Supplemental Information

KDM2B controls tumorigenesis in pancreatic cancer via Polycomb-dependent and -independent transcriptional programs

Alexandros Tzatsos, Polina Paskaleva, Francesco Ferrari, Vikram Deshpande, Svetlana Stoykova, Giamarco Contino, Kwok-Kin Wong, Fei Lan, Patrick Trojer, Peter J. Park, Nabeel Bardeesy

INVENTORY OF SUPPLEMENTARY MATERIALS:

SUPPLEMENTARY FIGURE LEGENDS

FIGURE S1, related to Figure 1

A) Validation of KDM2B IHC. **Left panels**: Analysis of mouse developing kidney with KDM2B sc-69477 antibody at E16.5 dpc. From the top to bottom, the scale bar represents 100, 50, and 20 microns, respectively. **Middle Panels:** In situ hybridization (ISH) analysis of KDM2B in the developing mouse embryo at E14.5dpc obtained through the http://www.eurexpress.org/ee/. The inset shows magnification of the positively stained developing kidney. Note the comparable profile to that seen by IHC. **Right panels:** Comparison of the IHC staining of a human PDAC metastasis specimen using the indicated KDM2B antibodies.

B) Validation of anti-KDM2B antibody Ab64920 for IHC. Pancreatic sections at age 12-weeks from a control mouse (left) or a transgenic mouse strain with a Doxycycline-inducible KDM2B allele (right) were stained for KDM2B.

C) Western blot analysis for the relative levels of KDM2B in non-transformed human pancreatic ductal epithelial (HPDE; left; KDM2B antibody described in (1)), and human pancreatic nestin-expressing (HPNE; right; KDM2B antibody from Millipore; catalog Number 09-864) cells compared to the indicated human PDAC cell lines.

FIGURE S2, related to Figure 2

A) qRT-PCR (left) and Western blotting (right) analysis of the knockdown efficiency of two different short hairpins against KDM2B in PANC1 and MiaPaca cells.

B) KDM2B overexpression rescues the growth of PDAC cells expressing shKDM2B. LZ10.7 cells expressing shKDM2B were superinfected pLVX-IRES-dtTomato lentiviruses encoding (i) empty vector, (ii) KDM2B ΔJmjC, and (iii) wild type KDM2B. The senescence-like appearance induced by shKDM2B was rescued by the concomitant overexpression of wild type (panel iii, red arrows) but not by KDM2B ΔJmjC (panel ii) or empty vector (panel i).

Panel iii shows that restoration of high levels of KDM2B expression fully rescued growth (red arrows) whereas non-infected cells remained senescent (white arrow).

C) LZ10.7 cells expressing shKDM2B were superinfected pLVX-IRES-dtTomato lentiviruses encoding empty vector, wild type KDM2B, or KDM2B ΔJmjC. $1.2x10^6$ cells were injected subcutaneously into the flanks of SCID mice (N=4 injections per condition). Tumors were excised and photographed after two months. The results are shown in the representative images (left panels) and in the table (right panel). The western blot shows levels of endogenous and ectopically expressed KDM2B in these cells; the full length KDM2B band is indicated (red arrowhead) on the longer exposure image.

D) HPDE cells were infected with the indicated lentiviruses and assayed for proliferation for the indicated period of time.

E) Orthotopic injection of $1x10^6$ *Ptf1α-Cre; LSL-Kras^{G12D}* pancreatic ductal cells results in the formation of focal low grade PanINs (shown at higher magnification in the inset). Scale bar = 200 microns.

F) The indicated mouse cell lines were analyzed by Western blotting. Note that KDM2B is upregulated in mouse PDAC cell lines compared to *Kras^{G12D}*-expressing, non-transformed mouse pancreatic duct cells.

G) NB508 PDAC cells from the *Pdx1-Cre;LSL-Kras^{G12D};p53$^{Lox/+}$* mouse model were infected with a lentivirus to knock down KDM2B and assayed for proliferation (left graph) and for tumor formation after injection of $5x10^5$ cells in the flanks of SCID mice after two weeks (right graph).

FIGURE S3, related to Figure 3

A) The gene expression changes induced by KDM2B overexpression in *Ptf1α-Cre; LSL-Kras^{G12D}* primary pancreatic ductal cells were examined using Ingenuity transcription factor prediction analysis. The chart shows predicted alterations in the activity of transcription

factors induced by KDM2B overexpression. Transcription factors with z-scores > 3 and *p* overlap value <0.01 are shown.

B) Effect of KDM2B overexpression on the absolute levels of the indicated mRNA transcripts in *Ptf1α-Cre; LSL-Kras*^{G12D} duct cells as determined by direct RNA-sequencing (Left panel). PRC2 and PRC1 components and PcG target genes are indicated. RT-PCR analysis confirming that overexpression of KDM2B silences the *Cdkn2a* locus (Right panel).

C-D) Effect of KDM2B knockdown on the levels of EZH2. **(C)** Absolute levels of EZH2 mRNA transcripts in six PDAC cell lines as determined by direct RNA-sequencing. KDM2B knockdown results in reduction in *EZH2* transcript levels in Mia-Paca, SW1990, and to a lesser extent YAPC, whereas there are no significant changes in the other cell lines. **(D)** Knockdown of KDM2B downregulates the levels of EZH2 protein in MiaPaca cells, but not in TU-8988 cells. BMI levels are unaffected.

E) Effect of KDM2B knockdown on the absolute levels of *MYC* mRNA transcripts in six PDAC cell lines as determined by direct RNA-sequencing.

FIGURE S4, related to Figure 4

A, B) Validation of KDM2B antibodies for ChIP-seq experiments. **(A)** The diagram depicts a Luciferase reporter construct driven by the GAL4 promoter that is stably integrated into HEK293 cells. ChIP with a GAL4 antibody showed a clear enrichment at the promoter of the luciferase transgene in cells transfected with either GAL4 or GAL4-KDM2B fusion protein. ChIP with different commercial antibodies against KDM2B revealed that the KDM2B 5G1 clone (Abnova) gave strong enrichment at this promoter in cells transfected with GAL4-KDM2B fusion protein. The other commercially available antibodies failed to give enrichment. **(B)** Retroviral-mediated overexpression of KDM2B increases the fold enrichment of KDM2B at target loci. The graphs show the relative enrichment of KDM2B at

GATA6 and *JAG1* loci (left) and the fold enrichment over input (right) in PANC1 cells that stably express KDM2B.

C) The charts show the percent overlap of the EZH2 (top) and KDM5A ChIP-seq targets (middle) identified in PDAC cells with the targets of the same factors identified in other cell lines. The lower chart shows the overlap between KDM2B targets in PDAC cells with the targets of PRC2 and PRC1 components in ES cells.

D) Comparative plot showing the binding of KDM2B (green), EZH2 (blue), and KDM5A (red) in the promoters of genes encoding nucleolar and mitochondrial ribosomal proteins. Note, that KDM2B and KDM5A bind to many of these genes, whereas EZH2 does not.

FIGURE S5, related to Figure 5

A) ChIP-seq enrichment density profiles of the KDM2B-EZH2 and KDM2B-KDM5A binding modules around the TSS. The dark colored solid line shows average log scale enrichment. The lighter colored shaded area represents 95% confidence intervals. Note the virtual absence of EZH2 at KDM2B-KDM5A targets (left column) and of KDM5A at the KDM2B-EZH2 targets (right column).

B) Genes bound by KDM2B, EZH2, and KDM5A were sorted based on their expression (RNA-seq) into five different bins from low (light color) to high expression (dark color). Note that overall KDM2B and KDM5A binding is stronger at more active genes, whereas EZH2 binding is stronger at inactive genes.

FIGURE S6, related to Figure 6

A) ChIP analysis of the binding of PcG proteins at the *GATA6* locus.

B) qRT-PCR analysis of expression changes of selected genes co-bound by KDM2B-EZH2 in response to knockdown of KDM2B or EZH2 in PANC1 cells.

C) Relative fold-change in the expression of *GATA6* mRNA upon knockdown of PcG proteins in the indicated cell lines.

D) qRT-PCR analysis of the relative expression of *KDM2B* and *GATA6* in quasi-mesenchymal (QM) and classical PDAC cell lines.

E) Spearman *r* correlation of the expression of *KDM2B* and *GATA6* in human PDAC samples (data obtained from GSE17891 (2), and GSE15471 (3)).

F) *Left panel:* Western blot analysis of the indicated proteins in quasi-mesenchymal (QM) and classical PDAC cell lines. *Right panel:* MiaPaca cells were infected with the indicated lentiviruses to knock down KDM2B. Cells were analyzed after seven days by Western blotting (right panel).

G) ChIP analysis to determine the relative fold enrichment of the indicated histone marks on selected target genes that define the different KDM2B modules in PANC1 cells. The blue and red bars represent KDM2B-EZH2 co-bound genes and KDM2B-KDM5A co-bound genes, respectively.

H-J) PDAC cell lines were infected with the indicated lentiviruses and analyzed by Western blotting for the indicated proteins.

FIGURE S7, related to Discussion

A) Binding of KDM2B and EZH2 at the genomic loci of *MIRLET7BHG* and *hsa-miR-101-1* in PANC1 cells. KDM2B binds at the promoter of *MIRLET7BHG*, a non-coding RNA that hosts *LET7B* (red arrow). The same non-coding RNA hosts additional miRNAs of the same family, such as *LET7A3*, as well as *miR4763*, and *miR3619*. KDM2B also binds at the promoter of *hsa-miR-101-1* along with EZH2. The color graphs at the bottom show the functional annotation of chromatin in human embryonic stem cells as described in (4).

B) ChIP analysis showing the binding of KDM2B, RING1B, and BMI1, to the promoters of *MIRLET7BHG* and *has-miR-101-1*.

C) MiaPaca cells were treated with shKDM2B and analyzed for the relative levels of *MIRLET7BHG*.

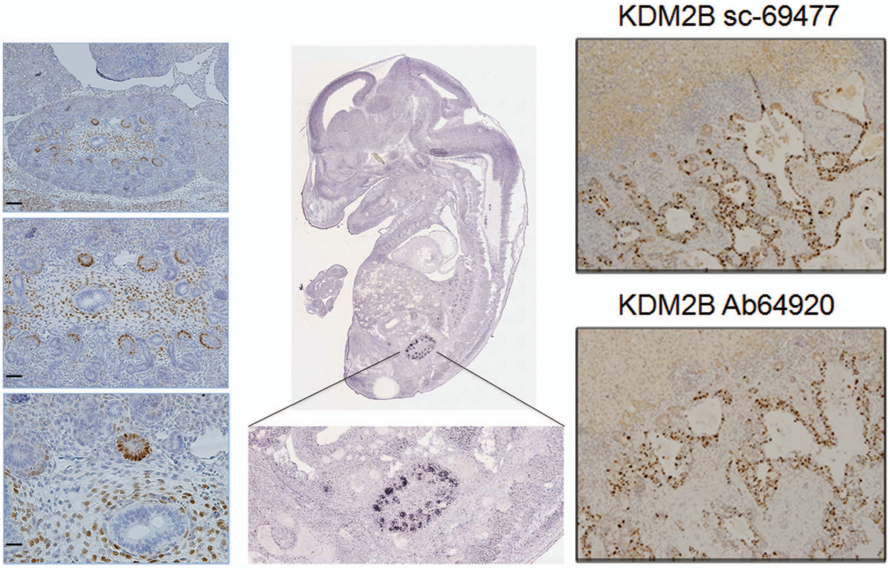
D) Spearman *r* correlation between the expression of KDM2B and EZH2 in 2158 human tumor samples with data obtained through GSE2109.

E) Spearman *r* correlation of the expression of KDM2B and EZH2 in human PDAC samples (data obtained from GSE15471 (3)).

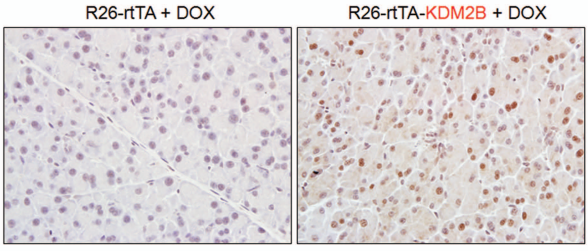
F) Unsupervised hierarchical clustering showing the expression of HDM in 2158 human tumor samples (GSE2109). KDM2B exhibits the highest degree of coordinate regulation with EZH2 when compared to all of the HDM family members (red line indicates the co-clustering of KDM2B and EZH2).

Figure S1, related to Figure 1

A



B



C

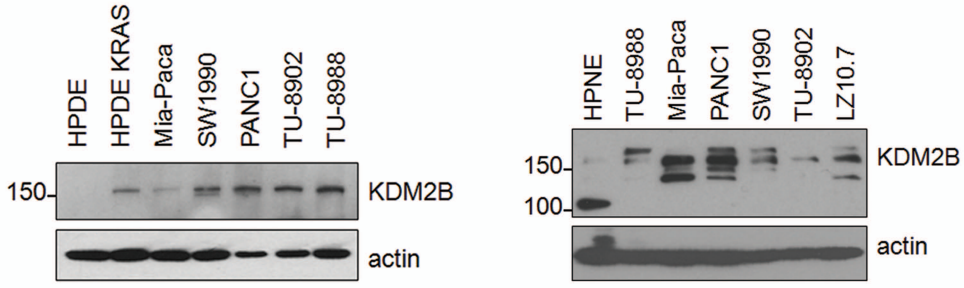


Figure S2, related to Figure 2

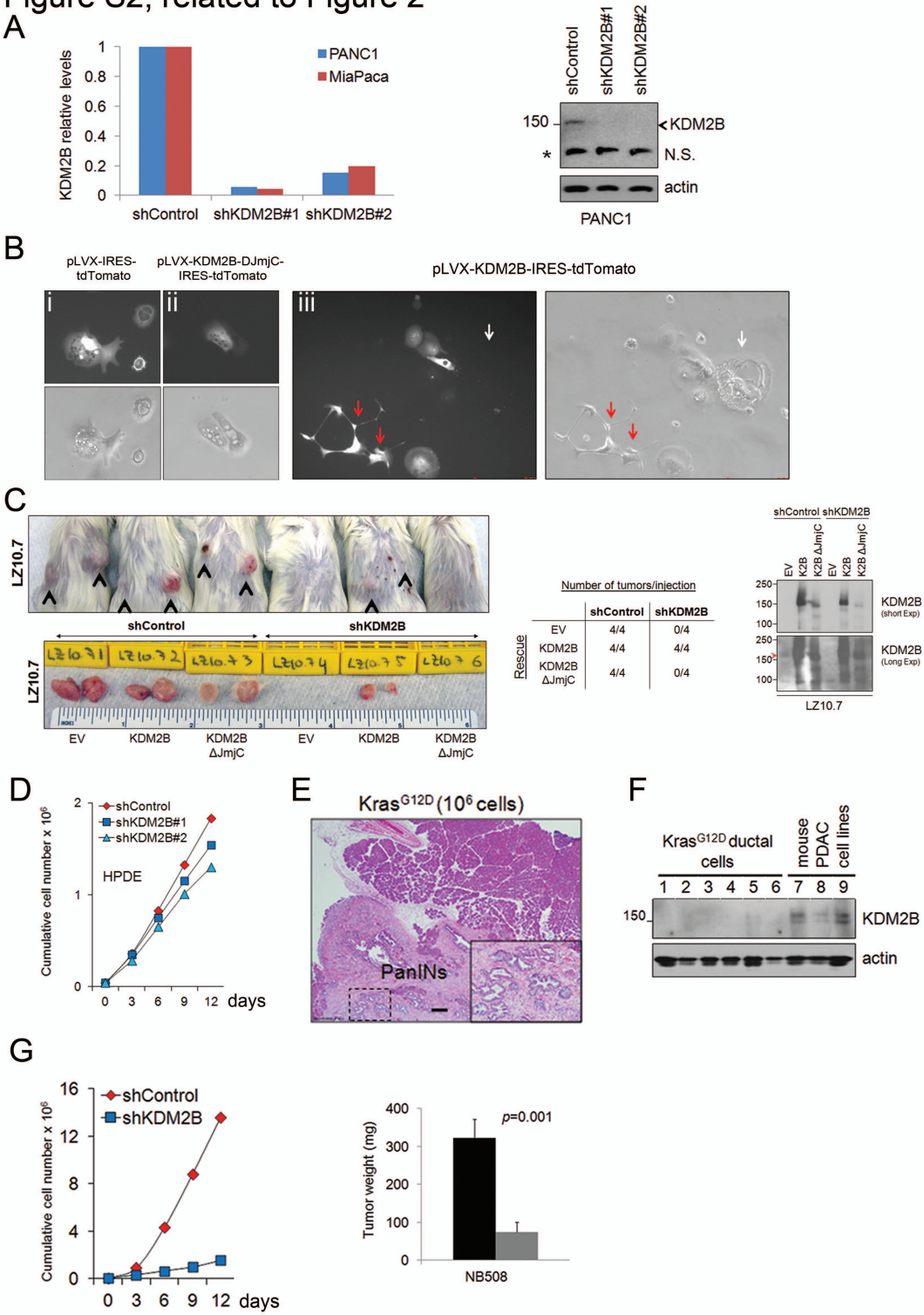
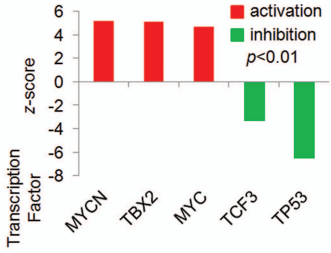
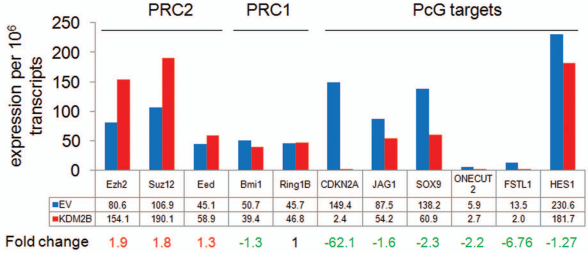


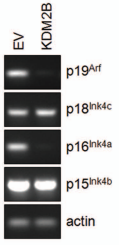
Figure S3, related to Figure 3

A

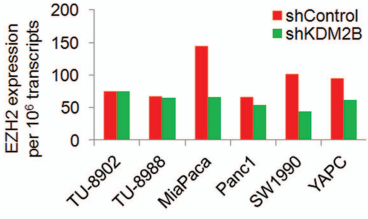


B

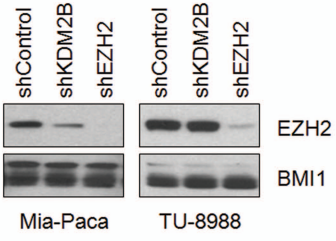




C



D



E

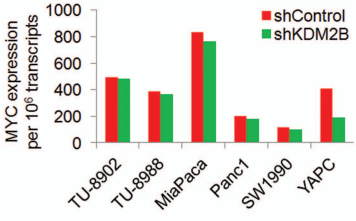
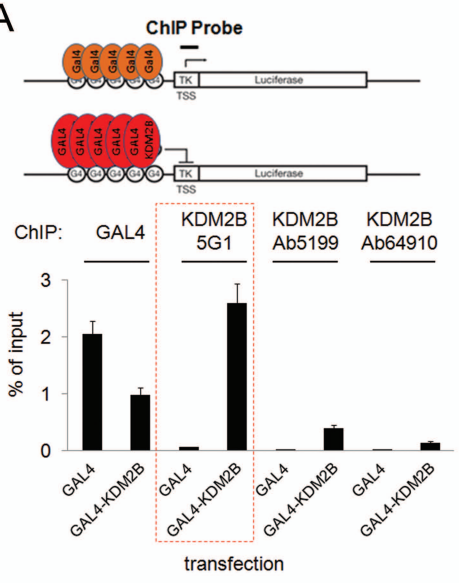
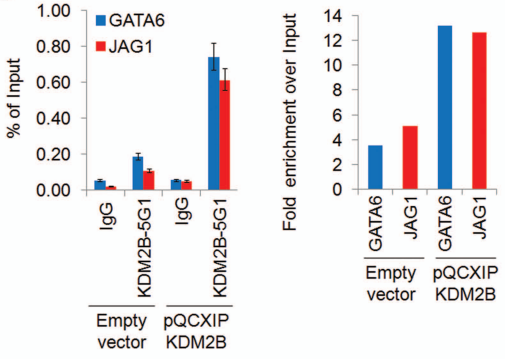


Figure S4, related to Figure 4

A



B



C

EZH2	mES cells/GSE13084 (2235 genes)
PANC1 (2113 genes)	1017 (**45.5%**)

KDM5A	U937/GSE10636 (276 genes)	LF2/GSE18776 (2295 genes)
PANC1 (3502 genes)	227 (**82.2%**)	1444 (**62.9%**)

Study in ES cells	KDM2B_PANC1 3502 genes
SUZ12/GSE19365 4566 gene	1117 (**24.4%**)
SUZ12/GSE13084 1776 genes	564 (**31.7%**)
EZH2/GSE13084 2235 genes	659(**29.5%**)
RING1B/GSE13084 1178 genes	417 (**35.3%**)

D

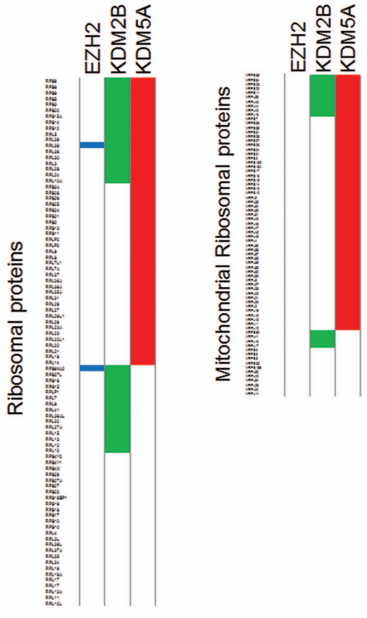
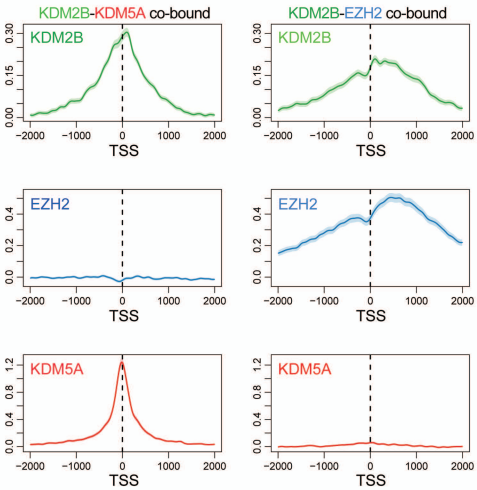


Figure S5, related to Figure 5

A



B

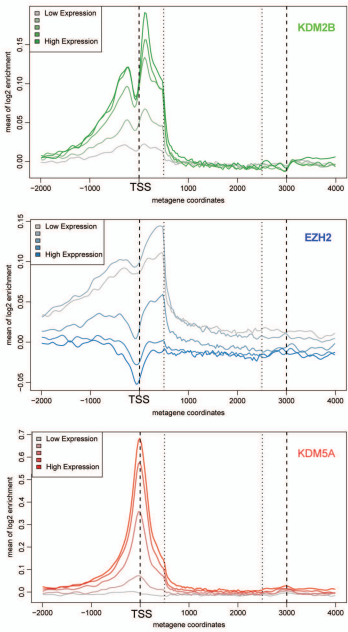


Figure S6, related to Figure 6

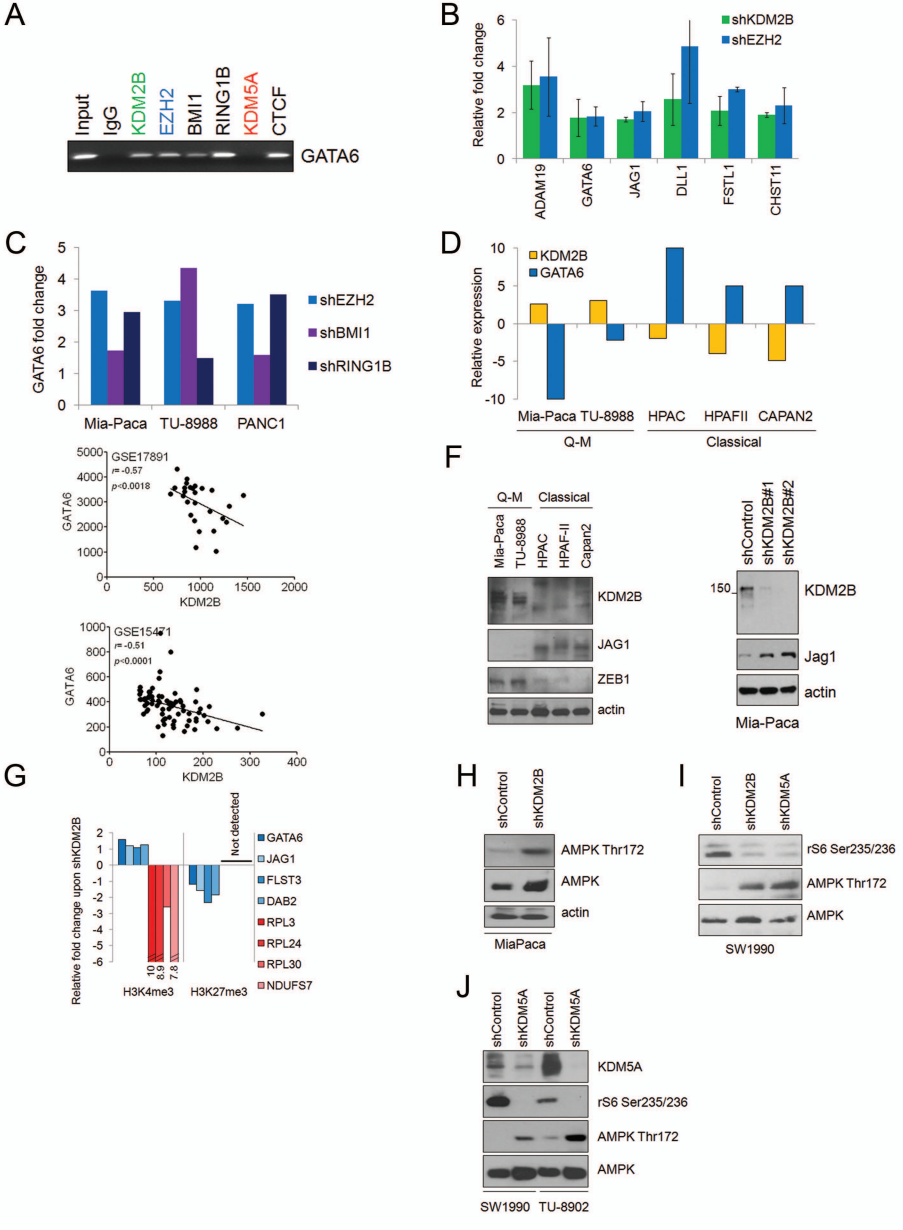
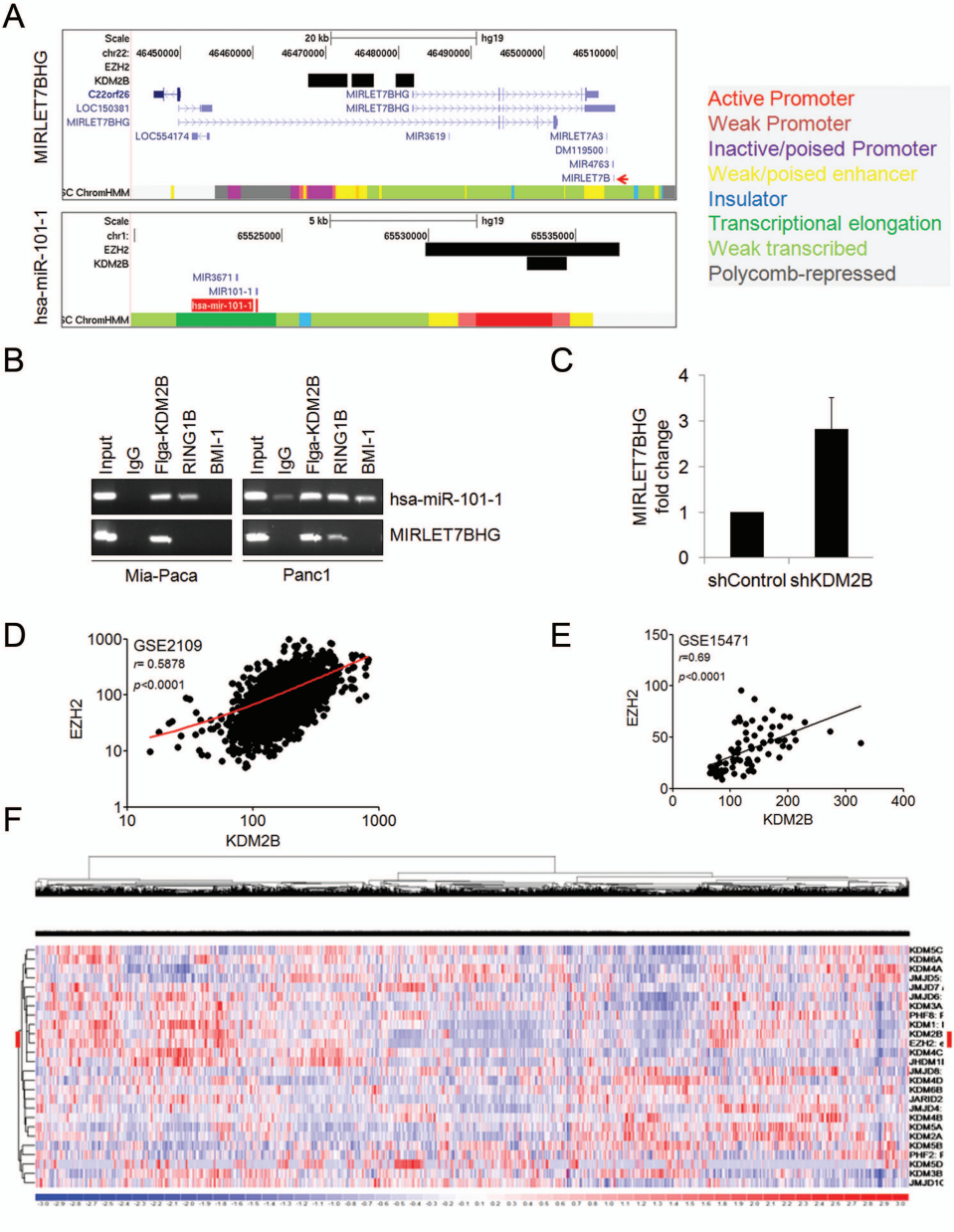


Figure S7, related to discussion



SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Direct RNA-sequencing (DRS). RNA was isolated with the RNeasy mini kit (#74104; Qiagen) and subjected to Digital Gene Expression (DGE) sample preparation and analysis on the HeliScope™ Single Molecule Sequencer from Helicos BioSciences (http://helicosbio.com/) as previously described (5).

Read alignment. RNA seq data were obtained with DRS protocol for Helicos sequencing technology (5). Sequencing reads were quality filtered and aligned to reference genome (hg18) using basic analysis pipeline of Helicos software (Helisphere) with default parameters (except for global ambig parameter set to "none" and filter align min score set to 4.3). Raw sequencing data are available in SRA data base (http://www.ncbi.nlm.nih.gov/sra/) with accession number SRA050208.

Gene expression estimation. DRS protocol yields only one strand specific read per transcript, starting at the 3'-end polyadenylation site. Therefore transcript abundance was evaluated as reads counts per million (RPM), which for DRS are equivalent to transcripts per million (TPM). RefSeq annotations tables were retrieved from UCSC Genome Browser (hg18 genome version) and read alignments were compared with position of annotated RefSeq transcripts. Each read was counted as associated with a specific transcript if aligned within a +/-2kb window around the transcript 3'-end, as in some genes there are alternative poly-A sites or 3'-end annotations are not certain, and reads may be located at some distance from the annotated poly-A site. This +/-2kb window is truncated upstream if the transcript is actually shorter than 2kb, or downstream if another annotated transcript start is less than 2kb apart downstream along the same strand. Then read counts were summarized at gene level, counting only once the reads mapped to multiple transcripts isoforms.

Normalization and differential expression analysis. RPM values were normalized using the TMM normalization procedure as described in (6) and as implemented in the edgeR Bioconductor package. Each shRNA RNA-seq sample is paired with matched control sample, therefore to identify differentially expressed genes following KDM2B silencing across all of the cell lines we used edgeR functions for paired samples analysis based on negative binomial distribution (7, 8). P-values obtained from this analysis taking into account all cell lines were used to rank genes for GSEA analysis presented in Figure 3A. The normalized gene expression levels (reads per million; RPM) are presented in Supplementary Table S1.

Gene set functional analysis (GSEA). As we noted some variability in response to KDM2B silencing observed in each individual cell line, we performed functional association analysis with GSEA (9) by analyzing each cell line independently, then comparing the selected differential functional classes. For each cell type, genes with less than 4 RPM expression in both shRNA and paired control sample are filtered out, to reduce noise from genes with always low expression, then log2 ratios of normalized expression in shRNA treated and paired control are used to rank genes for GSEA analysis. Gene sets significant at 0.05 *p*-value are selected for comparing results obtained across different cell lines (Figure 3A).

Mouse duct cells. For mouse pancreatic duct cell samples the same procedure was used to align DRS RNA-seq reads and quantify expression level, using mm9 as reference genome.

ChIP-seq. 10^7 cells were fixed with 1% formaldehyde for 20 minutes and cross-linking was terminated by the addition of glycine at a final concentration of 125mM for 5 min. Cells were washed twice with ice-cold PBS and lysed in ice cold PBS 0.5% NP-40 plus protease inhibitors. Nuclei were peleted via centrifugation for 15 minute at 800 x *g* at 4°C. Chromatin extraction, DNA sonication, immunoprecipitation, and library construction and sequencing took place according to Active's Motif pipeline (http://www.genpathway.com/). Previously validated antibodies were used to ChIP EZH2 (#07-689; Millipore) and KDM5A (3076; Cell Signaling).

Validation of a KDM2B antibody suitable for ChIP-seq. To validate a KDM2B antibody suitable for ChIP-seq, we generated a GAL4-KDM2B fusion protein and subsequently transfected HEK293 cells that carry a luciferase transgene driven by the *GAL4* promoter. Figure S4A shows that the 5G1 mouse monoclonal antibody (H00084678-M09; Abnova) efficiently ChIPs GAL4-KDM2B, but not GAL4, at the promoter of the luciferase transgene and was chosen to be used in the ChIP-seq experiment. In pilot experiments we also noticed that although this antibody efficiently ChIP endogenous KDM2B at the promoters of *GATA6* and *JAG1*, overexpression of KDM2B greatly increased the fold enrichment and improved the signal-to-noise ratio (Fig. S4B). For this reason we proceeded to ChIP-seq KDM2B in PANC1 cells that stably express KDM2B after infection with pQCXIP-KDM2B retroviral vector.

Read alignment. Illumina sequencing reads were aligned to the reference genome (hg19) using bwa aligner (10), with default parameters (except for "–q=10", "-l=28"). Only uniquely aligned reads were selected Raw sequencing data are available in SRA data base (http://www.ncbi.nlm.nih.gov/sra/) with accession number SRA050209.

Peak calling. For genome-wide identification of target proteins binding from ChIP-seq data (peak calling), we used spp package (11), which includes both algorithms for sharp peaks identification (precise binding positions) and a procedure detecting broad regions of enrichment. Duplicated reads, i.e. potential PCR artifacts, were filtered using a conservative approach by keeping no more than one read per genomic position. Unsupervised analysis of ChIP-seq reads distribution around gene bodies suggests that KDM5A has sharp peaks at Transcription Start Site (TSS) position whereas KDM2B and EZH2 have broader regions of enrichment around TSS. Therefore the spp WTD method for precise binding site location was used for KDM5A peak calling (with e-value ≤10, corresponding to FDR ≤0.0025) whereas enrichment analysis for broad regions was used for KDM2B and EZH2 (4.42 z-score threshold, i.e. approx 1e-5 p-value).

Binding association to genes. RefSeq transcripts annotations were downloaded from the UCSC genome browser. For each RefSeq transcript a window spanning 2kb upstream and 500bp downstream the annotated TSS was considered. This window was truncated downstream of the TSS if the RefSeq transcript is less than 2kb long, and is also truncated upstream of the TSS if another RefSeq transcript ending on the same genomic strand is located less than 2kb upstream. Each RefSeq TSS was associated with ChIP-seq target binding if any overlap existed between the ChIP-seq peak the TSS window. Then these overlaps were summarized at the gene level. The entire dataset is presented in Supplementary Table S2. In the Venn diagrams (Figure 4), the list of associated genes were compared across different ChIP-seq targets. Venn diagrams with proportional overlapping areas were plotted using the Vennerable R package.

Metagene profiles. To obtain an unbiased estimation of binding distribution near gene loci we plotted metagene profiles. These plots show average ChIP-seq enrichment signal over a large number of genes. For each RefSeq transcript a window -2kb upstream the TSS and 1kb downstream the Transcription End site (TES) was considered. RefSeq transcripts overlapping each other in the range of such windows were discarder to avoid interferences in ChIP-seq

signal distribution. Only genes 2.5 kb or longer were used, then the average ChIP-seq profiles is obtained by stretching or compressing real gene coordinates so as to fit a common gene size of 3kb. The regions outside the gene body are reported in real scale, as well as the first and last 500bp in the gene body.

Binding distribution at TSS. To show ChIP-seq signal at the TSS of selected genes we used instead a +/-2kb window centered at the TSS (using the most upstream TSS when multiple annotated TSS are present).

ENCODE data analysis. ENCODE ChIP-seq data. Publicly available ChIP-seq data were obtained from ENCODE database for H3K4me3 (Panc1 cell line) and MYC (HeLa cell line). To allow better comparison of results, raw reads for the selected ENCODE ChIP-seq samples were processed using the same analysis pipeline adopted for our data, including re-alignment of reads over the reference genome. Data from replicates were merged. As we are dealing with different ChIP-targets and data coming from different laboratories, to allow better direct comparability of enrichment profiles at the TSS (Figure 5D), ChIP-seq enrichment profiles around TSS of all genes were rescaled: for each sample the smoothed enrichment profiles computed at 20bp bin resolution were transformed to the corresponding percentile of enrichment signal over input control.

ENCODE methylation data. DNA methylation data for PANC1 cell line are available in ENCODE database as RRBS data. These data provide a non-continuous signal with variable coverage of individual promoter regions, therefore a continuous profile directly comparable to other ChIP-seq data is not available. Thus we report instead the average percentages of methylation in the promoter region. Read alignment results as provided in ENCODE database were used. Reads from replicates were merged and combined to estimate methylation level. Accordingly to ENCODE guidelines, only individual cytosines with at least 10 reads coverage were used to estimate the methylation percentage at individual positions in the merged replicates data.

ENCODE datasets analyzed in the current study and included in Supplementary Figure S2:

File_ID	Sample
wgEncodeUwHistonePanc1H3k04me3StdRawDataRep1	Panc1_H3K4me3
wgEncodeUwHistonePanc1H3k04me3StdRawDataRep2	Panc1_H3K4me3
wgEncodeUwHistonePanc1InputStdRawDataRep1	Panc1_INPUT
wgEncodeYaleChIPseqRawDataHelas3Input	HeLa_INPUT
wgEncodeYaleChIPseqRawDataRep1Helas3Cmyc	HeLa_MYC
wgEncodeYaleChIPseqRawDataRep2Helas3Cmyc	HeLa_MYC
wgEncodeHaibMethylRrbsPanc1HaibSitesRep1	Panc1_RRBSmethylation
wgEncodeHaibMethylRrbsPanc1HaibSitesRep2	Panc1_RRBSmethylation

Bioinformatics analysis**.** Raw data for GSE2109 (http://www.intgen.org/expo/); GSE17891 (2); GSE15471 (3); were downloaded from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geoprofiles) and analyzed with the dChIP analyzer (www.dchip.org)(12). The Invariant Set Normalization method was used to normalize arrays, and the model-based method was used for probe-selection and computing expression values. Background subtraction was performed with the mismatch probe (PM/MM difference) algorithm. Statistical analysis took place with the GraphPad PRISM Version 5.01.

Ingenuity Pathway Analysis (IPA). IPA Software (Ingenuity® Systems, www.ingenuity.com) is a knowledge-based approach for interpreting genome-wide expression profiles through an experts

curated database (Ingenuity® Knowledge Base) containing up-to-date information on thousand of mammalian genes and proteins and their biological interactions, structured into biological concepts related to pathways and diseases. The experimental data sets were used to query the IPA and to compose a set of interactive networks taking into consideration canonical pathways, relevant biological interactions, and cellular and disease processes. The significance of the association between the experimental data sets and the Ingenuity® Knowledge Base concepts was measured in two ways: (a) a ratio of the number of molecules from the data set that map to a particular Ingenuity® Knowledge Base concept, and (b) a Fisher's exact test was used to estimate the statistical significance of overlap between the two gene sets. To predict alterations in the activity of transcription factors that can explain the observed gene expression changes in PDAC cell lines expressing shKDM2B we considered transcription factors with z-scores >3 and a *p* overlap value of <0.01. The regulation z-score represents a statistical metric of the direction of the transcriptional changes (>1.5 fold) caused by KDM2B in at least 4 PDAC cell lines compared to the changes regulated by the transcription factor of interest in the Ingenuity® Knowledge Base. The positive or negative sign of z-score indicates activation or repression, respectively, of a particular transcription factor; a z-score higher than 2 is considered significant. The overlap *p*-value, calculated using Fisher's Exact Test, measures whether there is a statistically significant overlap between the dataset genes assigned for a particular transcription factor in Ingenuity® Knowledge Base and the genes that are regulated by KDM2B.

Chromatin Immunoprecipitation (ChIP). 10^6 cells were fixed with 1% formaldehyde for 15 minutes, followed by 2 washes with ice-cold PBS. Cells were lysed in ice cold PBS 0.5% NP-40 plus protease inhibitor mixture I from Roche (no. 11836170001) and nuclei were pelleted via centrifugation for 15 minute at 800 x *g* at 4°C. Fixed nuclei were resuspended in 500 µL of SDS lysis buffer [50 mM Tris-HCl (pH 8.0), 1% SDS, 150 mM NaCl, and 5 mM EDTA], and they were incubated on ice for 10 min. Lysates were sonicated in a Bioruptor sonicator 3 times, at 8 minutes each, with 30 second-off interval times at maximum output setting to achieve a chromatin size of 200–700 bp. The sonicated lysates were centrifuged at 14,000 x *g* for 30 min at 4°C and diluted ten times with dilution buffer [16.7 mM Tris (pH 8.0), 167 mM NaCl, 0.01% SDS, 1.1% (vol/vol) Triton X-100, and 1.2 mM EDTA]. DNA was recovered from immune complexes on protein A- or G- agarose beads with antibodies against H3K27me3 (9733), H3K4me3 (9751), H3K36me2 (9758), CTCF (4318), EZH2 (5246), RING1B (5694), BMI1 (5865), KDM5A (3876) bought from Cell Singaling, c-MYC (sc-42 and sc-764; Santa Cruz), and KDM2B (09-864; Millipore or H00084678-M09; clone 5G1 Abnova) overnight at 4°C on a rocking platform. Subsequently, the beads were washed with low-salt immune complex wash buffer [20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.1% SDS, 1% Triton X-100, and 2 mM EDTA], high-salt immune complex wash buffer [20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 0.1% SDS, 1% Triton X-100, and 2 mM EDTA], LiCl immune complex wash buffer [10 mM Tris-HCl (pH 8.0), 0.25 M LiCl, 1% Nonidet P-40, 1% sodium deoxycholate, and 1 mM EDTA] , and twice with Tris-EDTA. DNA-protein cross-links were eluted with elution buffer (0.1 M NaHCO3 and 1% SDS) at room temperature for 30 min. After adjusting NaCl concentration, cross-linking was reversed with overnight incubation at 65°C, followed by proteinase K treatment for 1 h at 45 °C. The immunoprecipitated DNA was recovered by a PCR purification kit (no. 28106; Qiagen) and analyzed either by real-time PCR with FastStart Universal SYBR Green (#04931850001; Roche) in an Stratagene MX3005P continuous fluorescence detector or by PCR in a 2% agarose gel. The sequence of the primers used in ChIP assays are provided as a supplementary excel file (Supplementary Table S3).

Cell Culture. Human PDAC cell lines were obtained from the MGH Center for Molecular Therapeutics, except for LZ10.7 cells which were a kind gift of Anirban Maitra (Johns Hopkins, Baltimore MD). All cell lines were cultured in DMEM supplemented with 10% (vol/vol) fetal

bovine serum, penicillin, and streptomycin Pancreatic ductal cells where isolated from 9-week-old *Ptf1α-Cre; LSL-Kras*^{G12D} mice as previously (13) and propagated in laminin (354239; BD Biosciences). Cells were infected with lentiviruses to overexpress KDM2B, expanded, and then subjected to fluorescence-activated cell sorting.

Retroviral and lentiviral mediated gene expression. Lentiviral constructs to overexpress KDM2B have been generated by subcloning the cDNA of KDM2B into the EcoRI sites of the pLVX-IRES-tdTomato (Clontech Cat# 631238) lentiviral vector. The retroviral constructs to overexpress wild-type and mutant forms of KDM2B have been previously described and were a kind gift from Dr. P. Tsichlis (Tufts Medical Center) (1, 14, 15). To avoid off-target effects of RNAi we validated at least two different hairpins per target gene based on the pLKO1 lentiviral system obtained through the RNAi Consortium (Broad Institute, http://www.broadinstitute.org/rnai/trc). The accession numbers of the hairpins used in this study are: human EZH2 (TRCN0000040076 and TRCN0000010475), human KDM5A (TRCN0000014629 and TRCN0000014632), human RING1B (TRCN0000033697), human BMI1 (TRCN0000020157), and human MYC (TRCN0000174055 and TRCN0000010390). The mouse hairpins against Kdm2b have been previously described (14). To knock down human KDM2B we used two different short hairpins that target either the Jumonji domain (shKDM2B#1TRCN0000118440/targets exon8) or the 3'UTR (shKDM2B#2/TRCN0000118437/targets 3'UTR). The human and mouse KDM2B loci encode multiple isoforms of KDM2B (1), and some of them lack the Jumonji domain. To distinguish between the KDM2B isoforms two different set of primers have been used: (a) KDM2B_LF (Long Form) amplifies KDM2B mRNA encoded by exons 5,6 and 7 and detects all the KDM2B isoforms that harbor the Jumonji domain (F-5'-actgtggtagacctggtggact-'3 and R-5'-gaagtcggtgaaacaacctttc-'3), and (b) KDM2B_LF/SF (Long Form/Short Form) amplifies KDM2B mRNA encoded by exons 24 and 25 and detects the majority of protein coding KDM2B isoforms (F-5'-tccttaaccgagatcaacctgt-'3 and R-5'-cagacatctcggctatgaactg-'3). In the RNA-seq experiments described herein we have used shKDM2B#1/TRCN0000118440 that exclusively targets the long form of KDM2B which harbors the Jumonji domain. Thus, those cells still express short isoforms of KDM2B that lack the Jumonji domain. The knock down efficiency of the long isoform by both shKDM2B#1 and shKDM2B#2 short hairpins is >80% as determined by qRT-PCR (Fig. S2A).

Quantitative real-time PCR. RNA was isolated with the RNeasy mini kit (#74104; Qiagen). cDNAs were synthesized with QuantiTect Reverse Transcription kit (#205310; Qiagen). PCR reactions were carried out in duplicate with FastStart Universal SYBR Green (#04913850001; Roche) in a Stratagene MX3005P continuous fluorescence detector.

Protein Extraction and Western Blotting. Cells were washed twice in ice-cold PBS and solubilized in lysis buffer [50 mM Tris (pH 7.5), 200 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.5% Na.Deoxycholate, 10 mM Na3VO4, 50 mM NaF, 1 mM b-glycerophosphate, 1 mM sodium pyrophosphate, 1 mM EDTA, 1 mM EGTA, and 1 mM PMSF supplemented with a mixture of protease inhibitors (#11836170001; Roche)]. The lysates were briefly sonicated and subsequently centrifuged for 20 min at 16,000x*g*. The supernatant (soluble whole-cell lysate) was analyzed by Western blotting with antibodies listed in the ChIP section.

Cell proliferation and viability assays. $4x10^4$ cells were plated in duplicate in twelve-well plates and passaged at a 1:2 or 1:3 ratio. Cells were counted with an automatic cell counter (Countess, Automated Cell Counter; Invitrogen) according to the manufacturer's instructions. The graphs show the cumulative cell number upon cell passaging for the indicated period of

time. Cell viability was evaluated by the trypan blue dye exclusion method and was quantified with an automatic cell counter (Countess, Automated Cell Counter; Invitrogen).

Histology and immunohistochemistry. Human pancreatic tumor tissues were obtained by V. Deshpande from the Massachusetts General Hospital tissue bank under approved protocols for the use of human specimens (authorized by the Office for the Protection of Research Subjects at the Dana Farber Cancer Institute (20 Overland Street Boston, MA 02115), and the Partners Healthcare Institutional Review Board IRB (116 Huntington Ave, Suite 1002, Boston MA, 02116). For immunohistochemistry, slides were deparaffinized in xylenes (two treatments, 6 min each), rehydrated sequentially in ethanol (5 min in 100%, 3 min in 95%, 3 min in 75%, and 3 min 40%), and washed for 5 min in 0.3% Triton X-100/PBS, and 3 min in water. For antigen unmasking, specimens were boiled for 20 min to 2 hours in 10 mM sodium citrate pH 6.0, rinsed three times with 0.3% Triton X-100/PBS, incubated for 10 min with 3% H_2O_2 at room temperature to block endogenous peroxidase activity, washed three times with 0.3% Triton X-100/PBS and blocked with 5% normal horse serum in 0.3% Triton X-100/PBS for one hour. To detect KDM2B we used a previous validated antibody (Ab64920; Abcam, (16), and we further confirmed its specificity in genetically engineered mouse models where KDM2B is overexpressed (manuscript in preparation). We also validated an additional goat anti-KDM2B antibody (sc-69477; Santa Cruz and) as shown in Fig. S1A. Primary antibodies were diluted 1:300 in blocking solution and incubated with the tissue sections at 4°C overnight. For KDM2B detection, specimens were washed 3 times for 3 min each with 0.3% Triton X-100/PBS and incubated with biotynilated anti-goat IgG antibody (1:200 dilution in blocking solution) for 1 hour at room temperature. Specimens were washed 3 times in 0.3% Triton X-100/PBS and treated with ABC reagent (Vectastain ABC kit #PK-6100) for 30 min, followed by three washes 3 min each. Finally, specimens were stained for peroxidase for 3 min with the DAB (Di-amine-benzidine) substrate kit (Vector Laboratories, cat no. SK-4100), washed with water and counterstained with hematoxylin. Slides were photographed with an Olympus DP72 microscope.

SUPPLEMENTAL REFERENCES

1. Pfau, R., Tzatsos, A., Kampranis, S.C., Serebrennikova, O.B., Bear, S.E., and Tsichlis, P.N. 2008. Members of a family of JmjC domain-containing oncoproteins immortalize embryonic fibroblasts via a JmjC domain-dependent process. *Proc Natl Acad Sci U S A* 105:1907-1912.
2. Collisson, E.A., Sadanandam, A., Olson, P., Gibb, W.J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G.E., Jakkula, L., et al. 2011. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* 17:500-503.
3. Badea, L., Herlea, V., Dima, S.O., Dumitrascu, T., and Popescu, I. 2008. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepatogastroenterology* 55:2016-2027.
4. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43-49.
5. Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., and Milos, P.M. 2009. Direct RNA sequencing. *Nature* 461:814-818.
6. Robinson, M.D., and Oshlack, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11:R25.

7. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140.
8. McCarthy, D.J., Chen, Y., and Smyth, G.K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*.
9. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545-15550.
10. Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
11. Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26:1351-1359.
12. Li, C., and Wong, W.H. 2001. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 98:31-36.
13. Schreiber, F.S., Deramaudt, T.B., Brunner, T.B., Boretti, M.I., Gooch, K.J., Stoffers, D.A., Bernhard, E.J., and Rustgi, A.K. 2004. Successful growth and characterization of mouse pancreatic ductal cells: functional properties of the Ki-RAS(G12V) oncogene. *Gastroenterology* 127:250-260.
14. Tzatsos, A., Paskaleva, P., Lymperi, S., Contino, G., Stoykova, S., Chen, Z., Wong, K.K., and Bardeesy, N. 2011. Lysine-specific demethylase 2B (KDM2B)-let-7-enhancer of zester homolog 2 (EZH2) pathway regulates cell cycle progression and senescence in primary cells. *J Biol Chem* 286:33061-33069.
15. Tzatsos, A., Pfau, R., Kampranis, S.C., and Tsichlis, P.N. 2009. Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus. *Proc Natl Acad Sci U S A* 106:2641-2646.
16. Ge, R., Wang, Z., Zeng, Q., Xu, X., and Olumi, A.F. 2011. F-box protein 10, an NF-kappaB-dependent anti-apoptotic protein, regulates TRAIL-induced apoptosis through modulating c-Fos/c-FLIP pathway. *Cell Death Differ* 18:1184-1195.