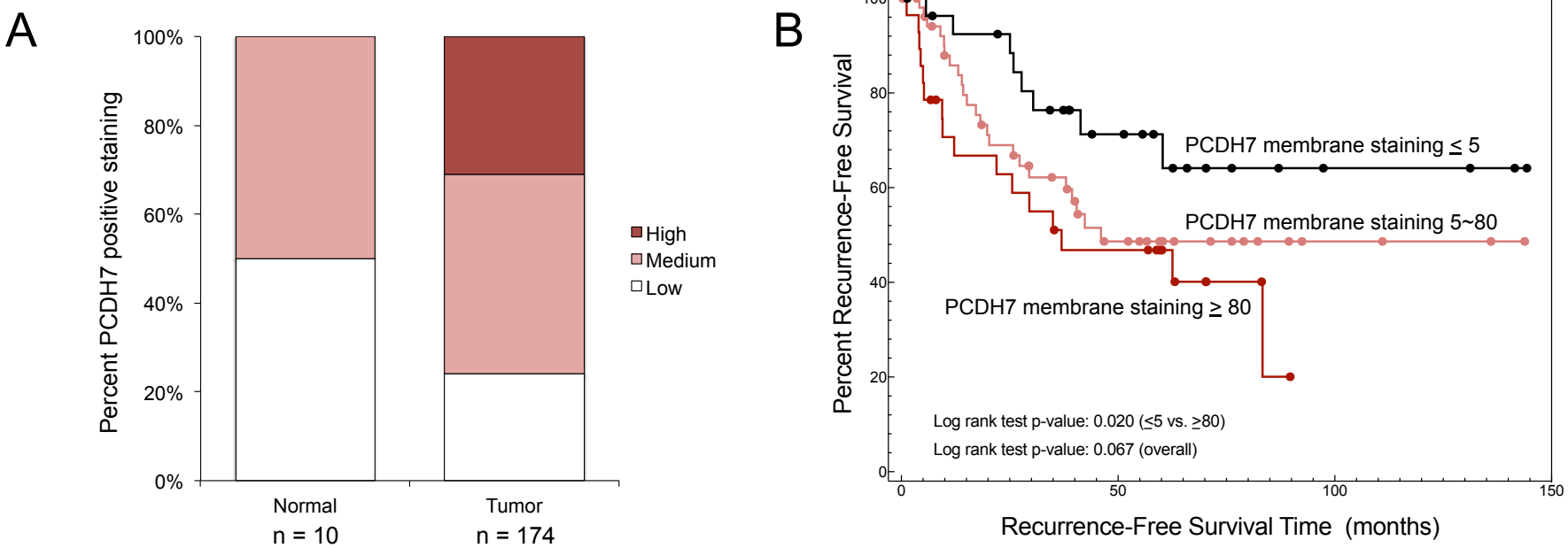
Supplementary Table S7. List of antibodies utilized in this study.

Supplementary Table S8. List of primer sequences utilized in this study.

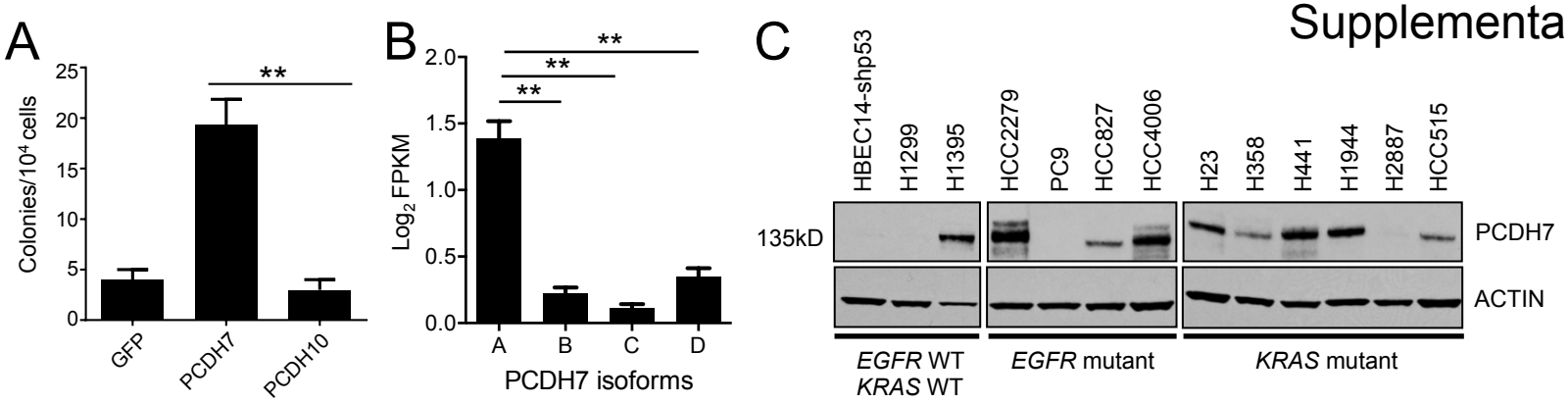
Supplementary Figure S1



Supplementary Figure S2



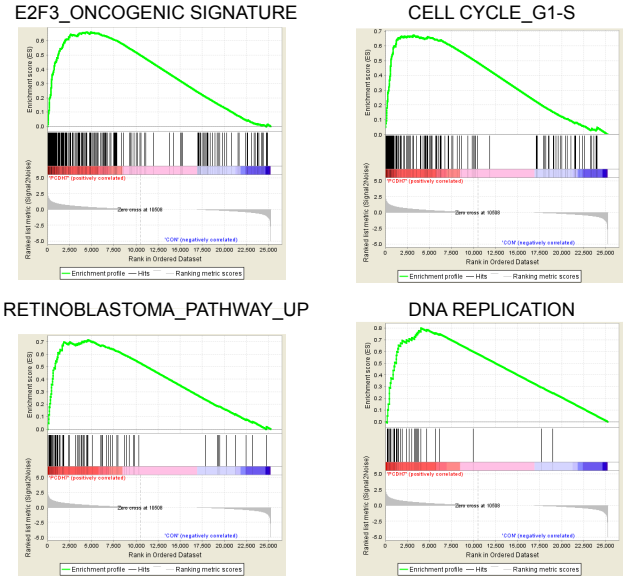
Supplementary Figure S3

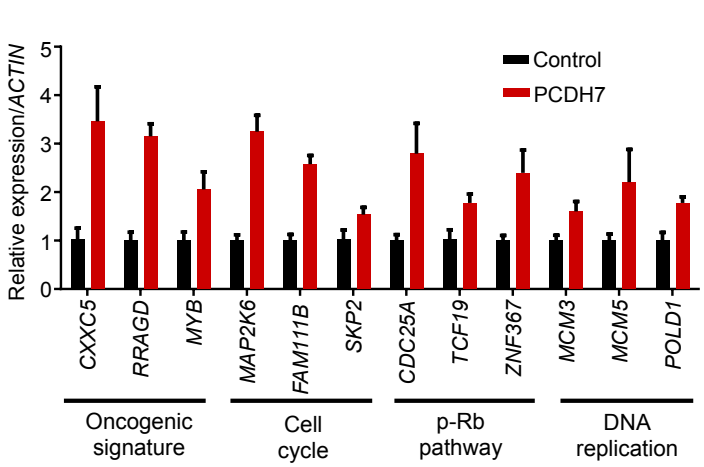


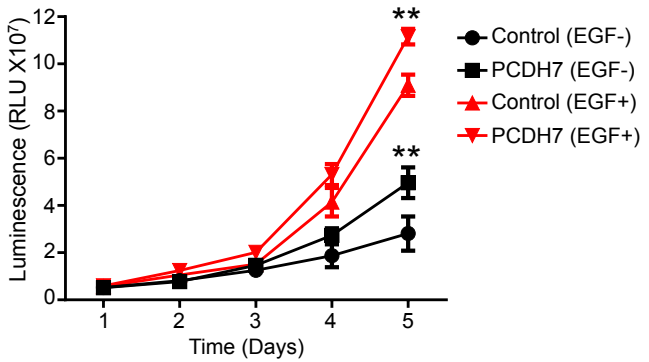
D

NAME	ES	NOM p-val	FDR q-val
BILD_E2F3_ONCOGENIC_SIGNATURE	0.660759	0	0
WHITFIELD_CELL_CYCLE_G1_S	0.67187285	0	0
VERNELL_RETINOBLASTOMA_PATHWAY_UP	0.71441567	0	2.35E-04
KEGG_DNA_REPLICATION	0.8006159	0	3.75E-04
REACTOME_DNA_STRAND_ELONGATION	0.8212658	0	3.13E-04
REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX	0.80395985	0	5.21E-04

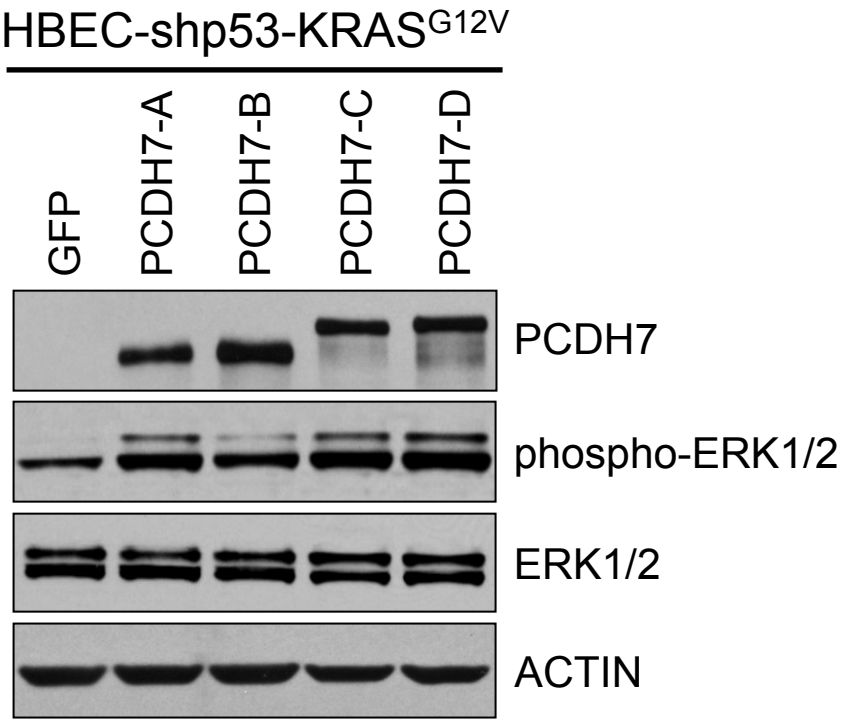
E



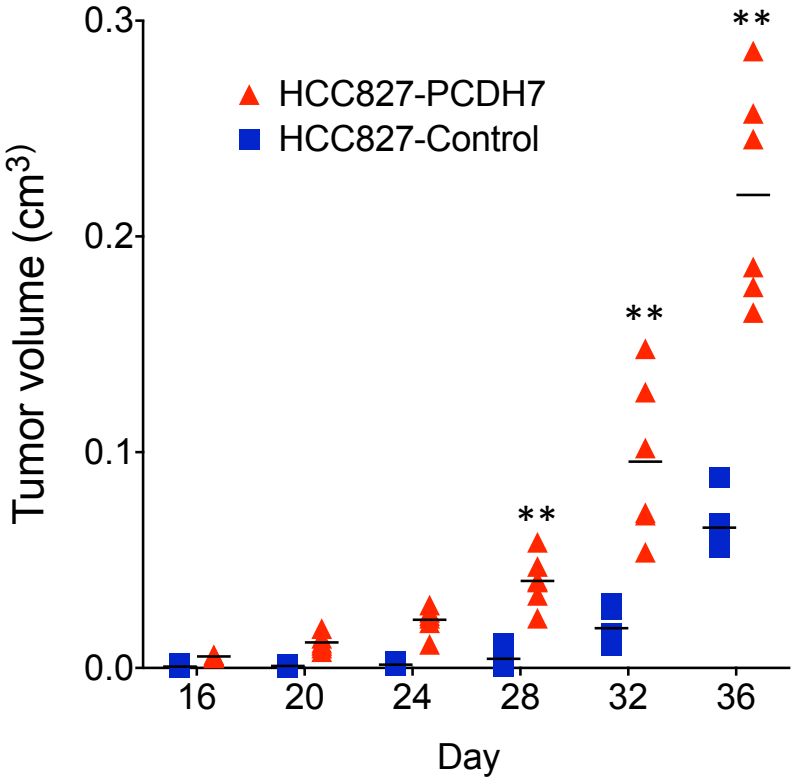




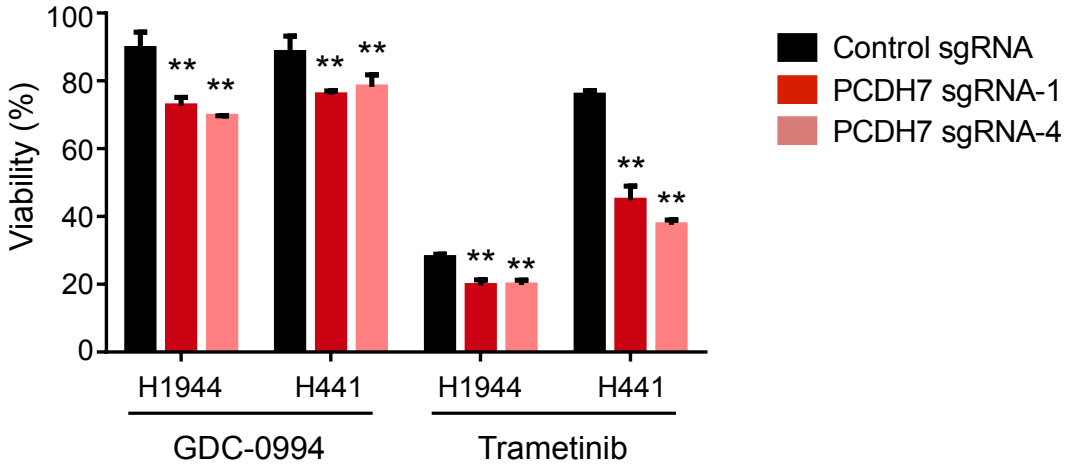
Supplementary Figure S4



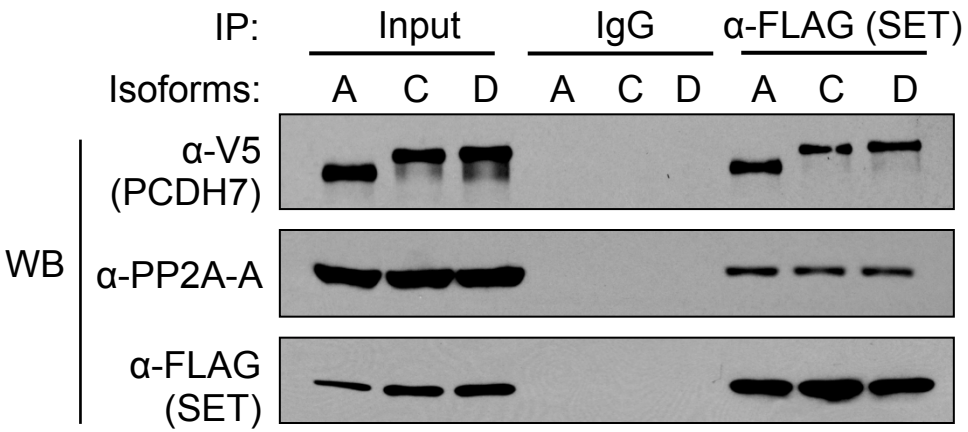
Supplementary Figure S5



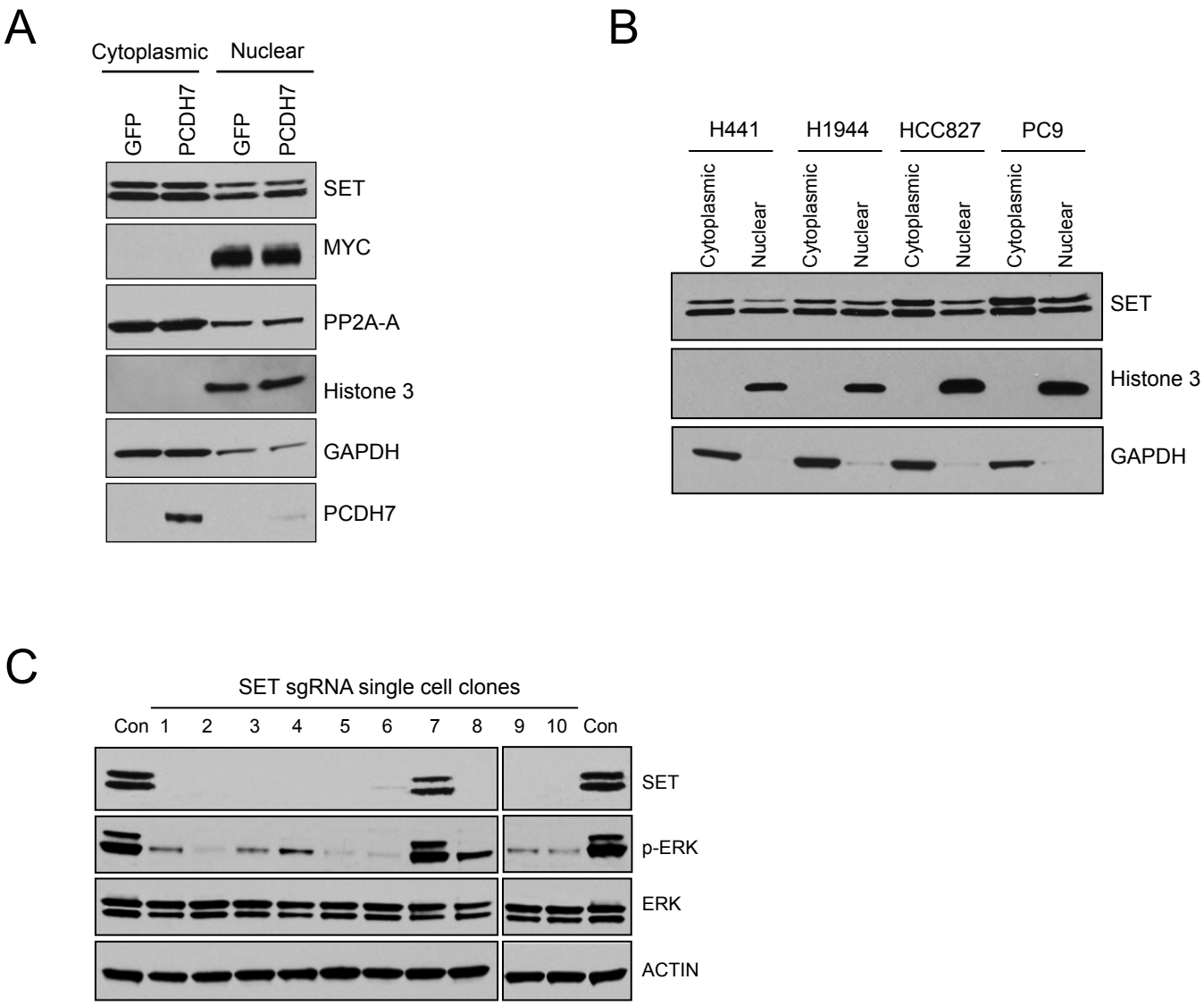
Supplementary Figure S6



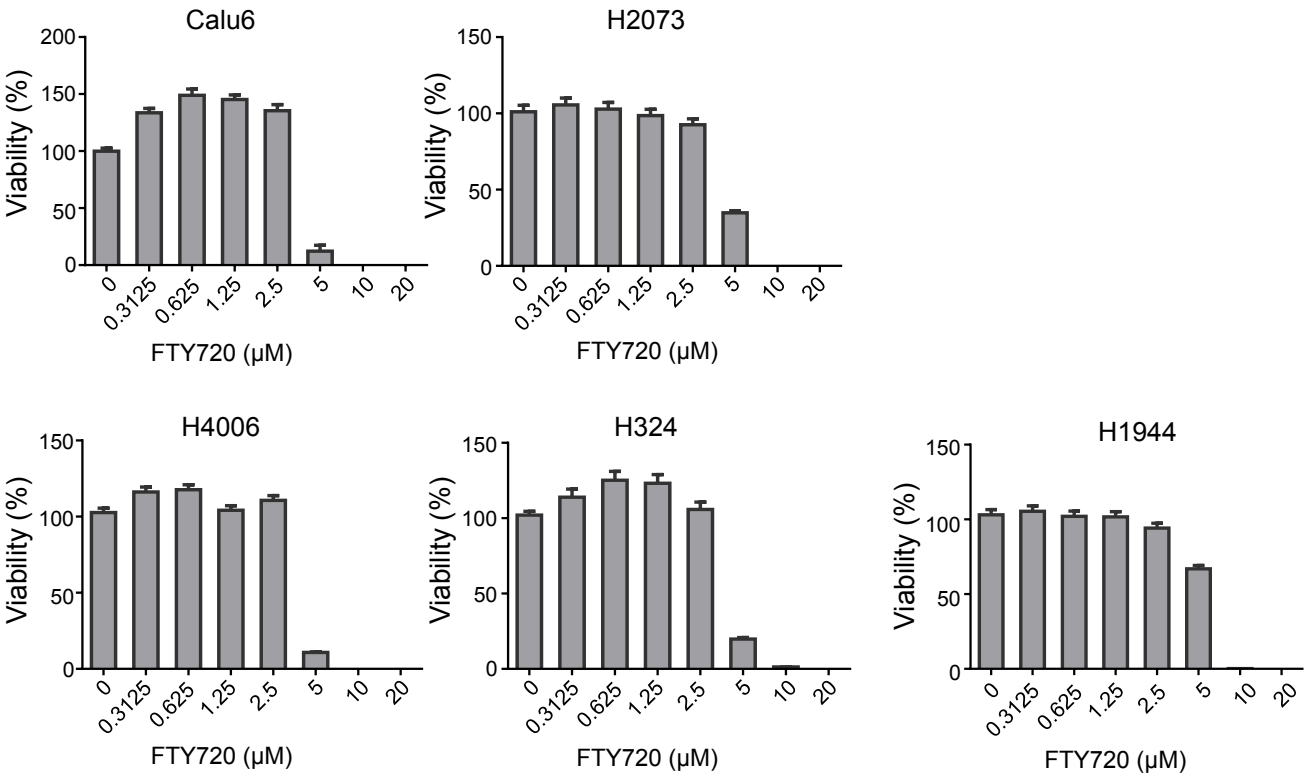
Supplementary Figure S7



Supplementary Figure S8



Supplementary Figure S9



Supplementary Table S1. Expression and survival analysis of PCDH family members in lung adenocarcinoma.

		Survival (p value / significance)										Expression (p value / significance)									
		Shedden *et al*, 2008 *Nat Med* (Adeno)		Bild *et al*, 2006 *Nature* (Adeno)		Tang *et al*, 2013 *Clin Cancer Res* (Adeno)		Bhattacharjee *et al*, 2001 *PNAS* (Adeno)		TCGA provisional (Adeno)		Landi MT *et al*, 2008 *PloS One* (Adeno vs. normal)		Lu TP *et al*, 2010 Cancer *Epidemiology, Biomarkers & Prevention* (NSCLC vs. normal)		Sanchez A *et al*, 2011 *International Journal of Cancer* (NSCLC vs. normal)		Su LJ *et al*, 2007 *BMC Genomics* (Adeno vs. normal)	
	Platform	AffyU133A		dAffyU133Plu2		IlluminaHuman6v3		AffyU95A		RNA seq		AffyU133A		AffyU133Plus2		AffyU133Plus2		AffyU133A	
δ-0 group	*PCDH20*	N/A		0.581369		N/A		N/A		0.72		N/A		0.0593043		9.30052E-09	L	N/A	
δ-1 group	*PCDH1*	0.47569		0.894702		0.605228		0.578388		0.446		0.0402992	L	0.1288085		0.4484851		0.4031197	
	PCDH7	0.00058	H	0.004278	H	0.00534145	H	0.03206	H	4.16E-06	H	7.676E-08	H	5.427E-13	H	1.28122E-11	H	5.0202E-06	H
	PCDH9	0.55668		0.443747		0.992703		N/A		0.133		8.562E-13	L	1.585E-23	L	5.41337E-19	L	9.9125E-07	L
	PCDH11X	0.60889		0.413519		N/A		0.28207		0.0799		0.2469483		5.371E-22	L	0.9569247		0.8457428	
δ-2 group	*PCDH8*	0.18967		0.199774		0.09725293		0.790945		0.663		0.0228569	H	0.0755945		0.1289002		0.1854034	
	PCDH10	N/A		0.173208		0.993564		N/A		0.398		N/A		1.026E-13	L	0.2544597		N/A	
	PCDH17	0.16394		0.984827		0.579395		0.426656		0.0297	L	1.407E-23	L	8.226E-19	L	5.44996E-11	L	6.7854E-06	L
	PCDH18	N/A		0.275691		0.61423549		N/A		0.0967		N/A		0.1205536		0.7048791		N/A	
	PCDH19	N/A		0.394025		0.70665671		N/A		0.773		N/A		0.0297184	L	2.5961E-07	H	N/A	

Survival
H: High expression is associated with poor survival (p<0.05)
L: Low expression is associated with poor survival (p<0.05)
N/A: no data available
cut off value: Mean

Expression
H: Higher in tumor compared to normal tissues (p<0.05)
L: Lower in tumor compared to normal tissues (p<0.05)
N/A: no data available

Supplementary Table S2. Affymetrix Probe IDs analyzed for each PCDH family member.

	Gene	EntrezID	Probe ID
δ-0 group	PCDH20	64881	232054_at
δ-1 group	PCDH1	5097	203918_at, 215277_at
	PCDH7	5099	205534_at, 205535_s_at, 210273_at
	PCDH11	27328	208366_at, 210292_s_at, 211227_s_at, 217049_x_at
	PCDH9	5101	219737_s_at, 219738_s_at
δ-2 group	PCDH8	5100	206935_at
	PCDH17	27253	205656_at, 227289_at, 228863_at
	PCDH18	54510	225975_at, 225977_at
	PCDH19	57526	227282_at
	PCDH10	57575	228635_at

Supplementary Table S3. Summary of clinical and pathologic information of the cohort of NSCLC patient tissue blocks used for TMA generation.

Cases		
	No. of Patients	218
Age		
	Average age	65
	Age Range	41-86
Gender		
	Males	115
	Females	103
Race		
	African American	12
	Asian	6
	Caucasian	194
	Hispanic	6
Smoking		
	Smokers	199
	Non-smokers	19
Pathologic Stage		
	IA	45
	IB	68
	IIA	12
	IIB	26
	IIIA	40
	IIIB	22
	IV	5
Pathologic Diagnosis	Adenocarcinoma	152
	Squamous	66
PCDH7 IHC Expression	Positive	136
	Negative	38
	Not Available (N/A)	44

Supplementary Table S4. Analysis of tissue microarray correlations

Univariate Recurrence Free Survival (Cox Regression)

PCDH7 membrane expression

		Recurrence		Hazard ratio and 95% CI				Overall
		N	Y	HR	Lower	Upper	p-value	p-value
Membr exp	<= 5	20	8	Reference				0.0765
	5 ~ 80	33	23	1.878	0.839	4.204	0.1253	
	>= 80	12	16	2.670	1.142	6.246	0.0235	

Multivariate Recurrence Free Survival (Cox Regression)

Initial model:

The variables with p<0.2 in univariate regression will enter the initial model.

They are: Tumor size, Final T stage, N stage, Final stage mountain, and PCDH7 membrane expression

Model selection method: backwards

Final model:

They are: Tumor size, Final stage mountain, PCDH7 membrane expression

		Hazard ratio and 95% CI				Overall
		HR	Lower	Upper	p-value	p-value
Tumor size	per cm	1.172	1.004	1.368	0.0446	0.0446
Final stage mountain	I-II	Reference				0.0007
	III-IV	2.893	1.564	5.350	0.0007	
Membr exp	<= 5	Reference				0.0411
	5 ~ 80	2.272	1.007	5.126	0.0481	
	>= 80	2.978	1.270	6.987	0.0121	

Multivariate Overall Survival (Cox Regression)

Initial model:

The variables with p<0.2 in univariate regression will enter the initial model.

They are: Gender, Age, Tumor size, Final T stage, N stage, Final stage mountain, EGFR.

Model selection method: backwards

Final model:

They are Tumor size, Final stage mountain

		Hazard ratio and 95% CI			
		HR	Lower	Upper	p-value
Tumor size	per cm	1.236	1.100	1.390	0.0004
Final stage mountain	I-II	Reference			
	III-IV	2.634	1.439	4.821	0.0017

Supplementary Table S5. Correlation between PCDH7 upregulation and alterations in *KRAS* or *EGFR* in CCLE lung cancer cell lines.

Case ID	PCDH7	KRAS	EGFR	Status (KRAS or EGFR alteration)	Status (KRAS only)	Status (EGFR only)
NCIH1975_LUNG	UP;	DOWN;	MUT: L858R,T790MAMP;	Yes	Yes	Yes
SQ1_LUNG	UP;		UP;	Yes	No	Yes
HCC827_LUNG	UP;		MUT: ELREA746delAMP;UP;	Yes	No	Yes
HCC4006_LUNG	UP;		MUT: ELR746delAMP;UP;	Yes	No	Yes
RERFLCKJ_LUNG	UP;			No	No	No
NCIH1435_LUNG	UP;			No	No	No
HS229T_LUNG	UP;			No	No	No
NCIH3255_LUNG	UP;		MUT: L858RAMP;UP;	Yes	No	Yes
KNS62_LUNG	UP;		UP;	Yes	No	Yes
NCIH1581_LUNG	UP;	DOWN;		Yes	Yes	No
ABC1_LUNG	UP;		UP;	Yes	No	Yes
NCIH661_LUNG	UP;	AMP;		Yes	Yes	No
NCIH1568_LUNG	UP;		AMP;UP;	Yes	No	Yes
NCIH146_LUNG	UP;			No	No	No
CALU6_LUNG	UP;	MUT: Q61KAMP;UP;		Yes	Yes	No
NCIH596_LUNG	UP;	DOWN;	AMP;UP;	Yes	Yes	Yes
SW900_LUNG	AMP;UP;	MUT: G12VAMP;UP;		Yes	Yes	No
NCIH2347_LUNG	UP;	MUT: L19F		Yes	Yes	No
NCIH2291_LUNG	UP;	MUT: G12F	MUT: V592FUP;	Yes	Yes	Yes
NCIH2405_LUNG	UP;		UP;	Yes	No	Yes
NCIH23_LUNG	UP;	MUT: G12CAMP;UP;		Yes	Yes	No
HCC2279_LUNG	UP;		MUT: ELREA746delAMP;UP;	Yes	No	Yes

	KRAS Alteration	EGFR Alteration	KRAS or EGFR Alteration	No Alteration for KRAS or EGFR
Lung PCDH7 UP: 22 CCLE cell lines	9	12	18	4

Supplementary Table S7. List of antibodies used in this study.

Antigen	Description	Company	Cat. No.	Applications
PCDH7	Mouse monoclonal	Abcam	ab139274	WB, IF, TMA
PCDH7	Rabbit polyclonal	Sigma	HPA011866	IP
SET	Mouse monoclonal	Santa Cruz	sc-133138	WB
SET	Rabbit polyclonal	Abcam	ab85389	IP
FLAG	Mouse monoclonal	Sigma	F3165	WB, IP
V5	Rabbit monoclonal	Sigma	V8137	WB
V5	Mouse monoclonal	Invitrogen	46-0705	WB
PP2A-A	Mouse monoclonal	Santa Cruz	sc-374264	WB
PP2A-A	Rabbit polyclonal	Millipore	07-250	IP
PP2A-C	Rabbit polyclonal	Millipore	06-222	IP
PP2A-C	Mouse monoclonal	Santa Cruz	sc-80665	WB
ERK	Rabbit monoclonal	Cell Signaling Technology	4695	WB
p-ERK	Rabbit monoclonal	Cell Signaling Technology	4370	WB
EGFR	Rabbit monoclonal	Cell Signaling Technology	4267	WB
p-EGFR	Rabbit monoclonal	Cell Signaling Technology	3777	WB
Pan-AKT	Rabbit monoclonal	Cell Signaling Technology	4685	WB
p-AKT	Rabbit monoclonal	Cell Signaling Technology	4060	WB
HSP27	Mouse monoclonal	Cell Signaling Technology	2402	WB
p-HSP27	Rabbit polyclonal	Cell Signaling Technology	2405	WB
IgG Isotype	Mouse monoclonal	Cell Signaling Technology	5415	IP, IF
IgG Isotype	Rabbit monoclonal	Cell Signaling Technology	3900	IP, IF
β-actin	Rabbit monoclonal	Cell Signaling Technology	4970	WB

Supplementary Table S8. List of primer sequences used in this study.

Gene	Primer sequence (5'-3')	Application
PCDH7	caccATGCTGAGGATGCGGACCGCGGGATG	PCDH7 isoform A cloning forward
	TCAGCCAAACACAGTAATGTATGGATGTAGACG	PCDH7 isoform A cloning reverse
	TTAGCCCTCCCTGGGATATTTAAATATATTTGGGATACATCTTACCTGTTTGCTGTACTTGT	PCDH7 isoform B cloning reverse
	CTACAGGTAAACTTCTCTTCTAGTGAGAG	PCDH7 isoform C/D cloning reverse
PCDH7 sgRNA-1	caccgCGACGTCCGCATCGGCAACG	Knock out PCDH7 with Lenti-CRISPR
	aaacCGTTGCCGATGCGGACGTCGc	
PCDH7 sgRNA-3	caccgCCTGGGCATCGTGACCGGAT	Knock out PCDH7 with Lenti-CRISPR
	aaacATCCGGTCACGATGCCCAGGc	
PCDH7 sgRNA-4	caccgCATCGTGACCGGATCGGGTG	Knock out PCDH7 with Lenti-CRISPR
	aaacCACCCGATCCGGTCACGATGc	
SET	cggggtaccgaATGTCGGCGCCGGCGGCCAAAGTCAG	FLAG tagged SET cloning
	ccgctcgagTTAGTCATCTTCTCCTTCATCCTCCTC	
SET sgRNA	caccgATGTTGTTACCCAAAAATTT	Knock out SET with Lenti-CRISPR
	aaacAAATTTTTGGGTAACAACATc	
SET mutant	TCCTAACTTCTGGGTTACCACTTTTGTCAACCATCCACAAGTGTCTG	CRISPR resistant SET cloning
	ATTTTGGCGATCAATTCTGACCTCTTCTG	
MCM5	TCACCTTCAAATACAGGGATGAACTC	qPCR
	GTACAAGTAGTCGGCCAGGTCCTCAT	
MYB	AGCACCGATGGCAGAAAGTACTAAAC	qPCR
	TTAAGTGCTTGGCAATAACAGACCAA	
SKP2	CTAAATCGAGAGAACTTTCCAGGTGT	qPCR
	CTTACAAACACCAGAGACCTTTAGCA	
MCM3	ATGATCCCAACTTTAGCCAGGAAGAT	qPCR
	CTGCACTCACCATCTTCTCCTTTTTC	

TCF19	AGGCTGGAATTGAGTGATGGAGAC	qPCR
	GTCCTGAGGCTTGACTCGTACTTGTT	
CXXC5	CCTCAGTGGCAGATGACACACCAC	qPCR
	GCCAAAAGAAGAGTAGTGGGAGAGC	
ZNF367	ATCCAGCAGAATCCGTTGTAACATCT	qPCR
	TTTGAACAAAGGCTTTTCCACAGTCT	
RRAGD	CTATGACAAGGAATCCACAGCCATC	qPCR
	CCCTTTTCTTTCAAAGCTTTCCTCTC	
MAP2K6	GTTCCACACCACCTCGAGATTTAGAC	qPCR
	CAGTTCCATTATAGGCTCCAGGTCAT	
CDC25A	CCAAACTCCACTACCCTGAGCTGTAT	qPCR
	TCTTCTTTAAAGTCCTCGTGGTGCAT	
FAM111B	AATGGGACAATAATGGAAACACAGGT	qPCR
	TGGATGTGTGTTTTTACCCACCATAA	
POLD1	TCTGAGCTGTATCAGAAGGAGGTATCC	qPCR
	CGCATGTAGAAGATGGGGCAGTCC	