## The Epigenetic Characteristics of Pancreatic Adenocarcinoma Revealed by a Global DNA Methylation Analysis by MethylCap-Seq

Yangxing Zhao1,6, Jinfeng Sun 2,6, Hongyu Zhang 3,6, Shicheng Guo 4,6, Jun Gu1, Wei Wang1, Ning Tang3, Xiaoyu Zhou5,Jian Yu1

1 State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

2 Zhongshan Hospital, Fudan University, 180 Fenglin Road, Shanghai 200032, China

3 Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

4 State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China

5 Key Laboratory of Contraceptive Drugs and Devices of NPFPC, Shanghai Institute of Planned Parenthood Research, Shanghai, China

6 These authors contributed equally to this work

Run title: Epigenetic Profile of Pancreatic Adenocarcinoma

Figures counts:

Table counts:

Abstract counts: 202

Mainbody counts:

Abbreviations:

PC: pancreatic cancer

MethylCap-Seq:

CpGIs: CpG islands

MSRE-qPCR: Methylation-sensitive restriction enzyme-based quantitative PCR

DMRs: differentially methylated regions

TSS: transcription start site

MSP:

BSP:

EST: expressed sequence tag

Correspondence:

Yu Jian, State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, LN 2200/25, Xietu Road, Shanghai 200032, China. Phone: 021-64032388, Fax: 021-64032388, Email:yujian@shsci.org

Abstract

**Purpose:** Extensive dysregulation and reprogramming of DNA methylation is an important feature of pancreatic cancer (PC). The present study aimed to characterize the genome methylation patterns in PC and investigate the clinical potential of the DNA methylation in PC diagnosis. **Methods:** Genome-wide DNA methylation profile was established with MethylCap-seq technique in PC tissues and adjacent non-tumor tissues (PN), respectively. Differentially methylated regions (DMRs) were identified by MACS and BALM methods. DMRs were validated in the independent set of PC and PN samples using methylation specific PCR (MSP), Bisulfite sequencing PCR (BSP) and Methylation-sensitive restriction enzyme-based quantitative PCR (MSRE-qPCR). Transcript expression regulation which include mRNA and expressed sequence tag (EST) caused by DNA methylation were investigated in A, B, C cell lines before and after 5-Aza-2’-deoxycytidine treatment and in clinical samples by MSRE-qPCR and RT-qPCR. **Results:** A total of 1,132 hyper-DMRs and 627 hypo-DMRs were identified in PC genome. 2,955 hyper-DMRs and 2,386 hypo-DMRs were associated with gene promoters. Moreover, 1,852 hyper-DMRs and 1,545 hypo-DMRs were found to be associated with CGIs as well as CGI shores. 80% pre-selected candidate DMRs were successfully validated with BSP and MSP technique in Internal and external sample sets which indicated the high accuracy and precision. In present study, the significant distribution difference were found for CpGI related hypo-DMR and hyper-DMR in genome elements, which CpGI related hyper-DMR preferred located in promoter region (P-value=3.32×10-4) while CpGI related hypo-DMRs were preferred located in intragenic regions (Pvalue= 5.73×10-5). Transcript analysis showed the aberrant DNA methylation in PC tissue and in PC cell lines was associated with gene or related EST expression. KEGG and gene ontology analysis showed that methylation based pathway analysis would be usefulness in cancer research. **Conclusions:** Our study characterized the genome-wide DNA methylation patterns in PC and identified DMRs that were distributed among various genomic contexts that might influence the expression of related genes or transcripts to promote PC. These DMRs might be potential diagnostic biomarkers and therapeutic targets in PC.

### Key word:

Pancreatic adenocarcinoma, Methylome, DNA methylation, MethylCap-seq

### Introduction

Pancreatic cancer (PC), a highly malignant tumor of the digestive system, is currently one of the solid tumors with the worst prognosis, with a postoperative 5-year survival rate of less than 25%. Nearly 100,000 people die from PC every year in American and Europe, where PC ranks fourth in cancer mortality both in male and female population [1]. The incidence of PC in China is also demonstrating an increasing trend [2].

From a biological point of view, genetics is predominantly responsible for the stable transfer of hereditary information between generations, whereas the different somatic phenotypes in different tissues and cells are influenced by epigenetics. Because most tumors develop with particular acquired biological phenotypes, changes surely play important roles [3]. DNA methylation, a well-studied epigenetic phenomenon, has already been extensively studied in PC.

At the gene level, genes such as p14ARF, p16INK4a and RASSF1A [4, 5] were found to have aberrant promoter methylation, leading to abnormalities in gene transcription. At the genomic level, by combining the techniques of methylated CpG island amplification with Agilent 244K Human Promoter ChIP-on-chip microarrays, Michael Goggins mapped the genome-wide methylation abnormalities in PC [6-7].

The study by Michael Goggins was based on a methylation array technology platform, which is more focused on CpG islands (CGIs), presumably with lower coverage of the whole genome than next-generation sequencing (NGS) technology [8].

MethylCap-seq is a robust procedure for genome-wide profiling of DNA methylation. The approach consists of the capture of methylated DNA using the MBD domain of MeCP2, and subsequent next-generation sequencing of eluted DNA.

Therefore, their description of the genome-wide methylation profile of PC could be incomplete, especially in regions such as CGI shores (2 kb flanking regions nearby CGI), non-CGI promoter regions and orphan CGIs. Methylation changes in these particular regions have already been suggested to be associated with certain tumor phenotypes or with tissue specificity [9, 10].

This study aimed to identify methylation-regulated genes that correlated with the development of PC; 10 PC tissue samples were pooled, as were 10 PC adjacent tissue samples(PN),and sequenced using the MethylCap-seq technique to characterize the genome-wide methylation profile of PC and to identify the various genomic regions with a high frequency of aberrant methylation, including regions of gene-associated CGIs, non-gene-associated CGIs (orphan CGIs), CpG shores and gene promoters without CGIs.

### Materials and Methods

#### Clinical samples

PC tissue samples were collected from 18 patients who had undergone surgical treatment without receiving preoperative chemotherapy or radiotherapy from May 2009 to March 2011 in Renji Hospital, School of Medicine, Shanghai Jiao Tong University. A diagnosis of PC was confirmed by histological examination. Resected tumor tissues and matched normal tissues at least 2 cm away from the tumor tissues were collected during the operation, labeled and stored at -80°C. The sixth edition of “Tumor, node and metastases (TNM) staging system proposed by the International Union Against Cancer (UICC)” was utilized to stage the tumor tissue samples (Table I). The study was approved by the medical ethics committee at Renji Hospital, School of Medicine, Shanghai Jiao Tong University. All the patients signed the informed consent form. DNA was isolated from frozen tissues or cell lines using the conventional proteinase K/organic extraction method as previously described [11].

#### Genome-wide methylation profiling by MethylCap-seq

Genomic DNA was extracted from 10 PC tissues and 10 matched normal tissues. Equal amounts of DNA were mixed to form the PC and PN groups. Pooled DNA (1.2 µg) from each group was used to generate the library for MethylCap-seq as previously described [11]. We used BWA alignment tools [12] with the default settings to map the 36 bp unpaired reads to the hg19 human genome reference assembly [13]. After removing PCR duplicates using Picard, the aligned data were converted, sorted, and indexed using Samtools [14] and Picard (http://picard.sourceforge.net). Methylation peaks (hypermethylated regions) were identified using MACS in the PC and PN samples as previously described [11]. The differentially methylated regions (DMRs) between PC and PN were identified using two methods, MACS [15] and BALM [16], to increase the detective power of MethylCap-seq. To decrease the false positive detection of DMR using BALM, the dual-threshold strategy was applied. A high-confidence threshold (0.975) was utilized in the PC hypermethylated region screening, and a low-confidence threshold (0.950) was utilized in the PN hypermethylated region screening. Then, cancer-specific methylation peaks were defined as hypermethylated regions. Similarly, normal tissue-specific methylation peaks were defined as hypomethylated regions with the reverse settings. Whole genome methylation (methylation of each CpG) was inferred using BALM, which was processed to create a Pearson correlation analysis among all the samples in the R environment. The refSeq genes (UCSC genes) and corresponding CpG islands (CGIs) were downloaded from the Table Browser of the UCSC database [11]. CGI shores were defined as the flanking region of 2K up and down of the CpG islands. The definiction of other elemetns were same with UCSC which included downstream, enhancer, exon, intron, miRNA, promoter and 5’ UTR. The BED file operations were performed using bed tools [17] and other Perl scripts. All the scripts are available upon email request. The generated genomic methylation profile was uploaded to a public database (Gene Expression Omnibus: GSE54854). Gene ontology analysis was performed using DAVID Bioinformatics Resources.

#### Methylation analysis

In the present study, BSP was utilized to determine the methylation status at single CpG resolution of DMRs identified by genomic methylation profiling. MSP was performed for qualitative methylation screening in a small set of PC samples. Using MSRE-qPCR, the DNA methylation status in orphan CGIs was quantitatively analyzed in PC cell lines before and after 5-Aza-2’-deoxycytidine treatment and in small samples of clinical PC tissues, as described previously [18]. Approximately 1.0 µg of genomic DNA extracted from PC or PN samples or PC cell lines was bisulfate-treated using EpiTect Kit (Qiagen, Hilden, Germany). Primers for MSP (Methylation-Specific PCR) and BSP (Bisulfite Sequencing PCR) were designed using MethylPrimer (Ref: Li LC and Dahiya R. MethPrimer: designing primers for methylation PCRs. Bioinformatics. 2002 Nov; 18(11):1427-31. PMID: 12424112). The methylation-sensitive restriction enzyme-based quantitative PCR (MSRE-qPCR) primers were designed using primer3(Ref: Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Research 40(15):e115). The sequences of the primers utilized in this study are listed in Table S1, S2. Jumpstart Taq (Sigma-Aldrich, St. Louis, MO, USA) was used in BSP and MSP with a 20 µl reaction volume per tube. The BSP and MSP reaction conditions were as follows: 94°C for 3 min; 35 cycles of 94°C for 20 s, annealing for 20 s, and 72°C for 20 s; and 72°C for 5 min. The PCR products were analyzed by electrophoresis in 1.5% agarose gels. The PCR products were TA cloned and verified by sequencing. At least 5 clones were sequenced for each BSP reaction.

#### Cell culture and 5-Aza-2-deoxycytidine treatment

Three pancreatic adenocarcinoma cell lines were used: BxPC-3 (ATCC, CRL-1687), PANC-1 (ATCC, CRL-1469), and CFPAC-1 (ATCC, CRL-1918). All the cell lines were cultured in RPMI1640supplemented with 10% FBS, 100 U/ml penicillin, and 100 U/ml streptomycin. All the cell lines were maintained at 37°C in a humidified atmosphere with 5% CO2.

The restoration of gene expression by de-methylation treatment was evaluated in the BxPC-3, CFPAC-1 and CFPAC-1 cell lines. For the CpG de-methylation analysis, exponentially growing cells were seeded at a density of 1.5x106 cells/100 mm dish and allowed to attach overnight. The cells were then treated with freshly prepared 5-aza-dC (5.0 µM; Sigma-Aldrich, St. Louis, MO, USA) for 3 days.

#### RNA isolation and Real-time PCR

Total RNA was prepared from cultured cells using Trizol reagent according to the manufacturer’s instructions (Invitrogen, USA) and then reverse transcribed using an oligo (dT) primer and SuperscriptII RNase H-Reverse Transcriptase (Invitrogen, USA). Real-time PCR was performed with primer pairs for the EST expression assay, and GAPDH was used as the internal control. Real-time PCR was performed as follows: 94°C for 3 min followed by 40cycles of 94°C for 10 sec,62°Cfor 10 sec and 72°C for 15 sec. Real-time qPCR was performed to detect GAPDH expression with a SYBR Green PCR Kit( Applied Biosystems, Foster city, CA, USA) on a ROTOR-GENE 6000 Real-Time PCR System (ROTOR-GENE).

#### Statistical analysis

Statistical calculations were performed using the SPSS statistical software package (Version 13.0; SPSS, Inc. Chicago, IL) and R [Ref]. The measurement data were analyzed using one-way ANOVA. P<0.05 was considered statistically significant. Permutation test with 10,000 iterations was conducted when the distribution of the statistic was unknown. Chi-square test was used for the statistic inference for the categorical variables. Bonferroni correction was applied when multiple test were conducted. *t*-test was used to detect the mean difference for continuous variables in two conditions.

### Results

#### Genome-wide DNA methylation profile evealed by MethylCap-seq in PC genome

The whole genome methylation profiles of the PC and PN samples were successfully established using the MethylCap-seq method. 33,784,358 and 30,868,151 raw reads of 37 base pair were acquired in PC and PN groups, respectively. By alignment with the human genome (hg19) sequence, 16,267,025 (48.15%) raw reads in the PC group and 15,033,135 (48.70%) raw reads in the PN group were uniquely positioned. The reads mapping to 28,691 CGIs defined by UCSC were investigated. 3.57% of the reads in the PC group and 4.25% of the reads in the PN group were positioned at CGI and the CGI cover rates in the PC group and PN group were 64.31% and 64.36%, respectively. These data indicated that our experiment provided considerable and unbiased information regarding genomic CG islands.

In total, 276,442 and 255,743 hypermethylation peaks were found in the PC and PN groups, respectively, which indicating aberrant DNA methylation were occurred cross the whole PC genome (Figure 1A). Significant larger methylation regions were found in PC genomes than control normal genomes (permutation test, P<10-12) which is coincident with genome-wide hypomethylation in cancer genome [Ref]. An analysis of the hypermethylation peaks approximately 5 kb from the TSSs revealed that methylation peaks accumulated near TSSs, and more of these peaks were present in the PC group than in the PN group (Figure 1B). After deleting the peaks common to both PC and PN groups (approximately 209,000), 66,807 hypermethylated and 46,815 hypomethylated PC-specific DMRs were identified (Figure 1C). Of these PC-specific DMRs, 36,959 hyper-DMRs and 25,605 hypo-DMRs were overlapped with genes (Figure 1C), and 1,131 hyper-DMRs and 727 hypo-DMRs were associated with CGIs. The distribution of the DMRs for hypermethylation regions and hypo-DMR regions were significantly different in TSS flanking region, intragenic and intergenic region (permutation test, P-value<1.25×10-10, Figure 1D). Hyper-DMRs were more likely occurred in TSS flanking regions while hypo-DMRs were more likely presented in intragenic regions (permutation test, P-value < 0.62×10-7). No significant difference was found for the distribution of reference gene related DMRs in the elements of the human genome which included enhancer, promoter, exon, intron, miRNA, UTR3 and UTR5 (chi-square test, P-value = 0.2303, Figure 1E, up). However, the distribution of DMRs which located both in CGI and overlapped with reference gene had significantly higher frequency in promoter for hyper-DMRs than that for hypo-DMRs (chi-square test, P-value = 3.32×10-4). Hypo-DMRs were more likely located in miRNA region than that for hyper-DMRs (chi-square test, P-value = 5.73×10-5, Figure 1E, below). These evidences showed similar phenomena of hypomethylation in the genome-wide cancer methylome and locally hypermation in promoter regions. In present study, seven miRNAs were found significantly hypermethylated in PC which included mir-9-3 (P-value=\*\*), mir-9-1 (P-value=\*\*), mir-124-3 (P-value=\*\*), mir-10b (P-value=\*\*), mir-124-2 (P-value=\*\*), mir-718 , (P-value=\*\*) and mir-203 (P-value=\*\*) while six miRNAs were found significantly hypomethylated in PC which included mir-210 (P-value=\*\*), mir-1469 (P-value=\*\*), mir-130b (P-value=\*\*), mir-149 (P-value=\*\*), mir-1224 (P-value=\*\*), and mir-564 (P-value=\*\*, Table 3).

We could find that more than 88% and 87% of gene related hyper-DMR and hypo-DMR were derived from CpGI or CpGI shores in PC genomes. The methylation of CpGI and CpGI shore regions have been considered as the most important factors for gene expression regulation [Ref]. In present study, the methylation pattern of orphan CpG islands was focused. We found that orphan CGIs only accounted for 12% (133 in 1132, Table S3) or 13% (96 in 727, Table S4) of the CGI-related DMRs (Fig 1F). When promoter-related DMRs were considered, 5,341 such DMRs were identified; we found that the proportion of promoters with CGIs to promoters without CGIs was significantly higher (P-value=1.24×10-9) in the hyper-DMR group (21%; 609 in 2955, Table S5) than in the hypo-DMR group (13%, 312 in 2386, Table S6, Figure 1F).

Regarding the genes and the manner in which they were disturbed by hyper-DMRs, 527 genes were affected by a CGI (483 DMRs, Table S7) and a CGI shore (502 DMRs, Table S8), 120 genes (Table S9) were affected by a CGI only (108 DMRs), and 1,278 genes were affected by a CGI shore only (1242 DMRs, Table S10). In contrast, in analyzing the genes affected by hypo-DMRs, 333 genes were affected by both a CGI (305 DMRs, Table S11) and a CGI shore (315 DMRs, Table S12), 51 genes were affected by a CGI only (47 DMRs, Table S13), and 1,254 genes were affected by CGI shore only (1183 DMRs, Table S14 and Figure 1G).

Gene Ontology and KEGG pathway analysis of the aberrantly methylated genes to discover the network, interaction of the DMRs in the development of the PC. It would provide useful information for cancer research with the interview of the gene ontology and KEGG pathway map based on DNA methylation. We conducted a GO analysis of the genes that showed promoter hyper- or hypo-methylation in the PC group. Statistical significance of P<0.05 indicated gene enrichment in various GO categories (Table 2). We determined that 615 hypermethylated genes were enriched in the sequence-specific DNA binding (GO:0043565), neuron differentiation (GO:0030182), regulation of transcription and DNA-dependent (GO:0006355), or cell morphogenesis involved in differentiation (GO:0000904); 383 hypomethylated genes were enriched in plasma membrane part (GO:0044459), channel regulator activity (GO:0016247), positive regulation of BMP signaling pathway (GO:0030513), protein homo oligomerization (GO:0051260), or neuron differentiation (GO:0030182). We also identified 111 genes with hypermethylated promoters that were enriched in regulation of transcription term; among these genes, the methylation status of DLX4, ELAVL2, IRX1, PITX2, SIM2, TBX5, and TFAP2Cwassubsequently validated in tissue samples. KEGG pathway analysis of the above mentioned genes revealed that both hyper- and hypo- methylated genes were enriched in neuroactive ligand-receptor interaction (hsa04080, Table 2). These pathway and gene ontology has been reported to be highly related to the carcinogenesis [Ref,Ref, Ref], which indicating the methylation based KEGG and gene ontology analysis would be usefullness in the cancer basis research.

#### Verification of PC-specific DMRs identified by MethylCap-seq

The accuracy and precision of the profile were validation in two sample sets which included the samples that were pooled in the Methycap-seq library construction and another independent sample set. In the first validation vignette, the DNA methylation status of the 10 most significant DMRs which located in the promoter region (P-values<10-15, Supplementary Table \*\*) were evaluated in the same set of PC and PN samples used to create the MethylCap-seq libraries with BSP and MSP techniques, respectively. Four candidate DMR genes which included C5orf38 (P-value=\*\*), EMX1 (P-value=\*\*), NPR3 (P-value=\*\*), VSTM2B (P-value=\*\*), ELAVL2 (P-value=\*\*) and TFAP2C (P-value=\*\*) were validated to be significant hyper-methylated in PC than that in PN samples with BSP or MSP technique. Another two genes which included, AGAP2 (P-value=\*\*) and TRADD (P-value=\*\*) were found significantly hypo-methylated in PC groups with MSP technique. The failed of the gene A and gene B (see supplementary) showed the false positive differential methylation analysis which might be caused by biased sampling process in the library construction or the DNA sequencing.

In the second validation vignette, the methylation status of 20 gene-associated DMRs scattered in various genetic elements, such as promoters, miRNAs, introns, exons or CGI shores, were analyzed by MSP in the PC and PN testing group. The results revealed that 8 gene-associated DMRs which included \*\*,\*\*,\*\*,\*\*,\*\*and \*\*, exhibited hypermethylation in at least one of the 8 PC samples but in none of the 5 PN samples. An additional 9 gene-associated DMRs exhibited hypermethylation in at least 2 of the 8 PC samples but in only 1 of the 5 PN samples. DMRs in the promoter regions of TRADD, AGAP2, and FAM115A exhibited a loss of methylation in PC. The MSP results are presented in Table S15. We also found gene A, gene B, and gene c has been reported to be differential methylated in PC which were coincident with our results.

1, 把pvalue信息都加上; 2, 选择这些位点的原因可以加上

#### MSRE-qPCR and RT-qPCR validation of the methylation of orphan CGIs and related EST expression in PC cell lines treated with

To evaluate the role of the DNA methylation on the transcription regulation in PC genome, the methylation status of 10 hypermethylated DMRs (Figure 3A) in promoter CGIs and orphan CGIs was quantitatively analyzed by MSRE-qPCR in the 3 PC cell lines before and after 5-Aza-2’-deoxycytidine treatment. Result showed that the methylation levels of 4 orphan CGIs which located in 7p3q21, 8p3q21, 9p3q21, 10p3q21 and one CGI which located in the promoter region of Gene D were significantly decreased after the treatment of 5-Aza-2’-deoxycytidine in cell line A, B and C with the P-value of 0.005, 0.01, 0.01, 0.03, 0.02, respectively (Figure 3B and Supplementary Table or Figure). The expression levels of these transcripts were analyzed by RT-qPCR to ascertain the correlation between the methylation status and the corresponding mRNA expression. We found the transcription levels were significantly increased for the genes whose methylation level were decreased after 5-Aza-2’-deoxycytidine treatment (Figure 3C). A quantitative analysis of the methylation status of the above DMRs in an independent set of samples which included 8 PC samples, 5 PN samples and 3 PC cell lines, confirmed the significant differences between PC and PN samples, indicating the preliminary diagnostic role of the DNA methylation biomarker identified by MethyCap-seq technique (Figure 3D)

1, 把pvalue信息都加上

2, 选择这些位点的原因可以加上

### Discussion

Pooling strategies have been utilized innumerous previous genomic studies to investigate the phenotypic similarities in specific models, with the advantage of conserving samples [19,20]. In the present study, the genome-wide methylation profiles of PC vs.PN tissues were established using two pooled samples.

In the present study, we identified 5,280 hyper-DMRs and 3,488 hypo-DMRs that were closely associated with genes and CGIs in PC compared with PN. The GO analysis of the genes associated with these DMRs revealed that the aberrantly hypermethylated genes primarily belonged to categories related to nucleic acid binding, DNA binding and activation of transcription factors, suggesting that the methylation of the affected genes coupled with the down regulation of RNA expression resulted in the decreased expression of other genes. The research by Helman Eon NSCLC and that by Zhao M demonstrated that methylation-enriched genes exhibited aberrant methylation and RNA expression in multiple tumor types, and these genes were referred to as tumor suppressor genes [21]. The methylation-enriched genes associated with “cell morphogenesis involved in differentiation” (GO: 0000904) may participate in the morphological changes and late-stage differentiation of PC. In contrast, the hypomethylated genes were mostly enriched in plasma membrane part (GO:0044459) and channel regulator activity (GO:0016247), and these genes have been confirmed to be enriched in other tumor types in genomic methylation studies, suggesting that the abnormal formation of the plasma membrane may be a common feature in tumor differentiation and maintenance[22,23].

We also conducted a KEGG pathway analysis of the genes with DMRs in their promoter; surprisingly, both hyper- and hypo-methylated genes were enriched in the same pathway, “neuroactive ligand-receptor interaction”, which mainly participates in the endocrine and exocrine functions of cells. Functional abnormalities in these genes have been demonstrated in studies on meningioma and PC [24, 25]. Our present study revealedbi-directional methylation changes in this pathway in PC, but the detailed mechanism of action of this pathwayin PC development requires additional research (Table 2).

As an epigenetic factor, miRNAs play an important part in the regulation of cellular biophysical functions, and miRNAs have been shown to be correlated with the apoptosis, invasion, metastasis, recurrence and drug resistance of tumor cells. The regulation of miRNAs by DNA methylation has been extensively studied. In this study, we identified particular miRNAs that were aberrantly methylated in PC. Among these miRNAs, hsa-mir-124-3 has been shown to be hypermethylated in PC and is involved in the progression, metastasis and recurrence of PC [26, 27]. The mir-130b and mir-210 are two hypomethylated miRNAs that were discovered in this study, and bothof these miRNAs were found to be over expressed in PC and have been correlated with the proliferation and invasion of PC cells [28, 29]. Certain miRNAs that we identified as aberrantly methylated in PC have already been studied in other tumor types and are regulated by DNA methylation. For example, hypermethylated miRNAs in PC, such as miR-9-3,mir-9-1,miR-124 and miR-203, are hypermethylated in NSCLC, breast cancer,cervical cancer, and hematological cancer, respectively; hypermethylation decreases miRNA expression, which promotes tumor development and tumor cell proliferation [30,31,32,33].mir-10b, which was found to be hypermethylated in PC in this study, was thought to be an inhibitor of tumor metastasis in animal models [34] and was found to be over-expressed in PC[35]. Therefore, further studies of the remaining miRNAs that were aberrantly methylated in PC will shed light on the mechanisms of pancreatic carcinogenesis, although the related research is currently very limited.

The top 40 genes, based on p-value, that exhibited methylation changes in the promoter region were selected and examined by MSP in the testing group samples. Of these genes, 18 showed significant differences in methylation between the PC and PN tissues (and/or PC cell lines). Interestingly, among these genes, 7 (DLX4, ELAVL2, IRX1, PITX2, SIM2, TBX5 and TFAP2C) were enriched for the annotation of regulation of transcription (GO: 0006355), which corresponds with the results obtained in our previous GO analysis. All of the genes discussed above have been investigated in PC and other tumor types, and their involvement in carcinogenesis has been confirmed.

On both the gene and genome levels, many hypermethylated genes previously reported in PC studies were identified as hyper-DMRs in the present study, including LHX1, FOXE1, PAX6, BNIP3[36], ALPP, CEBPA[37], CACNA1G[38], CCND2[39], BAI1,NRN1, PENK, FAM84A,and ZNF415 [40]. In addition, our study also identified certain other genes that are frequently hypermethylated in different types of cancer, such as RASSF1a, CDKN2A, hHML1, and CDH1 [41, 42]. Hence, we have established a relatively intact database of abnormal methylation in PC. We also compared our data with those by Goggins [6], who analyzed 9 pairs of PC vs. PN samples using Human CpG Island Microarray 244k chips and obtained (after data filtering with proper thresholds) 1658 differentially methylated known loci. The comparison revealed the following: 1) Regarding the ability to capture aberrantly methylated gene targets, MethylCap-seq identified more targets than the array method in terms of both hypermethylated genes (1983 vs. 1206) and hypomethylated genes (1692 vs. 379) (Tables S16 and S17), indicating that methylated DNA fragment enrichment plus deep sequencing can uncover more aberrant gene loci, even though it is more labor-intensive and time-consuming. 2) Regarding the genes revealed by both MethylCap-seq and array, there were 737 total genes discovered by both methods, which accounts for 46.7% of the total genes recovered. This high recovery rate when comparing the two methods reflects the reliability of the methods for this purpose. However, the unique genes that were identified suggest that these two methods each have their own particular advantages. 3) The high percentage of commonality among the hyper-DMR genes (46.2%) and low percentage of commonality among the hypo-DMR genes (30 genes, 2.9% of all the hypo-DMR genes) between the studies by Goggins’ and us might indicate that during PC development, hypermethylation is a more definite and destined process, whereas hypomethylation is somewhat more random or perhaps stochastic. Adrian Bird divided CGIs into three categories, TSS, intragenic and intergenic, and the latter two categories were defined as orphan CGIs [43]. Despite little understanding of the functions of orphan CGIs until recently, studies have shown that orphan CGIs are involved in the regulation of gene transcription, gene printing and non-coding RNA transcription and that they might display tissue-specific methylation profiles [44]. In this study, we investigated the methylation status of particular orphan CGIs in PC. Hypermethylated orphan CGIs have been found in PC. The methylation status of orphan CGIs was closely associated with the transcription levels of nearby un-annotated ESTs. Further studies should be conducted to clarify whether the methylation-regulated ESTs containing orphan CGIs are potential genes or gene elements. It is well known that merely 6.8% of CpGs are located in CGIs. The methylation status and biological functions of the other 93.2% of CpGs have not been adequately studied. The study by Yu indicated that the methylation of CpGs in CGI shores was involved in regulating gene transcription or establishing tissue-specific methylation patterns, and changes in the methylation status of CpGs at these sites might take place at an earlier stage in carcinogenesis than the changes that occur in promoter CGIs [45]. The methylation changes in all the above regions in PC are indispensable parts of the PC genomic methylation profile and may influence the transcription of related genes and non-coding RNAs and affect tissue-specific cell differentiation, ultimately leading to carcinogenesis.

In summary, the genome-wide methylation profiles of PC vs. PN tissues were established using MethylCap-seq in the present study, revealing globally reprogrammed and deregulated DNA methylation in PC. Compared with PN tissues, there were massive PC-specific changes in terms of aberrant hypermethylation or hypomethylation of CpGs in TSS CGIs, orphan CGIs, CGI shores and promoter regions without CGIs. The above findings will be helpful for elucidating the mechanisms of pancreatic carcinogenesis related tothe DNA methylation-regulated expression of genes and non-coding RNAs. Furthermore, the aberrantly methylated genes in PC identified in this study may be potential biomarker resources for the early diagnosis and treatment of this deadly disease.

### Competing interests

The authors have declared that no competing interests exist.

### Author Contributions

Conceived and designed the experiments: YZ JY. Performed the experiments: YZ, JS, JG, WW, NT, XZ, JY. Analyzed the data: SG, JY, HZ. Contributed reagents/materials/analysis tools: JG, WW. Wrote the paper: YZ, SG, JS, JY.

### Reference:

[1] Jemal A, Siegel R, Ward E, Murray T, Xu J, et al. (2007) Cancer statistics, 2007. CA Cancer J Clin 57(1): 43-66．

[2] [Ma C](http://europepmc.org/search;jsessionid=RfZz6OEh4vvYuKopeE7p.20?page=1&query=AUTH:%22Ma+C%22), [Jiang YX](http://europepmc.org/search;jsessionid=RfZz6OEh4vvYuKopeE7p.20?page=1&query=AUTH:%22Jiang+YX%22), [Liu SZ](http://europepmc.org/search;jsessionid=RfZz6OEh4vvYuKopeE7p.20?page=1&query=AUTH:%22Liu+SZ%22), [Quan PL](http://europepmc.org/search;jsessionid=RfZz6OEh4vvYuKopeE7p.20?page=1&query=AUTH:%22Quan+PL%22), [Sun XB](http://europepmc.org/search;jsessionid=RfZz6OEh4vvYuKopeE7p.20?page=1&query=AUTH:%22Sun+XB%22), et al. (2013) Trend and prediction on the incidence of pancreatic cancer in China. Chinese Journal of Epidemiology 34(2): 160-163.

[3] Kanwal R, Gupta S. (2012) Epigenetic modiﬁcations in cancer. Clin Genet 81: 303–311.

[4] Klump B, Hsieh C J, Nehls O, Dette S, Holzmann K, et al. (2003) Methylation status of p14ARF and p16INK4a as detected in pancreatic secretions. Br. J. Cancer 88: 217–222.

[5] Dammann R, Schagdarsurengin U, Liu L, Otto N, Gimm O, et al. (2003) Frequent RASSF1A promoter hypermethylation and K-ras mutations in pancreatic carcinoma. Oncogene 22: 3806–3812.

[6] Omura N, Li C P, Li A, Hong S M, Walter K, et al. (2008) Genome-wide profiling of methylated promoters in pancreatic adenocarcinoma. Cancer Biol Ther 7: 1146–1156.

[7] Vincent A, Omura N, Hong SM, Jaffe A, Eshleman J, et al. (2011) Genome-wide analysis of promoter methylation associated with gene expression profile in pancreatic adenocarcinoma. Clin Cancer Res 17(13): 4341-4354.

[8] Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, et al. (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. Nature Biotechnology 28: 1106–1114.

[9] Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41: 178–186.

[10] Doi A, Park IH, Wen B, Murakami P, Aryee MJ, et al. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 41: 1350–1353.

[11] Zhao Y, Guo S, Huang Z, Sun J, Zhu T, et al. (2012) Methylcap-seq Reveals Novel DNA Methylation Markers for the Diagnosis and Recurrence Prediction of Bladder Cancer in a Chinese Population. PLoS One DOI: 10.1371/journal.pone.0035175

[12] Li H, Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14): 1754-1760.

[13] Kuhn RM, Haussler D, Kent WJ. (2012) The UCSC genome browser and associated tools. Briefings in bioinformatics 14(2): 144-161.

[14] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-2079.

[15] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9: R137.

[16] Lan X, Adams C, Landers M, Dudas M, Krissinger D, et al. (2011) High resolution detection and analysis of CpG dinucleotides methylation using MBD-Seq technology. PloS one 6(7):e22226.

[17] Quinlan AR, Hall IM. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6): 841-842.

[18] Huang ZH, Hu Y, Hua D, Wu YY, Song MX, et al. (2011) Quantitative analysis of multiple methylated genes in plasma for the diagnosis and prognosis of hepatocellular carcinoma. Exp Mol Pathol 91(3): 702-707.

[19] Li LM, Hu ZB, Zhou ZX, Chen X, Liu FY, et al. (2010) Serum microRNA profiles serve as novel biomarkers for HBV infection and diagnosis of HBV-positive hepatocarcinoma. Cancer Res 70(23): 9798-9807.

[20] Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, et al. (2011) Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res 21 (7): 1074-1086.

[21] Zhao M, Sun J, Zhao Z. (2013) TSGene: a web resource for tumor suppressor genes. Nucleic Acids Res 41(Database issue):D970-6. doi: 10.1093/nar/gks937.

[22] Cicek MS, Koestler DC, Fridley BL, Kalli KR, Armasu SM, et al. (2013) Epigenome- wide ovarian cancer analysis identifies a methylation profile differentiating clear-cell histology with epigenetic silencing of the HERG K+ channel. Hum Mol Genet 22(15): 3038-47. doi: 10.1093/hmg/ddt160.

[23] Sriraksa R1, Zeller C, Dai W, Siddiq A, Walley AJ, et al. (2013) Aberrant DNA methylation at genes associated with a stem cell-like phenotype in cholangiocarcinoma tumors. Cancer Prev Res (Phila) 6(12): 1348-55. doi: 10.1158/1940-6207.CAPR-13-0104. Epub 2013 Oct 2.

[24] Wei P, Tang H, Li D. (2012) Insights into pancreatic cancer etiology from pathway analysis of genome-wide association study data. PLoS One 7(10): e46887. doi: 10.1371/ journal.pone.0046887.

[25] Keller A, Ludwig N, Backes C, Romeike BF, Comtesse N, et al. (2009) Genome wide expression profiling identifies specific deregulated pathways in meningioma.Int J Cancer 124(2): 346-351.

[26] Wang P, Chen L, Zhang J, Chen H, Fan J, et al. (2013) Methylation-mediated silencing of the miR-124 genes facilitates pancreatic cancer progression and metastasis by targeting Rac1. Oncogene doi: 10.1038/onc.2012.598.

[27] Gebauer K, Peters I, Dubrowinskaja N, Hennenlotter J, Abbas M, et al. (2013) Hsa-mir-124-3 CpG island methylation is associated with advanced tumours and disease recurrence of patients with clear cell renal cell carcinoma. Br J Cancer 108(1):131-138. doi: 10.1038/bjc.2012.537

[28] Zhao G, Zhang JG, Shi Y, Qin Q, Liu Y, et al. (2013) MiR-130b Is a Prognostic Marker and Inhibits Cell Proliferation and Invasion in Pancreatic Cancer through Targeting STAT3. PLoS One doi: 10.1371/journal.pone.0073803.

[29] Takikawa T, Masamune A, Hamada S, Nakano E, Yoshida N, et al. (2013) miR-210 regulates the interaction between pancreatic cancer cells and stellate cells. Biochem Biophys Res Commun 437(3): 433-9. doi: 10.1016/j.bbrc.2013.06.097.

[30] Heller G, Weinzierl M, Noll C, Babinsky V, Ziegler B, et al. (2012) Genome-wide miRNA expression profiling identifies miR-9-3 and miR-193a as targets for DNA methylation in non-small cell lung cancers. Clin Cancer Res 18(6): 1619-1629.

[31] Lehmann U, Hasemeier B, Christgen M, Müller M, Römermann D, et al. (2008) Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. J Pathol 214(1): 17-24.

[32] Wilting SM, van Boerdonk RA, Henken FE, Meijer CJ, Diosdado B, et al. (2010) Methylation-mediated silencing and tumour suppressive function of hsa-miR-124 in cervical cancer. Mol Cancer 9: 167. doi: 10.1186/1476-4598-9-167.

[33] Chim CS, Wong KY, Leung CY, Chung LP, Hui PK, et al. (2011) Epigenetic inactivation of the hsa-miR-203 in haematological malignancies. J Cell Mol Med 15(12): 2760-2767. doi: 10.1111/j.1582-4934.2011.01274.x.

[34] Ma L, Reinhardt F, Pan E, Soutschek J, Bhat B, et al. (2010) Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. Nat Biotechnol 28: 341-347. doi: 10.1038/nbt.1618.

[35] Nakata K, Ohuchida K, Mizumoto K, Kayashima T, Ikenaga N, et al. (2011) MicroRNA-10b is overexpressed in pancreatic cancer, promotes its invasiveness, and correlates with a poor prognosis. Surgery 150(5): 916-922. doi: 10.1016/j.surg.2011.06.017.

[36] Sato N, Fukushima N, Maitra A, Matsubayashi H, Yeo CJ, et al. (2003) Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. Cancer Res 63:3735–3742.

[37] Kumagai T, Akagi T, Desmond JC, Kawamata N, Gery S, et al. (2009) Epigenetic regulation and molecular characterization of C/EBPalpha in pancreatic cancer cells. Int J Cancer 124: 827–833.

[38] Ueki T, Toyota M, Sohn T, Yeo CJ, Issa JP, et al. (2000) Hypermethylation of multiple genes in pancreatic adenocarcinoma. Cancer Res 60:1835–1839.

[39] Matsubayashi H, Sato N, Fukushima N, Yeo CJ, Walter KM, et al. (2003) Methylation of cyclin D2 is observed frequently in pancreatic cancer but is also an age-related phenomenon in gastrointestinal tissues. Clin Cancer Res 9: 1446–1452.

[40] Omura N, Li CP, Li A, Hong SM, Walter K, et al. (2008) Genomewide profiling of methylated promoters in pancreatic adenocarcinoma. Cancer Biol Ther 7: 1146–1156.

[41] Omura N, Goggins M. (2009) Epigenetics and epigenetic alterations in pancreatic cancer. Int J Clin Exp Pathol 2: 310–326.

[42] Schutte M, Hruban RH, Geradts J, Maynard R, Hilgers W, et al. (1997) Abrogation of the Rb/p16 tumor-suppressive path way in virtually all pancreatic carcinomas. Cancer Res 57: 3126–3130.

[43] Deaton AM and Bird A. (2011) CpG islands and the regulation of transcription. Genes & Dev 25: 1010-1022.

[44] Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, et al. (2010) Orphan CpG islands identify numerous conserved promoters in mammalian genome. PLoS Genet DOI: 10.1371/journal.pgen.1001134

[45] Yu W, Jin C, Lou X, Han X, Li L, et al. (2011) Global analysis of DNA methylation by Methyl-Capture sequencing reveals epigenetic control of cisplatin resistance in ovarian cancer cell. PLoS One 6(12): e29450. doi:10.1371/ journal. pone.0029450.

### Figure legends

**Figure 1. Data mining of MethylCap-seq libraries.** A) Chromosomal view of genome-wide hyper- and hypo-methylated DNA in PC compared with PN. Red bar: hypermethylation; Green bar: hypomethylation. B)Hypermethylated peaks around the TSS site in PC compared with PN. Peaks were surveyed in a broad region (from 5 kb downstream to 5 kb upstream of the TSS). C)Differentially methylated regions (DMRs) that were specific for PC. The DMRs are shown according to their inclusion in refGene or CGI definitions. D)The genomic distribution of hyper-DMRs and hypo-DMRs in TSSs, intragenic regions, and intergenic regions. The total number of DMRs is presented at the top of each graph.E) Pie chart of the DMR distribution over the various gene structures based on sole refGene involvement versus both CGI and refGene involvement in PC-specific hyper- and hypo-methylated regions. The genomic context is defined as those found in the UCSC database.F) Pie chart of the PC-selective hyper- and hypo-DMR distribution over orphan CGIs or refGene-related CGIs and over CGI-containing promoters vs. no-CGI promoters. G)Hyper- and hypo-DMRs in PC and their related genes, considering the involvement of various CGI features (CGI, CGI shore, or both CGI and CGI shore).

**Figure 2. Representative results of the BSP and MSP validation of the MethylCap-seq data.** For each gene, the UCSC scheme of the gene locus and the examined promoter regions are shown. A) BSP results. PC, pancreatic cancer; PN, adjacent non-tumor tissue. B) MSP results. 3 PC cell lines (BxPc-3, CFPAC-1, and PANC-1), 8 PC samples (1C, 2C, 5C, 6C, 7C, 8C, 10C, and 311C) and 5 PN samples (1N, 2N, 7N, 10N, and 311N) were evaluated. C: GAPDH-BSP quality and quantity control for the confirmation of bisulfite-converted DNA templates. P: positive control; N: negative control.

**Figure 3. Methylation of CGIs (orphan CGIs or regular CGIs) might influence the expression of putative ESTs or mRNAs.** A)The UCSC scheme of CGIs and the nearby putative ESTs or mRNAs analyzed in this study. B)DNA methylation changes in PC cell lines after treatment with 5-Aza. The GAPDH-BSP product serves as a quality and quantity control for the bisulfite-converted DNA templates.C) EST expression after 5-Aza treatment determined by RT-qPCR. GAPDH mRNA expression was the loading control. D)Box-plot of the quantitative analysis of DNA methylation by MSRE-qPCR in 8 PC and 5 PN samples. P, positive control; N, negative control. The box is defined by 25%/75% quantiles. The methylation levels in the PC and PN samples were compared by one-way ANOVA, and the p-values are indicated.

### Tables

### Table 1.Clinical profile of the PC patients recruited in the present study

****PC: pancreatic carcinoma. PN: PC adjacent non-tumor tissue.

**Table 2. Gene ontology enrichment analysis of aberrant methylation in gene promoters in PC**

****

**Table 3. Aberrantly methylated miRNAs in PC**

****

\* logarithmic transformation of the P-values to show the significance levels of the differentially methylated regions (In DMR estimation) or methylation blocks (In MACS). Ahigher value indicates a higher probability of a DMR or a methylation block.