New DNA Methylation Markers for Pancreatic Cancer: Discovery, Tissue Validation, and Pilot Testing in Pancreatic Juice

John B. Kisiel1, Massimo Raimondo2, William R. Taylor1, Tracy C. Yab1, Douglas W. Mahoney3, Zhifu Sun3, Sumit Middha3, Saurabh Baheti3, Hongzhi Zou4, Thomas C. Smyrk5, Lisa A. Boardman1, Gloria M. Petersen6, and David A. Ahlquist1

Abstract

**Purpose:** Discriminant markers for pancreatic cancer detection are needed. We sought to identify and validate methylated DNA markers for pancreatic cancer using next-generation sequencing unbiased by known targets.

**Experimental Design:** At a referral center, we conducted four sequential case–control studies: discovery, technical validation, biologic validation, and clinical piloting. Candidate markers were identified using variance-inflated logistic regression on reduced-representation bisulfite DNA sequencing results from matched pancreatic cancers, benign pancreas, and normal colon tissues. Markers were validated technically on replicate discovery study DNA and biologically on independent, matched, blinded tissues by methylation-specific PCR. Clinical testing of six methylation candidates and mutant KRAS was performed on secretin-stimulated pancreatic juice samples from 61 patients with pancreatic cancer, 22 with chronic pancreatitis, and 19 with normal pancreas.

**Results:** Sequencing identified >500 differentially hyper-methylated regions. On independent tissues, AUC on 19 selected markers ranged between 0.73 and 0.97. Pancreatic juice AUC values for CD1D, KCNK12, CLEC11A, NDRG4, IZKF1, PKRCB, and KRAS were 0.92, 0.88, 0.85, 0.84, 0.83, and 0.75, respectively, for pancreatic cancer compared with normal pancreas and 0.92, 0.73, 0.76, 0.85, 0.73, 0.77, and 0.62 for pancreatic cancer compared with chronic pancreatitis (P = 0.001 vs. KRAS).

**Conclusions:** We identified and validated novel DNA methylation markers strongly associated with pancreatic cancer. On pilot testing in pancreatic juice, best markers (especially CD1D) highly discriminated pancreatic cases from controls. *Clin Cancer Res; 21(19); 4473-81. ©2015 AACR.*

Introduction

Incidence and mortality rates of pancreatic cancer continue to rise in the face of declining trends for other major cancers (1). Some forecast that pancreatic cancer will become the second most fatal cancer in the United States before 2020 (2). Undermining its extraordinarily high lethality, more than 46,000 Americans will be diagnosed with pancreatic cancer and nearly 40,000 will succumb this year (1). Better approaches to pancreatic cancer control are urgently needed.

Although population screening is currently not practiced, there are strong biologic and clinical justifications to explore early detection. Recent studies on the molecular epidemiology of pancreatic carcinogenesis suggest slow rates of progression from premalignant neoplasms to cancer and from earliest stage cancer to metastatic disease (3). Such long latency periods provide a window of opportunity for detection and curative treatment of presymptomatic precursor lesions or earliest stage pancreatic cancer. Indeed, incidentally discovered early-stage pancreatic cancers have the best reported cure rates (4, 5). Precursor lesions, including pancreatic intraepithelial neoplasm (PanIN), intraductal papillary mucinous neoplasm (IPMN; ref. 6), and pancreatic cancer are associated with molecular alterations (7–9) that could potentially serve as markers for early detection and screening. Pancreatic neoplasms exfoliate cells and DNA into local effluent and ultimately stool. We and others have detected both genetic and epigenetic markers in pancreatic juice (10, 11) and stool (12–15) from patients with pancreatic cancer and precursor lesions. A limitation with mutation markers relates to the unwieldy process of their detection; typically, numerous mutations across several genes must be assayed separately to achieve high sensitivity. In addition, some mutations common in pancreatic cancer may not be sufficiently specific; for example, mutant KRAS is frequently observed in chronic pancreatitis (16). Methylation of DNA at cytosine–phosphate–guanine (CpG) island sites provides marker candidates that are more broadly informative and sensitive than individual DNA mutations and may offer excellent specificity, as we have seen with stool DNA testing for colorectal cancer (17, 18).
Identification of screening markers which are both highly sensitive and highly specific can be challenging. We have observed that methylation markers discriminant in primary tumor tissues often fail when assayed in an intended medium, such as stool (15). Ideal candidate markers for pancreatic cancer screening would be universally present in pancreatic neoplasms, be absent in normal gastrointestinal mucosa, and have high signal strength.

Several methods are available to search for novel methylation markers. Microarray-based interrogation of CpG methylation is a high-throughput approach, but is biased toward known regions of interest, mainly the promoters of established tumor suppressor genes (19). Alternative methods for genome-wide analysis of DNA methylation have been developed in the last decade (20). Next-generation sequencing has provided important insights into the epigenetic regulation of gene expression in various cancers (21–23). Although whole-exome sequencing has been used to study mutations in pancreatic neoplasms (24), we are unaware of any methylome-wide search for pancreatic cancer screening markers using a next-generation sequencing approach.

We hypothesized that (1) a whole-methylome search by reduced representation bisulfite sequencing (RRBS; ref. 25) would identify novel methylation markers that would discriminate pancreatic cancer from benign pancreatic tissues and have low background levels in other gastrointestinal epithelia and (2) discovered markers would accurately detect pancreatic cancer by assay of pancreatic juice.

Materials and Methods

Study overview

Four sequential case–control studies were conducted. In the first three tissue-based studies, we aimed to (1) discover novel and highly discriminant methylation markers for pancreatic cancer using RRBS; (2) technically confirm these findings using methylation-specific PCR (MSP), a more agile assay system; and (3) biologically validate top candidate markers in an independent, matched tissue sample set. In the fourth study, we clinically pilot-tested selected candidates on archival pancreatic juice samples using quantitative MSP (qMSP) and quantitative real-time allele-specific target and signal amplification (QuARTS) assay of mutant KRAS. All components of this investigation were approved by our Institutional Review Board.

Translational Relevance

Pancreatic cancer mortality is rising; screening tests are urgently needed. Assay of molecular markers in distant media, such as pancreatic juice, stool, urine, or blood, is a rational but nascent approach to early detection. Next-generation sequencing, an unbiased marker discovery technique, is largely unexplored in pancreatic cancer. From >6 million CpGs genome-wide, top markers achieved high discrimination in pancreatic cancer tissues and were validated in independent samples. In pancreatic juice samples, methylated DNA markers were highly sensitive and specific, even against chronic pancreatitis controls, and superior to mutant KRAS. Known tumor suppressors were among methylated genes discovered but, more importantly, RRBS revealed novel candidates without previously reported roles in cancer biology. Methylation DNA markers held promise in noninvasive tools for pancreatic cancer detection from stool or blood. Assay of these markers from pancreatic juice by duodenal aspiration at esophagogastroduodenoscopy could complement imaging in evaluation of pancreatic masses or cystic neoplasms.
Clinical pilot testing. Selected methylated candidates and mutant KRAS were assayed by qMSP and QuARTS, respectively, on DNA extracted from blinded pancreatic juice samples, collected via simple duodenal luminal aspiration following a 16-microgram intravenous dose of secretin (ChirBoClin), as previously described (27). Pancreatic juice samples were prospectively collected at Mayo Clinic, Jacksonville, Florida, from March 1, 2012 to November 1, 2012. Patients were enrolled prospectively at the time of routine endoscopic ultrasound (EUS) or esophagogastroduodenoscopy (EGD) into one of three groups: those with pain suggestive of pancreatic disease; those suspected of having pancreatic cancer; and, those undergoing diagnostic EGD without suspicion of pancreatic disease or cancer. The latter group received EUS for research purposes. Patients were excluded if they could not provide informed consent or for prior gastric, pancreatic, or duodenal resection. Pancreatic cancer or main-duct IPMN diagnoses were confirmed by histopathology; chronic pancreatitis and normal-appearing pancreas diagnoses were confirmed by magnetic resonance imaging and EUS. Juice was rapidly placed in 2-mL vials, immediately snap-frozen in liquid nitrogen, and stored at −80°C.

Reduced representation bisulfite sequencing

Library preparation. Genomic DNA (300 ng) was fragmented by digestion with 10 units of MSPI, a methylation-specific restriction enzyme that recognizes CpG-containing motifs, to enrich sample CpG content and eliminates redundant areas of the genome (25). Digested fragments were end-repaired and A-tailed with 5 units of Klenow fragment (3′–5′ exo−), and ligated overnight to methylated TruSeq adapters (Illumina) containing barcode sequences (to link each fragment to its sample ID.) Size selection of 160 to 340 bp fragments (40–220 bp inserts) was performed using Agencourt AMPure XP SPRI beads/buffer (Beckman Coulter). Buffer cutoffs were 0.7 × 1.1 × sample volumes of beads/buffer. Final elution volume was 22 μL (EB buffer; Qiagen); qPCR was used to gauge ligation efficiency and fragment quality on a small sample aliquot. Samples then underwent bisulfite conversion (twice) using a modified EpiTect protocol (Qiagen). qPCR and conventional PCR (PfuTurbo Cx hotstart; Agilent) followed by Bioanalyzer 2100 (Agilent) assessment on converted sample aliquots determined the optimal PCR cycle number before final library amplification. The following conditions were used for final PCR: (i) each 50-μL reaction contained 5 μL of 10x buffer, 1.25 μL of 10 mmol/L each deoxyribonucleotide triphosphate (dNTP), 5 μL primer cocktail (~5 μmol/L), 15-μL template (sample), 1 μL PfuTurbo Cx hotstart and 22.75 water; temperatures and times were 95°C for 5 minutes; 98°C for 30 seconds; 16 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds, 72°C for 5 minutes, and 4°C hold, respectively. Samples were combined (equimolar) into 4-plex libraries based on the randomization scheme and tested with the bioanalyzer for final size verification, and with qPCR using phiX standards and adaptor-specific primers.

Sequencing and bioinformatics. Samples were loaded onto flow cells according to a randomized lane assignment with additional lanes reserved for internal assay controls. Sequencing was performed by the Next Generation Sequencing Core at the Mayo Clinic Medical Genome Facility on the Illumina HiSeq 2000. Reads were unidirectional for 101 cycles. Each flow cell lane generated 100 to 120 million reads, sufficient for a median coverage of 30- to 50-fold sequencing depth (read number per CpG) for aligned sequences. Standard Illumina pipeline software called bases and sequenced read generation in the fastq format. As described previously, (28) SAAP-RRBS, a streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing, was used for sequence alignment and methylation extraction.

MSP primer design. Primers for each marker were designed to target the bisulfite-modified methylated sequences of each target gene (IDT) and a region without CpG sites in the ACTB gene, a reference of bisulfite treatment and DNA input. The design was done by either Methprimer software (University of California, San Francisco, CA) or by semi-manual methods. Assays were tested and optimized by qPCR with SYBR Green (Life Technologies) dyes on dilutions of universally methylated and unmethylated genomic DNA controls.

Methylation-specific PCR. Quantitative MSP reactions were performed on tissue-extracted DNA as previously described (15). Additional specifications are provided in the Supplementary Methods.

Quantitative allele-specific real-time target and signal amplification. QuARTS assays were used for KRAS assays as previously published (29). Briefly, KRAS was PCR amplified with primers flanking codons 12/13 using 10-μL captured KRAS DNA templates. QuARTS assays then evaluated seven mutations at codons 12 and 13. Each QuARTS reaction incorporated primers, detection probes, an invasive oligo, FAM (Hologic), Yellow (Hologic), Quasar 670 (BioSearch Technologies) fluorescence energy transfer reporter cassettes (FRET), Cleavase 2.0 (Hologic), GoTaq DNA polymerase (Promega), MOPS buffer, MgCl₂, and deoxyribonucleotide triphosphate (dNTP). Plates contained standards made of engineered plasmids, ± controls, and water blanks, and were run in a LightCycler 480 (Roche).

Both methylated candidates and mutant KRAS copy numbers per sample were calculated in reference to standard curves. In the qMSP and QuARTS reactions, any sample for which at least 50 copies of ACTB were measured was included in the analysis. Any PCR product which amplified in reactions with primers and probes directed at the methylated or mutant target sequence was quantified by fluorescence values in relationship to the 1:5 serially diluted reference standards. The following conditions were used: 5,000, 1,000, 200, 40, 8 and 1.6 copies per well, respectively. For values below the analytical threshold, the copies per sample were assigned a value of 1 copy to normalize results for all samples by ACTB copy number or concentration, respectively.

Statistical analysis

Overall approach. Candidate CpGs were filtered by a priori read-depth and variance criteria, significance of differential methylation percentages between cases and controls and discrimination of cases from controls based on AUC and target-to-background ratio.

For the RRBS discovery phase, the primary comparison of interest was the methylation difference between cases, pancreatic controls, and colon controls at each mapped CpG. CpG islands are biochemically defined by an observed to expected CpG ratio >0.6 (30). However, for this model, tiled units of CpG analysis “differentially methylated region (DMR)” were created based on distance between CpG site locations for each chromosome.
islands with only single CpGs were excluded. Individual CpG sites were considered for differential analysis only if the total depth of coverage per disease group was $\geq$ 200 reads (an average of 10 reads/subject) and the variance of %-methylation was $>0$ (non-informative CpGs were excluded). To estimate the sample size required per group for DNA sequencing, we assumed a minimum read depth of 10 reads per sample and that the primary comparison is between normal tissues (pancreatic and colon) and cancer tissues. The highest background for the average normal tissue methylation was assumed to be 5% and a 3-fold increase in the odd ratio was deemed as the minimum effect difference that is biologically relevant. At the minimum depth of cover of 10 reads, a minimum of 18 samples per group was required to achieve 80% power with a two-sided test at a significance level of 5% and assuming binomial variance inflation factor of 1. As the estimated variance inflation factor increases, the power drops with only 18 subjects per group. However, we accepted sites with a minimum read depth of 20 to maintain sufficient power, across all inflation factors, for the given sample size under these assumptions.

Statistical significance was determined by logistic regression of the methylation percentage per DMR, based on read counts. To account for varying read depths across individual subjects, an overdispersed logistic regression model was used, where dispersion parameter was estimated using the Pearson $\chi^2$ statistic of the residuals from fitted model. DMRS, ranked according to their significance level, were further considered if methylation in benign pancreas and colon controls, combined, was $\leq$ 1% but $\geq$ 10% in pancreatic cancer.

For the validation and feasibility studies, the primary outcome was AUC for each marker, as calculated from the concentration-corrected copies per sample of each marker with pancreatic cancer in comparison with normal pancreas and normal colon. For the technical and biologic validation phases, 17 patients per group provided 80% power to distinguish an AUC of 0.85 from a null hypothesis of 0.50 in a one-sided test at the 0.05 level of significance. After technical validation confirmed AUC $>0.85$, a quantitative difference in median values of candidate copies per sample between cases and controls of at least 10-fold was used to select markers for biologic validation and pancreatic juice testing.

Pancreatic juices were convenience samples from an existing archive from which all samples were analyzed. AUC values of the each methylation marker in juice were compared with that of mutant KRAS in the same samples. The method of Delong, DeLong and Clarke-Pearson (31) was used to compare AUCs and measure significance of differences. A Bonferroni correction was used to avoid bias from multiple comparisons, establishing a significance threshold $P$ value of $<0.008$. Samples included 61 cancers and two control groups of approximately 20. Using the normal approximation of the AUC, the sample number per group was used to determine the variance of the statistic. With this approximation and assuming a one-sided significance level of 0.05 and 80% power, the minimum detectable AUC for the feasibility study was 0.70 in comparisons with the null hypothesis of 0.5. When comparing any two markers the paired variance was estimated assuming a low correlation between markers of 0.3 and a moderate correlation of 0.6. Assuming a one-sided significance level of 0.008 with 80% power, the minimum detectable difference between any paired markers was 0.32 (0.60 vs. 0.92) for a correlation of 0.3 between markers and 0.29 (0.60 vs. 0.89) for a correlation of 0.6 between markers. Regression models also tested the influence of age, sex, clinical tumor stage (T1 and 2 vs. T3 and 4, determined by endoscopic ultrasound) and tumor location (head vs. body/tail) on the strength of association between marker levels and case or control status.

The point-value, in copies per sample, for each marker was identified at the false positive rate-based cutoffs of 5% and 10% among normal pancreas controls and used to estimate marker sensitivity and 95% confidence intervals (CI) for pancreatic cancer in separate comparisons with normal pancreas and chronic pancreatitis.

## Results

### RRBS marker discovery

DNA extracts from 54 tissue samples (18 pancreatic cancer tumors, 18 benign pancreatic control tissues, and 18 normal colonic epithelial) were sequenced by RRBS (Fig. 1). Median age was 61 (interquartile range, 52–65), 61% were women, and 44% were current or former smokers. A total of 6,101,049 CpG sites were captured in any of the samples with at least 10× coverage. After selecting CpG sites where group coverage and variance criteria were met, a total of 1,217,523 CpG sites were further analyzed. Approximately 500 DMRs met significance criteria. Among these, we identified 87 candidate regions with sufficient methylation signatures for MSP primer design. Methylation signatures ranged from 3 to 52 neighboring CpGs. Methylation levels in pancreatic cancer samples were typically below 25%, reflecting the common contamination by stromal cells. The degree of stromal cell contamination could be quantified indirectly by KRAS testing; among pancreatic cancer specimens that harbored a heterozygous KRAS base change, the frequency of the mutant allele was generally four times less than the corresponding wild-type allele (Supplementary Fig. S1).

### Technical validation

After primer design, MSP assayed the 87 candidates in samples of DNA from an additional 20 unblinded pancreatic cancer lesions, 10 additional normal colonic epithelial samples (biologic replicates) as well as remaining DNA samples from the 18 sequenced pancreatic cancer lesions, 15 of the sequenced benign pancreatic tissues and 10 of the sequenced normal colon samples (technical replicates). ACTB was amplified in all samples. With either first or second-pass MSP, 38 of 87 candidate markers had an AUC $>0.85$ (Fig. 2; Supplementary Table S1). RRBS-identified candidates were compared with two published reports of pancreatic cancer methylation measured by microarray (8, 32). RRBS candidate pool was corroborated and comparably informative; however, 10 of the 38 top candidates were novel genes, not identified by hybridization array methods.

### Biologic validation

Based on the magnitude of difference in median copies per sample between cases and controls for each candidate marker, ABCB1, ADCY1, AK055957, BMP3, C13ORF18, CAGNA1C, CD1D, CLEC11A, ELOMO1, FOX2, GRIN2D, HKF1, KCNK12, KCN2, NDRG4, PRKC8R, RSPO3, SCARF2, SHH, SLC38A3, TWIST1, VWC3, and WT1 were selected for validation in independent, matched, blinded, randomly allocated DNA from 72 tissue samples. These included 18 pancreatic cancers, 18 benign pancreas tissues, and 36 normal colon epithelia. The median age of this subset was 60 (interquartile range, 54–64). The majority (55%) of samples came from men and 61% were current or
former smokers. ACTB was amplified in all samples. As shown (Fig. 3), candidates were strongly associated with pancreatic cancer in comparison with benign pancreatic and colonic controls, combined. The individual AUC values (and 95% CI) for AK055957, WT1, GRIN2D, CACNA1C, ELMO1, KCNN2, CD11D, TWIST1, C13ORF18, and CLEC11A were outstanding at 0.97 (0.92–1), 0.97 (0.93–1), 0.97 (0.93–1), 0.95 (0.91–1), 0.95 (0.9–1), 0.94 (0.88–1), 0.94 (0.86–1), 0.93 (0.83–1), 0.93 (0.84–1), and 0.93 (0.84–1), respectively. Excellent association was seen with nine other candidates with AUC values for RSPO3, PRKCB, KCNK12, SLC38A3, SHH, VWC2, SCARF2, and ADCY1 of 0.92 (0.85–0.98), 0.91 (0.81–1), 0.91 (0.83–0.98), 0.89 (0.78–1), 0.88 (0.77–0.99), 0.87 (0.73–1), 0.86 (0.74–0.98), and 0.85 (0.69–1).

The majority of novel candidates showed excellent signal to noise ratios. Specifically, for 10 candidate markers, methylated copy numbers were more than 30-fold higher among cases, compared with controls. For AK055957, KCNK12, ADCY1, ELM01, and PRKCB, copy numbers of methylated candidates were more than 100-fold greater in cases, compared with controls (Fig. 3; Supplementary Table S2). The biologically validated DMRs were compared with an open-access published dataset from the International Cancer Genome Consortium (ICGC). This set included 167 pancreas cancer and 29 control tissues in which DNA methylation was interrogated by Infinium Human Methylation 450K BeadChips (Illumina; ref. 32). All RRBS-derived biologically validated genes were corroborated by the ICGC results. The published sequences of the CpG probe sets for each annotated gene were compared with the coordinates and sequences for the RRBS-derived DMRs (Supplementary Table S2). Of 19 RRBS-derived DMRs, 13 had no sequence overlaps with the 450K probes. Of the remaining six, the RRBS-derived DMRs had at least one novel CpG, not contained in the list of significant probes reported for the hybridization array method.

**Pilot testing in pancreatic juice**

At the time of the pancreatic juice pilot, the full biologic validation analysis had not been completed. Six candidate methylation markers reflecting a range of AUC values and signal to noise ratios of at least 10 were chosen for feasibility testing in pancreatic juice. All 102 pancreatic juice samples from a preexisting freezer archive were tested and included 61 patients with pancreatic cancer, 22 with chronic pancreatitis, and 19 with normal pancreas (Table 1). ACTB was amplified in all samples.
Samples were available on only 3 patients with main duct IPMN; due to insufficient power, these were not included in regression analyses. Median age (range) for pancreatic cancer patients was 67 (IQR, 61–76), slightly older than those for chronic pancreatitis and normal pancreas patients at 64 (IQR, 53–72) and 60 (IQR, 49–70), respectively (P = 0.02). Whereas the majority of pancreatic cancer and chronic pancreatitis patients were men (58% and 68%, respectively), most normal pancreas patients (79%) were women (P = 0.007). A higher percentage of normal pancreas patients (68%) were never smokers, compared with pancreatic cancer (50%) and chronic pancreatitis (32%) groups (P = 0.04). Of pancreatic cancer cases, 16 (26%) were EUS T-stage 1 and 2 and 43 (71%) were located in the head of the pancreas.

For the detection of pancreatic cancer in comparison with normal pancreas, the AUC values for CD1D, KCNK12, CLEC11A, NDRG4, IZF1, PKRCB, and KRAS were 0.92 (0.86–0.98), 0.73 (0.61–0.86), 0.76 (0.64–0.87), 0.85 (0.76–0.94), 0.73 (0.61–0.86), 0.77 (0.65–0.89), and 0.62 (0.49–0.74), respectively. CD1D was the most discriminant individual marker for detection of pancreatic cancer in comparison with normal pancrea or chronic pancreatitis (Fig. 4, Supplementary Figs. S2–S7) and was significantly more discriminant than mutant KRAS (P = 0.001). From specificity cutoffs determined in normal pancreas patients, CD1D detected 75% of pancreatic cancer at 95% specificity, whereas falsely positive in only 9% of chronic pancreatitis patients (P < 0.0001, Fisher exact). In contrast, mutant KRAS was only positive in 55% of pancreatic cancer samples and falsely positive in 41% of chronic pancreatitis (P = 0.3). At 100% specificity CD1D falsely detected only 5% chronic pancreatitis patients (P < 0.0001), whereas KRAS was false positive in 32% of chronic pancreatitis (P = 0.4).

Age, sex, or current smoking did not significantly influence the strength of association between methylated marker levels and pancreatic cancer. There were no significant differences when patients were stratified for T-stage 1 and 2 compared with T3 and 4 or for tumor location in the head of the pancreas compared with body and tail.

**Discussion**

Methylome sequencing, without a priori bias to known CpG islands, yielded novel highly discriminant methylation markers for pancreatic cancer. Importantly, these findings were confirmed using an independent sample set of tumor and control tissues, showing that the RRBS process can successfully identify pancreatic cancer markers with low background levels in normal pancreatic parenchyma and colonic epithelial tissues. Many of the markers with the strongest association to pancreatic cancer also showed greater than 30-fold increases in the median copies per sample compared with controls; this observation is critical to the application of these markers in diagnostic test development where assays must detect tumor signal against the background biologic milieu. Novel candidates identified by this method were clinically piloted by assay from pancreatic juice, demonstrating utility for the detection of pancreatic cancer in blinded comparisons, even to diseased controls with chronic pancreatitis.

In the present study, a single marker, methylated CD1D was sensitive and specific for pancreatic cancer. Candidate marker

For the detection of pancreatic cancer in comparison with chronic pancreatitis, the AUC values for CD1D, KCNK12, CLEC11A, NDRG4, IZF1, PKRCB, and KRAS were 0.92 (0.86–0.98), 0.73 (0.61–0.86), 0.76 (0.64–0.87), 0.85 (0.76–0.94), 0.73 (0.61–0.86), 0.77 (0.65–0.89), and 0.62 (0.49–0.74), respectively. CD1D was the most discriminant individual marker for detection of pancreatic cancer in comparison with normal pancreas or chronic pancreatitis (Fig. 4, Supplementary Figs. S2–S7) and was significantly more discriminant than mutant KRAS (P = 0.001). From specificity cutoffs determined in normal pancreas patients, CD1D detected 75% of pancreatic cancer at 95% specificity, whereas falsely positive in only 9% of chronic pancreatitis patients (P < 0.0001, Fisher exact). In contrast, mutant KRAS was only positive in 55% of pancreatic cancer samples and falsely positive in 41% of chronic pancreatitis (P = 0.3). At 100% specificity CD1D falsely detected only 5% chronic pancreatitis patients (P < 0.0001), whereas KRAS was false positive in 32% of chronic pancreatitis (P = 0.4).

Age, sex, or current smoking did not significantly influence the strength of association between methylated marker levels and pancreatic cancer. There were no significant differences when patients were stratified for T-stage 1 and 2 compared with T3 and 4 or for tumor location in the head of the pancreas compared with body and tail.

**Discussion**

Methylome sequencing, without a priori bias to known CpG islands, yielded novel highly discriminant methylation markers for pancreatic cancer. Importantly, these findings were confirmed using an independent sample set of tumor and control tissues, showing that the RRBS process can successfully identify pancreatic cancer markers with low background levels in normal pancreatic parenchyma and colonic epithelial tissues. Many of the markers with the strongest association to pancreatic cancer also showed greater than 30-fold increases in the median copies per sample compared with controls; this observation is critical to the application of these markers in diagnostic test development where assays must detect tumor signal against the background biologic milieu. Novel candidates identified by this method were clinically piloted by assay from pancreatic juice, demonstrating utility for the detection of pancreatic cancer in blinded comparisons, even to diseased controls with chronic pancreatitis.

In the present study, a single marker, methylated CD1D was sensitive and specific for pancreatic cancer. Candidate marker

**Table 1. Patients who submitted pancreatic juice for clinical validation**

<table>
<thead>
<tr>
<th>Age, median (IQR), y</th>
<th>Pancreatic cancer (n = 61)*</th>
<th>Chronic pancreatitis (n = 22)</th>
<th>Normal pancreas (n = 19)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR), y</td>
<td>67 (61–76)</td>
<td>64 (53–72)</td>
<td>60 (49–70)</td>
<td>0.02</td>
</tr>
<tr>
<td>Men (%)</td>
<td>34 (58)</td>
<td>15 (68)</td>
<td>4 (21)</td>
<td>0.007</td>
</tr>
<tr>
<td>Smoking (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>11 (18)</td>
<td>10 (45)</td>
<td>2 (10)</td>
<td>0.03</td>
</tr>
<tr>
<td>Former</td>
<td>19 (32)</td>
<td>5 (23)</td>
<td>4 (21)</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>30 (50)</td>
<td>7 (32)</td>
<td>13 (68)</td>
<td></td>
</tr>
<tr>
<td>Diabetic (%)</td>
<td>15 (25)</td>
<td>6 (27)</td>
<td>1 (5)</td>
<td>0.4</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head (%)</td>
<td>(7)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Body (%)</td>
<td>(8)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Tail (%)</td>
<td>(21)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>EUS tumor stage, T1 or 2, %</td>
<td>16 (26)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Smoking history, diabetes diagnosis, and endoscopic ultrasound (EUS) stage were missing on a single pancreatic cancer patient.
performance was superior to mutant KRAS, which was poorly specific in patients with chronic pancreatitis. The methylation marker levels in pancreatic juice were unaffected by age, gender, cancer stage or site, similar to our observations of methylated DNA in pancreatic cancer when assayed from stool (15).

Some of the methylated DNA markers that we found to be highly discriminant for pancreatic neoplasia have been previously identified in array-based studies (8, 32, 33). Several of our RRBS-discovered markers were found on genes known to be important generally in tumorigenesis, cell signaling, and epithelial-to-mesenchymal transition (Supplementary Table S3), whereas others have no apparent or reported tumor-related role. Some of the identified markers may prove to be organ specific. Because DNA methylation is a highly conserved regulator of tissue development (34), the identification of unreported candidates raises optimism for the existence of DNA methylation events potentially unique to tumor type and site. Indeed, our preliminary observations suggest site specificity of various methylated DNA tumor markers (35). To our knowledge, 10 of the top technically validated markers have not been previously described and are novel to pancreatic cancer. Among 19 biologically validated markers, all contain CpG sites which are not captured by the Infinium 450K hybridization chip. These comparisons demonstrate the value of genome-wide scanning without bias to known DMRs or established gene promoters.

Our results also add significantly to the emerging body of data on next-generation sequencing in human cancer biology. Studies directly comparing these discovery techniques are limited but several recent reports highlight important differences. Among several genome-wide DNA methylation technologies, we selected RRBS for comparatively deeper genomic coverage than methylated DNA immunoprecipitation sequencing (MeDIP-seq) and methylated DNA capture by affinity purification (MethylCap-seq), although the latter two approaches may cover a wider range of genomic CpGs (19). Whereas MethylCap-seq may identify a greater number of hyper-methylated DMRs (21), RRBS data output has the advantage of single-nucleotide sequence resolution, which permits optimal design of secondary clinical assay platforms. As the study of genome-wide DNA methylation mapping progresses, modifications to sequencing-based platforms are likely to further improve marker yield and accuracy (22). However, with any discovery strategy, DMRs must be found in a sufficient majority of samples to permit tests of statistical and clinical significance (23).

There are several limitations to the present study. First, the sample size for RRBS is small but was determined by the power needed to detect a region with at least a 10% differential methylation rate in cases from among controls with the lowest background. With 18 subjects in each group, the overall sample size was similar to or larger than other available genome-wide studies (8, 21–24). Statistical power was further augmented by analysis of only CpGs with sufficient read depth and group coverage. Samples in the RRBS and the biologic validation experiments were tightly matched and randomly allocated to blinded flow cell lane assignment and well assignments, respectively. Tissues from patients with chronic pancreatitis were deliberately excluded from the marker discovery process to ensure the greatest homogeneity of observations within groups (36) and also to exclude possible field effects (37) from undiagnosed pancreatic cancer or precursor lesions. Furthermore, the inclusion of samples from individuals with chronic pancreatitis in clinical piloting controlled for the

**Table 2. Discrimination of DNA markers in pancreatic juice for pancreatic cancer**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Area under ROC curve (95% CI)</th>
<th>Sensitivity at 90% specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pancreatic cancer vs. normal pancreas</td>
<td>Pancreatic cancer vs. chronic pancreatitis</td>
</tr>
<tr>
<td>Methylated</td>
<td>CD1D</td>
<td>0.92 (0.86–0.98)</td>
</tr>
<tr>
<td></td>
<td>KCNJK12</td>
<td>0.88 (0.80–0.95)</td>
</tr>
<tr>
<td></td>
<td>CLEC11A</td>
<td>0.85 (0.76–0.95)</td>
</tr>
<tr>
<td></td>
<td>NDNG4</td>
<td>0.85 (0.77–0.94)</td>
</tr>
<tr>
<td></td>
<td>IKZF1</td>
<td>0.84 (0.75–0.95)</td>
</tr>
<tr>
<td></td>
<td>PRKCB</td>
<td>0.83 (0.74–0.92)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker</th>
<th>Area under ROC curve (95% CI)</th>
<th>Sensitivity at 90% specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutant</td>
<td>KRAS</td>
<td>0.75 (0.64–0.86)</td>
</tr>
</tbody>
</table>

Abbreviation: ROC, receiver-operating characteristics curve.

1P = 0.001 vs. KRAS.

2P = 0.03 vs. KRAS.

3P = 0.06 vs. KRAS.

Figure 3.

AUC plotted against the log ratio of median case to control copy numbers per sample of each marker shows both accuracy and signal strength. Control samples combined benign pancreas and normal colon. Confidence intervals for the AUC estimates and the signal-to-noise ratios are provided in Supplementary Table S2.
exclusion of this group in the discovery process. Inclusion of chronic pancreatitis controls may also partially explain why not all markers discriminant in tissues performed equally as well in pancreatic juice. When markers were selected for the pancreatic juice pilot, the full analysis of the biologic validation was not yet completed. In the full and completed biologic validation, several candidates emerged that might have superior performance. At this time, the DNA from the pancreatic juice analysis had been exhausted, prohibiting testing of those candidates; however, these markers will be of great interest in analysis of new samples, to be collected in a planned prospective clinical trial. Second, samples for the pilot study were from a prospectively enrolled convenience sample and were not matched. This resulted in several significant differences in baseline variables across groups, notably in age, sex and smoking history. However, adjusted analyses did not demonstrate any significant influence of those clinical variables on DNA markers. Greater statistical power may also facilitate the study of marker combinations for improved discrimination.

Third, the pancreatic juice sample collection method was also designed to study protein markers (27) and may not have been optimal for DNA recovery. Despite the inclusion of a protease inhibitor in the sample preparation and the use of nonoptimized first-pass primer designs, we were able to recover and assay sufficient marker DNA to make highly significant observations. In addition, the use of secretin stimulation in the collection protocol minimized potential for background contamination during duodenal luminal sampling and avoided the risks of pancreatic duct cannulation, as reviewed by Mastsubayashi and colleagues (10). Limited by total pancreatic juice DNA quantity, not all biologically validated markers were assessed in pancreatic juice; it is therefore likely that additional highly discriminant markers remain in the initial dataset and deserve further analysis.

Two of 3 patients with IPMNs containing high-grade dysplasia had substantially elevated marker levels in pancreatic juice. Although corroboration in larger sample size studies are clearly needed, this interesting finding suggests a potential future role for pancreatic juice testing to help guide management of cystic pancreatic lesions.

In this translational investigation from discovery to clinical application, our genome-wide search with RRBS identified novel DNA methylation markers which highly discriminated early stage pancreatic from normal tissue. Initial pilot studies on pancreatic juice both validate the biologic discrimination of these new markers and demonstrate their clinical feasibility for use in minimally invasive biologic media, such as pancreatic juice, blood, stool, or urine. Moving forward, we are compelled to corroborate these findings in expanded patient populations, validate additional candidates, and further assess tumor site-specificity of DNA methylation markers.

Disclosure of Potential Conflicts of Interest

J.B. Kisiel reports receiving a commercial research grant from Exact Sciences. W. Taylor has ownership interest (including patents) in and reports receiving a commercial research grant from Exact Sciences. T.C. Yab and D.W. Mahoney have ownership interest (including patents) in Exact Sciences. H. Zou is an employee of and has ownership interest (including patents) in Exact Sciences. D. Ahlquist is a consultant/advisory board member for Exact Sciences. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: J.B. Kisiel, M. Raimondo, W.R. Taylor, T.C. Yab, D.W. Mahoney, D.A. Ahlquist
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J.B. Kisiel, M. Raimondo, W.R. Taylor, T.C. Yab, T.C. Smyrk, L.A. Boardman, G.M. Petersen
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J.B. Kisiel, W.R. Taylor, T.C. Yab, D.W. Mahoney, Z. Sun, S. Middha, S. Baheti, T.C. Smyrk, D.A. Ahlquist
Writing, review, and/or revision of the manuscript: J.B. Kisiel, M. Raimondo, T.C. Yab, D.W. Mahoney, Z. Sun, S. Middha, H. Zou, T.C. Smyrk, L.A. Boardman, D.A. Ahlquist
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): T.C. Yab, G.M. Petersen, D.A. Ahlquist
Study supervision: J.B. Kisiel, T.C. Yab, D.A. Ahlquist
Other (acquisition of funding): D.A. Ahlquist

Grant Support

This work was made possible by grants (to J.B. Kisiel) from the Jack and Maxine Zarrow Family Foundation of Tulsa Oklahoma and the Paul Calabresi Program in Clinical-Translational Research (NCI CA80628). Additional partial support was provided by the Carol M. Gatton endowment for Digestive Diseases Research. Biospecimens were provided by support from the Mayo Clinic SPORE in Pancreatic Cancer (P50CA102701), the Lustgarten Foundation for Pancreatic Cancer Research, and the Clinical Core of the Mayo Clinic Center for Cell Signalling in Gastroenterology (P30DK084567). Reagents for QuARTS assays were provided by Exact Sciences (Madison, WI).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 23, 2014; revised April 20, 2015; accepted May 12, 2015; published OnlineFirst May 28, 2015.

Figure 4.

A, copy numbers per sample of methylated CD1D, assayed from pancreatic juice samples of patients with normal pancreas, chronic pancreatitis, and pancreatic cancer were used to calculate (B) receiver-operating characteristic curves of methylated CD1D for the detection of pancreatic in comparison with normal pancreas (black) and chronic pancreatitis (gray).
References
