Published in final edited form as:

Gastroenterology. 2010 March ; 138(3): 949–957.e7. doi:10.1053/j.gastro.2009.11.010.

Salivary Transcriptomic Biomarkers for Detection of Resectable Pancreatic Cancer

Lei Zhang^{*,†}, James J. Farrell^{‡,†}, Hui Zhou^{*}, David Elashoff^{‡‡}, David Akin^{*}, No-Hee Park^{*,¶,§}, David Chia[§], and David T. Wong^{*,¶,##,§§}

^{*} School of Dentistry and Dental Research Institute, University of California-Los Angeles, Los Angeles, California

[‡] Department of Medicine, David Geffen School of Medicine, University of California-Los Angeles, Los Angeles, California

^{‡‡} Department of Biostatistics, School of Public Health, University of California-Los Angeles, Los Angeles, California

[¶] Jonsson Comprehensive Cancer Center, University of California-Los Angeles, Los Angeles, California

[#] Division of Hematology & Oncology, David Geffen School of medicine, University of California-Los Angeles, Los Angeles, California

^{##} Division of Head and Neck Surgery/Otolaryngology, David Geffen School of Medicine, University of California-Los Angeles, Los Angeles, California

[§] Department of Pathology and Laboratory Medicine, Jonsson Comprehensive Cancer Center, University of California-Los Angeles, Los Angeles, California

^{§§} Henry Samueli School of Engineering and Applied Science, University of California-Los Angeles, Los Angeles, California

Abstract

Background & Aims—Lack of detection technology for early pancreatic cancer invariably leads to a typical clinical presentation of incurable disease at initial diagnosis. New strategies and biomarkers for early detection are sorely needed. In this study, we have conducted a prospective sample collection and retrospective blinded validation to evaluate the performance and translational

Correspondence: David T. Wong DMD, DMSc, Felix & Mildred Yip Endowed Professor, Division of Oral Biology and Medicine, Associate Dean for Research, School of Dentistry, Director, Dental Research Institute, University of California at Los Angeles, 10833 Le Conte Ave, 73-017 CHS, Los Angeles, CA 90095, Phone: 310-206-3048, Fax: 310-825-7609, dtww@ucla.edu.

[†]These authors contributed equally to this work

Authors' Contributions: LZ, JF and DW supervised all aspects of this study including study design, execution, and data interpretation. DA, DC and NP contributed to the study design. LZ and JF conducted the experiments. LZ, HZ and DE contributed to data acquisition and data interpretation. JF provided human saliva samples. DA contributed to sample collection, process and storage. LZ wrote the final manuscript. All the authors reviewed the manuscript. JF, DE, HZ and DW revised the manuscript critically.

Disclosures: The authors have declared that no competing interests and potential conflicts exist.

Accession Numbers: All Affymetrix Human Genome U133 Plus 2.0 Array data generated in this study have been uploaded to the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). The access number is GSE14245.

All other gene names are listed in the Supplementary material Table S4.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

utilities of salivary transcriptomic biomarkers for the non-invasive detection of resectable pancreatic cancer.

Methods—The Affymetrix HG U133 Plus 2.0 Array was used to profile transcriptomes and discover altered gene expression in saliva supernatant. Biomarkers discovered from the microarray study were subjected to clinical validation using an independent sample set of 30 pancreatic cancer, 30 chronic pancreatitis and 30 healthy controls.

Results—Twelve mRNA biomarkers were discovered and validated. The logistic regression model with the combination of four mRNA biomarkers (*KRAS*, *MBD3L2*, *ACRV1* and *DPM1*) could differentiate pancreatic cancer patients from non-cancer subjects (chronic pancreatitis and healthy control), yielding a ROC-plot AUC value of 0.971 with 90.0% sensitivity and 95.0% specificity.

Conclusions—The salivary biomarkers possess discriminatory power for the detection of resectable pancreatic cancer, with high specificity and sensitivity. This report provides the proof of concept of salivary biomarkers for the non-invasive detection of a systemic cancer and paves the way for prediction model validation study followed by pivotal clinical validation.

Keywords

Salivary biomarker; resectable pancreatic cancer; salivary transcriptome

Pancreatic cancer is the second most frequent gastrointestinal malignancy. Overall, it is the fourth commonest cause of cancer related mortality, reflecting its advanced stage of presentation. Early detection of pancreatic cancer offers the promise of improved mortality rates through surgical resection^{1–4}. A significant obstacle towards early detection of pancreatic cancer is the development of methods that efficiently identify potentially affected individuals. Current strategies exist for early detection of pancreatic cancer^{1, 5–10}; however, they are either confined to a small number of patients at greater risk, often relying on invasive procedures, or they lack the necessary sensitivity and specificity to make their widespread screening applicable^{11–13}. The search for potential useful biomarkers of pancreatic cancer is further complicated by the existence of several benign pancreatic diseases such as chronic pancreatitis, which has phenotypic overlap with early pancreatic cancer. The lack of highly specific pancreatic cancer biomarkers is often due to their presence in patients with chronic pancreatitis^{5, 6}.

As a mirror of the body, saliva is readily accessible noninvasively. Salivary constituents including DNA, RNA, protein and bacteria have been extensively linked to forensic sciences^{14–16}, oral disease^{17, 18, 19, 20}, and systemic disease^{21, 22, 23, 24}. Here, we report the use of a high throughput discovery approach to identify discriminatory biomarkers in saliva for the non-invasive detection of pancreatic cancer. Our results demonstrate that the salivary transcriptome profiles are significantly different between patients with pancreatic cancer and healthy controls. The salivary biomarkers identified and validated demonstrate discriminatory power for the detection of pancreatic cancer, with high specificity and sensitivity.

Material and Methods

Patients and study design

This study, which was approved by the UCLA Institutional Review Board, started sample collection in February 2006. The study design followed the principle of PROBE design (prospective specimen collection and retrospective blinded evaluation)²⁵. All subjects were recruited from the UCLA Medical Center. The saliva bank of pancreatic disease at the UCLA Dental Research Institute has collected 283 saliva samples since 2006. Of these, 114 samples, including 42 pancreatic cancer patients, 30 chronic pancreatitis patients and 42 healthy control

individuals (Supplementary material Table S1), were selected for the discovery and validation phase of this study. Inclusion criteria of disease patients consisted of confirmed diagnosis of pancreatic cancer confined to the pancreas [either resectable or borderline resectable (due to superior mesenteric vein or portal vein involvement)] and chronic pancreatitis. Exclusion criteria included evidence of locally advanced pancreatic cancer due to arterial involvement or direct extension into adjacent organs, metastatic pancreatic cancer, chemotherapy or radiation therapy prior to saliva collection and a diagnosis of other malignancies within 5 years from the time of saliva collection. Written informed consents and questionnaire data sheets were obtained from all patients who agreed to serve as saliva donors. The information on individual characteristics, such as age, gender, ethnicity, smoking history and drinking history, is presented in Table 1. Healthy control individuals were matched for age, gender and ethnicity to the cancer group. Unstimulated saliva samples were consistently collected, stabilized and preserved as previously described^{14, 26}. The sample supernatants were reserved at -80 °C prior to assay. A protocol for the standardized acquisition of the saliva samples is elaborated in the Supplementary material.

This study consisted of a discovery phase, followed by an independent validation phase. Of the 114 samples, 12 pancreatic cancer samples and 12 healthy control samples were chosen for the discovery phase. The transcriptomic approach profiled the saliva supernatant samples from 12 pancreatic cancer patients and 12 healthy control subjects using the Affymetrix HG U133 Plus 2.0 Array. Biomarkers identified from the microarray study were first verified using the discovery sample set (12 cancers and 12 healthy controls). An independent sample set, including 30 pancreatic cancer patients, 30 chronic pancreatitis patients and 30 healthy control subjects, was used for the biomarker validation phase (Figure 1). The validated biomarkers were evaluated within three levels of clinical discrimination categories: pancreatic cancer vs. healthy control; pancreatic cancer vs. chronic pancreatitis and pancreatic cancer vs. non-cancer (healthy control + chronic pancreatitis).

Salivary transcriptomic profiling

RNA was isolated from 330 µL of saliva supernatant using the MagMax™ Viral RNA Isolation Kit (Ambion, Austin, TX). This process was automated using KingFisher mL technology (Thermo Fisher Scientific), followed by TURBO DNase treatment (Ambion). Extracted RNA was linearly amplified using the RiboAmp RNA Amplification kit (Molecular Devices, Sunnyvale, CA). After purification, cDNA was *in vitro* transcribed and biotinylated using GeneChip Expression 3'-Amplification Reagents for *in vitro* transcription labeling (Affymetrix, Santa Clara, CA). Chip hybridization and scanning were performed at the UCLA microarray core facility. Using the MIAME criteria²⁸, all Affymetrix Human Genome U133 Plus 2.0 Array data generated in this study have been uploaded to the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). The access number is GSE14245.

U133 Plus 2.0 Array data analysis

The analysis was performed using R 2.7.0 (<http://www.r-project.org>). The Probe Logarithmic Intensity Error Estimation (PLIER) expression measures were computed after background correction and quantile normalization for each microarray dataset. Probeset-level quantile normalization was performed across all samples to make the effect sizes similar among all datasets. Finally, for every probeset, the two-sample t-test was applied to identify differential expression between cancer and healthy control. After obtaining the estimates and the p-values of each probeset, we corrected the p-values for false discovery rate (FDR).

Validation of mRNA biomarkers using quantitative PCR (qPCR)

The selected mRNA biomarkers were first verified by qPCR using the discovery sample set (12 pancreatic cancer and 12 healthy control) as described previously^{18, 29}. qPCR primers

were designed using Primer Express 3.0 software (Applied Biosystems, FosterCity, CA) (Supplementary material Table S3). All primers were synthesized by Sigma-Genosys (Woodlands, TX). The amplicons were intron spanning whenever possible. qPCR was carried out in duplicate. Verified biomarkers were then assayed by qPCR in the set of 90 independent samples. The Wilcoxon test was used to compare the biomarkers between groups.

Predictive model building and evaluation

The logistic regression (LR) method was used in prediction model building. For each validated biomarker, we constructed the receiver operating characteristic (ROC) curve and computed the area under the curve (AUC) value by numerical integration of the ROC curve. Next, the validated salivary biomarkers were fit into logistic regression models (separately for each group comparisons) and stepwise backward model selection was performed to determine final combinations of biomarkers. For each of these models, the predicted probability for each subject was obtained and was used to construct ROC curves. The standard error of the AUC and the 95% confidence interval (CI) for the ROC curve was computed according to previous publications^{31, 32}. The sensitivity and specificity for the biomarker combinations were estimated by identifying the cutoff-point of the predicted probability that yielded the highest sum of sensitivity and specificity.

A simulation study was performed to determine the magnitude of the bias introduced by model selection using multiple biomarker models. Briefly, we first permuted the group identities for the subjects [using the cancer vs. non-cancer (chronic pancreatitis and healthy control) comparison]. For each marker we computed the t-statistics between the permuted groups, then constructed a logistic regression model with the permuted group identities as the outcome and using stepwise selection with the most significant 12 biomarkers (to be analogous to the 12 significant qPCR markers found in the original data). For each of the resulting multiple marker models, we estimated the prediction accuracy by computing the AUC. This process was iterated 1000 times. The set of AUC values form an unbiased permutation distribution for the true model AUC and correct for biases generated by the model selection and coefficient estimation process. The choice of using 12 markers in the selection process is fairly conservative since typically fewer than 3 markers out of the 35 originally considered will be statistically significantly ($p < 0.05$) between the permuted groups (and therefore eligible for the model selection process that was performed).

Cross-disease comparisons of salivary mRNA biomarkers based on microarray studies

The validated mRNA biomarkers for pancreatic cancer detection were checked in other microarray studies that have been conducted in our laboratory on different diseases, including oral cancer¹⁹, breast cancer and lung cancer. Briefly, t-test p-values were calculated for all validated genes of pancreatic cancer study in other microarray studies to check whether they are also significant varied between cancers and controls in those diseases. Variation is considered significant if p-value is less than 0.05.

Results

Variation of Salivary Gene Expression Profiles between Pancreatic Cancer Patients and Healthy Controls

In the discovery phase, microarrays and qPCR were used to examine gene expression profiles and levels in saliva samples from pancreatic cancer patients ($n = 12$) and healthy controls ($n = 12$). It is important to assess the quantity and quality of mRNA in saliva to ensure the sufficiency and accuracy for microarray profiling. On average, 146.6 ± 58.7 ng ($n = 24$) of total RNA was obtained from 330 μ L of saliva supernatant. There was no significant difference in total RNA quantity between pancreatic cancer and healthy controls (t test, $P = 0.35$, $n = 24$).

The qPCR results demonstrated that all saliva samples ($n = 24$) contained transcripts of four genes (*GAPDH*, *ANXA2*, *RPL37* and *RPS16*), which were used as quality controls for human salivary RNAs²⁹. There were no significant difference in the level of these four saliva internal reference (SIR) genes between pancreatic cancer and healthy controls (t test, $P = 0.47$ for *GAPDH*; $P = 0.67$ for *ANXA2*; $P = 0.79$ for *RPL37*; $P = 0.85$ for *RPS16*, $n = 24$). A consistent amplification magnitude (368 ± 37.3 , $n = 24$) was obtained after two rounds of RNA amplification. On average, the yield of biotinylated cRNA was $34 \pm 2.9 \mu\text{g}$ ($n = 24$). There were no significant differences in the yield of cRNA between pancreatic cancer and healthy controls (t test, $P = 0.32$, $n = 24$).

Transcriptomic profiling revealed that 958 genes exhibited >2 fold up-regulation and 691 genes exhibited >2 fold down-regulation in the saliva of pancreatic cancer patients, relative to the healthy controls ($n = 24$, $P < 0.05$). These transcripts identified were unlikely to be attributed to chance (χ^2 test, $P < 0.0001$), considering the false positive rate with $P < 0.05$. Using a predefined criterion of a change in regulation > 4 -fold, and a more stringent cutoff of p-value < 0.01 , 49 up-regulated and 21 down-regulated transcripts were identified in pancreatic cancer samples.

Identification and Validation of mRNA Biomarkers for Pancreatic Cancer

Quantitative PCR (qPCR) was performed to verify the microarray results on the discovery sample set ($n = 24$). All 49 up-regulated and 21 down-regulated transcripts were evaluated. The qPCR results confirmed that the relative RNA expression levels of 23 up-regulated and 12 down-regulated transcripts, were consistent with the microarray data. A heatmap of these 35 verified genes was built based on the microarray data (Figure 2). Hierarchical clustering and gene function enrichment was performed using Euclidean distance metric and Centroid linkage method (unsupervised clustering) included in dChip software³³. Pancreatic cancer patients ($n=12$) and healthy controls ($n=12$) could be classified into distinct groups, indicating the discriminatory power of salivary mRNA biomarkers. The biological functions of these genes and their products are presented in Supplementary material Table S4. These verified transcriptomic candidates were then subjected to validation by qPCR in an independent cohort of 30 pancreatic cancer patients (Supplementary material Table S2), 30 healthy control subjects and 30 chronic pancreatitis patients. As shown in Table 2, a total of 7 up-regulated and 5 down-regulated genes were validated. These 12 mRNA biomarkers all showed significant difference between pancreatic cancer and healthy controls ($P < 0.05$, $n = 60$), yielding ROC-plot AUC values between 0.682 and 0.823. The expression patterns of these mRNA biomarkers were consistent with those retrieved by microarray assay (up/down-regulation and fold change). Importantly, the expression levels of six up-regulated mRNAs (*MBD3L2*, *KRAS*, *STIM2*, *ACRV1*, *DMD*, *CABLES1*) and three down-regulated mRNAs (*TK2*, *GLTSCR2*, *CDKL3*) were also significantly different between pancreatic cancer and chronic pancreatitis ($P < 0.05$, $n = 60$). The expression level of all 12 up/down-regulated mRNAs were significantly different between pancreatic cancer ($n = 30$) and non-cancer subjects (chronic pancreatitis and healthy control, $n = 60$) ($P < 0.05$), yielding ROC-plot AUC values between 0.661 and 0.791 (Table 2).

Prediction Models using the Validated mRNA Biomarkers

To demonstrate the clinical utility of salivary mRNAs biomarkers for pancreatic cancer detection, logistic regression models were built based on different combinations of biomarkers for three levels of clinical discrimination: pancreatic cancer vs. healthy control; pancreatic cancer vs. chronic pancreatitis and pancreatic cancer vs. non-cancer (healthy control + chronic pancreatitis) (Table 3). For pancreatic cancer vs. healthy control, the logistic regression model with the combination of four mRNA biomarkers (*KRAS*, *MBD3L2*, *ACRV1* and *CDKL3*) yielded a ROC-plot AUC value of 0.973 (95% CI, 0.895 to 0.997; $P < 0.0001$) with 93.3%

sensitivity and 100% specificity in distinguishing pancreatic cancer patients from healthy control subjects. For pancreatic cancer vs. chronic pancreatitis, the logistic regression model with the combination of three mRNA biomarkers (*CDKL3*, *MBD3L2*, *KRAS*) yielded a ROC-plot AUC value of 0.981 (95% CI, 0.907 to 0.997; $P < 0.0001$) with 96.7% sensitivity and 96.7% specificity in distinguishing pancreatic cancer patients from chronic pancreatitis. Most importantly, for the discrimination of pancreatic cancer vs. non-cancer, the logistic regression model with the combination of four mRNA biomarkers (*KRAS*, *MBD3L2*, *ACRV1* and *DPM1*) could differentiate pancreatic cancer patients from all non-cancer subjects, yielding a ROC-plot AUC value of 0.971 (95% CI, 0.911 to 0.994; $P < 0.0001$). The four-mRNA-biomarker logistic regression model provided the highest discriminatory power for differentiating pancreatic cancer from non-cancer subjects. Using a cutoff of 0.433, a sensitivity of 90.0% and a specificity of 95.0% was obtained for this four-biomarker logistic regression model (Figure 3).

In order to evaluate if the four-mRNA-biomarker model could be the result of data overfitting, a simulation study for the cancer vs. non-cancer prediction model was performed and resulted in an empirical p-value for the AUC of the model of $p < 0.001$ as none of the simulated AUC values were greater than 0.85. Thus, even after accounting for model selection and model fitting with multiple markers, the observed marker set has significantly more discriminatory power for detecting pancreatic cancer than we would expect by chance.

The effects of age and smoking history on the validated biomarkers were examined within each of the three clinical categories (Table S5). Overall, we found that neither age nor smoking had effects on the biomarkers more than we would expect by chance (only 2 out of 90 [2 covariates $\times$ 15 markers $\times$ 3 groups] tests were significant at $\alpha = 0.05$).

Cross-Disease Comparisons of Salivary mRNA Biomarkers

The determination of specific profiles of molecular changes in a specific cancer types is important because it is possible that the different cancers may have overlapping signatures. We have evaluated the specificity of the 12 validated mRNA biomarkers against other microarray discovery studies that have been performed in our laboratory on diverse cancers, including oral cancer¹⁹, breast cancer, and lung cancer. With the exception of *TK2* that showed significant variation in lung cancer ($P = 0.007$), none of the other 11 mRNAs/transcripts were significantly altered in other cancer microarray studies ($P > 0.05$, Table 4). All these cross-disease comparisons indicated that the validated mRNA biomarkers in saliva are specific for pancreatic cancer.

Discussion

Current clinical blood-based tests for pancreatic cancer including CA19-9 lack sufficient sensitivity and specificity to be of use in screening for pancreatic cancer, especially preinvasive forms^{6, 10–13}. Combination of other proteomic biomarkers with CA19-9 has the ability to distinguish pancreatic cancer from healthy controls with high discriminatory power; however, the sensitivity is relatively low^{7, 34}. Additionally, the utility of these biomarkers for discriminating pancreatic cancer from chronic pancreatitis is limited by their low sensitivity and specificity⁷. Our study is the first systematic study profiling transcriptome in saliva samples of resectable pancreatic cancer patients. The salivary biomarkers that were identified are discriminatory for the detection of resectable pancreatic cancer, with high sensitivity and specificity. It is particularly notable that the validated biomarkers can also discriminate pancreatic cancer from chronic pancreatitis with high sensitivity and specificity, demonstrating that these salivary biomarkers are specific for the detection of pancreatic cancer without the complication of chronic pancreatitis. Our findings enhance the prospect of an important role for salivary diagnostics in the detection of systemic cancers and diseases. Not only are these

saliva-based diagnostic and detection tests for pancreatic cancer simple and non-invasive, they may also represent an improvement in specificity and sensitivity over currently used procedures for pancreatic cancer detection.

In this study, the profiles of molecular signatures in saliva and their changes between disease and controls have been successfully linked to the detection of pancreatic cancer. Consistent with previous studies, our high-throughput analysis indicates that the mRNA in saliva supernatant is relatively stable and informative, and is a suitable source of disease discriminatory biomarkers^{18, 19, 35–38}. The consistency between different mRNA analysis methods (microarray and qPCR) demonstrates that the alteration of the salivary mRNA signatures between cancer group and control group can serve as biomarkers for detection of pancreatic cancer. Of the 12 validated mRNA biomarkers, several genes, e.g. *MBD3L2*, *GLTSCR2* and *TPT1*, have been linked to carcinogenesis^{39–47}. Of particular interest is that *KRAS*, a frequently mutated molecular target in pancreatic cancer^{48, 49}, is a discriminatory biomarker in saliva.

It remains to be investigated whether the aberrant expressions of these genes are mediated by salivary glands or by other mechanisms. We have previously published parallel animal studies to address the mechanistic issues of salivary diagnostics, specifically looking at the changes in salivary transcriptomic biomarkers during the development of cancer.⁵⁰ Using the mouse models of melanoma and non-small cell lung cancer, we compared the transcriptome biomarker profiles of tumor-bearing mice to those of control mice. Microarray analysis showed that salivary transcriptomes were significantly altered in tumor-bearing mice vs. controls. The animal model studies support the conclusion that upon systemic disease development, significant and disease-specific changes can occur in the salivary biomarker profile. There is a cancer-specific profile change in salivary mRNA biomarkers using mouse models for systemic disease development. Tumors are known to produce mediators (hormones/cytokines/lymphokines) which can modulate the activities and gene expression patterns of distal organs (salivary glands) through the vasculature. We hypothesize that upon contact with the salivary gland, cancer-specific mediators elicit altered gene expression profiles and translation of associated proteins that are secreted into the saliva. Stimulation of salivary gland by mediators released from remote tumors plays an important role in regulating the salivary surrogate biomarker profiles. These constitute cancer-associated salivary gland surrogate biomarkers, which do not necessarily reflect the primary tumor's transcriptome or proteome.

As a single biomarker is unlikely to detect a specific cancer with high specificity and sensitivity, we used logistic regression to evaluate the combinations of the validated biomarkers. The combination of multiple biomarkers increased the predictive utility substantially as demonstrated by the high ROC AUC values, even after taking stringent precautions to avoid data overfitting. First, the prediction models use only 2–4 markers for the 60 or 90 cases under consideration. However, due to the number of markers considered it is possible that the prediction accuracy estimated from the models is overly optimistic. In order to obtain a more realistic estimate of the predictive utility of the biomarkers (and avoid the consequences of potential data overfitting) for our logistic regression model, we employed leave-one-out cross-validation (LOOCV). The cross validation rate (cv.err) reflects a more accurate estimate of the true prediction accuracy of the model (Table 3). As can be seen from Table 3, all comparisons have cross validation rate of <0.2, indicating that the biomarkers in general have high prediction accuracy and that despite the moderate sample size we appear to have identified biomarkers that correlate with the presence of pancreatic cancer. The combined 4 mRNAs model for distinguishing pancreatic cancer and non-cancer samples achieves 96.7 % (1–3.3%) average prediction accuracy based on leave-one-out cross-validation. Additionally, our simulation study demonstrated that the predictive ability of the marker combination is significantly higher than we would expect by chance even after accounting for the variable selection process.

We do understand that, due to the modest sample size, the results presented in this study have their limitations. The potential for salivary mRNA biomarkers to identify very early stage and even pre-invasive pancreatic cancer needs further study. We have not tested our salivary biomarkers in population based screening for pancreatic cancer, which requires confirmation in community-based screening cohorts, ideally in a multi-center setting. The current study is a feasibility study, meriting a future prospective study. We have built the prediction models for pancreatic cancer. Efforts are in progress to examine the performance of the models in a pivotal validation study using the P_{Ro}BE design ²⁵.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ali Ammar for collecting and processing saliva samples. We thank Dr. Hua Xiao for technique support. We also thank the UCLA microarray core facility for technique support.

Grant Support: Funding support was provided by the National Institute of Health (UO1DE016275 and R21CA126733). Funding organizations did not have any role in study design, data collection or analysis, decision to publish, or preparation of the manuscript.

Abbreviations

AUC	area under the curve
ROC	receiver operating characteristic
FDR	false discovery rate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase

References

1. Canto MI, Goggins M, Hruban RH, Petersen GM, Giardiello FM, Yeo C, Fishman EK, Brune K, Axilbund J, Griffin C, Ali S, Richman J, Jagannath S, Kantsevov SV, Kalloo AN. Screening for early pancreatic neoplasia in high-risk individuals: a prospective controlled study. *Clin Gastroenterol Hepatol* 2006;4:766–81. quiz 665. [PubMed: 16682259]
2. Canto MI, Goggins M, Yeo CJ, Griffin C, Axilbund JE, Brune K, Ali SZ, Jagannath S, Petersen GM, Fishman EK, Piantadosi S, Giardiello FM, Hruban RH. Screening for pancreatic neoplasia in high-risk individuals: an EUS-based approach. *Clin Gastroenterol Hepatol* 2004;2:606–21. [PubMed: 15224285]
3. Rulyak SJ, Kimmey MB, Veenstra DL, Brentnall TA. Cost-effectiveness of pancreatic cancer screening in familial pancreatic cancer kindreds. *Gastrointest Endosc* 2003;57:23–9. [PubMed: 12518126]
4. Hoimes, CJ.; Moyer, MT.; Saif, MW. Biomarkers for early detection and screening in pancreatic cancer. *JOP*; Highlights from the 45th ASCO annual meeting; Orlando, FL, USA. May 29–June 2, 2009; 2009. p. 352-6.
5. Farrell JJ, van Rijnsoever M, Elsaleh H. Early detection markers in Pancreas Cancer. *Cancer Biomark* 2005;1:157–75. [PubMed: 17192037]
6. Koopmann J, Rosenzweig CN, Zhang Z, Canto MI, Brown DA, Hunter M, Yeo C, Chan DW, Breit SN, Goggins M. Serum markers in patients with resectable pancreatic adenocarcinoma: macrophage inhibitory cytokine 1 versus CA19-9. *Clin Cancer Res* 2006;12:442–6. [PubMed: 16428484]
7. Koopmann J, Zhang Z, White N, Rosenzweig J, Fedarko N, Jagannath S, Canto MI, Yeo CJ, Chan DW, Goggins M. Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. *Clin Cancer Res* 2004;10:860–8. [PubMed: 14871961]

8. Rogers CD, Fukushima N, Sato N, Shi C, Prasad N, Hustinx SR, Matsubayashi H, Canto M, Eshleman JR, Hruban RH, Goggins M. Differentiating pancreatic lesions by microarray and QPCR analysis of pancreatic juice RNAs. *Cancer Biol Ther* 2006;5:1383–9. [PubMed: 17012836]
9. Whitcomb DC. Inflammation and Cancer V. Chronic pancreatitis and pancreatic cancer. *Am J Physiol Gastrointest Liver Physiol* 2004;287:G315–9. [PubMed: 15246966]
10. Kim JE, Lee KT, Lee JK, Paik SW, Rhee JC, Choi KW. Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population. *J Gastroenterol Hepatol* 2004;19:182–6. [PubMed: 14731128]
11. Goggins M, Canto M, Hruban R. Can we screen high-risk individuals to detect early pancreatic carcinoma? *J Surg Oncol* 2000;74:243–8. [PubMed: 10962454]
12. Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, Somerfield MR, Hayes DF, Bast RC Jr. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol* 2006;24:5313–27. [PubMed: 17060676]
13. Pleskow DK, Berger HJ, Gyves J, Allen E, McLean A, Podolsky DK. Evaluation of a serologic marker, CA19-9, in the diagnosis of pancreatic cancer. *Ann Intern Med* 1989;110:704–9. [PubMed: 2930108]
14. Ballantyne J. Validity of messenger RNA expression analyses of human saliva. *Clin Cancer Res* 2007;13:1350. author reply 1351. [PubMed: 17317848]
15. Nussbaumer C, Gharehbaghi-Schnell E, Korschineck I. Messenger RNA profiling: a novel method for body fluid identification by real-time PCR. *Forensic Sci Int* 2006;157:181–6. [PubMed: 16289614]
16. Zubakov D, Hanekamp E, Kokshoorn M, van Ijcken W, Kayser M. Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. *Int J Legal Med* 2008;122:135–42. [PubMed: 17579879]
17. Hu S, Arellano M, Boontheung P, Wang J, Zhou H, Jiang J, Elashoff D, Wei R, Loo JA, Wong DT. Salivary proteomics for oral cancer biomarker discovery. *Clin Cancer Res* 2008;14:6246–52. [PubMed: 18829504]
18. Hu S, Wang J, Meijer J, Jeong S, Xie Y, Yu T, Zhou H, Henry S, Vissink A, Pijpe J, Kallenberg C, Elashoff D, Loo JA, Wong DT. Salivary proteomic and genomic biomarkers for primary Sjogren's syndrome. *Arthritis Rheum* 2007;56:3588–600. [PubMed: 17968930]
19. Li Y, St John MA, Zhou X, Kim Y, Sinha U, Jordan RC, Eisele D, Abemayor E, Elashoff D, Park NH, Wong DT. Salivary transcriptome diagnostics for oral cancer detection. *Clin Cancer Res* 2004;10:8442–50. [PubMed: 15623624]
20. Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM. The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. *J Transl Med* 2005;3:27. [PubMed: 15987522]
21. Bigler LR, Streckfus CF, Copeland L, Burns R, Dai X, Kuhn M, Martin P, Bigler SA. The potential use of saliva to detect recurrence of disease in women with breast carcinoma. *J Oral Pathol Med* 2002;31:421–31. [PubMed: 12165061]
22. Goldenberg RL, Culhane JF. Preterm birth and periodontal disease. *N Engl J Med* 2006;355:1925–7. [PubMed: 17079769]
23. Streckfus CF, Mayorga-Wark O, Arreola D, Edwards C, Bigler L, Dubinsky WP. Breast cancer related proteins are present in saliva and are modulated secondary to ductal carcinoma in situ of the breast. *Cancer Invest* 2008;26:159–67. [PubMed: 18259946]
24. Streckfus C, Bigler L, Dellinger T, Dai X, Kingman A, Thigpen JT. The presence of soluble c-erbB-2 in saliva and serum among women with breast carcinoma: a preliminary study. *Clin Cancer Res* 2000;6:2363–70. [PubMed: 10873088]
25. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008;100:1432–8. [PubMed: 18840817]
26. Denny P, Hagen FK, Hardt M, Liao L, Yan W, Arellanno M, Bassilian S, Bedi GS, Boontheung P, Cociorva D, Delahunty CM, Denny T, Dunsmore J, Faull KF, Gilligan J, Gonzalez-Begne M, Halgand F, Hall SC, Han X, Henson B, Hewel J, Hu S, Jeffrey S, Jiang J, Loo JA, Ogorzalek Loo RR, Malamud D, Melvin JE, Miroshnichenko O, Navazesh M, Niles R, Park SK, Prakobphol A,

- Ramachandran P, Richert M, Robinson S, Sondej M, Souda P, Sullivan MA, Takashima J, Than S, Wang J, Whitelegge JP, Witkowska HE, Wolinsky L, Xie Y, Xu T, Yu W, Ytterberg J, Wong DT, Yates JR 3rd, Fisher SJ. The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. *J Proteome Res* 2008;7:1994–2006. [PubMed: 18361515]
27. Preza D, Olsen I, Willumsen T, Boches SK, Cotton SL, Grinde B, Paster BJ. Microarray analysis of the microflora of root caries in elderly. *Eur J Clin Microbiol Infect Dis*. 2008
  28. Edgar R, Barrett T. NCBI GEO standards and services for microarray data. *Nat Biotechnol* 2006;24:1471–2. [PubMed: 17160034]
  29. Hu Z, Zimmermann BG, Zhou H, Wang J, Henson BS, Yu W, Elashoff D, Krupp G, Wong DT. Exon-level expression profiling: a comprehensive transcriptome analysis of oral fluids. *Clin Chem* 2008;54:824–32. [PubMed: 18356245]
  30. Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, Levanos VA, Sahasrabudhe A, Dewhirst FE. Bacterial diversity in human subgingival plaque. *J Bacteriol* 2001;183:3770–83. [PubMed: 11371542]
  31. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36. [PubMed: 7063747]
  32. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561–77. [PubMed: 8472349]
  33. Cheng, Li; Wong, WH. DNA-Chip Analyzer (dChip). In: Parmigiani, G.; Garrett, ES.; Irizarry, R.; Zeger, SL., editors. *The analysis of gene expression data: methods and software*. New York: Springer; 2003. p. 120–141.
  34. Firpo MA, Gay DZ, Granger SR, Scaife CL, DiSario JA, Boucher KM, Mulvihill SJ. Improved diagnosis of pancreatic adenocarcinoma using haptoglobin and serum amyloid A in a panel screen. *World J Surg* 2009;33:716–22. [PubMed: 19082654]
  35. Li Y, Zhou X, St John MA, Wong DT. RNA profiling of cell-free saliva using microarray technology. *J Dent Res* 2004;83:199–203. [PubMed: 14981119]
  36. Park NJ, Li Y, Yu T, Brinkman BM, Wong DT. Characterization of RNA in saliva. *Clin Chem* 2006;52:988–94. [PubMed: 16601067]
  37. Park NJ, Yu T, Nabili V, Brinkman BM, Henry S, Wang J, Wong DT. RNAProtect saliva: An optimal room- temperature stabilization reagent for the salivary transcriptome. *Clin Chem* 2006;52:2303–4. [PubMed: 17138851]
  38. Park NJ, Zhou X, Yu T, Brinkman BM, Zimmermann BG, Palanisamy V, Wong DT. Characterization of salivary RNA by cDNA library analysis. *Arch Oral Biol* 2007;52:30–5. [PubMed: 17052683]
  39. Smith JS, Tachibana I, Pohl U, Lee HK, Thanarajasingam U, Portier BP, Ueki K, Ramaswamy S, Billings SJ, Mohrenweiser HW, Louis DN, Jenkins RB. A transcript map of the chromosome 19q-arm glioma tumor suppressor region. *Genomics* 2000;64:44–50. [PubMed: 10708517]
  40. Kim YJ, Cho YE, Kim YW, Kim JY, Lee S, Park JH. Suppression of putative tumour suppressor gene GLTSCR2 expression in human glioblastomas. *J Pathol* 2008;216:218–24. [PubMed: 18729076]
  41. Okahara F, Ikawa H, Kanaho Y, Maehama T. Regulation of PTEN phosphorylation and stability by a tumor suppressor candidate protein. *J Biol Chem* 2004;279:45300–3. [PubMed: 15355975]
  42. Okahara F, Itoh K, Nakagawara A, Murakami M, Kanaho Y, Maehama T. Critical role of PICT-1, a tumor suppressor candidate, in phosphatidylinositol 3,4,5-trisphosphate signals and tumorigenic transformation. *Mol Biol Cell* 2006;17:4888–95. [PubMed: 16971513]
  43. Yim JH, Kim YJ, Ko JH, Cho YE, Kim SM, Kim JY, Lee S, Park JH. The putative tumor suppressor gene GLTSCR2 induces PTEN-modulated cell death. *Cell Death Differ* 2007;14:1872–9. [PubMed: 17657248]
  44. Arcuri F, Papa S, Carducci A, Romagnoli R, Liberatori S, Riparbelli MG, Sanchez JC, Tosi P, del Vecchio MT. Translationally controlled tumor protein (TCTP) in the human prostate and prostate cancer cells: expression, distribution, and calcium binding activity. *Prostate* 2004;60:130–40. [PubMed: 15162379]
  45. Chung S, Kim M, Choi W, Chung J, Lee K. Expression of translationally controlled tumor protein mRNA in human colon cancer. *Cancer Lett* 2000;156:185–90. [PubMed: 10880768]

46. Zhu WL, Cheng HX, Han N, Liu DL, Zhu WX, Fan BL, Duan FL. Messenger RNA expression of translationally controlled tumor protein (TCTP) in liver regeneration and cancer. *Anticancer Res* 2008;28:1575–80. [PubMed: 18630514]
47. Li Y, Elashoff D, Oh M, Sinha U, St John MA, Zhou X, Abemayor E, Wong DT. Serum circulating human mRNA profiling and its utility for oral cancer detection. *J Clin Oncol* 2006;24:1754–60. [PubMed: 16505414]
48. Mulcahy HE, Lyautey J, Lederrey C, qi Chen X, Anker P, Alstead EM, Ballinger A, Farthing MJ, Stroun M. A prospective study of K-ras mutations in the plasma of pancreatic cancer patients. *Clin Cancer Res* 1998;4:271–5. [PubMed: 9516910]
49. Pellegata NS, Sessa F, Renault B, Bonato M, Leone BE, Solcia E, Ranzani GN. K-ras and p53 gene mutations in pancreatic cancer: ductal and nonductal tumors progress through different genetic lesions. *Cancer Res* 1994;54:1556–60. [PubMed: 8137263]
50. Gao K, Zhou H, Zhang L, Lee JW, Zhou Q, Hu S, Wolinsky LE, Farrell J, Eibl G, Wong DT. Systemic disease-induced salivary biomarker profiles in mouse models of melanoma and non-small cell lung cancer. *PLoS One* 2009;4:e5875. [PubMed: 19517020]
51. Michaud DS, Joshipura K, Giovannucci E, Fuchs CS. A prospective study of periodontal disease and pancreatic cancer in US male health professionals. *J Natl Cancer Inst* 2007;99:171–5. [PubMed: 17228001]

Figure 1.

Schematic of the Strategy Used for the Discovery and Validation of Salivary Biomarkers. PC: pancreatic cancer; H: healthy control; CP: chronic pancreatitis.

Figure 2.

Heatmap of 35 qPCR verified genes (23 up-regulated and 12 down-regulated) based on the microarray data. Hierarchical clustering and gene function enrichment was performed using Euclidean distance metric and Centroid linkage method (unsupervised clustering). Pancreatic cancer patients (n=12) and healthy controls (n=12) could be classified into distinct groups, indicating the discriminatory power of salivary mRNA biomarkers. The GEO database access number of all microarray experiments is GSE14245.

Figure 3.

ROC curve and Interactive dot diagram for the logistic regression model. (A) The logistic regression model using four biomarkers (*KRAS*, *MBD3L2*, *ACRV1* and *DPM1*) yielded an AUC value of 0.971 (cutoff 0.433). (B) Interactive dot diagram was based on the qPCR data of the non-cancer group (n = 60) and cancer group (n = 30).

Table 1
Demographic Information of Subjects in the Discovery and Validation Phases

Validation Phase									
Discovery Phase					Validation Phase				
Demographic Variable	Characteristics	Pancreatic cancer (n=12)	Healthy control (n=12)	p	Pancreatic cancer (n=30)	Healthy control (n=30)	Chronic pancreatitis (n=30)	p	
Age (y)	Mean + SD	69.42 ± 8.91	67.42 ± 12.44	0.77	69.6 ± 11.50	64.93 ± 9.74	54.3 ± 10.65	<0.001	
Gender	Male	8 (66.7%)	8 (66.7%)	1	19 (63.3%)	19 (66.7%)	16 (53.3%)	0.766	
	Female	4 (33.4%)	4 (33.4%)	1	11 (36.7%)	11 (33.3%)	14 (46.7%)		
Ethnicity	Caucasian	9 (75%)	9 (75%)	1	20 (60%)	20 (60%)	20 (66.6%)	1	
	African American	0 (0.0%)	0 (0.0%)	1	2 (13.3%)	2 (13.3%)	2 (6.7%)		
	Asian	1 (8.3%)	1 (8.3%)	1	4 (13.3%)	4 (13.3%)	4 (13.3%)		
	Hispanic	2 (16.7%)	2 (16.7%)	1	4 (13.4%)	4 (13.4%)	4 (13.4%)		
Smoking					5	2	12	0.008	
Drinking		1			2	3	3	1	

For the validation samples, p-value was calculated among three groups. The detailed information on individual characteristics, such as age, gender, ethnicity, smoking history and drinking history, is presented in Supplementary material Table S1.

Table 2

Quantitative PCR Results of Twelve Validated mRNA Biomarkers in Saliva

Gene symbol	Pancreatic cancer vs. Healthy control			Pancreatic cancer vs. Chronic pancreatitis			Pancreatic cancer vs. non-cancer		
	P	AUC	Fold change	P	AUC	Fold change	P	AUC	Fold change
<i>MBD3L2</i>	< 0.001	0.788	8.0 ↑	0.003	0.718	4.3 ↑	< 0.001	0.754	5.9 ↑
<i>KRAS</i>	< 0.001	0.823	6.1 ↑	0.001	0.759	4.2 ↑	< 0.001	0.791	5.1 ↑
<i>STIM2</i>	< 0.001	0.759	4.3 ↑	0.002	0.733	3.1 ↑	< 0.001	0.746	3.7 ↑
<i>DMXL2</i>	0.007	0.699	4.6 ↑	0.106	0.622	2.8 ↑	< 0.001	0.661	3.1 ↑
<i>ACRV1</i>	< 0.001	0.745	3.9 ↑	< 0.001	0.753	4.9 ↑	< 0.001	0.749	4.4 ↑
<i>DMD</i>	< 0.001	0.758	4.1 ↑	0.003	0.718	2.9 ↑	< 0.001	0.738	3.4 ↑
<i>CABLES1</i>	< 0.001	0.783	4.1 ↑	0.003	0.721	2.7 ↑	< 0.001	0.753	3.4 ↑
<i>TK2</i>	0.002	0.731	4.7 ↓	0.015	0.682	4.5 ↓	0.001	0.707	4.6 ↓
<i>GLTSCR2</i>	< 0.001	0.785	4.8 ↓	< 0.001	0.769	5.4 ↓	< 0.001	0.777	5.1 ↓
<i>CDKL3</i>	0.014	0.682	3.8 ↓	0.035	0.659	4.5 ↓	0.009	0.671	4.1 ↓
<i>TPT1</i>	0.003	0.720	2.0 ↓	0.061	0.641	0.7 ↓	0.005	0.681	1.9 ↓
<i>DPM1</i>	0.004	0.712	2.6 ↓	0.123	0.617	0.6 ↓	0.011	0.665	2.4 ↓

Quantitative PCR was used to validate the microarray findings on an independent clinical sample set, including saliva from 30 pancreatic cancer patients, 30 healthy control subjects, and 30 chronic pancreatitis patients. Wilcoxon* Signed Rank test: if $P < 0.05$, the marker is validated. ↑: Up-regulated in pancreatic cancer; ↓: Down-regulated in pancreatic cancer.

Table 3
Combination of Salivary Biomarkers for Pancreatic Cancer Selected by Logistic Regression Model

Biomarker combination	AUC (95% CI)	Sensitivity	Specificity	cv.err
Pancreatic cancer vs. Healthy control (<i>KRAS</i> , <i>MBD3L2</i> , <i>ACRV1</i> and <i>CDKL3</i>)	0.973 (0.895 to 0.997)	0.933	1	0.067
Pancreatic cancer vs. Chronic pancreatitis (<i>CDKL3</i> , <i>MBD3L2</i> , <i>KRAS</i>)	0.981 (0.907 to 0.997)	0.967	0.967	0.067
Pancreatic cancer vs. non-cancer (<i>KRAS</i> , <i>MBD3L2</i> , <i>ACRV1</i> and <i>DPM1</i>)	0.971 (0.911 to 0.994)	0.9	0.95	0.033

The logistic regression model was built based on the validated mRNA biomarkers for distinguishing pancreatic cancer from healthy controls, pancreatic cancer from chronic pancreatitis, and pancreatic cancer from the non-cancer group. The best models for each comparison, providing the highest discriminatory power with the simplest combination, are shown with the symbol of each biomarker. The sensitivity and specificity for each model was obtained by identifying the cutoff point in the predicted probabilities from the logistic regression that maximized the sum of the sensitivity plus specificity. In general, these cutoff points correspond well with the proportion of cancer patients evaluated in each model. Abbreviations: 95% CI: 95% Confidence interval; cv.err: cross validation error rate.

Table 4
Cross-disease Comparison of Microarray Profiles of 12 Validated mRNA Biomarkers

Gene symbol	Pancreatic cancer	Oral cancer	Lung cancer	Breast cancer
<i>MBD3L2</i>	0.011	0.391	0.770	0.419
<i>KRAS</i>	< 0.001	0.248	0.346	0.906
<i>STIM2</i>	0.013	0.160	0.479	0.963
<i>DMXL2</i>	0.009	0.869	0.056	0.226
<i>ACRVI</i>	0.004	0.946	0.304	0.397
<i>DMD</i>	0.008	0.633	0.979	0.558
<i>CABLES1</i>	0.002	0.574	0.096	0.473
<i>TK2</i>	0.014	0.966	0.007	0.311
<i>GLTSCR2</i>	0.006	0.417	0.336	0.073
<i>CDKL3</i>	< 0.001	0.107	0.227	0.190
<i>TPT1</i>	0.007	0.213	0.331	0.422
<i>DPM1</i>	0.005	0.135	0.082	0.428

Cancer specificity of the twelve validated mRNA biomarkers were evaluated across different microarray discovery studies that has been performed in our laboratory on diverse cancers, including pancreatic cancer, oral cancer, breast cancer and lung cancer. T-test p-values were calculated for each transcript between cancers and healthy controls in different microarray studies. Except *TK2* that also showed significant variation in lung cancer microarray study ($P = 0.007$), the rest mRNAs/transcripts that showed significant variations in pancreatic cancer study were not significantly altered in other cancer microarray studies ($P > 0.05$).