Integrative Survival-Based Molecular Profiling of Human Pancreatic Cancer

Timothy R. Donahue1,2,3,4,†, Linh M. Tran2,3,*, Reginald Hill2,3, Yunfeng Li2,3, Anne Kovochich5, Joseph H. Calvopina6, Sanjeet G. Patel4, Nanping Wu2,3, Antreas Hindoyan2,3, James J. Farrell6, Xinmin Li5, David W. Dawson4,5†, Hong Wu2,3,4,7†

* These authors contributed equally.
† Corresponding authors

1Departments of Surgery, Division of General Surgery, 2Institute for Molecular Medicine, 3Molecular and Medical Pharmacology, 4Jonsson Comprehensive Cancer Center, 5Pathology and Laboratory Medicine, 6Medicine, Division of Digestive Diseases, 7Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA (UCLA)

Running Title: Integrative Profile of Human Pancreatic Cancer

Keywords: pancreatic cancer, array analysis, micro RNAs, PI3K/AKT signaling, SRC signaling

Corresponding author:
Timothy R. Donahue, MD
Department of Surgery
72-256 Center for the Health Sciences
10833 LeConte Avenue
Los Angeles, CA 90095-6904
Phone: 310-206-7440
Fax: 310-206-2472

Conflicts of Interest: None

Other notes:
Word count:
Abstract: 243, Text: 5017, Figure legends: 359
Total number of figures and tables:
Figures: 5, Tables: 2.
Statement of Translational Relevance

Despite evidence of the heterogeneous genomic and molecular changes associated with pancreatic ductal adenocarcinoma (PDAC) tumorigenesis, little is known about biologic subsets of tumors to guide patient stratification and individualized therapies. Recent Phase III trials with molecular agents have neither resulted in robust survival benefits nor attempted to stratify patients prior to randomization. We utilize a novel integrative survival-based genomic and molecular array analysis of human PDACs to derive a composite score that ranks genes based on expression and regulatory mechanisms. We find that micro RNA regulation plays a critical role in the malignant phenotype of PDAC. This approach shows that dysregulated PI3K/AKT or SRC signaling is significantly associated with distinct patient subgroups with more aggressive disease. These pathway-specific genes and miRNAs represent potentially useful clinical biomarkers and targets of individualized therapy for well-defined patient subgroups.
ABSTRACT

Purpose: To perform an integrative profile of human pancreatic cancer (PDAC) to identify prognosis-significant genes and their related pathways.

Experimental Design: A concordant survival-based whole genome in silico array analysis of DNA copy number, and mRNA & micro RNA (miRNA) expression in 25 early stage PDAC was performed. A novel composite score simultaneously integrated gene expression with regulatory mechanisms to identify the signature genes with the most levels of prognosis-significant evidence. The predominant signaling pathways were determined via a pathway-based approach. Independent patient cohorts (n=148 and 42) were then used as in vitro validation of the array findings.

Results: The composite score identified 171 genes whose expressions were able to define two prognosis subgroups (p=3.8e-5). Eighty-eight percent (151/171) of the genes were regulated by prognosis-significant miRNAs. The PI3K/AKT pathway and SRC signaling were densely populated by prognosis significant genes and driven by genomic amplification of SRC and miRNA regulation of p85α and CBL. On TMA validation (n=148), p85α protein expression was associated with improved survival for all patients (p=0.02), and activated P-SRC (Y418) was associated shorter survival for patients with low grade histology tumors (p=0.04). Interacting P-SRC and p85α revealed that they define two distinct PDAC patient subgroups (p=0.0066). Furthering the importance of these pathways, CBL protein expression was associated with improved survival (p=0.03) on a separate cohort (n=42).
Conclusions: These pathways and related genes may represent putative clinical biomarkers and possible targets of individualized therapy in the distinct patient subgroups they define.
INTRODUCTION

Pancreatic cancer (PDAC) is the fourth leading cause of cancer-related deaths in the United States (1) and has an extremely poor overall five year survival rate of only 4 percent. Most patients present with advanced stage disease and have a median survival of less than one year (2). Cytotoxic chemotherapy is marginally effective with standard gemcitabine or 5-FU based regimens increasing PDAC median survival by less than 2 months in advanced disease(3-5). Published phase III clinical trials of targeted molecular agents in unselected PDAC populations have also not shown robust survival benefits (6-10). Ultimately, our evolving understanding of significant genomic diversity in PDAC must be utilized to better inform targeted drug design and delivery.

Recent in-depth exome sequencing showed individual PDAC tumors average more than 60 distinct alterations, the majority of which occur at low frequencies across all tumors. Only a few high prevalence genomic changes were detected, including expected mutations in KRAS and loss or inactivation of known tumor suppressor genes (e.g. TP53, SMAD4). Despite this genomic heterogeneity, all tumors had genetic alterations that were linked to 12 core signaling pathways (11). Follow-up work comparing patient-matched primary PDAC tumors and subsequent metastases revealed acquisition of further mutations that varied by metastatic site. Strikingly, founder mutations of each metastatic subclone could be traced back to sequenced geographic sub-regions of the primary tumor (12, 13), providing new insights into the genetic events and timing of PDAC initiation and malignant phenotype. Notably, SMAD4 deletion has been the only genetic alteration from this work that has been linked to patient survival (14).
Others have used gene expression microarray analyses to define molecular signatures associated with PDAC disease progression. Stratford et al. (15) identified a 6 gene signature in primary tumors that was associated with metastatic disease and predicted shorter survival in an independent set of 67 patients. Collisson et al. (16) analyzed primary PDAC from cell lines and a combination of clinical datasets to classify three distinct PDAC molecular subtypes that were able to predict clinical survival, as well as response to therapy in experimental models. While such molecular profiling has provided valuable information, the remarkable genomic diversity of PDAC and the small size of most patient cohorts has clearly hindered the discovery of additionally biologically important molecular changes.

As a means to effectively study diverse genomic alterations in a small patient dataset, we hypothesized that the identification and refinement of prognosis-related genes in PDAC should be improved by increasing the depth of analysis for each tumor using multiple array platforms. The potential for this type of multi-dimensional analysis was shown in a recent prostate cancer study where several pathways of known prognostic significance were validated and new ones were additionally implicated (17). For our own survival-based analysis of PDAC, individual gene expression changes associated with survival were matched to potential genomic or epigenetic modes of regulation by integrating microarray results of mRNA expression with DNA copy number variation and microRNA levels. This approach validated pathways implicated in pancreatic tumorigenesis and uncovered previously unrecognized molecular events associated
with poor prognosis. The expressions of many identified genes were found to have associated miRNA alterations linked to survival. These genes and their regulatory mechanisms represent promising candidates for future studies addressing their function and evaluating their efficacy as predictive biomarkers and/or targets for molecular-based therapies.
MATERIALS AND METHODS

Patients and samples

All work was performed with UCLA Institutional Review Board approval. Three independent, non-overlapping patient cohorts were used in this study. The initial test cohort of 42 PDAC tumors and 7 non-malignant pancreas samples snap-frozen at the time of surgery were used for microarrays. Of these, only samples with tumor cell content >30% were chosen for final multi-platform analysis (n=25) as determined on representative H&E sections by a practicing gastrointestinal pathologist (DWD). The second patient cohort (n=42) were tumors isolated from FFPE tissue blocks and used as a validation cohort for qPCR. The third patient dataset (n=148) were tumors on a tissue microarray used as an IHC validation cohort. All clinicopathologic and survival information for each patient cohort were extracted from a prospectively maintained UCLA surgical database of pancreatic patients. Disease recurrence was assessed based on biopsy, radiographic evidence or death. The electronic medical record was used to determine associated clinical and pathologic features, as well as disease-free and disease-specific survival. Search of the social security death index (SSDI) was used to determine overall survival. Survival analysis of the TMA cohort was limited to overall survival. Disease-free, disease-specific and overall survival times were examined for the microarray and qPCR validation cohorts. Survival intervals were calculated from date of surgery to date of confirmed death or last patient contact.

Gene expression analysis
The gene expression was investigated by Affymetrix HGU133 Plus 2 Array on which multiple probe sets might be used to measure expression of a single gene. Therefore, we analyzed the data based on probe set IDs, and used the highest absolute value/score among multiple probe sets for the gene-based interpretation. Details of the array procedure, and normalization and filtering are detailed in the Supplemental Information (SI). In brief, GCRMA was used for normalization, and probe sets having presence calls in less than 30% of samples were eliminated before further analyses. Cox scores(18) were used to determine the correlation between individual probe set-based expression and disease free survival (DFS) time. Probe sets were then sorted based on the absolute values of their Cox scores. Prediction analysis for microarrays (19), implemented by Bioconductor pamr package, was then used to determine 1) if expression of the top ranked probe sets could be used to predict the sample outcome and 2) the minimal probe sets to define two prognosis groups with the highest statistical significance.

**MicroRNA expression analysis**

The Exiqon miRNA arrays (miRCURY LNA™ microRNA Array v.11.0 -hsa, mmu & rno) were used for measuring genome miRNA expression. Array intensities were adjusted and normalized by variance stabilizing transformation implemented in the Bioconductor vsn packages. The expression profile of each miRNA was represented by the average expression of its multiple spots on the array. Statistical analyses for the association between miRNAs and survival was performed as detailed for gene expression analysis.
DNA copy number analysis

Affymetrix SNP 6.0 arrays were used to detect copy number aberrations (CNAs) in tumor samples. CEL files produced by Affymetrix GeneChip Command Console (AGCC) software were imported into Affymetrix GTC 3.0.1, and analyzed using the Copy Number Analysis workflow using HapMap270 as the reference model. Regions with CNAs were then annotated with gene symbols base on the annotation file from the UCSC genome browser (build hg18). Only those genes whose loci had CNA present in at least 20% of the tumors were included for survival analysis in which Cox proportional hazards model was used to determine if the group with CNA at a given locus was at higher risk than the group without CNA.

Integrating multiple dimension data to identify signature genes with multiple levels of evidence

In this meta-analysis, a composite score was generated to quantitatively measure gene prognosis significance based on the multiple array platforms, and is described in detail in the SI. In brief, the composite score was developed based on the assumptions that: besides being correlated to survival, expression of the gene of interest is (1) correlated to its copy number, (2) anti-correlated to its regulating miRNA, and (3) such changes in copy number and miRNA expression are also associated with prognosis.

Mathematically, the operators of the survival-based composite score for each gene included: 1) rank of gene expression Cox score; 2) binary CNA score (0 or 1) requiring both a hazard ratio p-value <0.2 and concordance with gene expression change; 3) rank of miRNA Cox score with required anti-concordance between miRNA and
corresponding gene expression. The miRNA operator was further strengthened if a prognosis-significant CNA overlapped with the miRNA coding region.

**Pathway and Gene Ontology Analysis**

Survival signature genes were annotated by databases in the public domain: the KEGG and Molecular Signatures Databases (20). Significant enrichment of pathway/gene set was determined by Fisher exact test.

**Validation with Tissue Microarray**

The PDAC TMA has been previously detailed (21) and represents a totally separate, non-overlapping cohort of patients. Immunohistochemistry was visualized using Vectastain ABC Elite Kit (Vector Laboratories) following heat-induced antigen-retrieval (0.01M citric acid buffer, pH 6.0) and overnight incubation with primary antibodies (1:100 dilution), including P-SRC (Y418) (Abcam, ab47411) or p85α (Epitomics, Cat: 1675-1). Three separate 1.0 mm cores for each tumor in the TMA were independently scored by two blinded observers using semi-quantitative histoscores (range 0-300). Histoscores were the product of staining intensity (0-3) and percentage of tumor cells staining at that intensity (0-100). If any core’s histoscore differed by more than 30 points between observers, a revised score was assigned by consensus evaluation. Median histoscore of both observers was used for analysis.

**Quantitative Real Time PCR of Pancreatic Cancer Resection Samples**
For qPCR validation, H&E slides from a separate non-overlapping cohort of 42 additional PDAC patients resected at UCLA between 2002 and 2009 were reviewed by a practicing gastrointestinal pathologist (D.W.D.) to target the extraction of three 3 mm cores from areas of viable tumor in the corresponding FFPE blocks. RNA was isolated using the RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (Ambion). Total RNA was reverse-transcribed using random hexamer primers (High Capacity cDNA Reverse Transcription Kit, Ambion). SYBR green real-time quantitative PCR assays were performed on generated cDNA using a Roche (Basel, Switzerland) LightCycler 480 Real-Time PCR System. Reaction parameters and primers sequences are available upon request and were optimized for FFPE-derived cDNA based on the use of short (<90 bp), intron-spanning amplicons. Relative gene expression was normalized to ACTB as housekeeping gene.

**Statistical Analysis**

Survival estimate for each subgroup was generated using the Kaplan-Meier method. Log-rank test was used to compare Kaplan-Meier curves. Multivariate Cox proportional hazards models were used to test statistical independence and significance of multiple predictors.
RESULTS

Survival based integrated genomic and gene expression analyses

The overall strategy of our survival-based study is outlined in Supplemental Figure 1. Integrative multi-platform array analysis was used to simultaneously examine gene expression and regulatory mechanisms from an initial cohort of 25 patients in order to identify and refine genes and pathways with biologic significance. Array analysis was conducted on snap frozen primary patient tissue samples in order to directly link gene expression to genomic and epigenomic changes in the in situ context of primary tumor, and to extract a sufficient source of high quality nucleic acid for simultaneous mRNA, miRNA and SNP microarray analysis. As our goal was to identify genomic alterations and expression changes in neoplastic ductal epithelial cells, we limited our analysis to primary tumor samples where estimated tumor cell content exceeded 30% (median 60%, ranging from 35% to 90%). This cutoff was chosen after initial unsupervised clustering of microarrays comparing mRNA expression from normal pancreas, chronic pancreatitis and pancreas tumor samples showed tumors with low tumor cell content (<30%) more frequently clustered with normal and chronic pancreatitis controls (data not shown), consistent with previous work showing a large stromal content can confound array readouts in PDAC (22). Specific molecular alterations identified through this integrated genomic analysis were then independently validated in separate, non-overlapping patient cohorts by immunohistochemistry (n=148) or qPCR (n=42), whose clinicopathologic characteristics and survival outcomes largely overlap (Supplemental Table 1 and Supplemental Figure 2).
Table 1 provides the clinical and histopathologic parameters for the 25 patients. All patients had early stage PDAC and received adjuvant chemotherapy after surgery. At the time of analysis, 16 patients had recurrent disease with a median disease-free survival time (DFS) of 13.3 months, and ten of them had died of disease with a median disease-specific survival (DSS) of 20.6 months. Given the number of deaths and relatively short follow-up time (median 12.4 months for survivors) in our cohort, DFS was chosen as the outcome measure for our analysis. DFS is considered an accurate metric of survival outcome, as most patients with recurrent PDAC will succumb to disease. The clinicopathologic characteristics and survival outcomes of our cohort was similar to other published large cohorts of early stage PDAC(23).

**Copy number variations and mRNA and miRNA expression alterations as predictors of PDAC survival**

We first individually analyzed mRNA expression, miRNA expression and DNA copy number to determine DFS-associated changes in each microarray platform. For mRNA analysis, we used a non-grouping-based approach addressing survival as a continuous variable. The advantage of this approach over a grouping-based approach is that it does not set an arbitrary survival time threshold for dichotomization and thus better accounts for tumors with a continuum of clinical behavior (18). A semi-supervised clustering methodology was used whereby a Cox score was first generated for each gene based on a direct correlation between its expression and DFS. Signature genes were then prioritized and selected based on the absolute value of their Cox scores. Using this method, the most accurate gene set is selected from a series of prediction
analyses whereby different cut-off points are iterated to establish an optimal panel of genes accurately separating different prognosis subgroups. The top 500 scored probe sets (Supplemental Table 2) from this approach included 186 up- and 314 down-regulated transcripts able to segregate two highly significant and distinct prognosis groups (median DFS = 7.7 vs. 25.3 months; log-rank test p=0.000038; Figure 1A). The non-grouping approach was also utilized for separate miRNA microarray analysis, which yielded a panel of 31 miRNAs (1.8% of the total) able to robustly segregate two prognosis groups (median DFS = 9 vs. >36 months; p=0.00047; Figure 1B and Supplemental Table 3).

For survival-based analysis of DNA copy number aberrations (CNA), CNAs for each tumor were determined by comparing Affymetrix SNP microarray to the human hapmap reference model using Affymetric GTC software. CNAs were then mapped to specific gene loci using UCSC genome build hg18. Because of the size of our cohort, we focused on CNAs occurring at higher frequency (≥ 20% of the samples) and relaxed the cut-off to include Cox values approaching significance (p-value < 0.2). High frequency CNAs were clustered on chromosomes 1, 7, 8, and 20 for amplifications and chromosomes 6, 9, 17, and 18 for deletions (Figure 1C, upper panel). However, among them, the prognosis significant CNAs (Figure 1C, bottom panel) were located on specific loci, including amplifications on chromosomes 7, 11, 19, 20, and 22, and deletions on chromosomes 6 and 9. We identified CNAs associated with a total of 68 genes that optimally segregated patients into two statistically significant prognosis
groups (median DFS = 9.7 vs. 25.0 months, log rank test p=0.0063, Figure 1D and Supplemental Table 4).

**Integrated Molecular Analyses Further Refines Genes of Prognostic Importance in PDAC**

Although our initial analysis demonstrated a large number and variety of molecular alterations correlated with DFS in PDAC, it was less clear which alterations mechanistically link to the malignant phenotype or represent clinically useful biomarkers. While analysis of larger patient datasets can help address these questions, the cost and availability of large PDAC patient datasets are limiting. As an alternative, we hypothesized that relevant associations could be strengthened and refined in our cohort through an integrated assessment of gene expression in conjunction with concordant changes in genetic and epigenetic regulation.

We merged results from all three microarray platforms to generate a single composite score that integrates CNA and levels of mRNA and miRNA expression (see Supplemental Figure 1 and Supplemental Information). To associate genes with miRNAs, a list of 2.1 million potential miRNA-mRNA seed-match pairs were first generated from data in the public domain, including experimental evidence in TarBase® (24) and predictive sequence analysis using TargetScanS® (25) and miRbase® (26). Next, each component in the composite score was given a weight based on its ability to independently identify two prognosis groups. Individually, mRNA expression was the most robust predictor of DFS in our analysis, followed by miRNA expression and CNA.
changes, as determined by the log rank p-values generated from each platform (Figure 1). The log_{10} transformations of these p-values were used to weight each platform’s contribution to survival in the final composite score calculation for each gene, whereby a gene’s mRNA expression was weighted most heavily, followed by its linked miRNA alterations and finally CNA (see Supplemental Methods). The composite score also took into account whether there was concordance between CNA-gene expression (e.g. amplification associated with higher gene expression) or anti-concordance between miRNA-gene expression (e.g. as miRNAs predominantly will repress target gene expression). If these conditions were not met, the component was not included. Based on these criteria, our integrated composite score yielded a refined list of 171 signature genes (represented by 200 probe sets) that accurately segregated patients into two statistically significant prognostic groups (median DFS = 8.6 vs. 20.6, p=0.001, Figure 2A-B and Supplemental Table 5). From this list, 134 genes had corresponding changes in miRNA expression, 20 genes had corresponding changes in CNA and 17 genes shared corresponding changes in both miRNA expression and CNA (Figure 2C). Importantly, these results highlight a strong link between miRNAs regulated mechanism and survival outcome in PDAC.

Pathway Based Analysis of Signature Genes Links PDAC Survival to PI3K and SRC Signaling Pathways

We next performed survival-based pathway and gene ontology analyses of our integrated composite gene signature. Pathways most highly populated with survival correlated genes included ERBB signaling, apoptosis, purine metabolism, focal
adhesion and insulin signaling (Supplemental Table 6). Therefore, our data offers survival-based correlations to support an existing literature that links these pathways to pancreatic tumorigenesis (27-31). Our data also offers potential insights into novel mechanisms regulating these pathways and their relationship to clinical disease progression. Several survival-correlated genes in our integrated composite signature uniformly link poor prognosis to changes that will result in upregulation of PI3K/AKT/mTOR and SRC signaling (Figure 3). For instance, the expression of EGFR, a potential activator of both AKT and SRC signaling, was associated with worse prognosis in our analysis (Cox score +2.77), while that of CBL, a ubiquitin ligase able to negatively impact EGFR (32) or SRC(33) expression, correlated with improved survival in our analysis (Cox score -2.4). PIK3R1, which encodes the class Ia PI3K regulatory subunit p85α that antagonizes PI3K/AKT signaling, also correlated with improved survival (Cox score -2.5). Prognosis-linked changes in gene expression were more frequently correlated with prognosis-significant miRNA changes, implicating miRNAs as a critical factor in the malignant phenotype of PDAC (see Supplemental Table 3). For example, our data suggested prognosis-linked expression PIK3R1/p85α mRNA could be mediated by miRNA 519d, which is 1) broadly conserved among vertebrates, 2) predicted to bind to the 3’ untranslated region of PI3KR1, 3) inversely correlated with PIK3R1 expression in our analysis, and 4) an independent predictor shorter DFS (miRNA expression Cox score +2.42).

Highlighting the value of the integrated composite score over gene expression alone, SRC was not identified on independent mRNA array analysis, as it showed only a weak
non-significant trend towards worse prognosis (Cox expression score +1.22). However, SRC was subsequently captured as a high ranking gene (#21) in our integrated composite signature when its prognosis-significant (Cox Hazard Ratio = 4.1, p-value = 0.016) genomic amplification was considered with mRNA expression. Prognosis-related dysregulation of SRC was also inferred from our subsequent pathway analysis. SRC activation can occur through integrin-FAK-dependent (34, 35) or FAK-independent (36) mechanisms facilitated by PTPRA, which was associated with worse prognosis in our integrated composite signature (Cox score +2.24). ARFGAP1, which also correlated with worse prognosis in our analysis (Cox Score +1.62), has been shown to potentate SRC downstream signaling via its regulation of the actin cytoskeleton and, leading to enhanced cell motility (37). Interestingly, SRC, PTPRA, and ARFGAP1 are all located on chromosome 20. While SRC is regulated by amplification, both PTPRA and ARFGAP1 can be regulated by amplification and by miR541, which was inversely correlated with DFS (Cox score -3.12), again revealing the strength of our integrated approach.

Validation of Molecular Signatures

To validate the predictive value of genes in our integrated composite signature, we chose to focus on SRC and the PI3K/AKT signaling given that 1) they were populated with centrally located prognosis significant genes, 2) they are known to be dysregulated in human PDAC (35, 38), and 3) clinically available small molecule inhibitors targeting each have shown promise in pre-clinical models and phase I and II clinical trials (39, 40). For validation, we performed immunohistochemistry (IHC) to determine levels of
p85α and activated SRC phosphorylated at tyrosine position (Y419) using a tissue microarray (TMA) consisting of a separate large cohort (n=148) of treatment-naïve, resected Stage I and II PDACs. This TMA is previously detailed (21) and is similar to other large patient cohorts of early stage PDAC where pathologic stage, lymph node status and histologic tumor grade were each significantly correlated with survival. The overall survival of patients in this cohort is similar to those in the in silico analysis (median 24.2 vs. 19.8 months respectively), given the difference in follow-up times (Table 1 and Supplemental Table 1). IHC staining for each tumor was determined semi-quantitatively by histoscore, with a broad staining distribution seen for each antibody across the TMA (Figure 4B&C). For each IHC marker, patients were dichotomized on a histoscore cutoff value of 150, chosen to identify groups sufficiently populated for statistical power analysis and that minimized the chance of assigning patients with small differences to opposite groups. Neither P-SRC nor p85α expression was associated with various clinicopathologic factors, with the exception that low (includes well and moderately differentiated) histologic grade tumors have high p85α protein expression ($p = 0.00062, \chi^2$ test).

By Kaplan-Meier survival analysis, high p85α protein expression significantly correlated with better survival in the TMA (median survival 28.7 vs. 19.4 months, Log Rank test $p = 0.02$, Figure 5A), while high P-SRC expression did not significantly correlate with prognosis (median survival 21.1 vs. 25.6 months, $p=0.45$). We next examined whether P-SRC expression significantly correlated with survival in subgroups of patients stratified first on the strong independent prognostic factors of node status (pN0 or pN1...
or tumor grade (low or high). High P-SRC staining significantly associated with worse survival in the subgroup of patients with low grade tumor histology (median survival 23.2 vs. 42.5, p = 0.04, Figure 5B), but not in the subgroup with high grade tumor histology or either subgroup based on node status (data not shown).

We next addressed a potential interaction between p85α and P-SRC status by examining various combinations of dichotomized groups. Tumors with combined low P-SRC and high p85α had significantly better survival relative to any of the other paired combinations when considered individually (Figure 5C) or in aggregate (median survival = 36.6 vs. 20.4 months, p = 0.0066, Figure 5D). Multivariate Cox proportional hazards analyses, controlling for LN and grade, showed that high p85α was a near significant independent predictor of improved overall survival (HR 0.690, p = 0.068), while the combination of low P-SRC and high p85α was a significant independent predictor of improved survival in the TMA (HR = 0.53, p = 0.02) (Table 2). Of note, further IHC staining of a whole tissue sections of recently resected PDACs found that either p85α or PTEN loss appeared to correlate with enhanced PI3K/AKT signaling, as detected by P-AKT and P-S6 (Supplemental Figure 3). Although in need of prospective validation and further mechanistic evaluation, these results suggest tumors with combined lower signaling activity for both SRC (as detected by the surrogate of reduced P-SRC) and PI3K-AKT (as detected by the surrogate of increased p85α) may define a subset of patients with more favorable clinical outcome. These or other surrogate markers of pathway activation may be especially useful in discriminating patients with more or less aggressive clinical disease and assessing the use of drugs targeting these pathways.
We finally sought to validate the finding that CBL was linked to prognosis in our integrated composite signature (Figure 3). In the absence of a reliable IHC assay, we pursued survival-based quantitative PCR analysis of CBL normalized to ACTB in a separate cohort of 42 PDAC samples; these patients had a similar DSS with the array and TMA cohorts (25.4 vs. 19.8 vs. 24.2 months respectively), given the differences in follow-up times (Table 1 and Supplemental Table 1). Patients were again dichotomized into groups with low versus high CBL expression, the latter of which significantly correlated with better survival (median survival 44.3 vs. 20.6, p = 0.03, Figure 5E).

DISCUSSION

The large number and wide range of genetic alterations that characterize pancreatic cancer present both significant challenges and opportunities for improving our understanding and treatment of this highly aggressive and lethal malignancy. Seminal large scale genomic sequencing studies of PDAC offer a tantalizing spatial and temporal picture of the genomic alterations occurring in both primary tumors and metastatic lesions, but must also now be examined in greater detail to establish their association with the malignant phenotype of PDAC (11-13). We have adopted an integrative approach to identify and prioritize genes of potential importance in PDAC. Our survival based approach involved multi-dimensional analysis of gene expression, and genomic and epigenomic regulatory mechanisms. This novel strategy allowed us to identify and refine prognosis significant genes, some of which would not have been identified based on expression alone (e.g. SRC). We also highlight several
observations based on our integrated composite gene signature and subsequent pathway and gene-ontology based analysis.

1. Many of the pathways enriched for survival-correlated genes in our analysis have already been implicated in pancreatic cancer including ERBB, focal adhesion, insulin signaling and MAPK pathways (Supplemental Table 6). Several of these pathways are linked by EGFR, which itself appears in our integrated composite signature. This is not unexpected as it has been shown that EGFR is overexpressed and associated with disease progression and poor prognosis in PDAC(30, 31). Apart from EGFR, our integrated composite signature contained several additional prognosis-associated genes linked to both SRC signaling and the PI3K/AKT pathway. These included SRC, PIK3R1/p85α and CBL, important regulatory components that we further linked to PDAC survival in separate validation cohorts by either protein or gene expression.

The PI3K/AKT pathway can promote both PDAC initiation and invasive cancer progression. AKT activity is enhanced in up to 60% of PDAC tissues and cell lines (38). More recently, we showed PI3K pathway activation is critical for the onset and acceleration of tumors in mice with conditional Kras activation and Pten deletion (41). The factors responsible for PI3K/AKT pathway dysregulation in PDAC remain unresolved. Activating mutations of the p110 subunit of PI3K are rare in PDAC (42), as are mutations or deletions of PTEN (43). To address this issue, Ying et al.(44) recently found that AKT activation is increased in 68% of PDAC, but only less than half could be
explained by genomic variation of AKT or PTEN. This leaves open other potential mechanisms of AKT activation in PDAC. While its expression may be silenced via DNA methylation (45), PTEN expression did not correlate with survival in our analysis. Instead, our data offers potential alternative mechanisms for PI3K/AKT/mTOR dysregulation in PDAC. These include increased expression of the upstream receptor tyrosine kinase EGFR and downregulation of the p110 regulatory subunit p85α encoded by PI3KR1, possibly through silencing mediated by prognostically-linked miR519d. As validation of these observations, we showed p85α protein expression to be inversely correlated with overall survival in our large PDAC TMA cohort. Likewise, either p85α or PTEN loss was found to further correlate with increased P-AKT and P-S6 status as measured in a small cohort of PDAC tumors by IHC (Supplemental Figure 3).

Our findings are consistent with previous studies reporting higher levels of SRC expression are associated with worse DFS in PDAC (35, 46). SRC is the signature member of a family of non-receptor tyrosine kinases able to mediate diverse effects on cellular proliferation, differentiation, survival, motility, and angiogenesis (47). It is activated through multiple mechanisms, including via integrins and membrane-bound receptor tyrosine kinases (e.g. EGFR) (36). SRC overexpression and its activation (as detected by Y419 phosphorylation) is seen in most PDACs (48) (35). Likewise, mice with conditional Kras activation and deletion of the SRC inhibitory kinase Csk develop invasive PDAC more rapidly and at higher prevalence than those with intact Csk (49). Our integrated composite signature and subsequent IHC validation study of P-SRC identify a potentially important association between SRC activation and clinical disease.
progression in PDAC. Our results here also provide potential mechanisms of SRC
dysregulation in PDAC. Genomic amplification of the SRC locus on chromosome 20
was found in 6 of the 25 patients (25%) and was independently correlated with worse
prognosis (HR=4.1, p-value = 0.016). In addition to genomic amplification or its
possible activation by EGFR, SRC signaling could presumably be dysregulated in
PDAC via changes in CBL (Figure 3), which was correlated with better prognosis in both
our initial analysis and separate validation cohort. While our data is correlative at this
point, it raises the intriguing possibility that CBL may act as linchpin molecule regulating
SRC and/or PI3K/AKT signaling based on its ability to act as ubiquitin ligase targeting
both EGFR and SRC (32, 33).

To our knowledge, this represents the first survival-based integrated analysis of
molecular changes in PDAC that considers the multiple dimensions of mRNA, miRNA
and CNA. Our approach offers a paradigm for future larger and more complex
multidimensional studies seeking to link clinical phenotype with the highly diverse
molecular alterations that define PDAC or other cancer types. While our study is a
preliminary and retrospective analysis of PDAC patients with resected disease, it
provides several candidate biomarkers with the potential to stratify risk for disease
progression or predict response to molecular targeted therapy. Further prospective and
mechanistic studies are not only needed to validate the prognostic or predictive value of
these markers following surgical resection, but also to establish their potential use in the
non-operative or neoadjuvant setting. Regardless, we have provided multiple lines of
correlative data showing dysregulation of the PI3K/AKT pathway and SRC signaling are
linked to PDAC clinical disease progression. These data are strong rationale for future studies seeking to link prognostically-significant signature genes mechanistically to PI3K/AKT or SRC dysregulation and explore their utility as predictive biomarkers and targets of molecular therapy in the subsets of PDAC patients they define.
ACKNOWLEDGMENTS:

We thank the UCLA Institute for Molecular Medicine and Hirshberg Foundation for Pancreatic Cancer Research for support of array analysis and the UCLA Pancreas Tissue Bank; the UCLA Department of Pathology Tissue Array Core Facility for TMA construction; the UCLA Dept. of Pathology Clinical Microarray Core Facility for array processing; and Howard Reber, Joe Hines, and James Tomlinson, for surgical specimens used in the study. We acknowledge support for TD from UCLA Scholars in Translational Medicine Program (TD); LT and RH from the UCLA Tumor Biology Program (USHHS Ruth L. Kirschstein Institutional NRSA # T32 CA009056), RH from the Damon-Runyon Cancer Foundation and DD from an American Association for Cancer Research/Pancreatic Action Network Career Development Award.

ACCESSION NUMBERS:

Study data are deposited in NCBI GEO under accession number GSE32688.
REFERENCES (ONLY 50 ALLOWED)


FIGURE LEGENDS

Figure 1: Independent mRNA, miRNA, and SNP array analyses reveal signature sets predicting prognosis. Heat map and Kaplan-Meier (KM) survival curves of stratified subgroups based on (A) mRNA expression of 500 transcripts and (B) 31 miRNA with highest absolute Cox scores. (C) Genomic location of high (> 20% of patients) frequency CNAs (upper panel) and the subset also associated with prognosis (Hazard ratio p<0.2) (lower panel). Heat map and survival curves for subgroups defined by an unsupervised clustering approach of signature CNAs (encoding 68 genes) using Ward’s method for agglomeration and Manhattan function for distance metric.

Figure 2. Unsupervised classification of the top 171 genes from the integrated composite score significantly predicts survival. (A) Heatmap of gene expression, their regulating miRNA, and CNA of local loci. The 2-mean method was used to stratify patients based on gene expression. (B) Kaplan-Meier survival curves of the two stratified groups (LRT p-value: 1e-3). (C) Venn Diagram illustrating the distribution of regulatory mechanisms (CNA and miRNA) controlling expression of the 171 composite score genes.

Figure 3. Pathway based analysis of high ranking composite score genes reveals key signaling pathways associated with PDAC clinical progression. The diagram depicts putative interactions of highly ranked poor (red) and good (green) prognosis-associated genes from the composite signature in relation to SRC signaling or the PI3K/AKT/mTOR pathway. Genes not highlighted in green or red are implied.
Figure 4: p85α and P-SRC immunohistochemistry of UCLA TMA. (A) Representative IHC from three tumors shows variable positivity for p85α and/or P-SRC. (B) Cumulative distribution of p85α and P-SRC histoscores across all tumors.

Figure 5. p85α and P-SRC expression predict PDAC survival in independent validation cohorts. Kaplan Meyer (KM) curves for dichotomized groups of low (blue) vs. high (red) IHC expression on the TMA for (a) p85α in the full cohort or (b) P-SRC for tumors of low grade histology. KM curves are also shown for (c) various combinations of both markers or (d) the single group with combined low P-SRC and high p85α versus all other combinations. (e) KM curves of dichotomized low (blue) vs. high (red) CBL mRNA expression in a separate qPCR validation cohort of 42 resected PDAC samples.
Figure 2

A  mRNA  miRNA  CNVs

B  

C  

miRNA

134

17  20

Heat map color code

Z-score of expression

Del  Norm  Amp

-4  -2  0  2

No. at risk

Good  16  15  11  7  5  1

Poor  9  8  1  1  1  0

p = 9e-3

Time (months)

0  6  12  18  24  30  36

DFS (%)

0  20  40  60  80  100  120
Figure 4

A

B

C

Cumulative fraction

P-SRC score

P-SRC

p85a

Tumor 1

Tumor 2

Tumor 3
FIGURE 5

A

![Graph A](image)

B

![Graph B](image)

C

![Graph C](image)
FIGURE 5

D

Overall survival (%)

Time (months)

No. at risk
Low P-SRC & high p85a 25 23 19 13 12 10
High P-SRC or/low p85a 113 79 46 29 22 18

E

DSS (%)

Time (months)

No. at risk
Low CBL 25 21 9 3 3 3
High CBL 17 15 11 7 5 2

p = 0.0066

p = 0.031
Table 1. Clinical, histopathologic, and survival information for the 25 patients used for the in silico array analysis

<table>
<thead>
<tr>
<th>Factor</th>
<th>subcategory</th>
<th>n(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>total samples</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Age, years</td>
<td>median (range)</td>
<td>67 (49 - 85)</td>
</tr>
<tr>
<td></td>
<td>&lt; 65</td>
<td>9 (36%)</td>
</tr>
<tr>
<td></td>
<td>&gt; 65</td>
<td>16 (64%)</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>12 (48%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>13 (52%)</td>
</tr>
<tr>
<td>Tumor location</td>
<td>Head</td>
<td>24 (96%)</td>
</tr>
<tr>
<td></td>
<td>Tail</td>
<td>1 (4%)</td>
</tr>
<tr>
<td>Operative technique</td>
<td>Whipple</td>
<td>22 (92%)</td>
</tr>
<tr>
<td></td>
<td>Distal pancreatectomy</td>
<td>1 (4%)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1 (4%)</td>
</tr>
<tr>
<td>Tumor diameter, cm</td>
<td>median (range)</td>
<td>3.4 (1.8 - 5.9)</td>
</tr>
<tr>
<td></td>
<td>&lt; 2.5</td>
<td>8 (32%)</td>
</tr>
<tr>
<td></td>
<td>&gt; 2.5</td>
<td>17 (68%)</td>
</tr>
<tr>
<td>T stage</td>
<td>2</td>
<td>2 (8%)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>23 (92%)</td>
</tr>
<tr>
<td>Resection margins</td>
<td>Positive</td>
<td>4 (16%)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>21 (84%)</td>
</tr>
<tr>
<td>Tumor differentiation</td>
<td>Well</td>
<td>1 (4%)</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>13 (52%)</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>11 (44%)</td>
</tr>
<tr>
<td>Lymph nodes</td>
<td>Negative</td>
<td>8 (32%)</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>17 (68%)</td>
</tr>
<tr>
<td>Lymphovascular invasion</td>
<td>Absent</td>
<td>5 (20%)</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>18 (72%)</td>
</tr>
<tr>
<td>Perineural invasion</td>
<td>Absent</td>
<td>2 (8%)</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>21 (84%)</td>
</tr>
<tr>
<td>AJCC Stage</td>
<td>1</td>
<td>2 (8%)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>23 (92%)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Disease Free Survival (DFS), months</td>
<td>median</td>
<td>13.2</td>
</tr>
<tr>
<td>Disease Specific Survival (DSS), months</td>
<td>median</td>
<td>20.6</td>
</tr>
<tr>
<td>Follow-up Survivors, months</td>
<td>median (range)</td>
<td>12.4 (3.8 - 32.5)</td>
</tr>
</tbody>
</table>
Table 2. Multivariate Analysis of p85α and P-SRC Protein Expression on the UCLA TMA (n = 148).

| Parameters                  | p85α-based |     | p85α & P-SRC -based |     |
|-----------------------------|------------|--|---------------------|--|--|
|                             | Hazard Ratio<sup>(a)</sup> | P-value<sup>(b)</sup> | Hazard Ratio<sup>(a)</sup> | P-value<sup>(b)</sup> |
| Positive pN                 | 1.75       | 0.005 | 1.6                | 0.02 |
| High grade                  | 1.73       | 0.007 | 1.8                | 0.003 |
| High p85α                   | 0.69       | 0.068 | -                  | -    |
| High p85α & low P-SRC       | -          | -    | 0.53               | 0.02 |

<sup>(a)</sup>HR < or > 1 indicates decreased or increased risk of death for the listed variable, respectively. <sup>(b)</sup>P-value from the Wald statistic whose null hypothesis is the corresponding coefficient $\beta =0$ (i.e. HR=1) in Cox proportional hazard model.
Clinical Cancer Research

Integrative Survival-Based Molecular Profiling of Human Pancreatic Cancer


Clin Cancer Res Published OnlineFirst January 18, 2012.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-11-1539

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2012/01/18/1078-0432.CCR-11-1539.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.