

RESEARCH ARTICLE

Prognostic Fifteen-Gene Signature for Early Stage Pancreatic Ductal Adenocarcinoma

Dung-Tsa Chen^{1*}, Ashley H. Davis-Yadley², Po-Yu Huang³, Kazim Husain², Barbara A. Centeno⁴, Jennifer Permut-Wey⁵, Jose M. Pimiento², Mokenge Malafa²

1 Department of Biostatistics and Bioinformatics, Moffitt Cancer Center & Research Institute, Tampa, FL, 33612, United States of America, **2** Department of Gastrointestinal Oncology, Moffitt Cancer Center & Research Institute, Tampa, FL, 33612, United States of America, **3** Department of Applied Mathematics and Institute of Statistics, National Chung Hsing University, Taichung City, Taiwan, **4** Department of Anatomic Pathology, Moffitt Cancer Center & Research Institute, Tampa, FL, 33612, United States of America, **5** Department of Cancer Epidemiology, Moffitt Cancer Center & Research Institute, Tampa, FL, 33612, United States of America

* dung-tsa.chen@moffitt.org



OPEN ACCESS

Citation: Chen D-T, Davis-Yadley AH, Huang P-Y, Husain K, Centeno BA, Permut-Wey J, et al. (2015) Prognostic Fifteen-Gene Signature for Early Stage Pancreatic Ductal Adenocarcinoma. PLoS ONE 10 (8): e0133562. doi:10.1371/journal.pone.0133562

Editor: Francisco X. Real, Centro Nacional de Investigaciones Oncológicas (CNIO), SPAIN

Received: March 12, 2015

Accepted: June 28, 2015

Published: August 6, 2015

Copyright: © 2015 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Microarray data are deposited in GEO (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57495>).

Funding: The study was supported in part by the National Institutes of Health (5P30CA076292 and 1R01CA129227), the DeBartolo Family Personalized Medicine Institute Pilot Research Awards in Personalized Medicine, Taiwan National Science Council (NSC 101-2118-M-005-002), and Taiwan Graduate Students Study Abroad Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The outcomes of patients treated with surgery for early stage pancreatic ductal adenocarcinoma (PDAC) are variable with median survival ranging from 6 months to more than 5 years. This challenge underscores an unmet need for developing personalized medicine strategies to refine the current treatment decision-making process. To derive a prognostic gene signature for patients with early stage PDAC, a PDAC cohort from Moffitt Cancer Center (n = 63) was used with overall survival (OS) as the primary endpoint. This was further evaluated using an independent microarray cohort dataset (Stratford et al: n = 102). Technical validation was performed by NanoString platform. A prognostic 15-gene signature was developed and showed a statistically significant association with OS in the Moffitt cohort (hazard ratio [HR] = 3.26; p<0.001) and Stratford et al cohort (HR = 2.07; p = 0.02), and was independent of other prognostic variables. Moreover, integration of the signature with the TNM staging system improved risk prediction (p<0.01 in both cohorts). In addition, NanoString validation showed that the signature was robust with a high degree of reproducibility and the association with OS remained significant in the two cohorts. The gene signature could be a potential prognostic tool to allow risk-adapted stratification of PDAC patients into personalized treatment protocols; possibly improving the currently poor clinical outcomes of these patients.

Introduction

Pancreatic cancer is the fourth leading cause of cancer death in the United States with an estimated 38,000 deaths in 2013[1]. Only twenty percent of patients with pancreatic ductal adenocarcinoma (PDAC) have stage I and II disease and are candidates for potential curative

Competing Interests: The authors have declared that no competing interests exist.

treatment that typically includes surgical resection and adjuvant chemotherapy with or without radiation treatment. While the five-year survival rate for curative-intent surgical resection for pancreatic cancer is 15 to 30% [2, 3], there is substantial individual variation. Currently, the only accepted prognostic factor guiding treatment decisions for both surgeons and oncologists is the AJCC TNM staging. However, the prognostic performance of AJCC TNM staging for more than 80% of patients with resected pancreatic cancer (Stages IB, IIA, and IIB) is very poor, with the survival curves being virtually identical[4]; therefore, the current practice is to uniformly treat all patients with stage I and II PDAC with surgical resection followed by adjuvant therapy. This approach results in potential undertreatment with surgical resection for patients who are at high risk of early recurrence and overtreatment with adjuvant therapy for patients who are at low risk of recurrence. The core obstacle to personalized management strategies is the lack of definitive prognostic biomarker(s) to identify stage I and II PDACs with a high probability of occult metastases and corresponding poorer clinical outcomes. Better prognostic tools are needed to identify patients predicted to be at high or low risk of recurrence to help guide treatment decisions for medical and radiation oncologists as well as pancreatic surgeons.

Recent advances in genomic cancer research have led to numerous biomarker discoveries; some have been validated and have become clinical assays to help improve patient care. For example, Oncotype DX, a 21-gene signature has been used to predict breast cancer recurrence in patients with node negative disease, such that patients with a high risk score are recommended for receiving adjuvant chemotherapy[5, 6]. In the management of advanced non-small cell lung cancer, epidermal growth factor receptor (EGFR) mutation testing is routinely used to guide treatment plans[7]. While there are successful stories of biomarker application for clinical use in breast cancer[5], lung cancer[8], colorectal cancer[9], and melanoma[10], sadly, the clinical application of biomarkers in PDAC is very limited. For example, serum CA 19-9[11], a FDA-approved biomarker (since 1980s), has been used for prognosis in PDAC even though at least 10% of patient do not express it and metabolic abnormalities such as hyperbilirubinemia can affect it significantly. As such, there is an unmet need to develop reliable and robust biomarkers such as gene signatures to better predict outcomes and to help tailor treatment plans (such as use of surgery versus systemic therapy).

In response to this unmet need, we have analyzed microarray data from a Moffitt cohort of 63 patients with early stage PDAC (stage IB, IIA and IIB) and developed a prognostic 15-gene signature to predict overall survival (OS). We hypothesize that high risk for poor clinical outcomes in early-stage PDAC are reflected by specific transcriptomic features from this 15-gene signature.

Materials and Methods

Study cohorts

The study aimed to utilize the Moffitt cohort to develop a prognostic gene signature for PDAC patients. Since the sample size was not very large ($n = 63$), an external cohort (Stratford et al; $n = 102$) was used for validation. We targeted a large effect size ($HR > 3$) in order to reach 80% power or higher for a dichotomized gene signature score. Specifically, for the Moffitt cohort, with 42 events and assumption of 1:1 ratio of low and high risk groups by the gene signature, the power could reach at least 80% to detect a HR of 2.6 with type I error controlled at 5%. For the Stratford et al cohort (66 events), a ratio of 1:1 for low and high risk groups by the gene signature could reach 98% to detect a HR of 3. For a ratio of 1:3, the power was 80%. Power calculation was based on 'powerSurvEpi' R package. Analysis of combination of the signature with TNM staging system was also performed, but mainly for exploration due to potential small sample size issue in subgroup. However, if the analysis showed consistent patterns in both

cohorts, the results would be reported and discussed to warrant further study. A flow chart of study cohorts and data analysis was provided in [S1 Fig](#).

Moffitt cohort

This is a retrospective microarray study of resected PDAC samples at Moffitt Cancer Center constituting the primary dataset to develop the gene signature. The sample cohort, consisting of fresh frozen macrodissected tumor tissues from 63 patients with early stage PDAC, was collected and profiled from 2006 to 2011 under the Moffitt Total Cancer Care protocol and the research protocol MCC 17779 approved by Moffitt’s Scientific Review Committee and the Institutional Review Board (IRB #10.07.0008). The inclusion criteria were patients who underwent surgery for early stage/resectable (stage I and II) pancreatic cancer, that were consented for the TCC protocol and that had complete gene expression analysis. Exclusion criteria were patients found to be metastatic at presentation or patients without gene expression data. Patient information was anonymized and de-identified. Gene expression data were generated using Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57495>) The median follow-up time was 21 months (21 samples were from patients who were alive and 42 samples were from patients who had died). In other words, two thirds (67% = 42/63) of patient died in the follow-up period. Of these 42 patients, 9 (21% = 9/42) died from other diagnosis not cancer recurrence (n = 8) or unknown disease status (n = 1). However, due to nature of the retrospective study and limitation of registration of recurrent disease, patient died caused by recurrent disease may not be diagnosed. Distribution of survival time was not significantly different between patients who died from other causes and patients who died by cancer recurrence (p = 0.71; mean of survival time: 16.2 (standard deviation (SD) = 12.37) versus 18.2 (SD = 8.62) months for other causes and cancer recurrence, respectively). Clinical predictors included TNM stage, T stage, N stage, gender, histology grade, and age at diagnosis ([Table 1](#)).

Table 1. Descriptive statistics of clinical predictors in the Moffitt PDAC cohort (n = 63).

TNM stage	IB	13
	IIA	17
	IIB	33
T stage	T1	2
	T2	14
	T3	47
N stage	N0	30
	N1	32
	NA	1
Gender	Female	30
	Male	33
Histology grade	Well-Differentiated	6
	Moderate-Differentiated	35
	Poor-Differentiated	18
	NA	4
Age at diagnosis	youngest	24 years
	1st quartile	60.5 years
	median	68 years
	3rd quartile	74 years
	oldest	86 years

doi:10.1371/journal.pone.0133562.t001

One independent microarray dataset in pancreatic cancer was included to validate the gene signature: *Stratford et al [12] localized PDAC study* (<http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE21501>). The Stratford et al cohort [12] examined 132 patients with PDAC for genomic profiling. Among these patients, there were 102 patients with OS available. These patients (n = 102) were used to validate our gene signature. Microarray data were generated using Agilent-014850 Whole Human Genome Microarray 4x44K G4112F Array. We evaluated if the signature could predict OS in this cohort.

NanoString validation

RNA samples of fresh frozen macrodissected tumor tissues from 53 PDAC patients in the Moffitt cohort (a subset of the 63 PDAC patients) were used to measure gene expression of the 15-gene signature in the NanoString platform. The NanoString Assays were performed with 150-ng aliquots of RNA using the NanoString nCounter Analysis system. Nineteen invariant genes were selected to serve as house-keeping genes for normalization (S1 Table). After codeset hybridization overnight, the samples were washed and immobilized to a cartridge using the NanoString nCounter Prep Station. Cartridges were scanned in the nCounter Digital Analyzer at 555 fields of view for the maximum level of sensitivity. Gene expression was normalized using NanoStringNorm R package. Specifically, background correction was performed using the negative control at the cutoff of mean + 2 standard deviation. House-keeping genes were used to for normalization based on geometric mean.

Statistical Analysis

Data normalization. Different normalization methods could affect the results. However, there is no gold standard of which normalization to be best used. The problem gets more complicated when both cohorts used different microarray platforms. Each platform favors certain types of normalization to account for its unique microarray design. For Affymetrix genechip, RMA [13] is a common approach for normalization while the Loess method [14] is often used for the Agilent microarray. For this reason, both normalizations were used for the study.

Development of a prognostic gene signature. The goal was to develop a prognostic gene signature. We used OS as the primary endpoint for analysis in the Moffitt cohort as the training cohort. Instead of performing supervised whole genome analysis, the first step was to use sparse PCA [15, 16] to filter out most genes with small variation. The remaining genes were analyzed using univariate analysis of Cox proportional hazards model to identify genes associated with OS at the 25% level of the false discovery rate [17]. Genes associated with OS were stratified into two subsets: (1) genes with higher expression associated with poor survival and (2) genes with lower expression associated with poor survival. The PC1 scoring system below was used for each gene subset for further analysis.

PC1 scoring system. An overall risk score was generated by principal component analysis (PCA) to reflect the combined effect of a gene signature. Specifically, we used the first principal component (PC1) to represent the overall expression level for the signature. That is, PC1, defined as $\sum w_i x_i$, is a weighted average expression among all genes in the signature, where x_i represents gene i expression level, w_i is the corresponding weight (PC1's loading coefficient for gene i) with $\sum w_i^2 = 1$, and the w_i values maximize the variance of $\sum w_i x_i$. Prior to PCA, data were standardized by centering the mean and scaled by the standard deviation for each gene in the Moffitt cohort (training cohort). The standardized expression data were then used in PCA to generate principal components (PCs). Total variation of PC1 was examined to ensure at least two fold increase (compared to PC2) or more than 30% (both were based on our experiences in previous cancer studies [18, 19]). The PC1's loading coefficients were then used (fixed)

for both Moffitt and Stratford cohorts using their standardized expression data. The uniqueness of the PC1 scoring system is that it explains the largest total variation, likely linked to biological effect. More importantly, it integrates all molecular features into one score (efficient data reduction) for each patient, simplifying clinical decision-making. This approach has been used to derive various gene signatures previously [18–21]. For example, in our studies of lung and breast cancer [18, 19], we used PC1 to capture most gene signature information and this PC1 scoring approach was able to demonstrate the clinical association (e.g., cancer risk, prognosis, and prediction of chemotherapy) of gene signatures.

Association with OS and other clinical predictors. The influence of each gene signature was tested to see if the overall survival of two risk groups (high PC1 and low PC1) formed by a median-split of the PC1 score were statistically significantly different. The two-sided log-rank test was used to calculate p values. Evaluation of the median-split PC1 score, as an independent factor predicting PDAC prognosis, was done by clinical predictors including TNM stage (IB, IIA, and IIB), gender, histology grade, and age at diagnosis using multivariable Cox proportional hazards regression analysis.

External validation. One independent cohort was used to validate each gene signature. Due to the difference of microarray platforms in each cohort, gene level data were used for evaluation (a gene expression level was defined as an average of the expression level for a set of probe sets for the same gene; any probeset with a missing value was excluded). Before analysis, data were standardized by centering the mean and scaled by the standard deviation for each gene. The predetermined PC1's loading coefficients (weights) derived from the Moffitt cohort were used to calculate the PC1 score for the independent cohort.

Evaluation of chance as a random gene signature. A resampling scheme was used to generate 100,000 random gene signatures to simulate chance significance for the two pancreatic cancer datasets. Specifically, this analysis was performed to evaluate whether a selected signature was not inferior to random signatures. Each random signature was generated by randomly selecting the same number of probesets as the selected signature from the Moffitt cohort; each random signature was then applied to the two datasets to determine the significance level. The p value, as a random gene signature for a targeted gene signature, was defined as the proportion of p values from the random signatures less than the observed p value from the targeted signature in both datasets. That is, it was defined as a joint probability of p values from the random signatures generated from the Moffitt cohort less than the observed p value in the Moffitt cohort and p values from the random signatures generated from the Stratford et al cohort less than the observed p value in the Stratford et al cohort. Instead of using all the probesets, only the probesets that passed sparse PCA filtering were used to evaluate the chance of the candidate signature as a random noise signature.

Results

A prognostic 15-gene signature

By analyzing early stage PDAC patients from the Moffitt cohort of 63 patients with PDAC, a prognostic 15-gene signature was developed to predict OS. Specifically, sparse principal component analysis [15, 16] screened out most genes (59,918 probesets) and yielded 689 probesets for 488 genes with non-zero coefficients using L1-penalty at the amount of 10^{-6} . Further study of these genes using univariate analysis of Cox proportional hazards model showed 38 probesets (32 genes) associated with OS at 25% false discovery rate. The gene set was then separated into two subsets: (1) genes with higher expression associated with poor OS (18 probesets for 15 genes: C6orf15, CAPN8, HIST1H3H, IGF2BP3, KIF14, KRT6A, PMAIP1, PPBP, RTKN2, SCEL, SERPINB5, SLC2A1, SLC45A3, TMPRSS3, UCA1; [S2 Table](#)) and (2) genes

with lower expression associated with poor OS (20 probesets for 17 genes). While both gene subsets showed the prognostic effect in the Moffitt cohort, the gene subset with lower expression associated with poor OS was unable to demonstrate the clinical association in later external validation (Stratford et al cohort [12]). Therefore, we focused on the gene subset with higher expression associated with poor OS, the 15-gene signature, for evaluation. Since the first principal component (PC1) explains most total variation (51%; S2 Fig), a scoring system was developed using the PC1 to derive a risk score to summarize gene expression for the signature.

An independent prognostic predictor

The continuous PC1 score for the 15-gene signature showed significant association with OS (hazard ratio [HR] = 1.23 and $p = 0.0007$). Inclusion of PC2 and PC3 did not improve the analysis appreciably (S3 Table). Since the ratio of negative versus positive lymph node or stage (IB-IIA) versus stage IIB was about 1:1, median of the PC1 score was used to dichotomize the PC1 score and therefore to classify patients into low and high risk groups (low and high PC1 score, respectively). Results showed the high PC1 group had a poorer survival than the low PC1 group (Fig 1: HR = 3.26 and $p = 0.0002$ by log-rank test; median survival time: 2.92 (35 months) and 1.25 (15 months) years for low and high PC1, respectively). Importantly, both continuous and dichotomized PC1 scores remained significantly associated with OS after adjustment for other covariates, including histology grade, AJCC stage, gender, and age at diagnosis (HR = 3.03 and $p = 0.0015$ for the median dichotomized PC1; HR = 1.19 and $p = 0.0056$ for the continuous PC1). In addition, the continuous PC1 score had a weak correlation with histology grade ($p = 0.59$), gender ($p = 0.97$), and AJCC TNM stage ($p = 0.52$) by one-way ANOVA or two-sample t-test (S3 Fig).

Integration with the AJCC TNM staging system

To evaluate if the PC1 score could improve the AJCC TNM staging system for predicting better prognosis, the staging variable and the median dichotomized PC1 score were used to classify patients into 6 groups: (stage IB with low PC1: $n = 5$, stage IB with high PC1: $n = 8$, stage IIA with low PC1: $n = 12$, stage IIA with high PC1: $n = 5$, stage IIB with low PC1: $n = 15$, stage IIB with high PC1: $n = 18$). This classification yielded significant difference among the 6 Kaplan-Meier (KM) survival curves ($p < 0.001$ by log-rank test; Fig 2A). Further analysis by stratification of the staging variable, the PC1 score was able to classify patients into low risk (low PC1) and high risk (high PC1) in each stage with a statistically significant difference at stage IIA ($p < 0.001$; S4 Fig). Moreover, these 6 survival curves formed 3 distinct clusters ($p < 0.001$; Fig 2B): (1) Low risk: Low PC1 with stage IB and IIA (median survival time: never reach), (2) Intermediate risk: Low PC1 with stage IIB and high PC1 in stage IB (median survival time: 2.08 years), (3) High-risk: High PC1 with stage IIA and IIB (median survival time: 1.25 years).

External evaluation of the prognostic effect in the Stratford et al [12] localized PDAC study

There were 29 probesets matching to the 15-gene signature. Due to the difference of microarray platforms, gene level data were used for evaluation. The predetermined PC1 model from the Moffitt cohort was used to calculate the PC1 score for each patient in the Stratford et al cohort.

Prior to determining the cutoff, we evaluated comparability of the two cohorts regarding patient characteristics (T and N staging variable for both cohorts). Results showed both cohorts had a comparable distribution of T stage (T1-2: 16 (25%) vs. T3: 47 (75%) in the Moffitt cohort; T1-2: 18 (18%) vs. T3-4: 80 (82%) in the Stratford et al cohort; $p = 0.33$ by Fisher exact test).

PC1 (Low versus High): logn-rank p= 2e-04

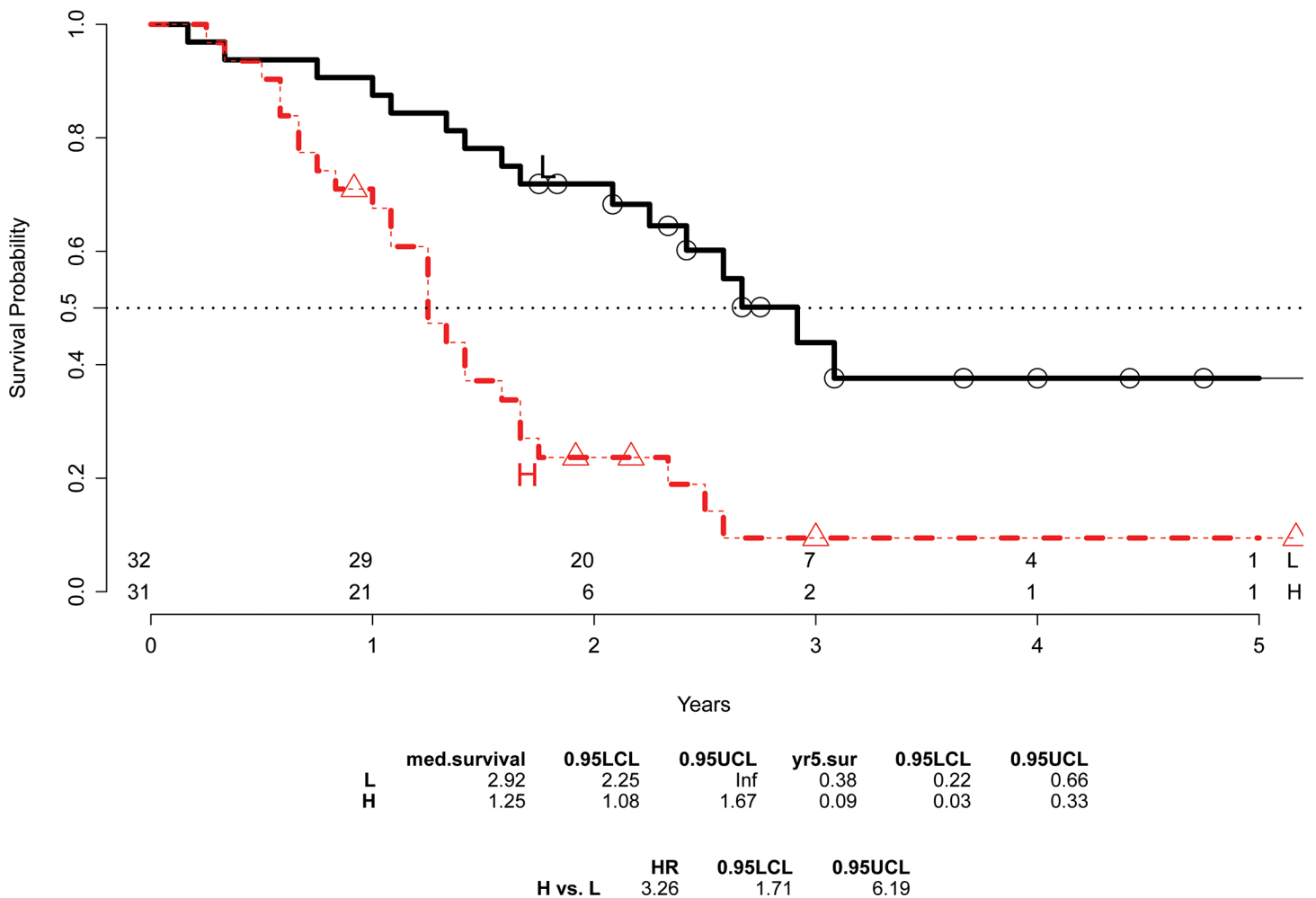


Fig 1. Association of the 15-gene signature with overall survival in the Moffitt cohort. A PC1 score was generated for each patient from the Moffitt cohort (n = 63) by principal component analysis to reflect the combined expression of the 15 genes. High and low PC1 groups were determined on the basis of a median split. Kaplan–Meier curves of overall survival are shown in the two groups. A statistically significant difference of the Kaplan–Meier survival curves between the high and low PC1 groups was determined by the two-sided log-rank test. The number of patients at risk is listed below the survival curves.

doi:10.1371/journal.pone.0133562.g001

However, distribution of the N staging variable was different between the two cohorts (N stage with negative lymph node: 28% in the Stratford et al cohort and 48% in the Moffitt cohort; $p = 0.01$). To adjust for the staging effect, the 1st quartile (25th percentile) of the PC1 score was used to classify patients into low and high PC1 in the Stratford et al cohort. A simulation study was conducted and showed that, even for an independent predictor, the cutoff guided by survival outcome-associated covariate (N staging) improves power compared to the cutoff derived from the training cohort (S4 Table). This classification yielded two significantly different survival curves with poor survival in the high PC1 group (HR = 2.07; $p = 0.02$; Fig 3) and maintained significance level (HR = 2.08; $p = 0.025$) after adjusting for the T and N staging variables. This PC1 classification had a comparable performance comparing to the N staging which also showed a significant association with OS ($p = 0.029$). There was a weak correlation

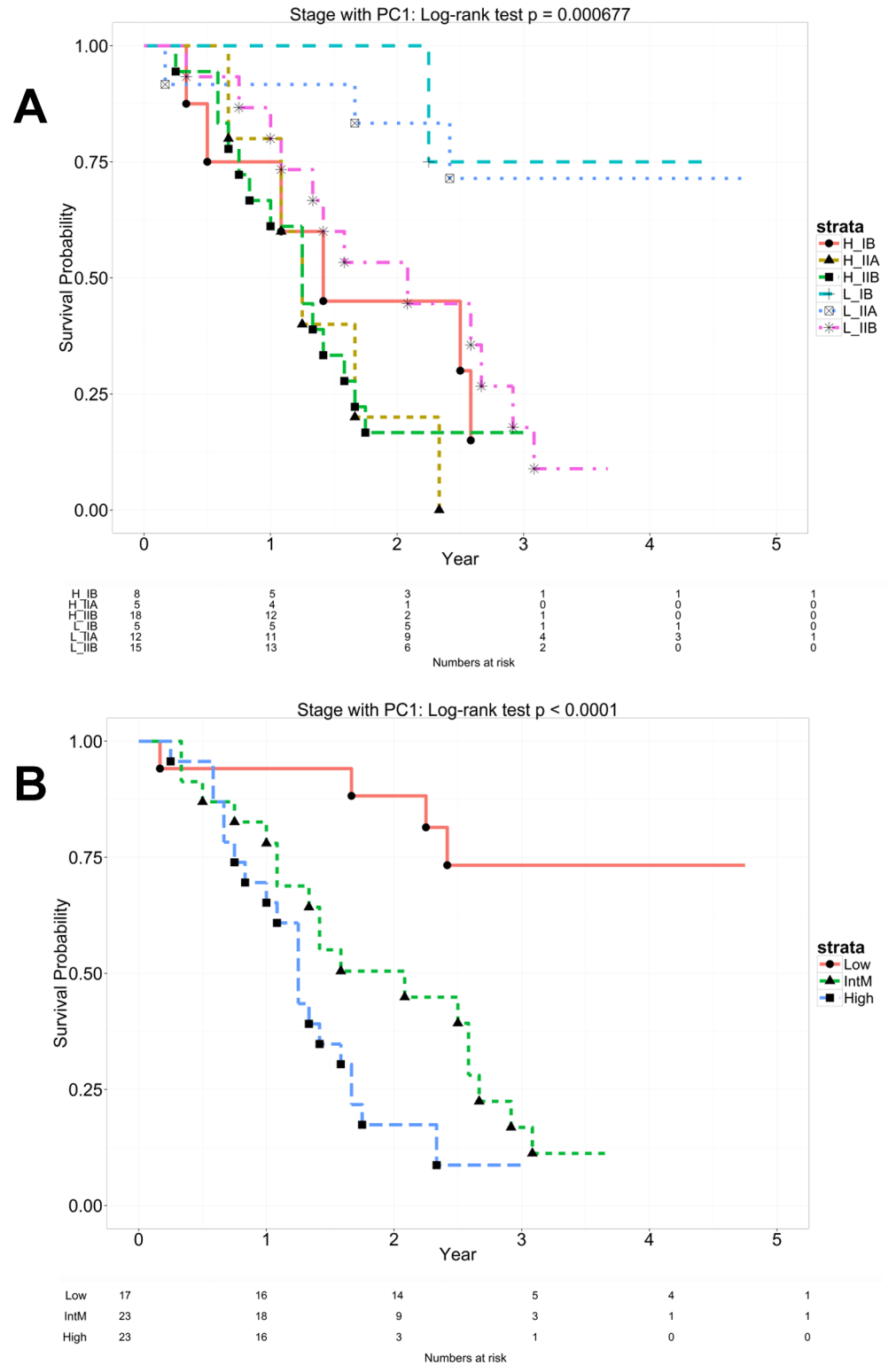


Fig 2. Integration of the 15-gene signature with the AJCC TNM staging system in the Moffitt cohort. A PC1 score was generated for each patient from the Moffitt cohort (n = 63) by principal component analysis to reflect the combined expression of the 15 genes. **(A)** The AJCC TNM staging variable and the median dichotomized PC1 score were used to classify patients into 6 groups: stage IB with low PC1: n = 5, stage IB with high PC1: n = 8, stage IIA with low PC1: n = 12, stage IIA with high PC1: n = 5, stage IIB with low PC1:

n = 15, stage IIB with high PC1: n = 18. Kaplan–Meier curves of overall survival were shown in these six groups. **(B)** Regrouping of the 6 survival curves into 3 distinct clusters: (1) Low risk: Low PC1 with stage IB and IIA (n = 17; MST = never reach), (2) Intermediate risk: Low PC1 with stage IIB and high PC1 in stage IB (n = 23; MST = 2.08 years), (3) High-risk: High PC1 with stage IIA and IIB (n = 23; MST = 1.25 years). A statistically significant difference of the Kaplan–Meier survival curves between the groups was determined by the two-sided log-rank test. The number of patients at risk is listed below the survival curves. MST = median survival time.

doi:10.1371/journal.pone.0133562.g002

of the PC1 score with T and N staging (p = 0.98 and 0.79, respectively). Further analysis using the N staging and the dichotomized PC1 to classify patients into 4 groups yielded significant differences among the 4 KM survival curves (p = 0.015 by log-rank test; [S5A Fig](#)) which formed

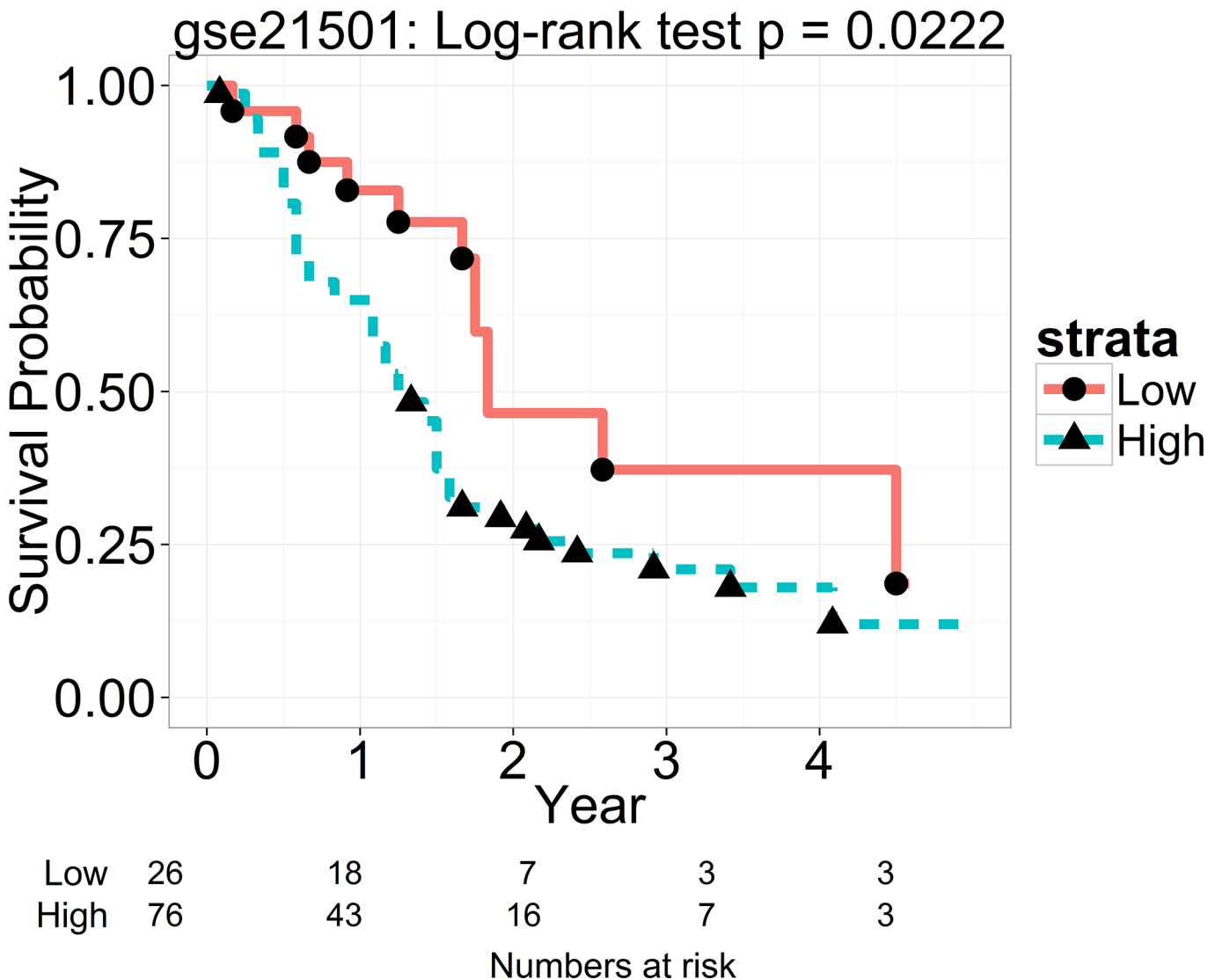


Fig 3. Prognostic value of the 15-gene signature in the Stratford et al cohort. A PC1 score was generated for each patient in the Stratford et al cohort (n = 102) using the loading coefficients of the first principal component from the Moffitt cohort. High and low PC1 groups were determined by the cutoff at the first quartile of the PC1 score to adjust for the distribution of the N staging. Kaplan–Meier curves of overall survival are shown in the two groups. A statistically significant difference of the Kaplan–Meier survival curves between the high and low PC1 groups was determined by the two-sided log-rank test. The number of patients at risk is listed below the survival curves.

doi:10.1371/journal.pone.0133562.g003

3 distinct clusters ($p = 0.005$; [S5B Fig](#)) similar to the ones in the Moffitt cohort: (1) Low risk: Low PC1 with N0 stage (median survival time: never reach), (2) Intermediate risk: Low PC1 with N1 stage and high PC1 in N0 stage (median survival time: 1.75 years), (3) High-risk: High PC1 with N1 stage (median survival time: 1.17 years).

Likelihood as a non-random gene signature

Since a study indicated many gene signatures were a random noise signature[22], we conducted a resampling approach to evaluate the likelihood of the 15-gene signature as a random signature. Results showed a joint p value of 0.00001 to be a random noise signature in both datasets for the 15 genes (18 probesets). In each cohort, the probability of p values from the random signatures less than the observed p value (0.0007) in the Moffitt cohort was 0.00001. For the Stratford et al cohort, the probability was 0.01.

NanoString validation

Gene expression by NanoString platform showed that the PC1 score of the 15-gene signature was associated with OS ($p = 0.03$; [Fig 4A](#)) in the Moffitt cohort. The predicted PC1 score also remained significantly associated with OS in the Stratford et al cohort ($p = 0.02$; [Fig 4B](#)). Pearson correlation analysis yielded 11 genes (73%) with a correlation (r) >0.6 , 1 gene with $r = 0.56$, 2 genes $r = 0.33$ – 0.39 , and 1 gene with $r = 0.02$ ([S5 Table](#)). Moreover, the PC1 loading coefficients were all positive with majority greater than 0.2 and comparable to the ones in the microarray data ([S6 Table](#)). The gene signature was further evaluated by sensitivity analysis based on different cutoff of correlation to select genes. Results showed that the PC1 based on the whole 15 genes or a subset of genes remained significant ($p < 0.05$) in most cases (the non-significant one had a marginal p value of 0.069; [S7 Table](#)).

Discussion

In the present study, we profiled resected primary tumors from early stage PDAC patients and developed a 15-gene signature based on overall survival. The signature was able to identify a low-risk subgroup of PDAC patients who were likely to survive longer than 2 years after surgery (median survival of 35 months) and a high-risk subgroup who likely survived less than 1.5 years (median survival of 15 months) ([Fig 1](#)). Importantly, the association of the signature with OS was independent of histology grade, AJCC TNM stage, gender, and age at diagnosis ($p = 0.0015$), indicating the potential clinical utility of the signature in predicting patients' outcomes. Since the signature was a predictor of OS independent of TNM staging, we evaluated if incorporation of the signature into the TNM staging system could improve survival prediction. Our evaluation results showed the signature could be a complementary prognostic adjunct to the AJCC TNM staging system ([Fig 2](#)). Specifically, the signature was able to differentiate patients with poor survival and good survival in each stage, especially at stage IIA ([S4B Fig](#)). Furthermore, combination of the signature and staging formed three distinct risk groups from low, intermediate, to high risk. This pattern was observed in both cohorts ([Fig 2B](#) and [S5B Fig](#)), especially for the two subgroups: (a) patients having low signature score and negative lymph nodes classified in the low risk group and (b) patients with high signature score and positive lymph node grouped in the high risk group. This consistent result indicates that a new classification system, by integrating the signature into the TNM staging system, could be a useful prognostic tool for early stage PDAC patients.

From a personalized medicine point of view, the signature offers one potential strategy. That is, there is an unmet need to identify high-risk PDAC patients for alternative treatment management given the unsatisfactory outcome of the standard of care (surgery followed by

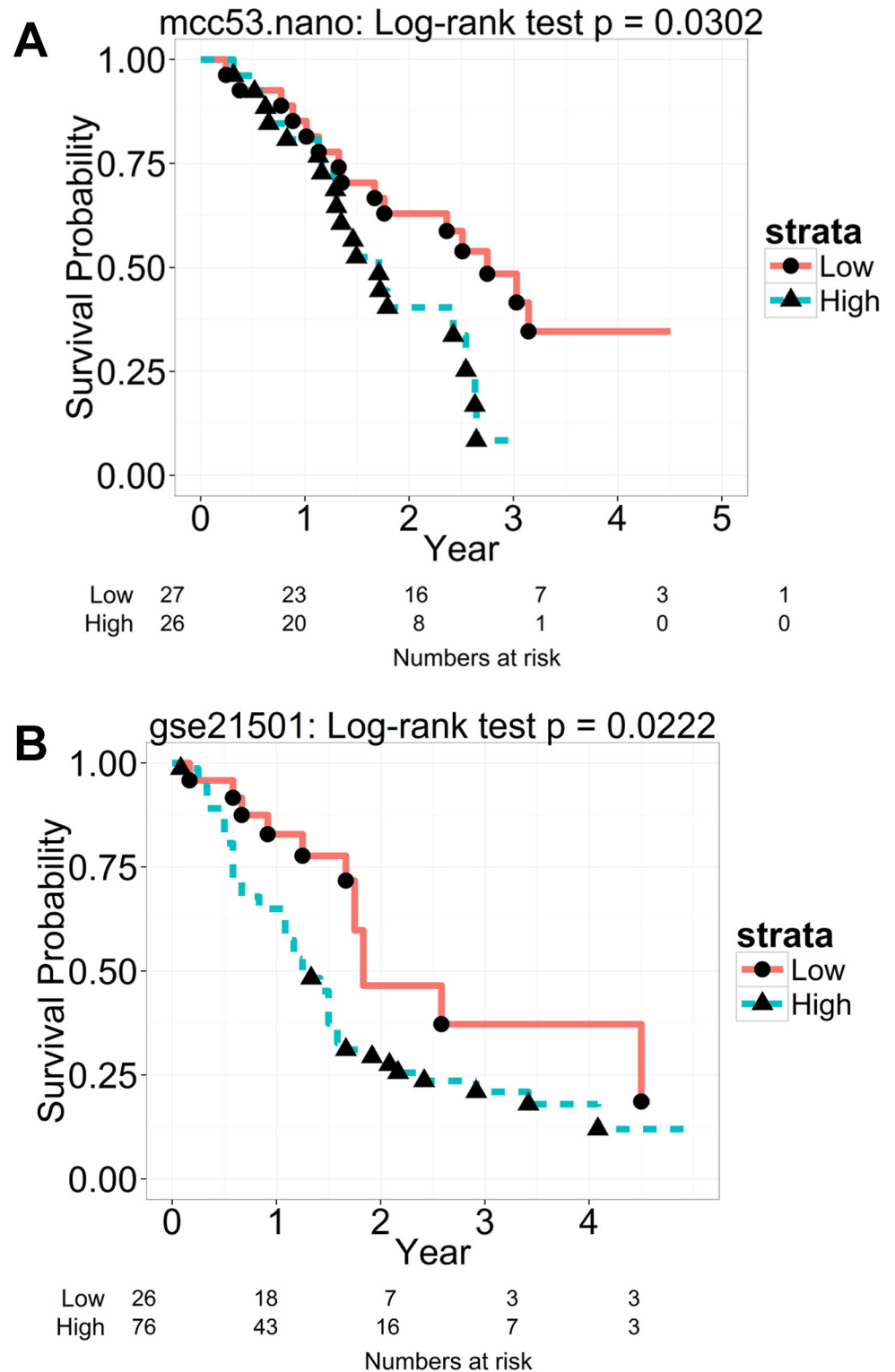


Fig 4. NanoString validation of the 15-gene signature. (A) By analyzing the NanoString gene expression data, a PC1 score was generated for each patient from the subset ($n = 53$) of the Moffitt cohort using principal component analysis to reflect the combined expression of the 15 genes. The median dichotomized PC1 score was used to classify patients into high and low PC1 groups. (B) For the Stratford et al cohort, the PC1 score was generated for each patient using the loading coefficients of the first principal component in the Moffitt cohort. High and low PC1 groups were determined by the cutoff at the first quartile of the PC1 score to

adjust for the distribution of the N staging. Kaplan–Meier curves of overall survival are shown in the two groups. A statistically significant difference of the Kaplan–Meier survival curves between the high and low PC1 groups was determined by the two-sided log-rank test. The number of patients at risk is listed below the survival curves.

doi:10.1371/journal.pone.0133562.g004

adjuvant therapy)[23–28]. It is possible that these high-risk patients have developed micrometastases before surgery. In this case, neoadjuvant therapy may offer a better solution to treat micrometastases to prevent disease recurrence, or just to allow the patient and their disease to declare themselves and avoid unnecessary surgery. In PDAC patients with clinically positive lymph nodes, the signature identifies a subgroup with a high signature score. This high-risk group has a median survival time less than 15 months (Fig 2B and S5B Fig) which could be a potential candidate for the neoadjuvant therapy. However, future prospective validation cohort may shed light on the issue as the survival differences may be impacted by the fact that our analysis for intermediate risk patients includes N1 and N0 disease, and the survival differences in this group may be attributable to the presence of nodal disease.

Our literature review indicates that the 15-gene signature may have clinical relevance. Forty percent of the 15 genes (6 genes) showed their association with various clinical outcomes in pancreatic cancer by other studies[29–39]: (1) IGF2BP3: Elevation of mRNA and protein expression levels was significantly increased in PDAC tissues (but not in benign pancreatic tissues) and also in PDAC patients with poor survival[29, 30]; (2) KIF14: Over-expression of mRNA and protein levels was observed in all highly nerve-invasive pancreatic tumor cells, suggesting an association with perineural invasion[31]; (3) PPBP: The plasma level of this protein was higher in pancreatic cancer patients than in healthy controls and was independent of CA19-9 serum levels. The combination of this protein with CA19-9 improved the detection of pancreatic cancer at an early stage[39]; (4) SERPINB5: Increased mRNA expression was correlated with increased metastasis[36] and to progression of PDAC from pancreatic intraepithelial neoplasias (PanINs)[34], as well as associated with poor survival[34]; (5) SLC2A1: Positive expression by immunohistochemistry was associated with higher-grade PanIN and IPMN[37, 38], suggesting its involvement in the preinvasive phase of pancreatic carcinogenesis; (6) TMRSS3: This gene was more highly expressed in pancreatic cancers than in non-neoplastic tissues[32], as well as associated with metastasis[33]. Moreover, the signature yielded a high degree of reproducibility from microarray to NanoString platform (S5 and S6 Tables) and the association with OS remained significant in the two cohorts ($p < 0.05$; Fig 4). In addition, the resampling test showed the likelihood to be a false positive signature was rare ($p = 0.00001$). By taking these considerations together, we believe the 15-gene signature is not a random-noise signature, but a signature specific to PDAC as a prognostic tool to improve treatment decisions.

While various gene signatures have been developed in PDAC patients, the majority of them focus on separating between PDAC and normal pancreas or on the progression of dysplastic lesions[32, 40–42]. Only a few prognostic gene signatures[43–46] were developed. Most of them were with a small number of genes in the signatures due to the challenge of uniform global gene expression in PDAC. None of these signatures were overlapped with the 15-gene signature, except two genes listed in one gene signature[46]. Two signatures[12, 43] we evaluated did not overlap with the 15-gene signature and failed to show significant association with OS using the Moffitt cohort. This non-overlapped result is not uncommon. Different studies often yield different gene signatures within the same disease type. Many factors contribute to the inconsistent result, such as study design, characteristics of study cohort, and analysis approaches. Thus, utilization of publically available microarray data as external validation

cohorts is not straightforward. We understand that the gene signature developed in our cohort is unlikely to be the best in other cohorts because it was not trained and optimized in external cohorts. However, if it shows an overall significant level in the external cohorts, the chance as a false positive gene signature becomes small. For the 15-gene signature, while it was not in the list of the optimal gene set in the Stratford et al cohort (data not shown), the p value showed an overall significant level ($p = 0.02$). In addition, it passes the noise signature test. Therefore, the 15-gene signature is unlikely a false positive signature.

Our study has some limitations. We have shown the prognostic value of the 15-gene signature using the Moffitt cohort and one publically available microarray dataset in PDAC. However, in order to be considered as a personalized medicine strategy for clinical decision making, validation of the 15-gene signature in a larger independent dataset is needed. Successful validation will advance the 15-gene signature to the next level for the analytical and clinical validity. We plan to evaluate the 15-gene signature and complete a large-scale validation using formalin-fixed paraffin embedded (FFPE) tissues from Total Cancer Care [47, 48] collected at the Moffitt Cancer Center. Since FFPE tissues are routinely collected and stored long-term on pancreatic cancer patients and Nanostring platform performs quite well with FFPE tissue [49], if the signature using the Nanostring platform could be validated in FFPE tissues, the signature would be a great clinical utility for broad application in personalizing treatment care.

Another issue is determination of the cutoff to define low and high risk groups in the validation cohort. Ideally, to avoid over-fitting, the cutoff for the validation cohort should be derived from the predetermined model in the training cohort, especially for an independent predictor. However, each cohort may have a different distribution of patient characteristics and the microarray experiments may not be implemented by the same procedure with the same platform. Therefore, it has been challenging to implement such an ideal formula. Various modifications were used to address the issue, such as the use of median cutoff in each cohort (training and validation cohorts) [50]. In this study, we used a survival-outcome-associated covariate to guide the cutoff of the signature score in both training and validation cohorts, rather than an arbitrary cutoff, such as median or quartile. Simulation results indicate, even for an independent predictor, the cutoff guided by outcome-associated covariate improves power compared to the cutoff derived from the training cohort (S4 Table)

In summary, this 15-gene signature could be useful to improve prediction of OS in PDAC patients. This is a potential prognostic tool to allow risk-adapted stratification of pancreatic cancer patients into personalized treatment protocols, thereby improving the currently poor clinical outcomes of these patients. Future prospective studies are needed to determine if the 15-gene signature can be used clinically to benefit early stage PDAC patients.

Supporting Information

S1 Fig. Flow chart of data analysis.

(PDF)

S2 Fig. Principal component analysis of the 15-gene signature.

(PDF)

S3 Fig. Association of the 15-gene signature with histology grade, gender, and TNM staging system in the Moffitt cohort.

(PDF)

S4 Fig. Analysis of the association between the 15-gene signature and overall survival by TNM stage in the Moffitt cohort.

(PDF)

S5 Fig. Integration of the 15-gene signature with the AJCC TNM staging system in the Stratford et al cohort.

(PDF)

S1 Table. Housekeeping genes (19 genes) for NanoString validation.

(PDF)

S2 Table. Univariate analysis of the 15-gene signature in microarray data at the Moffitt cohort.

(PDF)

S3 Table. Analysis of the first three principal components of the 15-gene signature (18 probesets) and 689 probesets after sparse PCA filtering for the association with OS using Cox proportional hazards model in the Moffitt cohort (n = 63).

(PDF)

S4 Table. Simulation study to evaluate the cutoff based on survival-outcome-associated covariate.

(PDF)

S5 Table. Correlation analysis of microarray and NanoString data at the Moffitt cohort (N = 53).

(PDF)

S6 Table. PC1 loading coefficients of the 15-gene signature.

(PDF)

S7 Table. NanoString validation by various cutoffs of Pearson correlation for gene selection.

(PDF)

Acknowledgments

The research was made possible through the Total Cancer Care Protocol at the H. Lee Moffitt Cancer Center & Research Institute. Total Cancer Care is enabled, in part, by the generous support of the DeBartolo Family, and we thank the many patients who so graciously provided tissue and data to the Total Cancer Care Consortium. Our study also received valuable assistance from the Information Shared Services, Tissue, Molecular Genomics, Biostatistics, and Cancer Informatics Core Facilities at the H. Lee Moffitt Cancer Center & Research Institute, an NCI designated Comprehensive Cancer Center, supported under NIH grant P30-CA76292.

Funding: The study was supported in part by the National Institutes of Health (5P30CA076292 and 1R01CA129227), the DeBartolo Family Personalized Medicine Institute Pilot Research Awards in Personalized Medicine, Taiwan National Science Council (NSC 101-2311-B-005-002), and Taiwan Graduate Students Study Abroad Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: None of the authors listed above has declared any conflicts of interest with this manuscript. There is a patent pending for the use of the 15-gene signature.

Author Contributions

Conceived and designed the experiments: DTC MM JMP. Performed the experiments: DTC AHD JMP. Analyzed the data: DTC PYH. Contributed reagents/materials/analysis tools: MM JMP KH BAC. Wrote the paper: DTC MM JWP JMP.

References

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA: a cancer journal for clinicians*. 2014; 64(1):9–29. Epub 2014/01/09. doi: [10.3322/caac.21208](https://doi.org/10.3322/caac.21208) PMID: [24399786](https://pubmed.ncbi.nlm.nih.gov/24399786/).
2. Ferrone CR, Brennan MF, Gonen M, Coit DG, Fong Y, Chung S, et al. Pancreatic adenocarcinoma: the actual 5-year survivors. *Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract*. 2008; 12(4):701–6. Epub 2007/11/21. doi: [10.1007/s11605-007-0384-8](https://doi.org/10.1007/s11605-007-0384-8) PMID: [18027062](https://pubmed.ncbi.nlm.nih.gov/18027062/).
3. Schnelldorfer T, Ware AL, Sarr MG, Smyrk TC, Zhang L, Qin R, et al. Long-term survival after pancreatoduodenectomy for pancreatic adenocarcinoma: is cure possible? *Annals of surgery*. 2008; 247(3):456–62. Epub 2008/04/01. doi: [10.1097/SLA.0b013e3181613142](https://doi.org/10.1097/SLA.0b013e3181613142) PMID: [18376190](https://pubmed.ncbi.nlm.nih.gov/18376190/).
4. Helm J, Centeno BA, Coppola D, Melis M, Lloyd M, Park JY, et al. Histologic characteristics enhance predictive value of American Joint Committee on Cancer staging in resectable pancreas cancer. *Cancer*. 2009; 115(18):4080–9. Epub 2009/07/25. doi: [10.1002/cncr.24503](https://doi.org/10.1002/cncr.24503) PMID: [19626671](https://pubmed.ncbi.nlm.nih.gov/19626671/).
5. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine*. 2004; 351(27):2817–26. Epub 2004/12/14. doi: [10.1056/NEJMoa041588](https://doi.org/10.1056/NEJMoa041588) PMID: [15591335](https://pubmed.ncbi.nlm.nih.gov/15591335/).
6. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*. 2006; 24(23):3726–34. doi: [10.1200/JCO.2005.04.7985](https://doi.org/10.1200/JCO.2005.04.7985) PMID: [16720680](https://pubmed.ncbi.nlm.nih.gov/16720680/).
7. Langer CJ. Epidermal growth factor receptor inhibition in mutation-positive non-small-cell lung cancer: is afatinib better or simply newer? *J Clin Oncol*. 2013; 31(27):3303–6. Epub 2013/08/28. doi: [10.1200/JCO.2013.49.8782](https://doi.org/10.1200/JCO.2013.49.8782) PMID: [23980079](https://pubmed.ncbi.nlm.nih.gov/23980079/).
8. Kratz JR, He J, Van Den Eeden SK, Zhu ZH, Gao W, Pham PT, et al. A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. *Lancet*. 2012; 379(9818):823–32. Epub 2012/01/31. doi: [10.1016/S0140-6736\(11\)61941-7](https://doi.org/10.1016/S0140-6736(11)61941-7) PMID: [22285053](https://pubmed.ncbi.nlm.nih.gov/22285053/); PubMed Central PMCID: [PMC3294002](https://pubmed.ncbi.nlm.nih.gov/PMC3294002/).
9. Salazar R, Bender RA, Bruin S, Capella G, Moreno Aguado V, Roepman P, et al. Development and validation of a robust high-throughput gene expression test (ColoPrint) for risk stratification of colon cancer Patients. *Gastrointestinal Cancers Symposium*. 2010; Orlando, FL, January 22–24, 2010 (abstr 295).
10. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*. 2011; 364(26):2507–16. Epub 2011/06/07. doi: [10.1056/NEJMoa1103782](https://doi.org/10.1056/NEJMoa1103782) PMID: [21639808](https://pubmed.ncbi.nlm.nih.gov/21639808/); PubMed Central PMCID: [PMC3549296](https://pubmed.ncbi.nlm.nih.gov/PMC3549296/).
11. Glenn J, Steinberg WM, Kurtzman SH, Steinberg SM, Sindelar WF. Evaluation of the utility of a radioimmunoassay for serum CA 19–9 levels in patients before and after treatment of carcinoma of the pancreas. *J Clin Oncol*. 1988; 6(3):462–8. Epub 1988/03/01. PMID: [3162513](https://pubmed.ncbi.nlm.nih.gov/3162513/).
12. Stratford JK, Bentrem DJ, Anderson JM, Fan C, Volmar KA, Marron JS, et al. A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma. *PLoS Med*. 2010; 7(7):e1000307. Epub 2010/07/21. doi: [10.1371/journal.pmed.1000307](https://doi.org/10.1371/journal.pmed.1000307) PMID: [20644708](https://pubmed.ncbi.nlm.nih.gov/20644708/); PubMed Central PMCID: [PMC2903589](https://pubmed.ncbi.nlm.nih.gov/PMC2903589/).
13. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003; 31(4):e15. Epub 2003/02/13. PMID: [12582260](https://pubmed.ncbi.nlm.nih.gov/12582260/); PubMed Central PMCID: [PMC150247](https://pubmed.ncbi.nlm.nih.gov/PMC150247/).
14. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*. 2002; 30(4):e15. Epub 2002/02/14. PMID: [11842121](https://pubmed.ncbi.nlm.nih.gov/11842121/); PubMed Central PMCID: [PMC100354](https://pubmed.ncbi.nlm.nih.gov/PMC100354/).
15. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Roy Stat Soc B*. 2005; 67:301–20.
16. Zou H, Hastie T, Tibshirani R. Sparse principal component analysis. *J Comput Graph Stat*. 2006; 15:265–86.
17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*. 1995; 57(1):289–300. PMID: [ISI:A1995QE45300017](https://pubmed.ncbi.nlm.nih.gov/ISI:A1995QE45300017/).
18. Chen DT, Hsu YL, Fulp WJ, Coppola D, Haura EB, Yeatman TJ, et al. Prognostic and predictive value of a malignancy-risk gene signature in early-stage non-small cell lung cancer. *Journal of the National Cancer Institute*. 2011; 103(24):1859–70. Epub 2011/12/14. doi: [10.1093/jnci/djr420](https://doi.org/10.1093/jnci/djr420) PMID: [22157961](https://pubmed.ncbi.nlm.nih.gov/22157961/); PubMed Central PMCID: [PMC3243673](https://pubmed.ncbi.nlm.nih.gov/PMC3243673/).
19. Chen DT, Nasir A, Culhane A, Venkataramu C, Fulp W, Rubio R, et al. Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. *Breast cancer research and*

- treatment. 2010; 119(2):335–46. Epub 2009/03/07. doi: [10.1007/s10549-009-0344-y](https://doi.org/10.1007/s10549-009-0344-y) PMID: [19266279](https://pubmed.ncbi.nlm.nih.gov/19266279/); PubMed Central PMCID: [PMC2796276](https://pubmed.ncbi.nlm.nih.gov/PMC2796276/).
20. Marchion DC, Cottrill HM, Xiong Y, Chen N, Bicaku E, Fulp WJ, et al. BAD phosphorylation determines ovarian cancer chemosensitivity and patient survival. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2011; 17(19):6356–66. Epub 2011/08/19. doi: [10.1158/1078-0432.CCR-11-0735](https://doi.org/10.1158/1078-0432.CCR-11-0735) PMID: [21849418](https://pubmed.ncbi.nlm.nih.gov/21849418/); PubMed Central PMCID: [PMC3186862](https://pubmed.ncbi.nlm.nih.gov/PMC3186862/).
 21. Hopewell EL, Zhao W, Fulp WJ, Bronk CC, Lopez AS, Massengill M, et al. Lung tumor NF-kappaB signaling promotes T cell-mediated immune surveillance. *The Journal of clinical investigation*. 2013. Epub 2013/05/03. doi: [10.1172/JCI67250](https://doi.org/10.1172/JCI67250) PMID: [23635779](https://pubmed.ncbi.nlm.nih.gov/23635779/).
 22. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS computational biology*. 2011; 7(10):e1002240. Epub 2011/10/27. doi: [10.1371/journal.pcbi.1002240](https://doi.org/10.1371/journal.pcbi.1002240) PMID: [22028643](https://pubmed.ncbi.nlm.nih.gov/22028643/); PubMed Central PMCID: [PMC3197658](https://pubmed.ncbi.nlm.nih.gov/PMC3197658/).
 23. Neoptolemos JP, Stocken DD, Bassi C, Ghaneh P, Cunningham D, Goldstein D, et al. Adjuvant chemotherapy with fluorouracil plus folinic acid vs gemcitabine following pancreatic cancer resection: a randomized controlled trial. *Jama*. 2010; 304(10):1073–81. Epub 2010/09/09. doi: [10.1001/jama.2010.1275](https://doi.org/10.1001/jama.2010.1275) PMID: [20823433](https://pubmed.ncbi.nlm.nih.gov/20823433/).
 24. Liao WC, Chien KL, Lin YL, Wu MS, Lin JT, Wang HP, et al. Adjuvant treatments for resected pancreatic adenocarcinoma: a systematic review and network meta-analysis. *Lancet Oncol*. 2013; 14(11):1095–103. Epub 2013/09/17. doi: [10.1016/S1470-2045\(13\)70388-7](https://doi.org/10.1016/S1470-2045(13)70388-7) PMID: [24035532](https://pubmed.ncbi.nlm.nih.gov/24035532/).
 25. Van Laethem JL, Hammel P, Mornex F, Azria D, Van Tienhoven G, Vergauwe P, et al. Adjuvant gemcitabine alone versus gemcitabine-based chemoradiotherapy after curative resection for pancreatic cancer: a randomized EORTC-40013-22012/FFCD-9203/GERCOR phase II study. *J Clin Oncol*. 2010; 28(29):4450–6. Epub 2010/09/15. doi: [10.1200/JCO.2010.30.3446](https://doi.org/10.1200/JCO.2010.30.3446) PMID: [20837948](https://pubmed.ncbi.nlm.nih.gov/20837948/); PubMed Central PMCID: [PMC2988636](https://pubmed.ncbi.nlm.nih.gov/PMC2988636/).
 26. Neoptolemos JP, Stocken DD, Friess H, Bassi C, Dunn JA, Hickey H, et al. A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. *N Engl J Med*. 2004; 350(12):1200–10. Epub 2004/03/19. doi: [10.1056/NEJMoa032295](https://doi.org/10.1056/NEJMoa032295) PMID: [15028824](https://pubmed.ncbi.nlm.nih.gov/15028824/).
 27. Oettle H, Post S, Neuhaus P, Gellert K, Langrehr J, Ridwelski K, et al. Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer: a randomized controlled trial. *Jama*. 2007; 297(3):267–77. Epub 2007/01/18. doi: [10.1001/jama.297.3.267](https://doi.org/10.1001/jama.297.3.267) PMID: [17227978](https://pubmed.ncbi.nlm.nih.gov/17227978/).
 28. Smeenk HG, van Eijck CH, Hop WC, Erdmann J, Tran KC, Debois M, et al. Long-term survival and metastatic pattern of pancreatic and periampullary cancer after adjuvant chemoradiation or observation: long-term results of EORTC trial 40891. *Annals of surgery*. 2007; 246(5):734–40. Epub 2007/10/31. doi: [10.1097/SLA.0b013e318156eef3](https://doi.org/10.1097/SLA.0b013e318156eef3) PMID: [17968163](https://pubmed.ncbi.nlm.nih.gov/17968163/).
 29. Nischalke HD, Schmitz V, Luda C, Aldenhoff K, Berger C, Feldmann G, et al. Detection of IGF2BP3, HOXB7, and NEK2 mRNA expression in brush cytology specimens as a new diagnostic tool in patients with biliary strictures. *PLoS One*. 2012; 7(8):e42141. Epub 2012/08/11. doi: [10.1371/journal.pone.0042141](https://doi.org/10.1371/journal.pone.0042141) PMID: [22879911](https://pubmed.ncbi.nlm.nih.gov/22879911/); PubMed Central PMCID: [PMC3413695](https://pubmed.ncbi.nlm.nih.gov/PMC3413695/).
 30. Schaeffer DF, Owen DR, Lim HJ, Buczkowski AK, Chung SW, Scudamore CH, et al. Insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3) overexpression in pancreatic ductal adenocarcinoma correlates with poor survival. *BMC Cancer*. 2010; 10:59. Epub 2010/02/25. doi: [10.1186/1471-2407-10-59](https://doi.org/10.1186/1471-2407-10-59) PMID: [20178612](https://pubmed.ncbi.nlm.nih.gov/20178612/); PubMed Central PMCID: [PMC2837867](https://pubmed.ncbi.nlm.nih.gov/PMC2837867/).
 31. Abiatari I, DeOliveira T, Kerkadze V, Schwager C, Esposito I, Giese NA, et al. Consensus transcriptome signature of perineural invasion in pancreatic carcinoma. *Mol Cancer Ther*. 2009; 8(6):1494–504. Epub 2009/06/11. doi: [10.1158/1535-7163.MCT-08-0755](https://doi.org/10.1158/1535-7163.MCT-08-0755) PMID: [19509238](https://pubmed.ncbi.nlm.nih.gov/19509238/).
 32. Iacobuzio-Donahue CA, Ashfaq R, Maitra A, Adsay NV, Shen-Ong GL, Berg K, et al. Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from three major technologies. *Cancer Res*. 2003; 63(24):8614–22. Epub 2003/12/26. PMID: [14695172](https://pubmed.ncbi.nlm.nih.gov/14695172/).
 33. Wallrapp C, Hahnel S, Muller-Pillasch F, Burghardt B, Iwamura T, Ruthenburger M, et al. A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer. *Cancer Res*. 2000; 60(10):2602–6. Epub 2000/05/29. PMID: [10825129](https://pubmed.ncbi.nlm.nih.gov/10825129/).
 34. Cao D, Zhang Q, Wu LS, Salaria SN, Winter JW, Hruban RH, et al. Prognostic significance of maspin in pancreatic ductal adenocarcinoma: tissue microarray analysis of 223 surgically resected cases. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*. 2007; 20(5):570–8. Epub 2007/03/31. doi: [10.1038/modpathol.3800772](https://doi.org/10.1038/modpathol.3800772) PMID: [17396143](https://pubmed.ncbi.nlm.nih.gov/17396143/).
 35. Ino Y, Yamazaki-Itoh R, Oguro S, Shimada K, Kosuge T, Zavada J, et al. Arginase II expressed in cancer-associated fibroblasts indicates tissue hypoxia and predicts poor outcome in patients with

- pancreatic cancer. *PLoS One*. 2013; 8(2):e55146. Epub 2013/02/21. doi: [10.1371/journal.pone.0055146](https://doi.org/10.1371/journal.pone.0055146) PMID: [23424623](https://pubmed.ncbi.nlm.nih.gov/23424623/); PubMed Central PMCID: PMC3570471.
36. Mardin WA, Petrov KO, Enns A, Senninger N, Haier J, Mees ST. SERPINB5 and AKAP12—expression and promoter methylation of metastasis suppressor genes in pancreatic ductal adenocarcinoma. *BMC Cancer*. 2010; 10:549. Epub 2010/10/14. doi: [10.1186/1471-2407-10-549](https://doi.org/10.1186/1471-2407-10-549) PMID: [20939879](https://pubmed.ncbi.nlm.nih.gov/20939879/); PubMed Central PMCID: PMC2966466.
 37. Basturk O, Singh R, Kaygusuz E, Balci S, Dursun N, Culhaci N, et al. GLUT-1 expression in pancreatic neoplasia: implications in pathogenesis, diagnosis, and prognosis. *Pancreas*. 2011; 40(2):187–92. Epub 2011/01/06. doi: [10.1097/MPA.0b013e318201c935](https://doi.org/10.1097/MPA.0b013e318201c935) PMID: [21206329](https://pubmed.ncbi.nlm.nih.gov/21206329/); PubMed Central PMCID: PMC3164314.
 38. Pizzi S, Porzionato A, Pasquali C, Guidolin D, Sperti C, Fogar P, et al. Glucose transporter-1 expression and prognostic significance in pancreatic carcinogenesis. *Histology and histopathology*. 2009; 24(2):175–85. Epub 2008/12/17. PMID: [19085834](https://pubmed.ncbi.nlm.nih.gov/19085834/).
 39. Matsubara J, Honda K, Ono M, Tanaka Y, Kobayashi M, Jung G, et al. Reduced plasma level of CXC chemokine ligand 7 in patients with pancreatic cancer. *Cancer Epidemiol Biomarkers Prev*. 2011; 20(1):160–71. Epub 2010/12/15. doi: [10.1158/1055-9965.EPI-10-0397](https://doi.org/10.1158/1055-9965.EPI-10-0397) PMID: [21148121](https://pubmed.ncbi.nlm.nih.gov/21148121/).
 40. Alldinger I, Dittert D, Peiper M, Fusco A, Chiappetta G, Staub E, et al. Gene expression analysis of pancreatic cell lines reveals genes overexpressed in pancreatic cancer. *Pancreatol: official journal of the International Association of Pancreatology*. 2005; 5(4–5):370–9. Epub 2005/06/29. doi: [10.1159/000086537](https://doi.org/10.1159/000086537) PMID: [15983444](https://pubmed.ncbi.nlm.nih.gov/15983444/).
 41. Morse DL, Balagurunathan Y, Hostetter G, Trissal M, Tafreshi NK, Burke N, et al. Identification of novel pancreatic adenocarcinoma cell-surface targets by gene expression profiling and tissue microarray. *Biochemical pharmacology*. 2010; 80(5):748–54. Epub 2010/06/01. doi: [10.1016/j.bcp.2010.05.018](https://doi.org/10.1016/j.bcp.2010.05.018) PMID: [20510208](https://pubmed.ncbi.nlm.nih.gov/20510208/); PubMed Central PMCID: PMC2914681.
 42. Chakraborty S, Baine MJ, Sasson AR, Batra SK. Current status of molecular markers for early detection of sporadic pancreatic cancer. *Biochimica et biophysica acta*. 2011; 1815(1):44–64. Epub 2010/10/05. doi: [10.1016/j.bbcan.2010.09.002](https://doi.org/10.1016/j.bbcan.2010.09.002) PMID: [20888394](https://pubmed.ncbi.nlm.nih.gov/20888394/); PubMed Central PMCID: PMC3014374.
 43. Zhang G, Schetter A, He P, Funamizu N, Gaedcke J, Ghadimi BM, et al. DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS One*. 2012; 7(2):e31507. Epub 2012/03/01. doi: [10.1371/journal.pone.0031507](https://doi.org/10.1371/journal.pone.0031507) PMID: [22363658](https://pubmed.ncbi.nlm.nih.gov/22363658/); PubMed Central PMCID: PMC3282755.
 44. Sergeant G, van Eijnsden R, Roskams T, Van Duppen V, Topal B. Pancreatic cancer circulating tumour cells express a cell motility gene signature that predicts survival after surgery. *BMC Cancer*. 2012; 12:527. Epub 2012/11/20. doi: [10.1186/1471-2407-12-527](https://doi.org/10.1186/1471-2407-12-527) PMID: [23157946](https://pubmed.ncbi.nlm.nih.gov/23157946/); PubMed Central PMCID: PMC3599097.
 45. Van den Broeck A, Vankelecom H, Van Delm W, Gremeaux L, Wouters J, Allemeersch J, et al. Human pancreatic cancer contains a side population expressing cancer stem cell-associated and prognostic genes. *PLoS One*. 2013; 8(9):e73968. Epub 2013/09/27. doi: [10.1371/journal.pone.0073968](https://doi.org/10.1371/journal.pone.0073968) PMID: [24069258](https://pubmed.ncbi.nlm.nih.gov/24069258/); PubMed Central PMCID: PMC3775803.
 46. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med*. 2011; 17(4):500–3. Epub 2011/04/05. doi: [10.1038/nm.2344](https://doi.org/10.1038/nm.2344) PMID: [21460848](https://pubmed.ncbi.nlm.nih.gov/21460848/); PubMed Central PMCID: PMC3755490.
 47. Yeatman TJ, Mule J, Dalton WS, Sullivan D. On the eve of personalized medicine in oncology. *Cancer Res*. 2008; 68(18):7250–2. Epub 2008/09/17. 68/18/7250 [pii] doi: [10.1158/0008-5472.CAN-08-1374](https://doi.org/10.1158/0008-5472.CAN-08-1374) PMID: [18794109](https://pubmed.ncbi.nlm.nih.gov/18794109/); PubMed Central PMCID: PMC2650840.
 48. Koomen JM, Haura EB, Bepler G, Sutphen R, Remily-Wood ER, Benson K, et al. Proteomic contributions to personalized cancer care. *Mol Cell Proteomics*. 2008; 7(10):1780–94. Epub 2008/07/31. R800002-MCP200 [pii] doi: [10.1074/mcp.R800002-MCP200](https://doi.org/10.1074/mcp.R800002-MCP200) PMID: [18664563](https://pubmed.ncbi.nlm.nih.gov/18664563/); PubMed Central PMCID: PMC2559938.
 49. Reis PP, Waldron L, Goswami RS, Xu W, Xuan Y, Perez-Ordóñez B, et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC biotechnology*. 2011; 11:46. doi: [10.1186/1472-6750-11-46](https://doi.org/10.1186/1472-6750-11-46) PMID: [21549012](https://pubmed.ncbi.nlm.nih.gov/21549012/); PubMed Central PMCID: PMC3103428.
 50. Tang H, Xiao G, Behrens C, Schiller J, Allen J, Chow CW, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res*. 2013; 19(6):1577–86. Epub 2013/01/30. doi: [10.1158/1078-0432.CCR-12-2321](https://doi.org/10.1158/1078-0432.CCR-12-2321) PMID: [23357979](https://pubmed.ncbi.nlm.nih.gov/23357979/); PubMed Central PMCID: PMC3619002.