Am J Cancer Res 2018;8(12):2548-2563
www.ajcr.us /ISSN:2156-6976/ajcr0087876

Original Article

Cellular 5-hydroxylmethylcytosine content determines tumorigenic potential and prognosis of pancreatic ductal adenocarcinoma

Yi-Lng Chen[1,2], Chun-Mei Hu[1], Jeh-Ting Hsu[3], Chin-Chun Chang[1], Ting-Yu Huang[1], Pei-Hsun Chiang[1], Wei-Yi Chen[4], Yu-Ting Chang[5], Ming-Chu Chang[5], Yu-Wen Tien[6], Eva YHP Lee[7], Yung-Ming Jeng[8], Wen-Hwa Lee[1,9]

[1]Genomics Research Center, Academia Sinica, Taipei, Taiwan; [2]Ph.D. Program in Translational Medicine, National Taiwan University, Taipei, Taiwan; [3]Department of Information Management, Hsing Wu University, New Taipei City, Taiwan; [4]Institute of Biochemistry and Molecular Biology, National Yang Ming University, Taipei, Taiwan; [5]Department of Internal Medicine, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan; [6]Department of Surgery, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan; [7]Department of Biological Chemistry, University of California, Irvine, CA, USA; [8]Department of Pathology, National Taiwan University Hospital, Graduate Institute of Pathology, College of Medicine, National Taiwan University, Taipei, Taiwan; [9]Drug Development Center, China Medical University, Taichung, Taiwan

Received November 4, 2018; Accepted November 29, 2018; Epub December 1, 2018; Published December 15, 2018

Abstract: We stratified pancreatic ductal adenocarcinoma (PDAC) based on the tumorigenic properties of cancer cells, and aimed to identify clinically useful immunohistochemical (IHC) markers with mechanistic insights. The tumorigenic properties of PDACs were determined using patient-derived xenograft in NOD/SCID/IL2Rγ^{null} mice. The success of tumor engraftment was significantly correlated to poor survival, and its predictive values were superior to clinicopathological parameters. To search IHC-based biomarkers as surrogate for high tumorigenicity with prognostic values, 11 candidates of potentially clinical useful prognostic markers were selected. Among them, 5hmC content of the cancer cells was validated. Elevated 5hmC content positively correlated with *in vivo* tumorigenicity and poor prognosis in both primary and validation cohorts. Enrichment of cancer-associated 5hmC in CDX2 and FOXA1 lineage-specific transcriptional factor genes further pointed out the potential role of 5hmC in modulating cellular differentiation to enhance tumor malignancy during PDAC progression. Tumor-associated 5hmC content defined a subpopulation of PDAC with high lineage plasticity and tumorigenic potential, and was a prognostic IHC marker that provided a clinical basis for future management of PDAC.

Keywords: 5hmC, pancreatic cancer, tumorigenicity, prognostic biomarker, lineage plasticity

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is a devastating disease with a 5-year survival rate less than 5% [1, 2]. The mortality rate of pancreatic cancer is nearly equal to its incidence rate, in contrary to the steady increase in survival for most cancers. The only therapeutic option to achieve long-term survival is surgical resection for a minority (15 to 20%) of the patients with stage I or II resectable pancreatic cancer [1]. However, a great variation in disease prognosis exists among the patients of the same stage. A significant fraction of patients has early tumor recurrence and distant metastasis with limited lifespan even after a complete resection [1].

Precise decision of patient treatment is based on the understanding of molecular features of cancer, which are stratified by biomarkers to improve patient selection for treatment options. Recently, three comprehensive transcriptomic analysis of microarray and RNA-seq data from patient specimens revealed 4 PDAC subtypes with distinct transcriptional network and histological characteristics, including squamous (also termed quasimesenchymal or basal), pan-

Table 1. Comparison of clinical characteristics between engraftment-positive and -negative groups in the primary PDAC cohort

Characteristics	Overall n = 50 No.	Engraftment n = 20 No.	No engraftment n = 30 No.	*P*
Age				
Mean (year, SD)	67.4 (11.3)	68.2 (13.7)	66.9 (9.5)	
Range	32-88	32-88	50-86	
< 60	13	4	9	0.65
≥ 60	37	16	21	
Gender				
Male	28	15	13	0.06
Female	22	5	17	
AJCC stage				
IB	2	0	2	0.18
IIA	23	12	11	
IIB	25	8	17	
Tumor grade				
G1	11	2	9	0.19
G2-G3	37	17	20	
Unknown	2	1	1	
Primary tumor extent (T)				
T1	0	0	0	0.42
T2	6	1	5	
T3	44	19	25	
T4	0	0	0	
Lymph node status (N)				
N0	25	12	13	0.39
N1	25	8	17	
Distant metastasis (M)				
M0	50	20	30	0.2
M1	0	0	0	
Resection margin status (R)				
R0	35	12	23	0.49
R1	14	7	7	
Unknown	1	1	0	
Lymphovascular invasion				
Negative	22	8	14	0.86
Positive	28	12	16	
Perineural invasion				
Negative	4	1	3	0.96
Positive	45	18	27	
Unknown	1	1	0	

creatic progenitor (also termed classical), immunogenic, and aberrantly differentiated endocrine exocrine (also termed exocrine-like) subtypes [3-5]. Additionally, deep whole-genome sequencing of PDACs identified 4 subtypes according to frequency and distribution of chromosomal structural rearrangements, including stable, locally rearranged, scattered, and unstable [6]. These studies suggested that PDAC stratification was an important process to address subtype specific molecular mechanisms that could be exploited to therapeutic benefit.

Cancer cell intrinsic tumorigenic property is highly associated with PDAC malignancy [7]. Thus, we aimed to stratify PDACs according to this feature and its underlying molecular mechanism. Here, we used patient-derived xenograft transplantation in immune deficient mice to stratify PDAC according to their *in vivo* tumorigenic ability. We found that the engraftment-positive group had significant worse prognosis than engraftment-negative group. The results suggested that xeno-engraftment defines two subtypes of PDAC with distinct tumorigenic potential and patient outcome. Since immunohistochemical (IHC) staining was a well-documented and highly reproducible method to be translated into routine clinical settings [8], we intended to find a clinical useful IHC marker to represent the properties for efficient xeno-engraftment. Among many potential biomarkers, elevated 5hmC content of cancer cells was the best, which correlated positively with *in vivo* tumorigenicity and poor prognosis in both primary and validation cohorts.

5hmC is a derivative of 5mC oxidation. Rather than a transient intermediate of DNA demethylation, 5hmC acts as a distinct epigenetic mark for transcriptional regulation of normal differentiation [9]. Global loss and redistribution of 5hmC is a common feature in various types of

cancers including pancreatic cancer [10]. We identified that cancer-associated 5hmC was enriched in CDX2 and FOXA1 lineage-specific transcriptional factor genes and pointed out the potential role of 5hmC in modulating cellular differentiation to enhance tumor malignancy during PDAC progression.

Materials and methods

Patient information in the primary prospective cohort study

Between Nov 2011 and Dec 2015, pancreatic cancer specimens were collected from 71 patients who underwent pancreaticoduodenectomy in National Taiwan University Hospital (NTUH). The diagnosis of the 71 patients were PDAC (n = 50, 70%), ampullary adenocarcinoma (n = 14, 20%), pancreatic neuroendocrine tumor (PNET) (n = 2, 3%), intraductal papillary mucinous neoplasms (IPMNs) (n = 3, 4%), intraductal papillary mucinous carcinoma (IPMC) (n = 1, 1%), and chronic sclerosing pancreatitis (n = 1, 1%). Based on the 7th edition of the American Joint Committee on Cancer (AJCC) criteria, 50 PDAC patients were staged as IB (n = 2), IIA (n = 23), and IIB (n = 25) (**Table 1**). 4 PDAC patients without follow-up information were excluded from the survival analysis. The study was approved by the Research Ethnic Committee of NTUH (approval number 201303029RINC and 201411085RINB).

Retrospective cohort study for validation of the prognostic target

For the independent validation of the IHC marker, an independent retrospective cohort of 73 patients who received surgical resection in NTUH from 2007 to 2011 was use. The pancreatic cancer specimens were also collected from patients who underwent pancreaticoduodenectomy, and had pathologically confirmed PDAC. There were 2, 5, 19, and 44 patients of stages IA, IB, IIA, and IIB, respectively, by the AJCC staging system (3 patients with unknown AJCC status). Clinical characteristics of patients in the validation cohort were listed in Supplementary Table 1.

Xeno-transplantation

Immediately after surgery, tumor fragments (F0) with a volume about 1000 mm^3 were transfer to Hank's Balanced Salt Solution (HBSS) and stored at 4°C for transportation to the lab. Next, the tumor specimens were kept in RPMI 1640 medium and cut into pieces of 1 mm^3 for subcutaneous and orthotopic implantation. Each patient specimen was delivered to 4 mice for subcutaneous (2 mice) and orthotopic (2 mice) implantation. Engraftment was done in 2~3 hours after resection. To evaluate the success of engraftment, tumor growth was measured once a week. The cases were determined no engraftment if there was no palpable tumor one year after xeno-transplantation. When the tumor grew to 1000 mm^3 in size, the mice were euthanized for collecting the F1 tumor. A small part of the F1 tumor was cut into pieces of 1 mm^3 for re-implantation to passage on NOD/SCID/IL2Rγ^{null} (NSG) mice. Most of the F1 tumors were cut into pieces of less than 1 mm^3 and underwent enzymatic digestion using collagenase P (1 mg/ml; Roche 11213865001), dispase (1 mg/ml; Roche 11097113001), and soybean trypsin inhibitor (0.1 mg/ml; Gibco 17075-029) in RPMI 1640 medium (supplemented with 10% fetal calf serum and 10 μg/ml antibiotics/antimycotics) at 37°C for 1 hour with intermittent pipetting every 30 minutes. After incubation, the dissociated sample was neutralized, filtered (100 μm pore size), and washed with culture medium before plating. The primary cells would be trypsinized and cryopreserved 24 to 48 hours after plating. Animal studies were approved by the Institutional Animal Care and Use Committee of the Academia Sinica, Taipei, Taiwan (Protocol #14-05-709).

Immunohistochemical (IHC) staining

The formalin-fixed, paraffin-embedded whole tissue sections (4 μm thick) were processed for IHC studies. Immunohistochemistry for each antigen was performed on an automated staining system (Ventana Benchmark LT, Ventana Medical Systems; Bond-max, Leica Biosystems) or manually. For Ventana Benchmark LT staining system, iView DAB detection kit (Catalog Number: 760-091; Ventana Medical Systems) was used. Briefly, tissue sections were dewaxed and rehydrated using the EZ Prep solution. Antigen retrieval was performed by incubating slides in Cell Conditioner 1 (pH 8.5) for 40 minutes (mild condition) or 100 minutes (standard condition). After being blocked with 3% H_2O_2,

the slides were incubated with primary antibodies, and then incubated with a biotinylated secondary antibody. The peroxidase activity was visualized using a diamino-benzidine tetrahydroxychloride solution. The sections were counter-stained with hematoxylin. For Bond-max staining system, Bond Polymer Refine Detection kit (DS9800) was used. Tissue sections were deparaffinized and treated with heat-induced epitope retrieval buffer (pH 9) for 40 minutes (Epitope Retrieval Solution 2 AR9640). Next, peroxidase blocking was carried out for 10 minutes following by primary antibody incubation. Subsequently, tissue sections were incubated with polymer and DAB-chromogen for development, and finally counter-stained with hematoxylin. Manual staining was used in a few primary antibodies if the automated machine was not proper to develop clear staining results. After deparaffinization and rehydration, antigen retrieval was performed by placing slides in pH 6 (target retrieval solution S1699; Dako) or pH 10 (Antigen retrieval AR-10 solution HK059-5KE; BioGenex) solutions and heating slides (autoclave) at 100°C for 10 minutes. Slides were cooled to room temperature and washed in PBS. Sections were blocked by incubation in 3% H_2O_2 for 10 minutes and 10% FBS for 60 minutes at room temperature, and then incubated overnight with primary antibody at 4°C in a humidified chamber. For the negative controls, the primary antibodies were replaced with normal mouse IgG and normal rabbit IgG. The information for the primary antibodies and individual staining protocols was listed in Supplementary Table 2.

IHC assessment

IHC staining of IL-17RB (membrane staining), Stathmin1 (cytoplasmic staining), and 5hmC (nuclear staining) in tumor regions were visually scored by a trained pathologist. The IHC scoring range for 5hmC, IL-17RB, and Stathmin1 is 0 to 3+, including: (1) a score of 0: staining undetectable in tumor cells, (2) a score of 1+: staining expressed in < 10% tumor cells, (3) a score of 2+: staining expressed in 10 to 50% tumor cells, and (4) a score of 3+: staining expressed in ≥ 50% tumor cells. The Ki-67 indexes in tumor regions were scored by Aperio ImageScope (Aperio, USA) using IHC nuclear image analysis algorithm. For each section, at least 10 randomly selected 400 × high-power field cancer regions were used, and the nucleus with positive staining/total nucleus in cancer cells were calculated. The IHC scoring ranges included 0 to 10%, 10 to 20%, 20 to 30%, and ≥ 30%. For statistical analyses, the score of 0 and 1+ (or 0 to 10% and 10 to 20%) was combined as the low expression group, and the score of 2+ and 3+ (or 20 to 30% and ≥ 30%) was combined as the high expression group. The association between 5hmC and TET1 expression was measured in 119 PDAC patient specimens collected at NTUH (2007-2015) in the primary and validation cohorts (5hmC 0: n = 12; 5hmC 1+: n = 37; 5hmC 2+: n = 50; 5hmC 3+: n = 20). For each case, at least 10 randomly selected 400 × high-power field cancer-cell only regions were outlined by pen tools and analyzed by nuclear algorithm of Aperio ImageScope software (Aperio, USA) to calculate the average percentage of 1+, 2+, and 3+ cells with 5hmC or TET1 expression. The association between 5hmC and CDX2 (or FOXA1) expression was measured in 10 PDAC patient specimens collected at NTUH (2011-2015) in the primary cohort (5hmC 0: n = 1; 5hmC 1+: n = 5; 5hmC 2+: n = 1; 5hmC 3+: n = 3). For each case, 4 randomly selected 400 × high-power field cancer-cell only regions with their corresponding regions in serial sections were outlined by pen tools and analyzed by nuclear algorithm of Aperio ImageScope software (Aperio, USA) to calculate the average percentage of 1+, 2+, and 3+ cells with 5hmC, CDX2, or FOXA1 expression. Regions of normal pancreatic duct, low grade PanIN, high grade PanIN, and PDAC were outlined by pen tools to analyze 5hmC expression (1+, 2+, and 3+) by nuclear algorithm of Aperio ImageScope software (Aperio, USA). H-score was assigned by [1 × (% of 1+ cells) + 2 × (% of 2+ cells) + 3 × (% of 3+ cells)].

Establishment of primary cultured pancreatic cancer cell lines

Since xenograft was a good source to increase the success rate and frequently used to generate pancreatic tumor cell lines [11], we dissociated and digested the xenografts for generating primary cell culture to perform functional studies.

TET1 knockdown experiments

To knockdown TET1 expression in patient-derived pancreatic cancer cells, the lentiviral vector carrying TET1 specific shRNA was

obtained from National RNAi Core Facility (Academia Sinica, Taiwan). The target sequences were 5'-GCAGCTAATGAAGGTCCAGAA-3' and 5'-CCTCCAGTCTTAATAAGGTTA-3' (clone #2) for human TET1. The negative control was shRNA against β-galactosidase (shLacZ). HEK-293T cells were co-transfected with shRNA containing lentiviral vector, packaging plasmid psPAX2, and envelope plasmid pMD2G. Virus-containing supernatant was collected 48 hours post transfection. Patient-derived PC080 and PC084 cells were infected with lentivirus and then selected with 1~2 μg/ml puromycin.

Soft agar colony formation assay

In a 12-well plate, cells (10^4 per well) were seeded in the culture medium containing 0.35% agar on top of a layer of the culture medium containing 0.5% agar. Cells were maintained in a humidified 37°C incubator for 14 days. The colonies were fixed with ethanol containing 0.05% crystal violet for quantification.

Dot blot

PC080/PC084 shLacZ and shTET1 cells were plated in 10 cm culture dish. After 3 days for the monolayer to reach 70~80% confluency, genomic DNA was collected from the trypsinized cells and isolated by the genomic DNA extraction kit (QIAGEN, #51306). Genomic DNA was denatured by 0.4 N NaOH and 0.01 M EDTA at 95°C for 10 min, and spotted onto Zeta probe blotting membrane (Bio-rad) with Bio-Dot microfiltration apparatus (Bio-rad). The membrane was washed with 2X Saline-sodium citrate (SSC) solution, air-dried, and UV-crosslinked (5 mJ/cm^2). The amount of each sample was measured by staining with 0.02% methylene blue in 0.3 M sodium acetate (pH 5.2). Blocking was performed in 5% skimmed milk in PBS-T (Tris-buffered saline containing 0.1% Tween 20) for 1 hr at room temperature. 5mC antibody (Eurogentec BI-MECY-0100; 1:500 dilution) or 5hmC antibody (Active motif #39770; 1:1000 dilution) was incubated overnight at 4°C. The membrane was then washed and probed with HRP-conjugated IgG secondary antibody for 1 hr at room temperature, and visualized with ECL reagent.

Orthotopic injection

NSG mice were anesthetized by using isoflurane. Local shaving and disinfection were performed around the left upper quadrant, and the abdominal cavity was opened by a 0.5~1 cm longitudinal incision. The spleen was lifted and the pancreatic tail was identified. 10^5 PC084 shLacZ/shTET1 and 2 × 10^5 PC080 shLacZ/shTET1 cells in pre-diluted Matrigel (10 μl Matrigel + 10 μl PBS) were then slowly delivered to the pancreas using a 1 ml insulin syringe with 29 G needle. The pancreas was placed back, the muscle layer was closed with 6-0 absorbable surgical suture, and the skin was closed with skin staples. NSAID drug meloxicam (25 μg per mouse) subcutaneous injection was used for postoperative analgesia.

Bioinformatics analysis of ChIP-seq data

The ChIP-seq data were obtained from GEO accession number GSE71839 (Bhattacharyya *et al.*, 2016). The ChIP-seq data were aligned with the human genome (hg19) using BWA aligner [12] with default parameters. Peak calling of enriched binding regions was performed using MACS [13] with default parameters. Relative enriched peaks around gene regions analyzed by PAVIS (Peak Annotation and Visualization) tool [14] in at least 6 xenografts, and not observed in normal cells were included as cancer-associated 5hmC loci (gene regions from -5000 upstream regions of the Transcription Start Site to +1000 downstream regions of Transcription Termination Site, P value ≤ 10^{-90}, fold enrichment ≥ 5). The results were visualized with IGV software [15].

Statistical analysis

To analyze the effect of tumor engraftment and clinicopathological factors on prognosis of pancreatic cancer patients, Kaplan-Meier curves were plotted and log rank tests were used to evaluate which factor was a prognostic indicator for progression-free survival (time from surgical resection to local recurrent or distant metastasis) and overall survival (time from surgical resection to death). Univariate and multivariate Cox regression survival analyses were performed to adjust the association between survival and 5hmC expression levels for varying clinical parameters, including tumor grade, primary tumor extent (T), lymph node status (N), resection margin status (R), perineural invasion, and lymphovascular invasion. Chi-square analysis was performed to determine the association between tumor engraftment and the

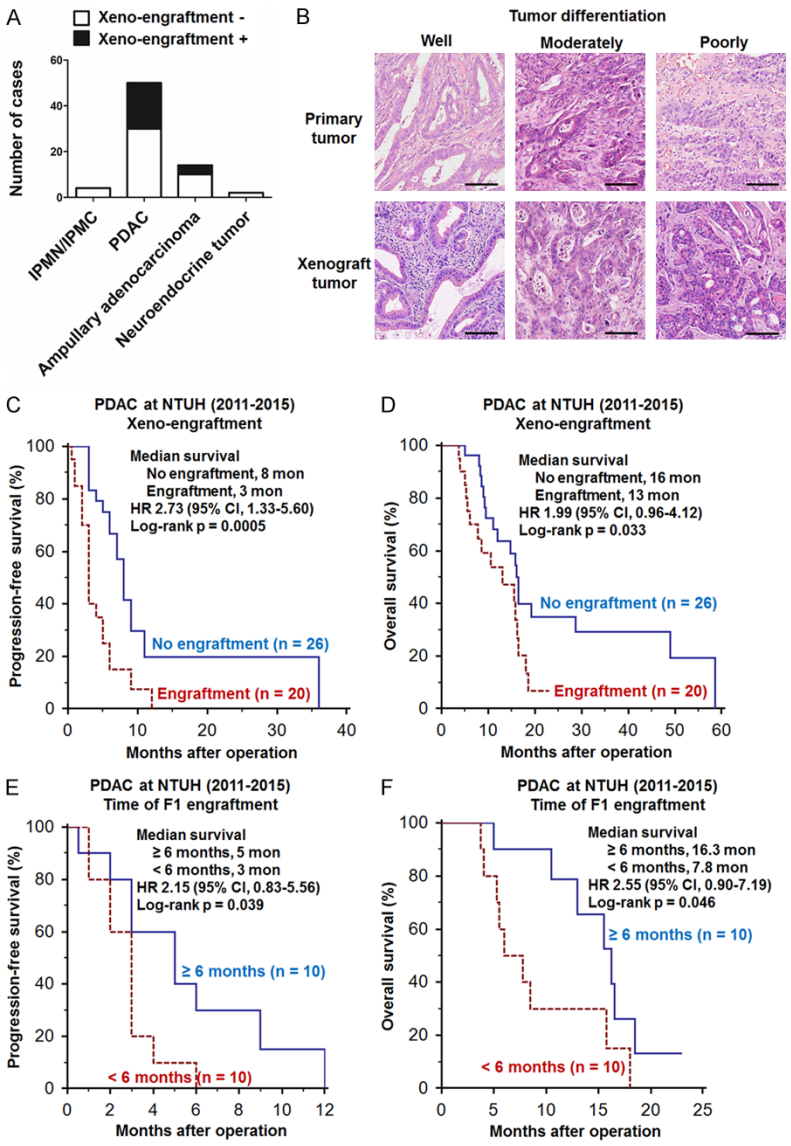


Figure 1. The success of engraftment is associated with poor clinical outcome. A. The engraftment case number of the pancreaticoduodenectomy specimens in NOD/SCID/IL2Rγ^{null} mice. White column: number of cases failed to form xenografts; Black column: number of cases succeeded in forming xenografts. B. Represented images of hematoxylin and eosin (H&E) stains for primary and xenograft tumors with three main histologic grades. Scale bar, 50 μm. C. Kaplan-Meier analysis of progression-free survival according to engraftment status (P = 0.0005, Log-rank test). D. Kaplan-Meier analysis of overall survival according to engraftment status (P = 0.033, Log-rank test). E. Kaplan-Meier analysis of progression-free survival according to time of F1 engraftment (*P* = 0.039, Log-rank test). F. Kaplan-Meier analysis of overall survival according to time of F1 engraftment (*P* = 0.046, Log-rank test).

expression levels of IHC biomarkers, and the correlation between 5hmC content and clinicopathological parameters. Student's *t*-test and Pearson's correlation coefficient were used for testing the difference and correlation between variables. The statistical analysis was performed using MedCalc 11.5.1.0 (MedCalc software, Belgium) and Prism 5 (GraphPad Software, USA).

Results

The success of engraftment is associated with poor clinical outcome

Previous studies found that early passages of xenograft tissues closely mirror the morphology and biology of primary human tumor [16]. Thus, we used xenograft model as a platform to differentiate tumorigenic ability. From 2011 to 2015, 71 patients who underwent pancreaticoduodenectomy at National Taiwan University Hospital were included in the study. Of these 71 xeno-transplantations, 24 led to the establishment of engraftment in NOD/SCID/IL2Rγ^{null} (NSG) mice (F1). PDAC specimens had a higher engraftment rate (20/50 = 40%) than ampullary adenocarcinoma (4/14 = 28.6%). The cases of pancreatic neuroendocrine tumor (PNET), intraductal papillary mucinous neoplasm (IPMN), and intraductal papillary mucinous carcinoma (IPMC) hadn't successfully engrafted, suggesting the tumor engraftment rate was dependent on cancer type (**Figure 1A**). Tumor grafts preserved the histopathologic characteristics of the original tumor from which they were derived (**Figure 1B**).

The clinical characteristics of the 50 patients with resected PDAC were described in **Table 1**. There were Stage IB (n = 2), Stage IIA (n = 23), and Stage IIB (n = 25) cases, according to the seventh edition of the AJCC criteria. The histopathological grades of most cases were moder-

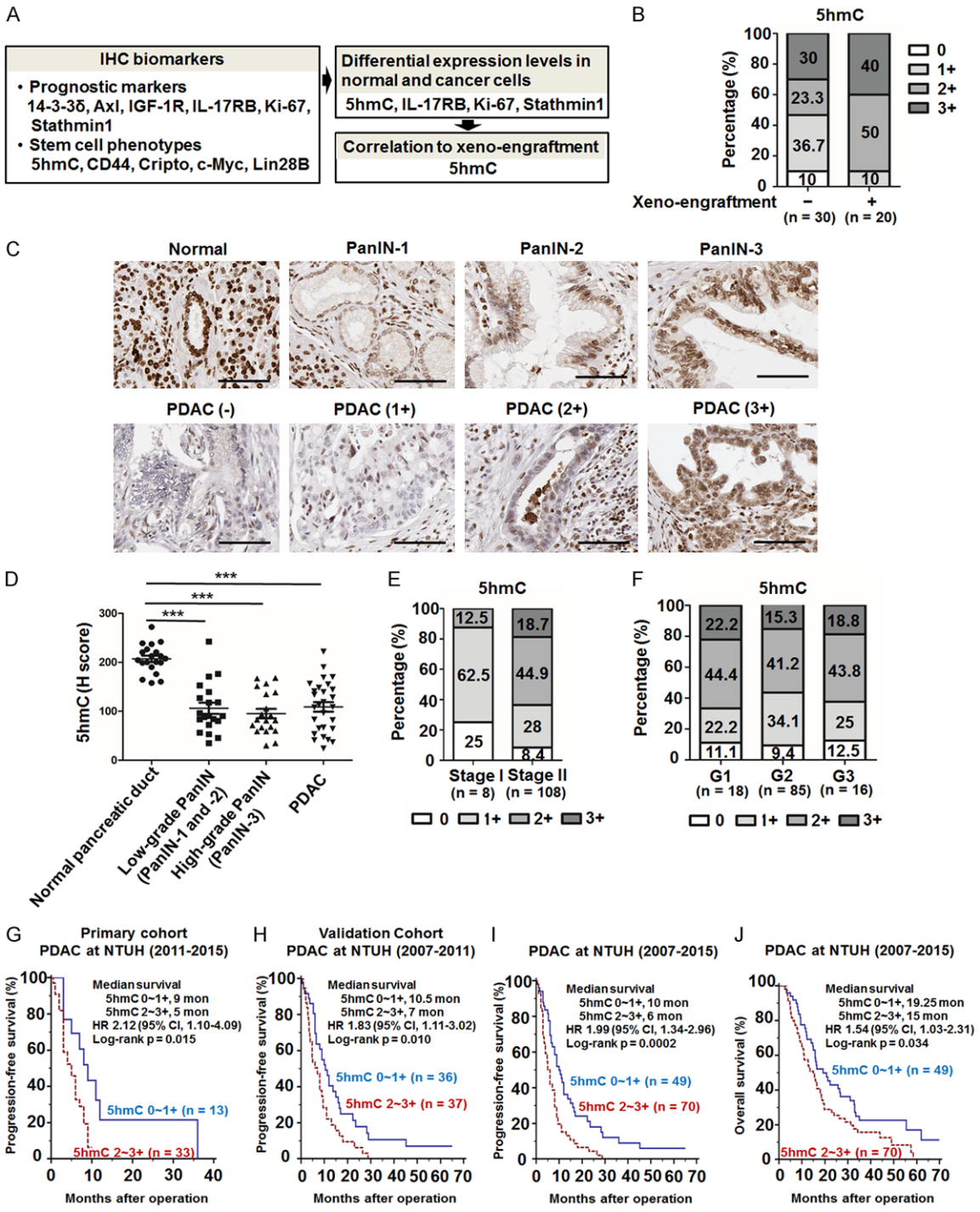


Figure 2. Cellular 5hmC content serves as a biomarker for tumorigenic potential and predicted prognosis in PDAC patients after curative resection. A. Algorithm to decipher candidate immunohistochemical (IHC) biomarkers as surrogates for the *in vivo* tumorigenic activity. IHC biomarkers with prognostic values for PDAC and related to stem cell phenotypes were included for further verification. Biomarkers with differential expression levels in cancer and normal cells were selected. Correlations between expression levels of biomarkers and engraftment status were examined. B. Comparisons of 5hmC expression levels in cancer regions of xeno-engraftment negative and positive patient specimens. C. Representative IHC images of 5hmC expression in normal pancreatic tissue, pancreatic intraepithelial neoplasia (PanIN), and PDAC. Scale bar, 100 μm. 5hmC expression levels were scored by 0 (no staining), 1+ (< 10% among cancer cells), 2+ (10 to 50% among cancer cells, and 3+ (≥ 50% among cancer cells). For statistical analyses, the score of 0 and 1+ was combined as the low expression group, and the score of 2+ and 3+ was combined as the high expression group. D. The expression levels of 5hmC in normal pancreatic duct (n = 20),

low-grade PanIN (n = 20), high-grade PanIN (n = 20), and PDAC (n = 20). H-score was assigned by [1 × (% of 1+ cells) + 2 × (% of 2+ cells) + 3 × (% of 3+ cells)]. E. Comparisons of 5hmC expression levels in cancer regions of stage I and stage II (AJCC 7th edition staging system) PDAC specimens including primary and validation cohorts. F. Comparisons of 5hmC expression levels in cancer regions of G1 (well differentiated), G2 (moderately differentiated), and G3 (poorly differentiated) PDAC specimens including primary and validation cohorts. G. Kaplan-Meier analysis of progression-free survival according to 5hmC expression levels in the primary cohort (P = 0.015, Log-rank test). H. Kaplan-Meier analysis of progression-free survival according to 5hmC expression levels in the validation cohort (P = 0.010, Log-rank test). I. Kaplan-Meier analysis of progression-free survival according to 5hmC expression levels in cohorts including primary and validation groups (P = 0.0002, Log-rank test). J. Kaplan-Meier analysis of overall survival according to 5hmC expression levels in cohorts including primary and validation groups (P = 0.034, Log-rank test).

Table 2. Correlations between the expression profiles of IHC biomarkers and engraftment status

Biomarker	Engraftment + N = 20, N (%)	Engraftment - N = 30, N (%)	Correlation to engraftment (Chi-squared test)
5hmC			
Low	2 (10)	15 (50)	P = 0.009
High	18 (90)	15 (50)	
Stathmin1			
Low	3 (15)	13 (43.3)	P = 0.073
High	17 (85)	17 (56.7)	
IL-17RB			
Low	9 (45)	22 (73.3)	P = 0.085
High	11 (55)	8 (26.7)	
Ki-67			
Low	9 (45)	20 (66.7)	P = 0.219
High	11 (55)	10 (33.3)	

ately to poorly differentiated (G2 to G3; 37/48 = 77%). For survival analysis, 46 patients were involved and 4 patients were excluded due to their lack of follow-up information. Kaplan-Meier plot demonstrated a significant difference in progression-free survival (8 vs. 3 months, P = 0.0005) and overall survival (16 vs. 13 months, P = 0.033) between patients whose tumors led to xenograft success and those without (**Figure 1C** and **1D**). Additionally, the growth rate of xenografts also provided a significant prediction of patient prognosis. Patients with F1 tumors taking less than 6 months to reach a tumor volume of 1000 mm^3 had inferior survival than patients with F1 tumor taking 6 months or longer to grow (**Figure 1E** and **1F**). Other clinicopathological parameters including age, gender, AJCC stage, tumor grade, primary tumor extent (T), lymph node status (N), distant metastasis (M), resection margin status (R), lymphovascular invasion, and perineural invasion had similar distribution between tumor engraftment-positive and -negative groups (**Table 1**). Thus far, there is no available clinicopathological parameter representing high tumorigenicity and prognostic values of the engraftment group (Supplementary Figure 1).

Cellular 5hmC content serves as a surrogate marker for tumorigenic potential of PDAC

Considering clinical application, we screened biomarkers that could be analyzed by IHC to serve as surrogates for the accurate, but not practical, tumorigenicity assay due to time issue. Since previous studies reported that the engraftment-related genes were associated with shorter disease-free survival [17], we focused on IHC prognostic markers for PDAC (14-3-3δ, Axl, IGF-1R, IL-17RB, Ki-67, and Stathmin1) [18-24]. It was noted that xeno-engraftment is a gold standard to identify cancer cells with stem cell-like properties [25, 26]. Thus, IHC biomarkers for stem cell phenotypes (5hmC, CD44, Cripto, c-Myc, and Lin28B) [9, 27-30] were also included (**Figure 2A**).

Staining procedures with the individual antibodies were listed in Supplementary Table 2. We excluded biomarkers without differential expression levels between cancer and normal acinar/ductal cells (Supplementary Figure 2A). 5hmC, Stathmin1, IL-17RB, and Ki-67 showed distinctively staining results (**Figure 2C** and Supplementary Figure 2B), and their expression levels in engraftment-positive and -negative groups were compared. Of the four biomarkers, high 5hmC expression level in the cancer regions was most significantly correlated with xenograft formation (P = 0.009) (**Table 2** and **Figure 2B**). Thus, 5hmC content in the cancer regions was identified as the closest surrogate biomarker for tumorigenic ability in xenograft to stratify PDAC.

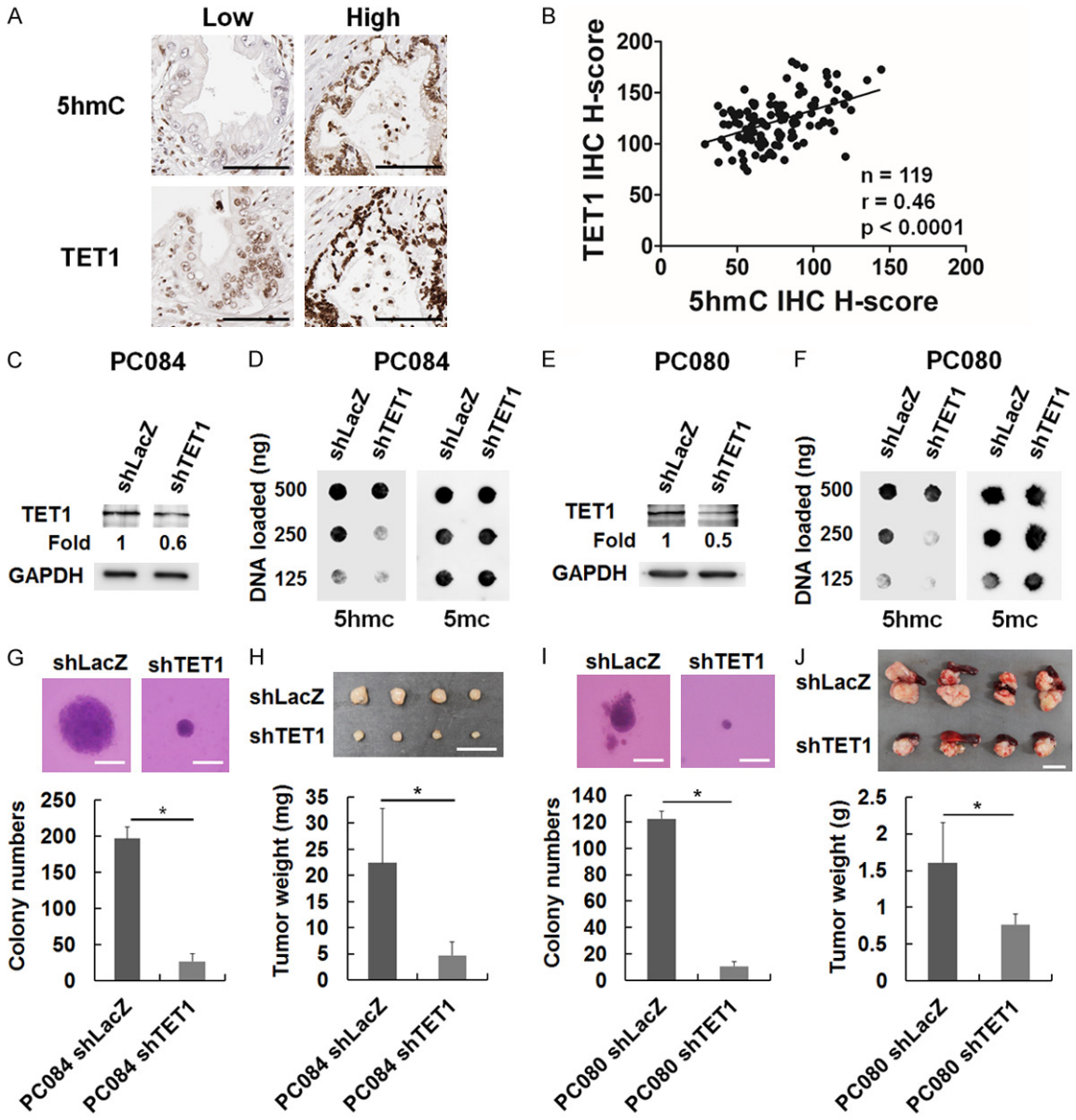


Figure 3. High 5hmC content of cancer cells plays roles in promoting tumorigenesis. A. Representative serial section images of 5hmC and TET1 IHC staining in PDAC specimens with low and high 5hmC contents. Scale bar, 100 μm. B. Pearson's correlation coefficient analysis of TET1 and 5hmC according to H-score (n = 119, r = 0.46, P < 0.0001). C and E. Western blotting analysis of proteins harvested from patient-derived cells (generated from PC084 and PC080 xenografts) stably expressing lentiviral-based $LacZ^{shRNA}$ or $TET1^{shRNA}$. Antibodies used were anti-TET1 (GTX124207, Genetex) and anti-GAPDH (GTX627408, Genetex). D and F. DNA dot blot analysis of genomic 5hmC and 5mC of PC084 and PC080 cells stably expressing lentiviral-based $LacZ^{shRNA}$ or $TET1^{shRNA}$. G and I. Soft agar colony formation assays of PC084 and PC080 cells stably expressing lentiviral-based $LacZ^{shRNA}$ or $TET1^{shRNA}$. Top panel: representative images of colony size. Bottom panel: colony number. Colonies with diameter ≥ 50 μm were counted. Scale bar, 50 μm. Values were presented as mean ± SD of four independent experiments. *, P < 0.05 (two-tailed t-test). H and J. Tumorigenic assay of PC084 and PC080 cells stably expressing lentiviral-based $LacZ^{shRNA}$ or $TET1^{shRNA}$. Top panel: images of tumor size. Scale bar, 1 cm. Bottom panel, quantification of tumor weight. Values were presented as mean ± SD of four independent experiments. *, P < 0.05 (two-tailed t-test).

Previous studies reported that global 5hmC was reduced in cancer cells comparing with adjacent normal cells [31]. Our immunohistochemical staining results also showed strong intensity of 5hmC in the nucleus of normal acinar/ductal cells, and significantly decreased in

PanIN and PDAC lesions (**Figure 2C** and **2D**). Although the intensity of 5hmC expression level in cancer regions was generally decreased compared to normal cells, the percentages of 5hmC positive cancer cells were in a broad range (**Figure 2C** and **2D**). 5hmC levels in cancer cells were significantly associated with tumor stage and gender (**Figure 2E** and Supplementary Table 3), but not other clinicopathological parameters including tumor grade (**Figure 2F** and Supplementary Table 3).

5hmC content in cancer cells predicts prognosis in PDAC patients after curative resection

To test whether 5hmC content is correlated with patient prognosis, we performed a Kaplan-Meier survival analysis and found that high 5hmC content (5hmC 2~3+) was significantly correlated with shorter progression-free survival (P = 0.015) compared to low 5hmC content (5hmC 0~1+) (**Figure 2G**). In consistent with our previous results, 5hmC content is the only biomarker among the four potential markers for the progression-free survival. The rest three had no significant P values (Supplementary Table 4). To further validate the clinical application of 5hmC as a prognostic marker in an independent cohort, we collected another 73 patients diagnosed at NTUH from 2007 to 2011. The basic characteristics of the patients were listed in Supplementary Table 1. Kaplan-Meier progression-free survival analysis found that high 5hmC expression (2 to 3+) was correlated to worse prognosis (P = 0.010) (**Figure 2H**). Combining the primary and validate cohort studies yielded a more precise estimate of the correlation between high 5hmC content and inferior progression-free survival (P = 0.0002) (**Figure 2I**). Importantly, high 5hmC content also predicted shorter overall survival in PDAC patients (P = 0.034) (**Figure 2J**). Cox multivariate analyses of progression-free survival and overall survival in total patients demonstrated that 5hmC was an independent prognostic factor (P = 0.0013), even superior to tumor grade, lymph node status, and resection margin status (Supplementary Table 5).

High 5hmC content-mediated by TET1 in patient-derived cancer cells promotes tumorigenesis

It was reported that TET1, but not TET2 or TET3, play a dominant role in 5hmC formation [32]. In addition, a significant positive correlation (r = 0.46, P < 0.0001) was observed between TET1 and 5hmC expression levels in PDAC patient specimens (**Figure 3A** and **3B**). To further validate whether high 5hmC content directly involved in tumorigenesis, we performed TET1 knockdown by two independent TET1 shRNAs to reduce 5hmC in patient-derived cancer cells (PC084 and PC080) with high levels of 5hmC (**Figure 3C**, **3E**, Supplementary Figures 3 and 5). TET1-depleted PC084 and PC080 cancer cells contained lower levels of 5hmC (**Figure 3D** and **3F**) and had diminished cell colony formation activity *in vitro* (**Figure 3G** and **3I**). Furthermore, we performed *in vivo* tumorigenesis assay by injecting PC084 and PC080 cells with high or low levels of 5hmC orthotopically into NSG mice, and found that cancer cells with lower 5hmC content had reduced ability to grow in NSG mice (**Figure 3H** and **3J**). These data suggested that high 5hmC content in cancer cells was required for tumor growth.

Transcriptional activations of CDX2 and FOXA1 correlate with gene specific enrichment of 5hmC modification in PDAC

To identify cancer-associated 5hmC target genes, we analyzed chromatin immunoprecipitation sequencing (ChIP-seq) data of 5hmC-enriched gene loci in 11 primary patient-derived xenografts and immortalized human pancreatic ductal epithelial cells (HPDE, HPNE) from Gene Expression Omnibus (GEO) database (GSE71839). Due to the importance of transcriptional reprogramming during cell state transitions, about 175 genes with characteristic of 5hmC-enriched transcription factor (peaks ranged from -5000 of Transcription Start Site to +1000 of Transcription Termination Site; P value ≤ 10^{-90}; fold enrichment ≥ 5) in at least 6 xenografts, and not observed in normal cells, were collected for pathway analysis by Ingenuity Pathway Analysis (IPA) software (**Figure 4A**). The top ranked pathways of cancer-associated hydroxymethylated transcription factors were involved in the regulation of embryonic stem cell pluripotency and differentiation, and the genes listed in these pathways included CDX2, FOXA1, GATA4, HNF4A, HOXB1, POU5F1, and PHC3 (**Figure 4B**). Genome browser view of cancer-associated 5hmC peaks revealed that 5hmC was enriched at gene body regions of CDX2, FOXA1, HNF4A, HOXB1, GATA4, and PHC3 (**Figure 4C** and Supplementary Figure

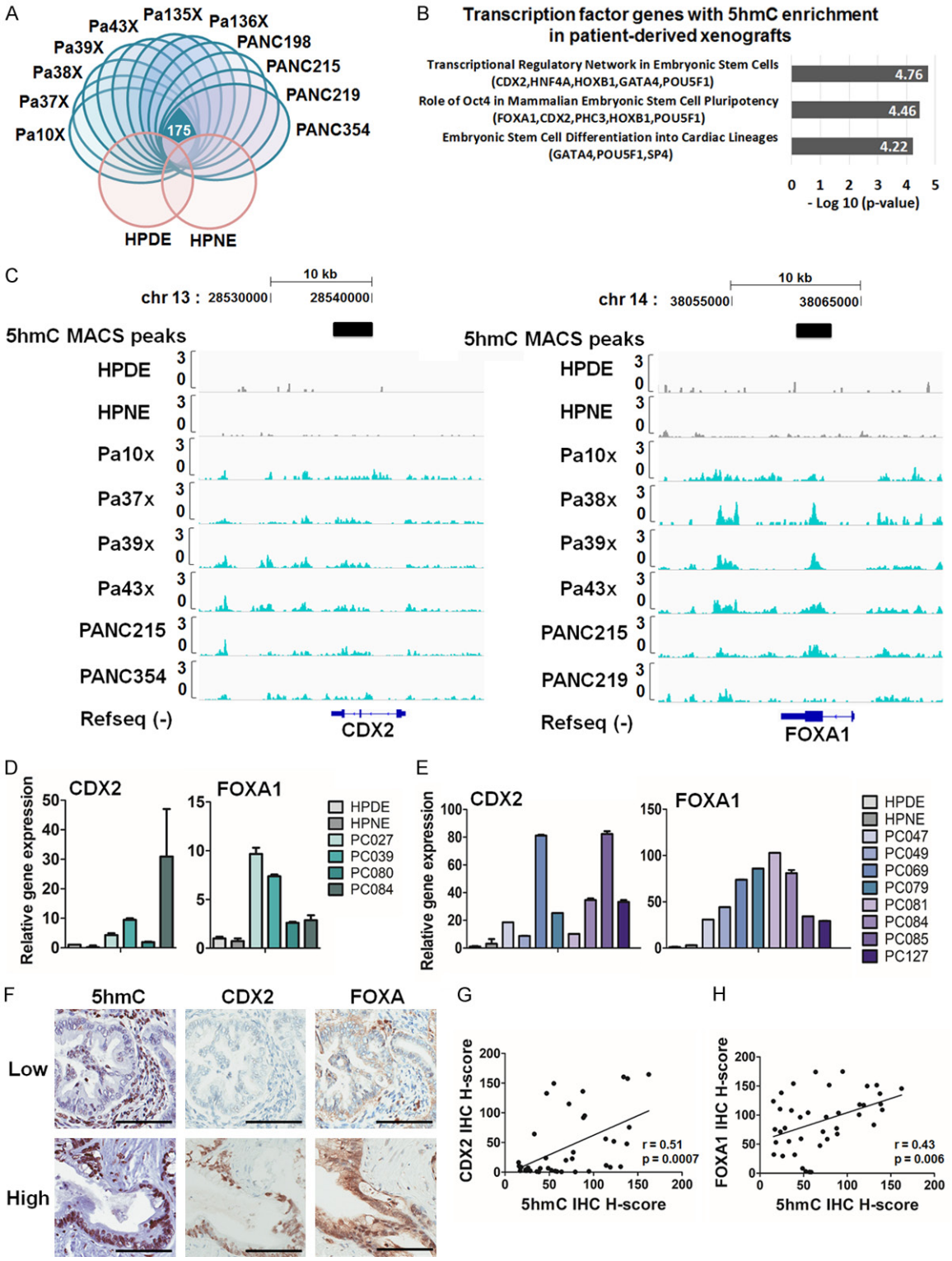


Figure 4. Transcriptional activations of CDX2 and FOXA1 correlate with gene-specific enrichment of 5hmC modification in PDAC tumor cells. A. Identification of transcription factor genes contained 5hmC around their gene regions (from -5000 upstream regions of the Transcription Start Site to +1000 downstream regions of Transcription Termination Site) by analyzing the ChIP-seq data from Gene Expression Omnibus (GEO) database (GSE71839). Transcription factors with their gene loci marked by 5hmC peaks (*P* value ≤ 10^{-90}, fold enrichment ≥ 5) in at least 6 xenografts (among Pa10x, Pa37x, Pa38x, Pa39x, Pa43x, Pa135x, Pa136x, PANC198, PANC215, PANC219, and PANC354), but not observed in immortalized human pancreatic ductal epithelial cells (HPDE and HPNE) were included. Total 175

transcription factor genes were found to have cancer-associated 5hmC enrichment around their gene regions. B. These 175 genes were categorized by pathway analysis using Ingenuity Pathway Analysis (IPA) software. C. Gene tracks of 5hmC occupancy at CDX2 and FOXA1 genes in normal pancreatic cells (HPDE and HPNE) and patient-derived xenografts (Pa10x, Pa37x, Pa38x, Pa39x, Pa43x, PANC215, PANC219, and PANC354). The 5hmc profiling data was obtained from GEO database (GSE71839). Model-based Analysis of ChIP-Seq (MACS) identified 5hmc peaks were denoted on the top of each panel. 5hmC MACS peaks on genes were visualized by Integrative genomics viewer (IGV) normalized with coverage reads. D. Quantitative real-time PCR analyses of CDX2 and FOXA1 expression were performed to compare gene expression levels in patient-derived cancer cells at an early passage number (PC027, PC039, PC080 and PC084) and immortalized human pancreatic ductal epithelial cells (HPDE and HPNE). Data were representative of three independent experiments and values were presented as mean ± SD (n = 3). E. CDX2 and FOXA1 were highly expressed in patient-derived xenografts compared to HPDE and HPNE cells by quantitative real-time PCR analysis. Data were representative of three independent experiments and values were presented as mean ± SD (n = 3). F. Representative IHC images of 5hmC, CDX2, and FOXA1 in PDAC specimens with 5hmC low and high contents. Scale bar, 100 μm. G and H. Pearson's correlation coefficient analysis of 5hmC and CDX2 (or FOXA1) according to H-score (n = 40; 5hmC vs. CDX2: r = 0.51, P = 0.0007; 5hmC vs. FOXA1: r = 0.43, P = 0.006).

4A). Since gene-body 5hmC was frequently correlated with transcriptional activation [33], mRNA levels of these genes were compared in early passaged patient-derived cancer cells (PC027, PC039, PC080, and PC084) and normal cells (HPDE and HPNE). Indeed, mRNA levels of HNF4A, HOXB1, GATA4, and PHC3 were over-expressed in cancer cells to different extents (**Figure 4D** and Supplementary Figure 4B), while CDX2 and FOXA1 were the only two genes over-expressed in all the four cancer cell lines (**Figure 4D**). Consistently, CDX2 and FOXA1 were up-regulated in patient-derived xenografts compared to normal cells (**Figure 4E**). Thus, these two transcription factors may play an oncogenic roles under the regulation of 5hmC modification. To confirm the correlation between 5hmC content and the expression levels of CDX2 or FOXA1, IHC staining was performed in serial sections (**Figure 4F**). Levels of 5hmC, CDX2, and FOXA1 in similar regions were quantified by Aperio Imagescope software. Positive correlations were observed between 5hmC content and CDX2 or FOXA1 expression levels (**Figure 4G** and **4H**). The results suggested that locus-specific distributions of 5hmC in the gene bodies of CDX2 and FOXA1 were correlated to transcriptional activation in cancer cells.

Discussion

Biomarkers to distinguish pancreatic cancer subtypes with prognostic values will improve therapeutic decisions for this highly heterogeneous disease. Since tumor intrinsic traits enable tumor malignancy [34], we used patient-derived xenograft models as a platform to stratify PDAC subgroups by intrinsic tumorigenic potential of PDAC. For clinical use, we identified an IHC biomarker, 5hmC content, as a surrogate for this property.

From PDAC patient specimens, almost all the normal pancreatic cells showed high intensity of 5hmC. On the other hand, 5hmC expression was significantly decreased in PanIN and cancer regions. This observation was consistent with the previous reports that cancer cells contained less 5hmC when compared with matched surrounding normal tissues [31]. However, in contrast to the reports in most other cancer types [35], we found that retained high 5hmC expression in cancer cells served as a significant predictor of worse progression-free survival in pancreatic cancer, independent of conventional clinicopathological factors. These results, therefore, suggested that specific oncogenic pathways were regulated by 5hmC in PDAC.

Genome-wide analysis of 5hmC-enriched loci by high throughput sequencing was performed in pancreatic cancer patient-derived xenografts and low-passage pancreatic cancer cell lines (data accessible at NCBI GEO database, accession GSE71839) (Bhattacharyya et al., 2016). They found that 5hmC redistributed and enriched at oncogenic enhancers to increase transcript expression for many oncogenic pathways [10]. Similarly, high 5hmC content was correlated with the tumorigenicity of glioblastoma cells [36]. Our observation that cancer-associated 5hmC modification promoted pancreatic cancer malignancy was in accordance with these results.

Interestingly, redistribution of 5hmC in cancer cells occurred in gene bodies of transcription factors for lineage differentiation, especially

CDX2 and FOXA1 which were commonly regulated. CDX2 and FOXA1 are transcription factors for lineage specification during normal development. CDX2 directly regulates several pro-intestinal transcription factors to drive intestinal differentiation, and is ectopically expressed in various extra-intestinal adenocarcinomas [37]. It is expressed in about 30% of PDACs, which have inferior survival compared to those with CDX2-negative tumors [37, 38]. FOXA1 is required for multiple lineage differentiation [39]. In pancreas, FOXA1 regulates PDX1 expression to control the expansion and differentiation of the pancreatic primordia. The expression of FOXA1 is silenced in adult pancreatic tissue and is re-expressed in primary tumor and further enhanced in metastatic PDACs [40].

5hmC content is crucial for controlling lineage commitment during normal cellular differentiation. An early increase and maintenance of global high 5hmC precedes chondrogenic differentiation [41]. In addition, 5hmC modification locates in the gene body of highly expressed genes in precursor cells during T cell development and differentiation [42]. During the differentiation process of human embryonic stem cells toward pancreatic endoderm, 5hmC marks are positively correlated to enhancer activity and binding of lineage-specific transcription factors [43]. Dynamic 5hmC is cell-type specific and high 5hmC may increase chromatin accessibility for transcription factor binding to facilitate multi-lineage potential [44]. Deregulation of lineage-specific pathways enhances cancer plasticity and heterogeneity [45], which is a driving force for tumor evolution, metastasis, and drug resistance [46].

In summary, 5hmC content is a clinical useful IHC marker to improve prediction of progression-free survival of PDAC patients after curative resection. Enriched cancer-associated 5hmC contents may increase lineage plasticity to aid in the process to differentiate cells *in vivo* for persistent engraftment and tumor heterogeneity.

Acknowledgements

This research work was supported by funds from Academia Sinica, and grants from Ministry of Science and Technology, Taiwan (MOST 102-2321-B-001-061-, MOST 104-0210-01-09-02, MOST 105-0210-01-13-01, and MOST 106-0210-01-15-02). This work was financially supported by the "Drug Development Center, China Medical University" from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. We thank the BioIT at IBMS, Academia Sinica, for the bioinformatics/Microarray analysis/mathematical statistics support.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Wen-Hwa Lee, Genomics Research Center, Academia Sinica, 128 Academia Road, Taipei 115, Taiwan. Tel: +886 22789 8777; Fax: +886 22789 8771; E-mail: whlee@uci.edu; Dr. Yung-Ming Jeng, Department of Pathology, National Taiwan University Hospital, National Taiwan University College of Medicine, 7 Chung-Shan South Road, Taipei 100, Taiwan. Tel: +886 22312 3456 Ext. 63561; Fax: +886 22393 4172; E-mail: mrna0912@gmail.com

References

[1] Ryan DP, Hong TS and Bardeesy N. Pancreatic Adenocarcinoma. N Engl J Med 2014; 371: 1039-1049.

[2] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018; 68: 7-30.

[3] Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, Miller DK, Christ AN, Bruxner TJ, Quinn MC, Nourse C, Murtaugh LC, Harliwong I, Idrisoglu S, Manning S, Nourbakhsh E, Wani S, Fink L, Holmes O, Chin V, Anderson MJ, Kazakoff S, Leonard C, Newell F, Waddell N, Wood S, Xu Q, Wilson PJ, Cloonan N, Kassahn KS, Taylor D, Quek K, Robertson A, Pantano L, Mincarelli L, Sanchez LN, Evers L, Wu J, Pinese M, Cowley MJ, Jones MD, Colvin EK, Nagrial AM, Humphrey ES, Chantrill LA, Mawson A, Humphris J, Chou A, Pajic M, Scarlett CJ, Pinho AV, Giry-Laterriere M, Rooman I, Samra JS, Kench JG, Lovell JA, Merrett ND, Toon CW, Epari K, Nguyen NQ, Barbour A, Zeps N, Moran-Jones K, Jamieson NB, Graham JS, Duthie F, Oien K, Hair J, Grutzmann R, Maitra A, Iacobuzio-Donahue CA, Wolfgang CL, Morgan RA, Lawlor RT, Corbo V, Bassi C, Rusev B, Capelli P, Salvia R, Tortora G, Mukhopadhyay D, Petersen GM, Munzy DM, Fisher WE, Karim SA, Eshleman JR, Hruban RH, Pilarsky C, Morton JP, Sansom OJ, Scarpa A, Musgrove EA, Bailey UM, Hofmann O, Sutherland RL, Wheeler DA,

Gill AJ, Gibbs RA, Pearson JV, Waddell N, Biankin AV and Grimmond SM. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016; 531: 47-52.

[4] Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, Feiler HS, Ko AH, Olshen AB, Danenberg KL, Tempero MA, Spellman PT, Hanahan D and Gray JW. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011; 17: 500-503.

[5] Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, Smyla JK, Anderson JM, Kim HJ, Bentrem DJ, Talamonti MS, Iacobuzio-Donahue CA, Hollingsworth MA and Yeh JJ. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet 2015; 47: 1168-1178.

[6] Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, Johns AL, Miller D, Nones K, Quek K, Quinn MC, Robertson AJ, Fadlullah MZ, Bruxner TJ, Christ AN, Harliwong I, Idrisoglu S, Manning S, Nourse C, Nourbakhsh E, Wani S, Wilson PJ, Markham E, Cloonan N, Anderson MJ, Fink JL, Holmes O, Kazakoff SH, Leonard C, Newell F, Poudel B, Song S, Taylor D, Waddell N, Wood S, Xu Q, Wu J, Pinese M, Cowley MJ, Lee HC, Jones MD, Nagrial AM, Humphris J, Chantrill LA, Chin V, Steinmann AM, Mawson A, Humphrey ES, Colvin EK, Chou A, Scarlett CJ, Pinho AV, Giry-Laterriere M, Rooman I, Samra JS, Kench JG, Pettitt JA, Merrett ND, Toon C, Epari K, Nguyen NQ, Barbour A, Zeps N, Jamieson NB, Graham JS, Niclou SP, Bjerkvig R, Grutzmann R, Aust D, Hruban RH, Maitra A, Iacobuzio-Donahue CA, Wolfgang CL, Morgan RA, Lawlor RT, Corbo V, Bassi C, Falconi M, Zamboni G, Tortora G, Tempero MA, Gill AJ, Eshleman JR, Pilarsky C, Scarpa A, Musgrove EA, Pearson JV, Biankin AV and Grimmond SM. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 2015; 518: 495-501.

[7] Walters DM, Stokes JB, Adair SJ, Stelow EB, Borgman CA, Lowrey BT, Xin W, Blais EM, Lee JK, Papin JA, Parsons JT and Bauer TW. Clinical, molecular and genetic validation of a murine Orthotopic Xenograft Model of Pancreatic Adenocarcinoma Using Fresh Human Specimens. PLoS One 2013; 8: e77065.

[8] Matos LL, Trufelli DC, de Matos MG and da Silva Pinhal MA. Immunohistochemistry as an important tool in biomarkers detection and clinical practice. Biomark Insights 2010; 5: 9-20.

[9] Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S and Reik W. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 2011; 473: 398-402.

[10] Bhattacharyya S, Yu Y, Suzuki M, Campbell N, Mazdo J, Vasanthakumar A, Bhagat TD, Nischal S, Christopeit M, Parekh S, Steidl U, Godley L, Maitra A, Greally JM and Verma A. Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. Nucleic Acids Res 2013; 41: e157-e157.

[11] Damhofer H, Ebbing EA, Steins A, Welling L, Tol JA, Krishnadath KK, van Leusden T, van de Vijver MJ, Besselink MG, Busch OR, Henegouwen MIvB, van Delden O, Meijer SL, Dijk F, Medema JP, van Laarhoven HW and Bijlsma MF. Establishment of patient-derived xenograft models and cell lines for malignancies of the upper gastrointestinal tract. J Transl Med 2015; 13: 115.

[12] Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25: 1754-1760.

[13] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W and Liu XS. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 2008; 9: R137-R137.

[14] Huang W, Loganantharaj R, Schroeder B, Fargo D and Li L. PAVIS: a tool for Peak Annotation and Visualization. Bioinformatics 2013; 29: 3097-3099.

[15] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G and Mesirov JP. Integrative genomics viewer. Nat Biotechnol 2011; 29: 24-26.

[16] Joo KM, Kim J, Jin J, Kim M, Seol HJ, Muradov J, Yang H, Choi YL, Park WY, Kong DS, Lee JI, Ko YH, Woo HG, Lee J, Kim S and Nam DH. Patient-specific orthotopic glioblastoma xenograft models recapitulate the histopathology and biology of human glioblastomas in situ. Cell Rep 2013; 3: 260-273.

[17] Moon HG, Oh K, Lee J, Lee M, Kim JY, Yoo TK, Seo MW, Park AK, Ryu HS, Jung EJ, Kim N, Jeong S, Han W, Lee DS and Noh DY. Prognostic and functional importance of the engraftment-associated genes in the patient-derived xenograft models of triple-negative breast cancers. Breast Cancer Res Treat 2015; 154: 13-22.

[18] Wu HH, Hwang-Verslues WW, Lee WH, Huang CK, Wei PC, Chen CL, Shew JY, Lee EY, Jeng YM, Tien YW, Ma C and Lee WH. Targeting IL-17B-IL-17RB signaling with an anti-IL-17RB antibody blocks pancreatic cancer metastasis by silencing multiple chemokines. J Exp Med 2015; 212: 333-349.

[19] Koorstra JB, Karikari CA, Feldmann G, Bisht S, Rojas PL, Offerhaus GJ, Alvarez H and Maitra

A. The Axl receptor tyrosine kinase confers an adverse prognostic influence in pancreatic cancer and represents a new therapeutic target. Cancer Biol Ther 2009; 8: 618-626.

[20] Hirakawa T, Yashiro M, Murata A, Hirata K, Kimura K, Amano R, Yamada N, Nakata B and Hirakawa K. IGF-1 receptor and IGF binding protein-3 might predict prognosis of patients with resectable pancreatic cancer. BMC Cancer 2013; 13: 392.

[21] Li J, Hu G, Kong F, Wu K, Song K, He J and Sun W. Elevated STMN1 expression correlates with poor prognosis in patients with pancreatic ductal adenocarcinoma. Pathol Oncol Res 2015; 21: 1013-1020.

[22] Karamitopoulou E, Zlobec I, Tornillo L, Carafa V, Schaffner T, Brunner T, Borner M, Diamantis I, Zimmermann A and Terracciano L. Differential cell cycle and proliferation marker expression in ductal pancreatic adenocarcinoma and pancreatic intraepithelial neoplasia (PanIN). Pathology 2010; 42: 229-234.

[23] Guweidhi A, Kleeff J, Giese N, El Fitori J, Ketterer K, Giese T, Buchler MW, Korc M and Friess H. Enhanced expression of 14-3-3sigma in pancreatic cancer and its role in cell cycle regulation and apoptosis. Carcinogenesis 2004; 25: 1575-1585.

[24] Li Z, Dong Z, Myer D, Yip-Schneider M, Liu J, Cui P, Schmidt CM and Zhang JT. Role of 14-3-3σ in poor prognosis and in radiation and drug resistance of human pancreatic cancers. BMC Cancer 2010; 10: 598.

[25] Nguyen LV, Vanner R, Dirks P and Eaves CJ. Cancer stem cells: an evolving concept. Nat Rev Cancer 2012; 12: 133-143.

[26] Stewart JM, Shaw PA, Gedye C, Bernardini MQ, Neel BG and Ailles LE. Phenotypic heterogeneity and instability of human ovarian tumor-initiating cells. Proc Natl Acad Sci U S A 2011; 108: 6468-6473.

[27] Wang T, Wu H, Li Y, Szulwach KE, Lin L, Li X, Chen IP, Goldlust IS, Chamberlain SJ, Dodd A, Gong H, Ananiev G, Han JW, Yoon YS, Rudd MK, Yu M, Song CX, He C, Chang Q, Warren ST, Jin P. Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency. Nat Cell Biol 2013; 15: 700-711.

[28] Fiorenzano A, Pascale E, D'Aniello C, Acampora D, Bassalert C, Russo F, Andolfi G, Biffoni M, Francescangeli F, Zeuner A, Angelini C, Chazaud C, Patriarca EJ, Fico A and Minchiotti G. Cripto is essential to capture mouse epiblast stem cell and human embryonic stem cell pluripotency. Nat Commun 2016; 7: 12589.

[29] Zhang J, Ratanasirintrawoot S, Chandrasekaran S, Wu Z, Ficarro Scott B, Yu C, Ross Christian A, Cacchiarelli D, Xia Q, Seligson M, Shinoda G, Xie W, Cahan P, Wang L, Ng SC, Tintara S, Trapnell C, Onder T, Loh YH, Mikkelsen T, Sliz P, Teitell Michael A, Asara John M, Marto Jarrod A, Li H, Collins James J and Daley George Q. LIN28 regulates stem cell metabolism and conversion to primed pluripotency. Cell Stem Cell 2016; 19: 66-80.

[30] Quintanilla RH Jr, Asprer JS, Vaz C, Tanavde V and Lakshmipathy U. CD44 is a negative cell surface marker for pluripotent stem cell identification during human fibroblast reprogramming. PLoS One 2014; 9: e85419.

[31] Yang H, Liu Y, Bai F, Zhang JY, Ma SH, Liu J, Xu ZD, Zhu HG, Ling ZQ, Ye D, Guan KL and Xiong Y. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene 2013; 32: 663-669.

[32] Putiri EL, Tiedemann RL, Thompson JJ, Liu C, Ho T, Choi JH and Robertson KD. Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells. Genome Biol 2014; 15: R81.

[33] Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE and Zhang Y. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev 2011; 25: 679-684.

[34] Hanahan D and Weinberg Robert A. Hallmarks of cancer: the next generation. Cell 2011; 144: 646-674.

[35] Orr BA, Haffner MC, Nelson WG, Yegnasubramanian S and Eberhart CG. Decreased 5-hydroxymethylcytosine is associated with neural progenitor phenotype in normal brain and shorter survival in malignant glioma. PLoS One 2012; 7: e41036.

[36] Takai H, Masuda K, Sato T, Sakaguchi Y, Suzuki T, Suzuki T, Koyama-Nasu R, Nasu-Nishimura Y, Katou Y, Ogawa H, Morishita Y, Kozuka-Hata H, Oyama M, Todo T, Ino Y, Mukasa A, Saito N, Toyoshima C, Shirahige K and Akiyama T. 5-Hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-methylosome complex. Cell Reports 2014; 9: 48-60.

[37] Werling RW, Yaziji H, Bacchi CE and Gown AM. CDX2, a highly sensitive and specific marker of adenocarcinomas of intestinal origin: an immunohistochemical survey of 476 primary and metastatic carcinomas. Am J Surg Pathol 2003; 27: 303-310.

[38] Xiao W, Hong H, Awadallah A, Zhou L and Xin W. Utilization of CDX2 expression in diagnosing pancreatic ductal adenocarcinoma and predicting prognosis. PLoS One 2014; 9: e86853.

[39] Yang YA, Zhao JC, Fong KW, Kim J, Li S, Song C, Song B, Zheng B, He C and Yu J. FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function. Nucleic Acids Research 2016; 44: 8153-8164.

[40] Roe JS, Hwang CI, Somerville TDD, Milazzo JP, Lee EJ, Da Silva B, Maiorino L, Tiriac H, Young CM, Miyabayashi K, Filippini D, Creighton B, Burkhart RA, Buscaglia JM, Kim EJ, Grem JL, Lazenby AJ, Grunkemeyer JA, Hollingsworth MA, Grandgenett PM, Egeblad M, Park Y, Tuveson DA and Vakoc CR. Enhancer reprogramming promotes pancreatic cancer metastasis. Cell 2017; 170: 875-888, e820.

[41] Taylor SE, Li YH, Smeriglio P, Rath M, Wong WH and Bhutani N. Stable 5-Hydroxymethylcytosine (5hmC) acquisition marks gene activation during chondrogenic differentiation. J Bone Miner Res 2016; 31: 524-534.

[42] Tsagaratou A, Aijo T, Lio CW, Yue X, Huang Y, Jacobsen SE, Lahdesmaki H and Rao A. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. Proc Natl Acad Sci U S A 2014; 111: E3306-3315.

[43] Li J, Wu X, Zhou Y, Lee M, Guo L, Han W, Mo W, Cao WM, Sun D, Xie R and Huang Y. Decoding the dynamic DNA methylation and hydroxymethylation landscapes in endodermal lineage intermediates during pancreatic differentiation of hESC. Nucleic Acids Res 2018; 46: 2883-2900.

[44] Ruzov A, Tsenkina Y, Serio A, Dudnakova T, Fletcher J, Bai Y, Chebotareva T, Pells S, Hannoun Z, Sullivan G, Chandran S, Hay DC, Bradley M, Wilmut I and De Sousa P. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. Cell Res 2011; 21: 1332-1342.

[45] Cheung WK and Nguyen DX. Lineage factors and differentiation states in lung cancer progression. Oncogene 2015; 34: 5771-5780.

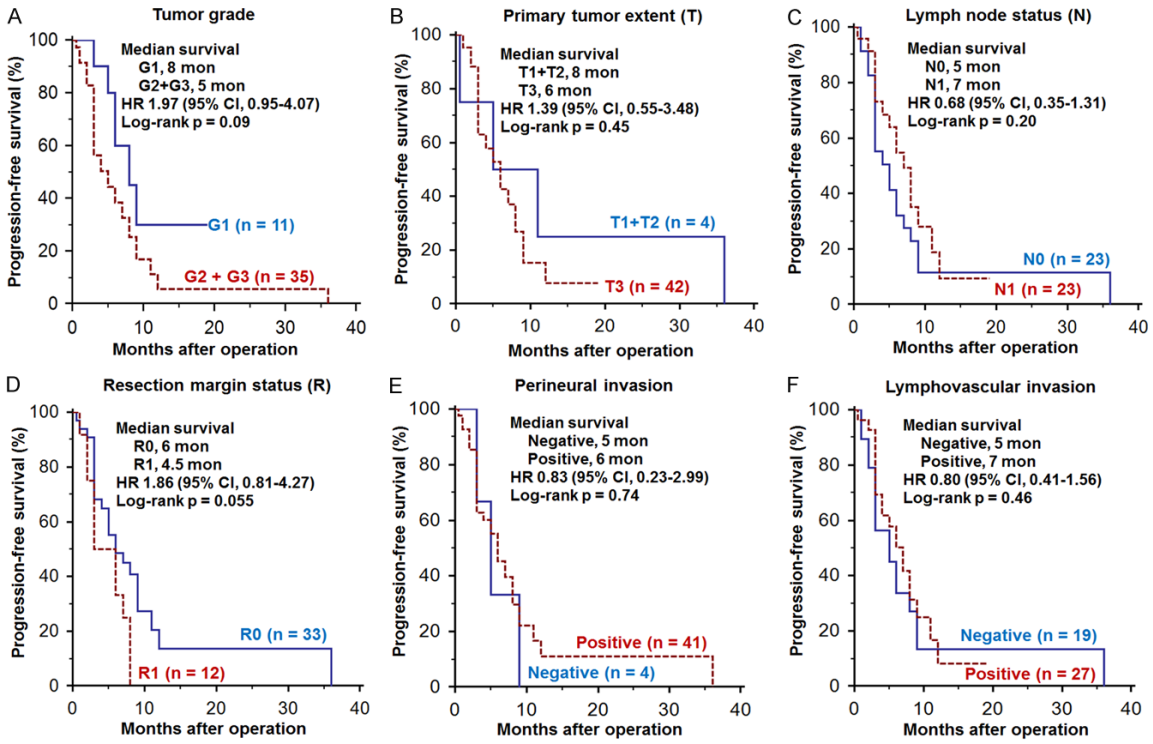
[46] McGranahan N and Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. Cell 2017; 168: 613-628.

Supplementary Table 1. Clinical characteristics of the PDAC patients in the validation cohort (n = 73)

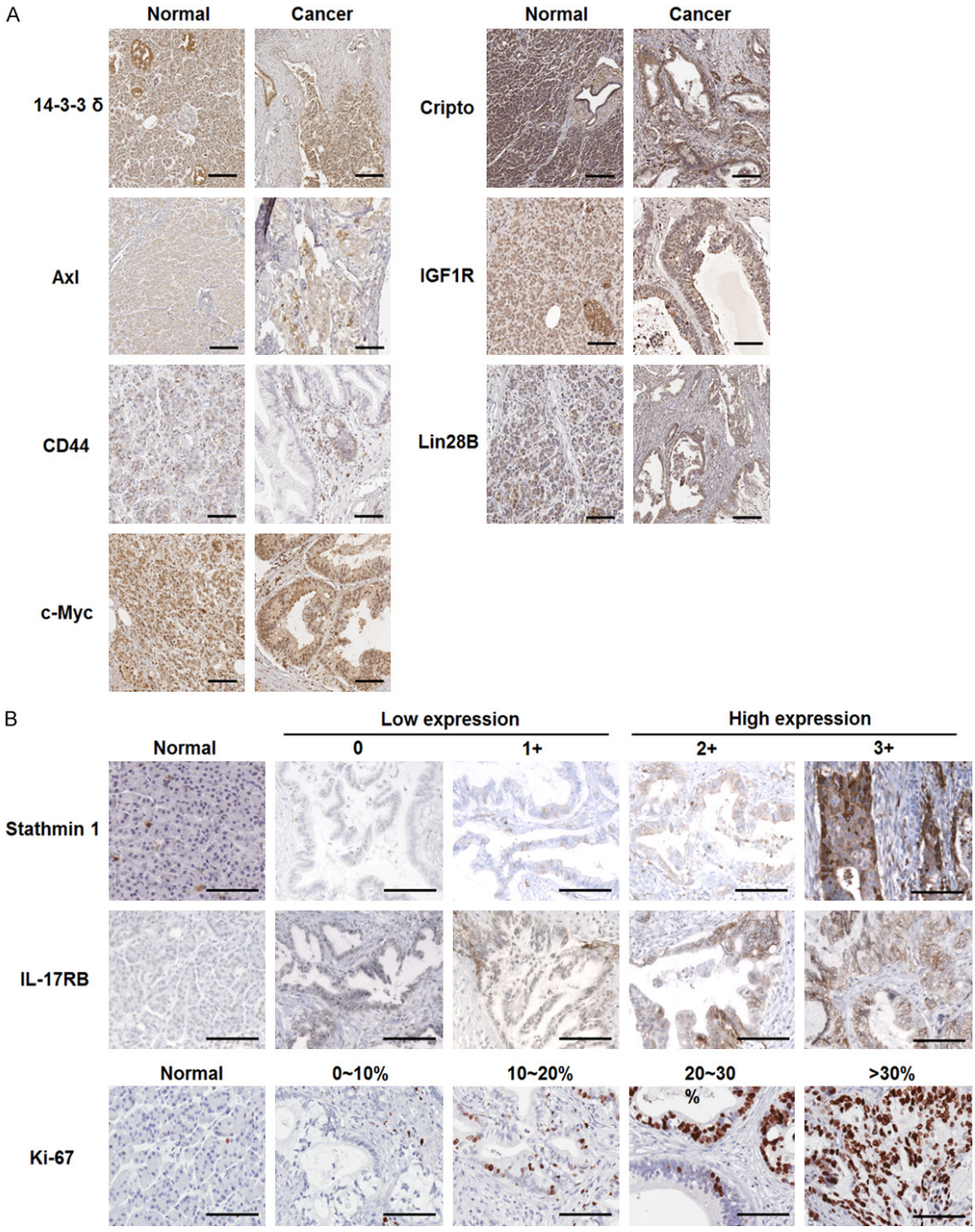
Characteristics	No.
Age	
Mean (year, SD)	64.3 (12.5)
Range	26-91.2
< 60	29
≥ 60	44
Gender	
Male	42
Female	31
AJCC stage	
IA	2
IB	5
IIA	19
IIB	44
Unknown	3
Tumor grade	
G1	7
G2-G3	66
Primary tumor extent (T)	
T1	2
T2	11
T3	60
Lymph node status (N)	
N0	26
N1	44
Unknown	3
Distant metastasis (M)	
M0	73
Resection margin status (R)	
R0	60
R1	12
Unknown	1
Lymphovascular invasion	
Negative	14
Positive	44
Unknown	15
Perineural invasion	
Negative	7
Positive	59
Unknown	7

Supplementary Table 2. Characteristics and staining procedures for the antibodies employed in IHC analysis

Antibody	Source	Clonality	Species	Method	Antigen retrieval	Antibody incubation	Dilution
5hmC	Active motif (#39770)	Polyclonal	Rabbit	Ventana	Mild condition	120 min, 25 °C	1:500
14-3-3δ	GeneTex (#112949)	Polyclonal	Rabbit	Ventana	Standard condition	60 min, 25 °C	1:500
Stathmin1	Abcam (ab11269)	Polyclonal	Rabbit	Ventana	Standard condition	60 min, 25 °C	1:1000
IL-17RB	Homemade (clone A81)	Monoclonal	Mouse	Ventana	Mild condition	120 min, 25 °C	1:500
Ki-67	BD (#550609)	Monoclonal	Mouse	Ventana	Standard condition	100 min, 37 °C	1:10
IGF1R	GeneTex (#111666)	Polyclonal	Rabbit	Ventana	Standard condition	120 min, 25 °C	1:100
c-Myc	GeneTex (#84065)	Monoclonal	Mouse	Manual stain	Heat-induced pH 6	Overnight, 4 °C	1:50
Axl	GeneTex (#108560)	Polyclonal	Rabbit	Manual stain	Heat-induced pH 10	Overnight, 4 °C	1:50
Cripto	Santa Cruz (sc-376448)	Monoclonal	Mouse	Ventana	Mild condition	60 min, 25 °C	1:400
Lin28B	Proteintech (16178-1-AP)	Polyclonal	Rabbit	Ventana	Mild condition	240 min, 25 °C	1:50
CD44	Abcam (ab51037)	Monoclonal	Rabbit	Ventana	Standard condition	120 min, 25 °C	1:100
TET1	GeneTex (#627420)	Monoclonal	Mouse	Ventana	Standard condition	120 min, 25 °C	1:1500
CDX2	Cell Marque (EPR2764Y)	Monoclonal	Rabbit	Ventana	Standard condition	30 min, 25 °C	1:500
FOXA1	Novus Biologicals (NBP2-45269)	Monoclonal	Mouse	Bond-max	Heat-induced pH 9	120 min, 25 °C	1:100



Supplementary Figure 1. Kaplan-Meier analysis of progression-free survival according to (A) tumor grade (P = 0.1, Log-rank test), (B) primary tumor extent (T) (P = 0.45, Log-rank test), (C) lymph node status (N) (P = 0.20, Log-rank test), (D) resection margin status (R) (P = 0.055, Log-rank test), (E) perineural invasion (*P* = 0.74, Log-rank test), and (F) lymphovascular invasion (*P* = 0.46, Log-rank test).



Supplementary Figure 2. Validation of the candidate biomarkers by IHC staining. A. IHC staining results of the potential biomarkers (14-3-3δ, Axl, CD44, c-Myc, Cripto, IGF1R, and Lin28B) in cancer and adjacent normal regions in PDAC sections. Scale bar, 100 μm. B. IHC staining of Stathmin1, IL-17RB, and Ki-67 in cancer and adjacent normal regions in PDAC sections. Scale bar, 100 μm. The expression levels were scored by 0 to 3+ in Stathmin1 and IL-17RB (0 for negative; 1+ for < 10% among cancer cells; 2+ for 10 to 50% among cancer cells; 3+ for ≥ 50% among cancer cells). Ki-67 expression levels were calculated by nuclear algorithm of Aperio ImageScope software, and the scoring ranges were including 0 to 10%, 10 to 20%, 20 to 30%, and ≥ 30% among cancer cells. For statistical analyses, the score of 0 and 1+ (or 0 to 10% and 10 to 20%) was combined as the low expression group, and the score of 2+ and 3+ (or 20 to 30% and ≥ 30%) was combined as the high expression group.

Supplementary Table 3. Correlations between 5hmC content and clinicopathological variables combining primary and validation cohort studies (n = 119)

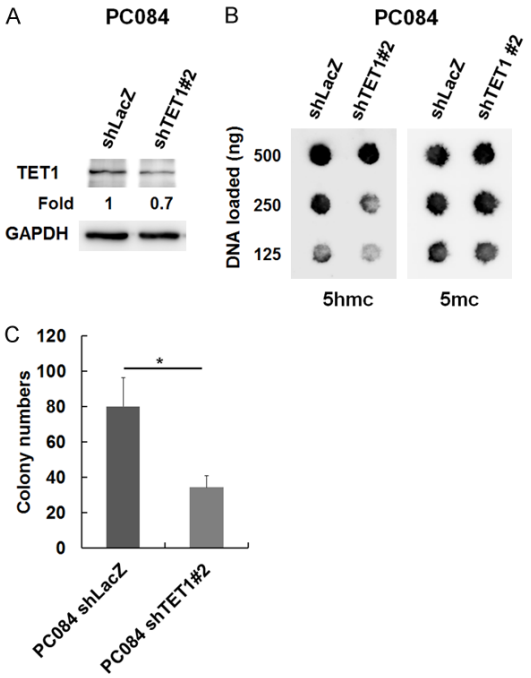
	n	5hmC expression in tumor		*P*
		0~1+	2~3+	
Age				
< 60	42	17	25	0.94
≥ 60	77	32	45	
Gender				
Male	69	22	47	0.03
Female	50	27	23	
AJCC stage				
IA/IB	8	7	1	0.02
IIA/IIB	108	40	68	
Unknown	3	2	1	
Tumor grade				
G1	18	6	12	0.64
G2-G3	101	43	58	
Primary tumor extent (T)				
T1/T2	17	10	7	0.18
T3	102	39	63	
Lymph node status (N)				
N0	49	24	25	0.16
N1	67	23	44	
Unknown	3	2	1	
Resection margin status (R)				
R0	93	43	50	0.1
R1	24	6	18	
Unknown	2	0	2	
Lymphovascular invasion				
Negative	33	15	18	0.61
Positive	71	27	44	
Unknown	15	7	8	
Perineural invasion				
Negative	11	5	6	0.97
Positive	100	41	59	
Unknown	8	3	5	

Supplementary Table 4. Univariate Cox regression analysis comparing IHC scores of 5hmC, Stathmin1, IL-17RB, and Ki-67 for progression-free survival (PFS) of the PDAC patients included in primary cohort (n = 46)

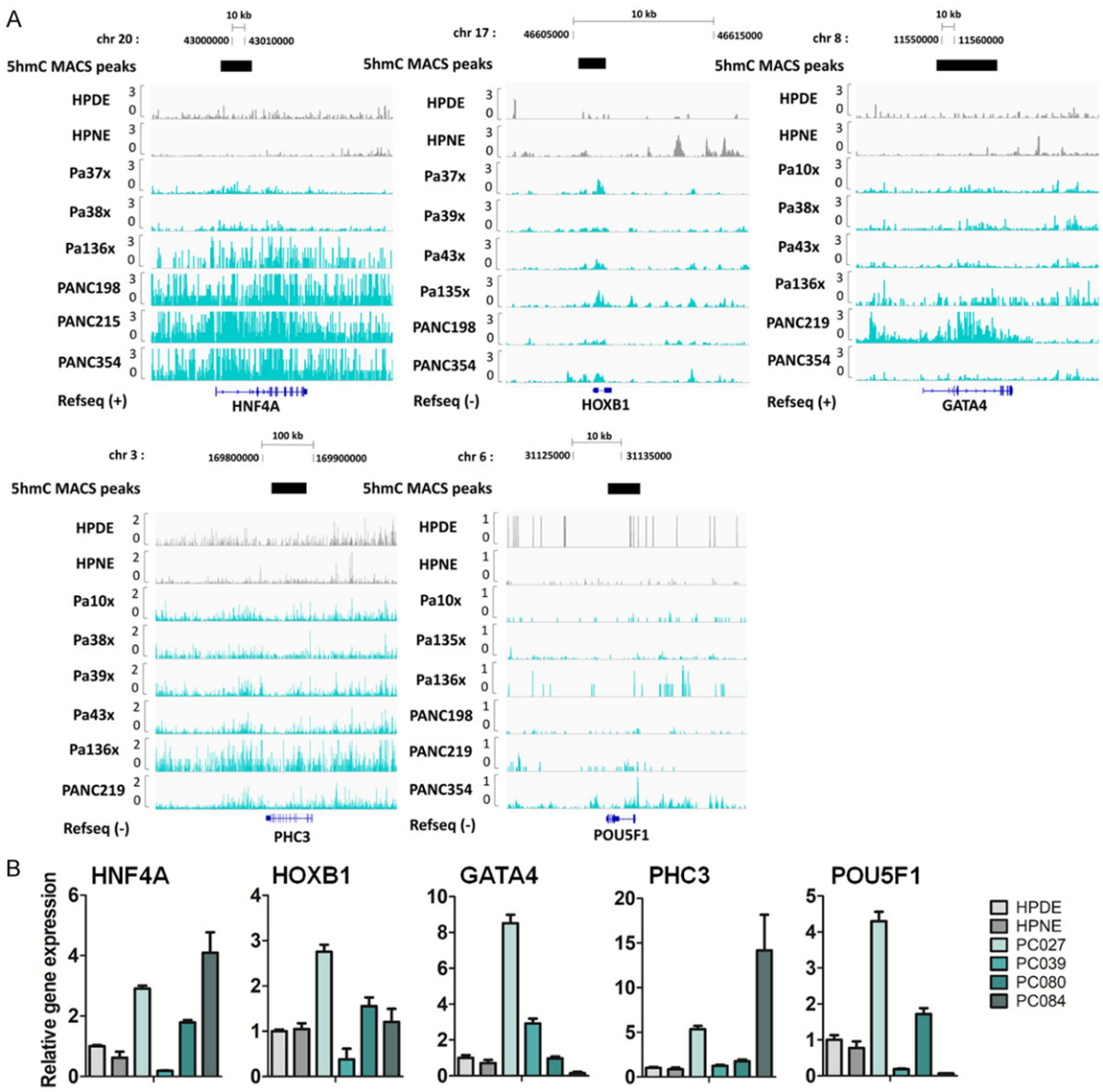
Biomarker	Univariate survival analysis of biomarkers (Cox regression analysis)	
	HR (95% CI)	*P*
5hmC (Low *vs.* High)	2.58 (1.11-6.04)	0.03
Stathmin1 (Low *vs.* High)	1.28 (0.62-2.64)	0.51
IL-17RB (Low *vs.* High)	1.28 (0.65-2.51)	0.48
Ki-67 (Low *vs.* High)	1.91 (0.93-3.90)	0.08

Supplementary Table 5. Univariate and multivariate cox regression analysis of 5hmC content and clinicopathological variables for progression-free survival (PFS) and overall survival (OS) of the PDAC patients included in primary and validation cohorts (n = 119)

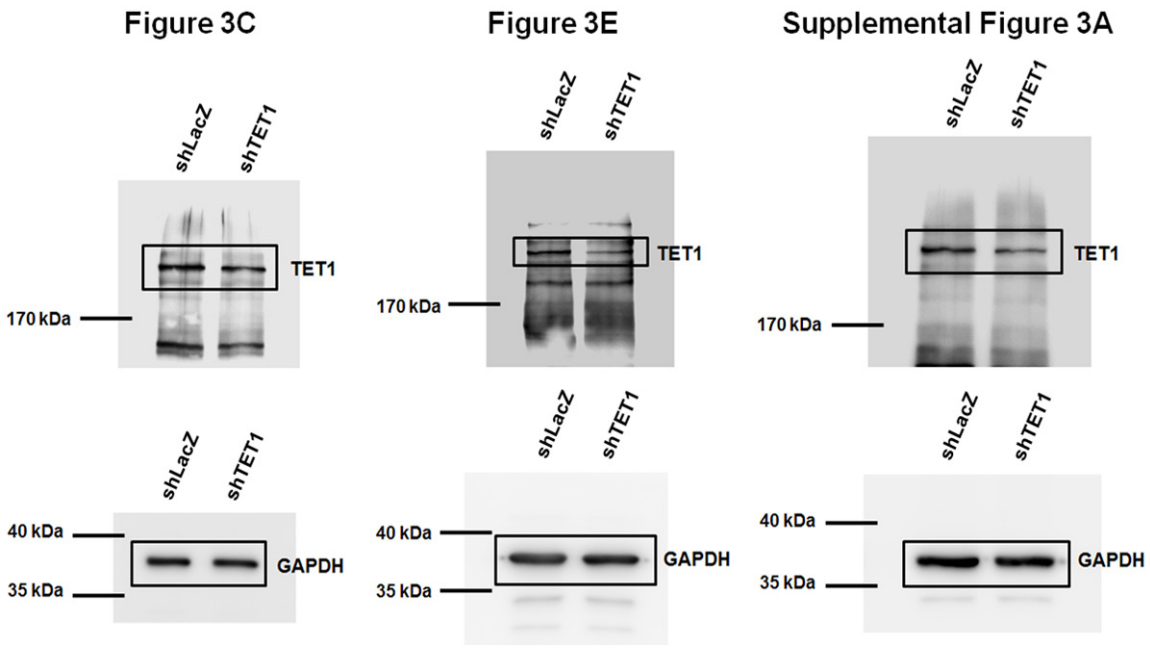
Factors	PFS				OS	
	Univariate analysis		Multivariate analysis		Univariate analysis	
	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
5hmC (Low *vs.* High)	2.13 (1.40-3.23)	0.0004	2.37 (1.40-3.99)	0.0013	1.56 (1.03-2.37)	0.04
Age (< 60 *vs.* ≥ 60)	1.41 (0.91-2.16)	0.12	1.34 (0.82-2.20)	0.25	0.99 (0.65-1.50)	0.97
Gender (Male *vs.* Female)	1.06 (0.71-1.57)	0.78	1.06 (0.66-1.71)	0.8	1.15 (0.76-1.74)	0.5
Tumor grade (G1 *vs.* G2+G3)	1.62 (0.89-2.95)	0.12	2.07 (1.01-4.24)	0.05	1.10 (0.63-1.95)	0.73
Primary tumor extent (T1+T2 *vs.* T3)	1.12 (0.66-1.92)	0.67	0.65 (0.30-1.44)	0.3	1.26 (0.73-2.19)	0.41
Lymph node status (N0 *vs.* N1)	0.98 (0.66-1.47)	0.94	0.92 (0.56-1.50)	0.74	1.05 (0.70-1.58)	0.82
Resection margin status (R0 *vs.* R1)	1.94 (1.21-3.13)	0.006	1.77 (1.02-3.08)	0.04	1.36 (0.80-2.31)	0.26
Lymphovascular invasion (Positive *vs.* Negative group)	1.09 (0.70-1.71)	0.72	0.79 (0.45-1.41)	0.43	0.91 (0.58-1.45)	0.7
Perineural invasion (Positive *vs.* Negative group)	1.00 (0.49-1.97)	0.97	0.98 (0.42-2.26)	0.96	1.22 (0.56-2.63)	0.62



Supplementary Figure 3. The role of 5hmC in tumorigenesis confirmed by depletion of TET1 expression. A. Patient-derived cells generated from PC084 xenografts stably expressing lentiviral-based LacZ^{shRNA} or TET1^{shRNA} (clone #2). Knockdown efficiency of TET1 was verified by Western blotting. B. DNA dot blot analysis of genomic 5hmC and 5mC of PC084 cells stably expressing lentiviral-based LacZ^{shRNA} or TET1^{shRNA} (clone #2). C. Tumorigenicity was assessed by soft agar colony formation assay. Quantification of colony numbers in PC084 cells stably expressing lentiviral-based LacZ^{shRNA} or TET1^{shRNA} (clone #2) were shown. Colonies with diameter ≥ 50 μm were included. Values were presented as mean ± SD of four independent experiments. Scale bar, 50 μm. *, $P < 0.05$ (two-tailed t-test).



Supplementary Figure 4. Comparisons of 5hmC occupancy profiles and mRNA expression of the genes including HNF4A, HOXB1, GATA4, PHC3, and POU5F1 between cancer and normal cells. A. Gene tracks of 5hmC occupancy at HNF4A, HOXB1, GATA4, PHC3, and POU5F1 genes in normal pancreatic cells and patient-derived xenografts. The 5hmc profiling data in normal (HPDE and HPNE) and patient-derived (Pa10x, Pa37x, Pa38x, Pa39x, Pa43x, Pa135x, Pa136x, PANC198, PANC215, PANC219, and PANC354) cancer cells were obtained from Gene Expression Omnibus database (GSE71839) and gene tracks were visualized by Integrative genomics viewer with normalized coverage reads. The MACS identified 5hmc peaks were denoted on the top of each panel. B. Quantitative real-time PCR analysis of HNF4A, HOXB1, GATA4, PHC3, and POU5F1 was performed to compare gene expression levels in patient-derived cancer cells at an early passage number (PC027, PC039, PC080 and PC084) and immortalized human pancreatic ductal epithelial cells (HPDE and HPNE). Data were representative of three independent experiments and values were presented as mean ± SD (n = 3).



Supplementary Figure 5. Full immunoblots shown in the manuscript.