

# Integrated Proteomic Profiling of Cell Line Conditioned Media and Pancreatic Juice for the Identification of Pancreatic Cancer Biomarkers<sup>§</sup>

Shalini Makawita<sup>‡§</sup>, Chris Smith<sup>§</sup>, Ihor Batruch<sup>¶</sup>, Yingye Zheng<sup>||</sup>, Felix Rückert<sup>\*\*</sup>, Robert Grützmann<sup>\*\*</sup>, Christian Pilarsky<sup>\*\*</sup>, Steven Gallinger<sup>‡‡</sup>, and Eleftherios P. Diamandis<sup>‡§¶§§</sup>

Pancreatic cancer is one of the leading causes of cancer-related deaths, for which serological biomarkers are urgently needed. Most discovery-phase studies focus on the use of one biological source for analysis. The present study details the combined mining of pancreatic cancer-related cell line conditioned media and pancreatic juice for identification of putative diagnostic leads. Using strong cation exchange chromatography, followed by LC-MS/MS on an LTQ-Orbitrap mass spectrometer, we extensively characterized the proteomes of conditioned media from six pancreatic cancer cell lines (BxPc3, MIA-PaCa2, PANC1, CAPAN1, CFPAC1, and SU.86.86), the normal human pancreatic ductal epithelial cell line HPDE, and two pools of six pancreatic juice samples from ductal adenocarcinoma patients. All samples were analyzed in triplicate. Between 1261 and 2171 proteins were identified with two or more peptides in each of the cell lines, and an average of 521 proteins were identified in the pancreatic juice pools. In total, 3479 nonredundant proteins were identified with high confidence, of which ~40% were extracellular or cell membrane-bound based on Genome Ontology classifications. Three strategies were employed for identification of candidate biomarkers: (1) examination of differential protein expression between the cancer and normal cell lines using label-free protein quantification, (2) integrative analysis, focusing on the overlap of proteins among the multiple biological fluids, and (3) tissue specificity analysis through mining of publically available databases. Preliminary verification of anterior gradient homolog 2, syncollin, olfactomedin-4, polymeric immunoglobulin receptor, and collagen alpha-1(VI) chain in

plasma samples from pancreatic cancer patients and healthy controls using ELISA, showed a significant increase ( $p < 0.01$ ) of these proteins in plasma from pancreatic cancer patients. The combination of these five proteins showed an improved area under the receiver operating characteristic curve to CA19.9 alone. Further validation of these proteins is warranted, as is the investigation of the remaining group of candidates. *Molecular & Cellular Proteomics* 10: 10.1074/mcp.M111.008599, 1–20, 2011.

Pancreatic cancer is the fourth leading cause of cancer-related deaths and one of the most highly aggressive and lethal of all solid malignancies (1). Because of the asymptomatic nature of its early stages, coupled with inadequate methods for early detection, the majority of patients (>75%) present with locally advanced and inoperable disease at the time of diagnosis (1). At these advanced stages, chemotherapy, radiation, and combinatorial therapies are largely anecdotal, and less than 5% of patients survive up to five-years postdiagnosis (1, 2).

One way to aid in the clinical management of cancer patients is through the use of serum biomarkers. Currently, the most widely used biomarker for pancreatic cancer is carbohydrate antigen 19.9 (CA19.9)<sup>1</sup>, a sialylated Lewis A antigen found on the surface of proteins (3, 4). Although CA19.9 is elevated mainly in late stage pancreatic cancer, it is also elevated in benign diseases of the pancreas and in other malignancies of the gastrointestinal tract (5). Other tumor markers such as members of the carcinoembryonic antigen

From the <sup>‡</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; <sup>§</sup>Department of Clinical Biochemistry, University Health Network, Toronto, ON, Canada; <sup>¶</sup>Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, ON, Canada; <sup>||</sup>The Fred Hutchinson Cancer Research Center, Seattle, Washington; <sup>\*\*</sup>Visceral, Thoracic and Vascular Surgery, University Hospital Carl Gustav Carus, Technical University of Dresden, Germany; <sup>‡‡</sup>Zane Cohen Familial Gastrointestinal Cancer Registry and Department of Surgery, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada

Received February 12, 2011, and in revised form, May 19, 2011

Published, MCP Papers in Press, June 7, 2011, DOI 10.1074/mcp.M111.008599

<sup>1</sup> The abbreviations used are: CA19.9, carbohydrate antigen 19.9; CM, conditioned media; 2D-LC-MS/MS, two dimensional liquid chromatography tandem mass spectrometry; AGR2, anterior gradient homolog 2; PIGR, Polymeric immunoglobulin receptor; OLFM4, Olfactomedin-4; SYCN, Syncollin; COL6A1, Collagen alpha-1(VI) chain; SCX, strong cation exchange; LTQ, linear ion trap; FDR, false discovery rate; GO, gene ontology; emPAI, exponentially modified protein abundance index; ROC, receiver operating characteristic; AUC, area under curve; KEGG, Kyoto Encyclopedia of Genes and Genomes; KLK, kallikrein; CV, coefficient of variation; TiSGeD, Tissue-Specific Genes Database; TiGER, Tissue-specific and Gene Expression and Regulation.

(CEA) (6, 7) and mucin (MUC) (8–10) families have also been associated with pancreatic cancer. When used in combination, with or without CA-19.9, some of these markers have shown enhanced sensitivity and specificity; however none have become a constant fixture in the clinic. The lack of a single highly specific and sensitive marker has led to a growing consensus in the field toward the development of multiparametric panels of biomarkers, whereby the combinatorial assessment of multiple molecules can likely achieve increased sensitivity and specificity for disease detection and management (11–13).

Protein-based biomarkers that can be detected in circulation are typically proteins that are secreted, shed, or cleaved from tumor cells, or ones that might leak out because of local tissue destruction during disease progression (14). As such, biological fluids in close proximity to tumor cells likely serve as enriched sources of potential biomarkers before they enter the circulation and become vastly diluted and potentially masked by proteins of high abundance (15–18). With respect to pancreatic cancer, proteomic analysis of biological fluids such as pancreatic juice, cyst fluids, and bile have been conducted (19–26). Protein numbers ranging from 22 to 170 have been identified in six pancreatic juice studies using a variety of different MS-based approaches (19, 20, 22, 24–26), as well as over 460 proteins identified in a cyst fluid study (23), and 127 proteins in the bile proteome from patients with bile duct stenosis (21).

Tissue culture supernatants or conditioned media (CM) is another relevant fluid, the use of which, for the identification of novel biomarkers, has been demonstrated in multiple cancer sites by our group (27–30), and others (31–36). What is lacking in the field is integrative analysis and mining of the proteomes from different biological sources pertaining to a disease type for biomarker discovery. The utility of using an integrative approach to biomarker discovery has been described recently (15, 16). Given that cancer is a highly heterogeneous disease, through integration and comparison of proteomes from multiple biological sample types, the advantages of one source might account for the shortcomings of others, resulting in more relevant and stronger candidates for verification in plasma.

As such, in the present study, we performed in-depth proteomic analyses, integrating and comparing the proteomes of CM from six pancreatic cancer cell lines (MIA-PaCa2, BxPc3, PANC1, CAPAN1, CFPAC1, and SU.86.86), the normal human pancreatic ductal epithelial cell line (HPDE) and two pools of pancreatic juice (each pool containing three samples). All samples were analyzed in triplicate using strong cation exchange chromatography followed by liquid chromatography (LC)-tandem MS (MS/MS) on an linear trap quadrupole (LTQ)-Orbitrap mass spectrometer, and a total of 3479 nonredundant proteins were identified with two or more peptides in our comparative proteomics analysis.

One of the challenges in high throughput proteomics-based discovery studies is in the selection of the most promising

candidates for verification (37). In the present study, we used three strategies for generation of candidates. First, given the relatively large number of cancer cell lines profiled, proteins overexpressed in the cancer cell lines in comparison to the HPDE cell line were examined through label-free quantification. Four-hundred and eighty-three proteins were found to be expressed over 5-fold in at least one cancer cell line in comparison to HPDE, 63 of which were annotated as extracellular or cell surface and overexpressed in at least three cancer cell lines. This included several proteins previously studied as pancreatic cancer serum biomarkers, helping to provide credence to our label-free quantitative approach. As a second strategy, we performed a comparative analysis, integrating the pancreatic juice proteome with that of the cell lines and selecting proteins that are common to the cancer cell lines and pancreatic juice. Comparison to a third biological fluid (pancreatic cancer ascites, Kosanam *et al.*, unpublished) and focusing on proteins also identified in the ascites proteome helped to further filter candidates. Tissue specificity is also a desired criteria for biomarkers (37, 38), and our third strategy entailed tissue specificity analysis of identified proteins based on mining of the Tissue-Specific Genes Database (TiSGeD) (39), Tissue-specific and Gene Expression and Regulation (TiGER) (40), Unigene (41), and Human Protein Atlas (42) databases, focusing on proteins specific to or highly expressed in the pancreas.

Of the derived candidates, initial verification studies using ELISAs in plasma samples from patients with established pancreatic cancer and controls resulted in the identification of five proteins—anterior gradient homolog 2 (AGR2), Polymeric immunoglobulin receptor (PIGR), Olfactomedin-4 (OLFM4), Syncollin (SYCN), and Collagen alpha-1(VI) chain (COL6A1)—that showed a significant increase ( $p < 0.01$ ) in plasma from pancreatic cancer patients. AGR2, PIGR, and COL6A1 were proteins overexpressed in three or more cell lines, OLFM4 was a protein consistently identified in the multiple biological fluids, and SYCN was a protein identified in our proteomic analysis that showed high tissue specificity to the pancreas. CA19.9 levels were also assessed in the plasma samples and the combination of the five proteins showed an improved area under the receiver operating characteristic (ROC) curve to CA19.9. Our verification studies in plasma provide evidence that our approach can identify candidates with potential diagnostic utility for pancreatic cancer. Further validation of AGR2, PIGR, OLFM4, SYCN, and COL6A1 in a larger number of plasma samples is warranted, as is the investigation of the remaining group of candidates.

### EXPERIMENTAL PROCEDURES

**Cell Lines**—Six pancreatic cancer cell lines (MIA-PaCa2 (CRL-1420), PANC1 (CRL-1469), BxPc3 (CRL-1687), CAPAN1 (HTB-79), CFPAC-1 (CRL-1918), and SU.86.86 (CRL-1837)) were obtained from the American Type Culture Collection (ATCC, Manassas, VA). The cell lines were derived from pancreatic ductal adenocarcinomas, which account for ~85–90% of all pancreatic cancers. The cell lines origi-

nated from primary tumors of the head or body of the pancreas (MIA-PaCa2, PANC1, BxPc3), or from metastatic sites (CAPAN1, CFPAC-1, SU.86.86) (43, 44). The cell lines were derived from individuals of similar ethnic background and age group (with the exception of CFPAC-1), and all of the cancer cell lines, except for BxPc3, are positive for K-ras mutations, which is found in 85–90% of pancreatic cancers. An HPV transfected “normal” HPDE (45), provided by Dr. Ming-Sound Tsao at Princess Margaret Hospital, Toronto, Ontario, Canada was also analyzed. Apart from a slightly aberrant expression of p53, molecular profiling of this cell line has shown that expression of other proto-oncogenes and tumor suppressor genes are normal (45).

Cell culture media specified by ATCC for each of the six pancreatic cancer cell lines were used and are as follows: Dulbecco’s modified Eagle’s medium (DMEM) (Catalog No. 30–2002 from ATCC) with 10% fetal bovine serum (Catalog No.10091–148; Invitrogen) was used for MIA-PaCa2 and Panc1; Roswell Park Memorial Institute (RPMI) 1640 medium modified to contain 2 mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate, 4500 mg/L glucose, 1500 mg/L sodium bicarbonate (ATCC Catalog No. 30–2001) with 10% fetal bovine serum was used for SU.86.86 and BxPc3; Iscove’s modified Dulbecco’s medium (Catalog No. 30–2005) with 10 and 20% fetal bovine serum was used for the CFPAC-1 and Capan1 cell lines, respectively. The HPDE cell line was grown in keratinocyte serum free media (Catalog No.17005–042; Invitrogen) supplemented with bovine pituitary extract and recombinant epidermal growth factor. All cells were cultured in an atmosphere of 5% CO<sub>2</sub> in air in a humidified incubator at 37 °C.

**Cell Culture**—Cells were cultured in T-175 cm<sup>2</sup> flasks at determined optimal seeding densities of  $\sim 10 \times 10^6$ , for MIA-PaCa2, Panc1, and Capan1,  $14 \times 10^6$  for BxPc3,  $3 \times 10^6$  for HPDE,  $13 \times 10^6$  for CFPAC1 and  $4 \times 10^6$  for Su.86.86 in three replicates per cell line. Cells were first cultured for 48 h in 40 ml of their respective growth media to obtain adherence to culture flasks. The media was then removed and the cells and flasks were subjected to two gentle washes with 30 ml of phosphate-buffered saline (Invitrogen). Forty millilitres of chemically defined Chinese hamster ovary serum-free medium (Invitrogen) supplemented with 8 mM glutamine (Invitrogen) was then added and the cells were left to culture for determined optimal incubation periods of 72 h for Capan1, CFPAC1, and SU.86.86, 96 h for BxPc3 and HPDE, and 144 h for MIA-PaCa2. The chemically defined Chinese hamster ovary media that the cells were grown in were subsequently collected and centrifuged at 1500 rpm for 10 min to remove cellular debris. Total protein concentration (as determined through a Coomassie (Bradford) total protein assay, (46)) was measured in each of the three replicates and a volume corresponding to 1 mg of total protein from each of the replicates was subjected to the sample preparation protocol below.

Before analysis of samples, the seeding density and incubation periods were optimized through a procedure described previously (28, 29). Briefly, cells were left to culture for a period of 5–7 days using different seeding densities. Each day, a 1 ml sample of media was collected and protein secretion was assessed by measuring the levels of proteins known to be secreted from cells (e.g. multiple kallikreins, KLK) through ELISAs. Cell death was assessed through measurement of LDH, an intracellular enzyme that is released into the media on cell death or lysis via an automated activity assay and an “optimal” incubation period and seeding density that supports increased protein secretion (*i.e.* increased levels of KLK) with minimal cell death (*i.e.* low LDH) were selected (28, 29).

**Pancreatic Juice**—Pancreatic juice samples were provided by Dr. Felix Rückert, Dresden, Germany. Approximately 50–500  $\mu$ l of pancreatic juice was collected from the main pancreatic duct of patients undergoing pancreatic surgery. The study was approved by the local ethics committee and all patients had given written informed consent

before operation. On collection, the samples were stored at  $-80$  °C and were not thawed until use in this study. Samples from patients with clinically confirmed cases of pancreatic ductal adenocarcinoma without visible signs of blood were selected for analysis. Six pancreatic juice samples met these criteria. The samples were centrifuged at 16,000 rpm for 10 min at 4 °C to remove tissue debris. Total protein concentration of each sample was measured using the Biuret method (47). Keeping in line with the cell line conditioned media analysis, it was desirable to use a total protein amount of 1 mg for analysis of each of the three replicates per sample. As a result, two pools of pancreatic juice (pools A and B) were made, containing three samples each, with total protein concentrations of 2.65 mg/ml and 2.32 mg/ml for pool A and B, respectively. A volume corresponding to 1 mg of total protein was retrieved from each pool, in triplicate, and subjected to the standardized sample preparation protocol below (with the exception of dialysis).

**Sample Preparation**—Samples were processed as described previously (28). Briefly, samples were dialyzed using a 3.5 kDa molecular weight cut-off membrane (Spectrum Laboratories, Inc., Compton, CA) in 5 L of 1 mM NH<sub>4</sub>HCO<sub>3</sub> buffer solution at 4 °C overnight and subsequently frozen and lyophilized to dryness to concentrate proteins using a ModulyoD Freeze Dryer (Thermo Electron Corporation). Proteins in each lyophilized replicate were denatured using 8 M urea and reduced with the addition of 200 mM dithiothreitol (final concentration of 13 mM) in 1 M NH<sub>4</sub>HCO<sub>3</sub> at 50 °C for 30 min. Samples were then alkylated with the addition of 500 mM iodoacetamide and incubated in the dark, at room temperature, for 1 h. Each replicate was then desalted using a NAP5 column (GE Healthcare), frozen and lyophilized. Last, samples were trypsin-digested (Promega, sequencing-grade modified porcine trypsin) through an overnight incubation at 37 °C using a ratio of 1:50 trypsin to protein concentration. Tryptic peptides were frozen in solution at  $-80$  °C to inhibit trypsin function and lyophilized.

**Strong Cation Exchange (SCX) on a High Pressure Liquid Chromatography (HPLC) System**—The tryptic peptides were resuspended in 510  $\mu$ l of a solution containing 0.26 M formic acid in 10% acetonitrile with pH 2–3 (mobile phase A) and loaded directly onto a 500  $\mu$ l loop connected to a PolySULFOETHYL A<sup>TM</sup> column (The Nest Group, Inc. Southborough, MA). The column has a silica-based hydrophilic, anionic polymer (poly-2-sulfoethyl aspartamide) with a pore size of 200 Å and a diameter of 5  $\mu$ m. The SCX chromatography and fractionation was performed on an HPLC system (Agilent 1100) using a 1-hour procedure with a linear gradient of mobile phase A. For elution of peptides, an elution buffer, which contained all components of mobile phase A with the addition of 1 M ammonium formate, was introduced at 20 min in the 60 min method, and a wavelength of 280 nm was used to monitor the eluent (28). Fractions were collected every min from the 20 min time point onwards resulting in the collection of 40 one-min fractions. Collected fractions were left unpooled or subsequently combined into 2, 3, or 5 min pools, according to the elution profile of the resulting SCX chromatogram. As a general strategy, in which the absorbance readings of the elution profile was greater (typically the first 10–15 min of elution), fractions were left unpooled or pooled every 2 min to keep sample complexity at a minimum. Where the absorbance readings were lower (toward the end of the method), fractions were pooled in 3 or 5 min pools. The same pooling method was used for the three replicates of each sample.

**Mass Spectrometry (LC-MS/MS)**—The SCX fractions or pools were purified through OMIX Pipette Tips C<sub>18</sub> (Varian Inc.) to further remove impurities and salts and eluted in 4  $\mu$ l of 70% MS Buffer B (90% acetonitrile, 0.1% formic acid, 10% water, 0.02% trifluoroacetic acid) and 30% MS Buffer A (95% water, 0.1% formic acid, 5% acetonitrile, 0.02% trifluoroacetic acid). Eighty microliters of MS Buffer A was added to the eluent, and 40  $\mu$ l of sample was loaded onto a 3 cm C<sub>18</sub>

trap column (with an inner diameter of 150  $\mu\text{m}$ ; New Objective), packed in-house with 5  $\mu\text{m}$  Pursuit  $\text{C}_{18}$  (Varian Inc.). A 96-well microplate autosampler was used for sample loading. Eluted peptides from the trap column were subsequently loaded onto a resolving analytical PicoTip Emitter column, 5 cm in length (with an inner diameter of 75  $\mu\text{m}$  and 8  $\mu\text{m}$  tip, New Objective) and packed in-house with 3  $\mu\text{m}$  Pursuit  $\text{C}_{18}$  (Varian Inc, Palo Alto, CA). The trap and analytical columns were operated on the EASY-nLC system (Proxeon Biosystems, Odense, Denmark), and this liquid chromatography setup was coupled online to an LTQ-Orbitrap XL hybrid mass spectrometer (Thermo Fisher Scientific, San Jose, California) using a nano-electrospray ionization (ESI) source (Proxeon Biosystems, Odense, Denmark). Samples were analyzed using a gradient of either 54 or 90 min (for 5 min pools, a 90 min gradient was used, and for 2min, 3min, and nonpooled samples, a 54 min gradient was used). Samples were analyzed in data dependent mode and whereas full MS<sup>1</sup> scan acquisition from 450–1450  $m/z$  occurred in the Orbitrap mass analyzer (resolution 60,000), MS<sup>2</sup> scan acquisition of the top six parent ions occurred in the linear ion trap (LTQ) mass analyzer. The following parameters were enabled: monoisotopic precursor selection, charge state screening, and dynamic exclusion. In addition, charge states of +1, >4, and unassigned charge states were not subjected to MS<sup>2</sup> fragmentation.

**Protein Identification**—XCalibur software (version 2.0.5; Thermo Fisher) was used to generate RAW files of each MS run. The RAW files were subsequently used to generate Mascot Generic Files (MGF) through `extract_msn` on Mascot Daemon (version 2.2.2). Once generated, MGFs were searched with two search engines, Mascot (Matrix Science, London, UK; version 2.2) and X!Tandem (Global Proteome Machine Manager; version 2006.06.01), to confer protein identifications. Searches were conducted against the nonredundant Human International Protein Index database (version 3.62), which contains a total of 167,894 forward and reverse protein sequences using the following parameters: fully tryptic cleavages, 7 ppm precursor ion mass tolerance, 0.4 Da fragment ion mass tolerance, allowance of one missed cleavage, fixed modifications of carbamidomethylation of cysteines, and variable modification of oxidation of methionines. The files generated from MASCOT (DAT files) and X!Tandem (XML files) for the three replicates of each cell line or pancreatic juice pool were then integrated through Scaffold 2 software (version 2.06; Proteome Software Inc., Portland, Oregon) resulting in a nonredundant list of identified proteins per sample. Results were filtered separately for each individual cell line (three replicates combined) and pancreatic juice pool (three replicates combined) using the X!Tandem LogE filter and Mascot ion-score filters on Scaffold to achieve a protein false discovery rate (FDR) <1.0%. FDR was calculated as  $[2\text{XFP}/(\text{TP} + \text{FP})] \times 100$ , in which FP (false positive) is the number of proteins that were identified based on sequences in the reverse database component and TP (true positive) is the number of proteins that were identified based on sequences in the forward database component. Protein isoforms and individual members of a protein family would be separately identified only if peptides that enable differentiation of isoforms were identified based on generated MS/MS data. If such peptides were not identified, Scaffold software would group all isoforms under the same gene name. Different proteins that contained similar peptides and which were not distinguishable based on MS/MS data alone were grouped to satisfy the principles of parsimony.

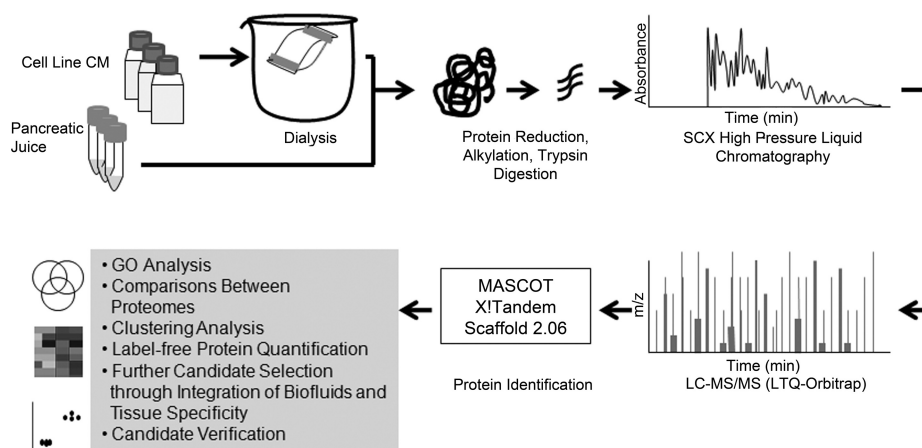
**Data Analysis**—Scaffold prot-XML reports were generated and uploaded onto Protein Center Professional Edition (version 3.5.2.1; Proxeon Bioinformatics, Odense, Denmark) to facilitate comparisons between cell line CM and pancreatic juice proteomes, and to obtain Genome Ontology information. Specifically, cellular localization, function, and process annotations were extracted by Protein Center from the Gene Ontology (GO) Consortium (<http://www.geneontology.org/>

[GO.tools.shtml](http://www.geneontology.org/)). Because of the large number of different GO annotations per localization, function, and process, Protein Center reduces terms to ~20 high-level terms that are used for filtering. Details can be found at <http://tgh.proteincenter.proxeon.com/ProXweb/Help/Manual/apd.html>. A Microsoft Excel Macro developed in-house by Dr. Irvin Bromberg, Mount Sinai Hospital, was also used for comparison of protein lists based on accession number or gene name. Hierarchical clustering analysis of proteomic data was performed using PermutMatrix, available freely online at <http://www.lirmm.fr/~caraux/PermutMatrix/EN/index.html>. PermutMatrix is a software originally developed for gene expression analysis (48). More recently it has been used and validated for proteomics (49). For clustering analysis, a standardization method described previously (32), was adapted, in which average exponentially modified protein abundance index (emPAI) values from the triplicate analysis of the samples were exported from Protein Center into a space delimited Microsoft Excel file. For visualization, comparison and data analysis purposes, cell line or pancreatic juice samples with missing emPAI values for a particular protein were assigned half the minimum emPAI value for that protein in the data set (32). The emPAI values were imported into PermutMatrix and transformed to Z score values for normalization. Two-way hierarchical clustering analysis was performed using the Pearson and Ward's minimum variance methods for distance and aggregation, respectively. Resultant dendograms with cell lines and pancreatic juice samples on the x axis and gene name on the y axis were exported. Last, Ingenuity Pathway Analysis (IPA, Ingenuity Systems, [www.ingenuity.com](http://www.ingenuity.com)), which uses a knowledge-base from literature, was used to obtain disease associations of groups of proteins and their associated pathway interactions.

**Label-free Protein Quantification**—Relative label-free protein quantification analysis was conducted between the cancer cell lines and the HPDE normal pancreatic ductal epithelial cell line to ascertain proteins over- or under-expressed in the cancer cell lines based on spectral counting (50–52). The “Quantitative Value” function of Scaffold 2.06 software, which provides normalized spectral counts based on the total number of spectra identified in each sample was used. One Scaffold file containing all of the normalized spectral counts of each of the three replicates from the seven cell lines was generated for proteins identified with two or more peptides. One-way ANOVA was conducted to determine proteins that show a significant difference among the seven cell lines ( $p < 0.05$ ). For proteins that showed a  $p$  value <0.05, the average spectral count for the three replicates was calculated and fold-change was determined by dividing the average counts from each of the cancer cell lines with that of the normal HPDE cell line and *vice versa*. Not all proteins were identified in all of the cell lines and unidentified proteins or missing values in a particular biological sample were assigned a normalized spectral count of 1 to keep from dividing by zero and to prevent overestimation of fold-changes. In addition, all proteins had to have been identified by ten or more spectra in at least one biological sample to be included. Percent coefficient of variance (%CV) was evaluated as a measure of variance for the triplicates of each sample, and proteins with ambiguous peptides were searched individually to ensure that normalization of spectral counts did not significantly alter assigned spectral count values. Quantification of multiple isoforms occurred only if the isoforms were identified as individual entries (*i.e.* with distinct gene names) through Scaffold Software.

**Plasma Samples**—Blood samples were collected from pancreatic cancer patients at the Princess Margaret Hospital GI Clinic in Toronto, Canada, or from kits sent directly to consenting patients recruited from the Ontario Pancreas Cancer Study at Mount Sinai Hospital following a standardized protocol (age range 55–86; median age 68; 10 female and 10 male). Samples were collected with informed consent, and with the approval of the institutional ethics board. All cases

FIG. 1. **Schematic outline of proteomic analysis methodology.** Sample processing, followed by strong cation exchange (SCX) and reverse-phase chromatography coupled online to an LTQ-Orbitrap mass spectrometer, and subsequent data analysis is outlined. CM, conditioned media; LC-MS/MS, liquid chromatography tandem mass spectrometry.



were histologically confirmed adenocarcinomas of the pancreas and all samples were from patients not receiving chemotherapy at the time of collection. Samples from healthy controls were obtained from the Familial Gastrointestinal Cancer Registry. The controls are non-blood relatives of patients in the Familial Gastrointestinal Cancer Registry studies (age range 46–84; median age 60; 9 female and 11 male). Blood was collected in ACD (anticoagulant) vacutainer tubes and plasma samples were processed within 24 h of blood draw. To pellet the cells, blood samples were centrifuged at room temperature for 10 min at  $1000 \times g$ . Immediately after centrifugation, the plasma samples were aliquoted into 250- $\mu$ l cryotubes and stored at  $-80^\circ\text{C}$  or liquid nitrogen until use in this study.

**ELISAs and Immunoassays**—Enzyme-linked immunosorbent assays for AGR2, SYCN, OLFM4, COL6A1, PIGR, NUCB2, and PLAT were purchased commercially and performed according to the manufacturer's instructions. Five ELISA kits were purchased from USCN LifeSciences (AGR2, Catalog # E2285Hu; SYCN, Catalog # E93879Hu; OLFM4, Catalog # E90162Hu; COL6A1, Catalog # E92150Hu; PIGR, Catalog # E91074Hu). The ELISA kits for NUCB2 and PLAT were purchased from Phoenix Pharmaceuticals (Belmont, CA) (Catalog # EK-003–26) and American Diagnostica Inc. (Greenwich, CT) (Catalog # 860), respectively. The ELECSYS CA 19–9 immunoassay by Roche was used to measure CA19.9 levels in plasma and kallikrein 6 and 10 internal control proteins were measured in CM using in-house developed ELISA assays, as described previously (53, 54).

**Statistical Analysis**—Mann-Whitney U-tests were applied to verification experiments in plasma to determine if differences in the medians were significant between cancer and control groups using Graph Pad Prism 4 Software. The five candidates that showed a statistically significant difference ( $p < 0.05$ ) were then assessed in combination, in comparison to CA19.9, through ROC curve analysis. The area under the curves (AUC) were calculated using ROC software and the corresponding variances were calculated with a bootstrap method.

## RESULTS

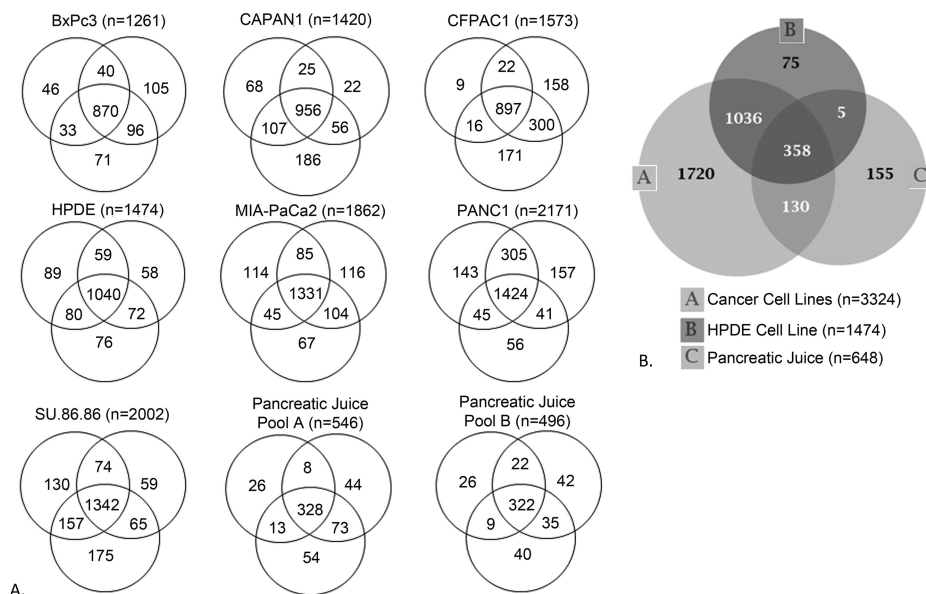
**Increasing Protein Yield**—Once the cell lines were grown and CM collected, the samples were subjected to a two-dimensional (2D)-LC-MS/MS analysis, which combined SCX liquid chromatography on an HPLC system, followed by LC-MS/MS. A schematic of the workflow is provided in Fig. 1. Guided by our previous experience (27–30), SCX fractions were initially collected at 5-min intervals during peptide elution, resulting in approximately eight fractions that were ana-

lyzed using a  $\sim 2$  h reverse-phase method on the LTQ-Orbitrap mass spectrometer. This resulted in the identification of 1305, 1468, and 1755 proteins ( $\geq 1$  peptide) in the triplicate analysis of the BxPc3, HPDE, and MIA-PaCa2 cell lines, respectively (supplemental Table S1). In some of the individual 5-min fractions analyzed (specifically the fractions that contained the highest absorbance readings during SCX peptide elution)  $>700$  proteins were identified per fraction (data not shown). Based on previous experience, this was a relatively large number of proteins to have been identified in individual fractions. Consequently, we opted to employ a different fraction collection and pooling strategy. By collecting fractions every minute from SCX and pooling fractions based on the intensity of peaks eluting on the SCX chromatogram (as described in the *Experimental Procedures*; supplemental Fig. S1), we identified 2017 proteins for the BxPc3 cell line, 2297 for HPDE, and 2756 for the MIA-PaCa2 cell line ( $\geq 1$  peptide) subjected to the same growth and sample processing conditions, in triplicate. To ensure that this increase in protein identification was not because of variation in cell growth or sample collection, an additional replicate using MIA-PaCa2 CM left over from the initial analysis, which had been stored in  $-80^\circ\text{C}$ , was also run and 2348 proteins were identified (a 52–56% increase from the individual replicates of the first run of MIA-PaCa2) (supplemental Table S1).

Over 90% of the proteins from the first analysis were re-identified and the new pooling strategy resulted in approximately a 54–58% increase in protein yield across the cell lines. This improved strategy was used for proteomic analysis of the remaining cell lines and the pancreatic juice samples.

**Protein Identification through LC-MS/MS**—Six human pancreatic cancer cell lines, one “near normal” HPDE and six pancreatic juice samples from ductal adenocarcinoma patients (in two pools) were analyzed in triplicate in this study (Fig. 2). Using both MASCOT and X!Tandem search engines, between 2017 and 3250 proteins were identified in the seven cell lines and 1018 and 957 proteins were identified from pool A and B of pancreatic juice, respectively (Table I). These numbers represent proteins identified in the three replicates

**FIG. 2. Total nonredundant proteins identified.** Venn diagrams depicting total proteins identified with  $\geq 2$  peptides in the three replicates of each cell line CM and pancreatic juice sample (A). Overlap of 3479 total nonredundant proteins identified in the conditioned media and pancreatic juice analysis is also depicted (B). HPDE, (normal) human pancreatic ductal epithelial cell line.



combined, with 1 or more peptides and with protein FDR of  $<1.0\%$ . FDR was calculated as described in the *Experimental Procedures* section (55–57). For increased stringency and assurance of protein identification, only proteins identified with two or more peptides were included in the remainder of the analysis, resulting in between 1261 and 2171 proteins for each of the cell lines and a total of 648 nonredundant proteins from the pancreatic juice analysis. This data is summarized in Table I and a complete list of proteins identified ( $\geq 2$  peptides), including number of unique peptides, International Protein Index accession number, number of assigned spectra and percent sequence coverage is presented in supplemental Table S2 for each cell line and pancreatic juice pool. Further and more specific details on peptide/protein identification are provided in supplemental Tables S3–S5 (supplemental Table S3: BxPc3, CAPAN1, and CFPAC1; supplemental Table S4: HPDE, MIA-PaCa2, and PANC1; supplemental Table S5: SU.86.86, pancreatic juice pool A, and pancreatic juice pool B).

**Protein Overlap Among Samples**—From our combined analysis, a total of 12,805 proteins were identified with  $\geq 2$  peptides. The majority of these proteins were common to multiple biological samples and our study resulted in the identification of 3479 nonredundant proteins (3324 in the cell lines and 648 in the pancreatic juice analysis; supplemental Table S6). Six-hundred and forty-four proteins (of 3324; 19.4%) were common to all cell lines and an average of 143 proteins were unique to each (Table II). Similar to previous findings (32), proteins common to an increasing number of cell lines were, on average, of higher abundance (Table II). From our preliminary studies of the three cell lines described in the “increasing protein yield” section, 81 additional nonredundant proteins ( $\geq 2$  peptides) were identified. These were not included in the remainder of the analyses but

have been included in supplemental Table S6 under a separate tab.

Significant overlap was noted between the pancreatic juice and CM proteins. Approximately 76% (493 of 648) of proteins identified in the pancreatic juice samples were also identified in the cell line analysis (Fig. 2B), which underscores the similarity in the proteomes among these biological fluids; however many proteins that are largely associated with exocrine pancreatic function were unique to the pancreatic juice and were not identified in the cell lines. Analysis of overrepresented Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways through Protein Center (supplemental Table S7), further revealed the KEGG pancreatic secretion pathway (hsa04972; supplemental Fig. S2) to be one of three pathways overrepresented in the pancreatic juice proteome in comparison to the combined cell line proteome ( $p = 3.611 \times 10^{-5}$ ) (58, 59).

**Gene Ontology—Function, Process, and Cell Localization Classifications**—Gene ontology classifications, which include function, process, and cell localization, were obtained for all identified proteins. Proteins that are secreted into the extracellular milieu or cleaved from the plasma membrane of cells have the highest chance of entering the circulation and serving as serological biomarkers. Between 34.1 and 42.6% of proteins in each of the cell lines and 57% of proteins in the pancreatic juice samples were annotated as belonging to the extracellular and cell surface compartments (Table I). In total, 1376 (40%) of 3479 proteins were associated with these two annotations. The cytoplasm received the greatest number of annotations in both biological fluids and  $\sim 2.9\%$  of the total contingent of proteins did not contain cell localization information and remained unannotated (Fig. 3A). It is important to note that proteins can be classified as belonging to multiple cellular localizations, processes, and functions and as a re-

TABLE I  
 Total number of proteins identified in triplicate analysis of cell line conditioned media and pancreatic juice

	Cell Lines							Pancreatic Juice	
	BxPc3	CAPAN1	CFPAC1	HPDE	MIA-PaCa2	PANC1	SU.86.86	Pool A	Pool B
Total Non-Redundant Proteins <sup>a</sup> [with $\geq 2$ peptides]	2017 [1261]	2182 [1420]	2427 [1573]	2297 [1474]	2756 [1862]	3250 [2171]	3010 [2002]	1018 [546]	957 [496]
Number of Proteins Identified with ...									
Only 1 peptide	756	762	854	823	894	1079	1008	472	461
Only 2 peptides	336	374	394	400	464	491	491	172	150
Only 3 peptides	226	230	290	252	281	347	322	96	79
Only 4 peptides	144	171	175	166	217	248	224	43	64
$\geq 5$ peptides	555	645	714	656	900	1085	965	235	203
Protein False Discovery Rate <sup>b</sup>	0.69%	0.82%	0.66%	0.87%	0.80%	0.62%	0.73%	0.79%	1.0%
Number [%] of Extracellular and Cell Surface Proteins with $\geq 2$ peptides <sup>c</sup>	511 [40.5%]	605 [42.6%]	665 [42.3%]	592 [40.2%]	635 [34.1%]	757 [34.9%]	749 [37.4%]	314 [57.5%]	281 [56.7%]

<sup>a</sup> All nonredundant proteins identified with  $\geq 1$  peptide; the number of total proteins identified with  $\geq 2$  peptides is enclosed in brackets.

<sup>b</sup> False discovery rate (FDR) calculated as  $[2XFP/(TP+FP)] \times 100$ , where FP (false positive) is the number of proteins that were identified based on sequences in the reverse database component and TP (true positive) is the number of proteins that were identified based on sequences in the forward database component. FDR pertains to total proteins identified with  $\geq 1$  peptide; False discovery rate = 0.0% for  $\geq 2$  peptide identifications.

<sup>c</sup> Pertains to proteins annotated using Genome Ontology (GO) Slim categories through Protein Center software.

 TABLE II  
 Protein overlap between cell line conditioned media

Number of Cell Lines	Number of Proteins <sup>a</sup>	Average emPAI values $\pm$ Standard Deviation <sup>c</sup>	% of the CM Proteins Identified in the Pancreatic Juice <sup>d</sup>
7	644	1.75 $\pm$ 1.26	42%
6	285	0.98 $\pm$ 0.78	14%
5	295	1.04 $\pm$ 1.01	14%
4	268	0.87 $\pm$ 0.83	14%
3	336	0.73 $\pm$ 0.76	10%
2	494	0.74 $\pm$ 0.85	5%
1	1001 <sup>b</sup>	0.70 $\pm$ 0.998	4%

<sup>a</sup> Indicates the number of proteins with two or more peptides that were commonly identified in the reported number of cell lines.

<sup>b</sup> A total of 1001 proteins were only identified in one of the seven cell lines; an average of 143 proteins were unique to each cell line.

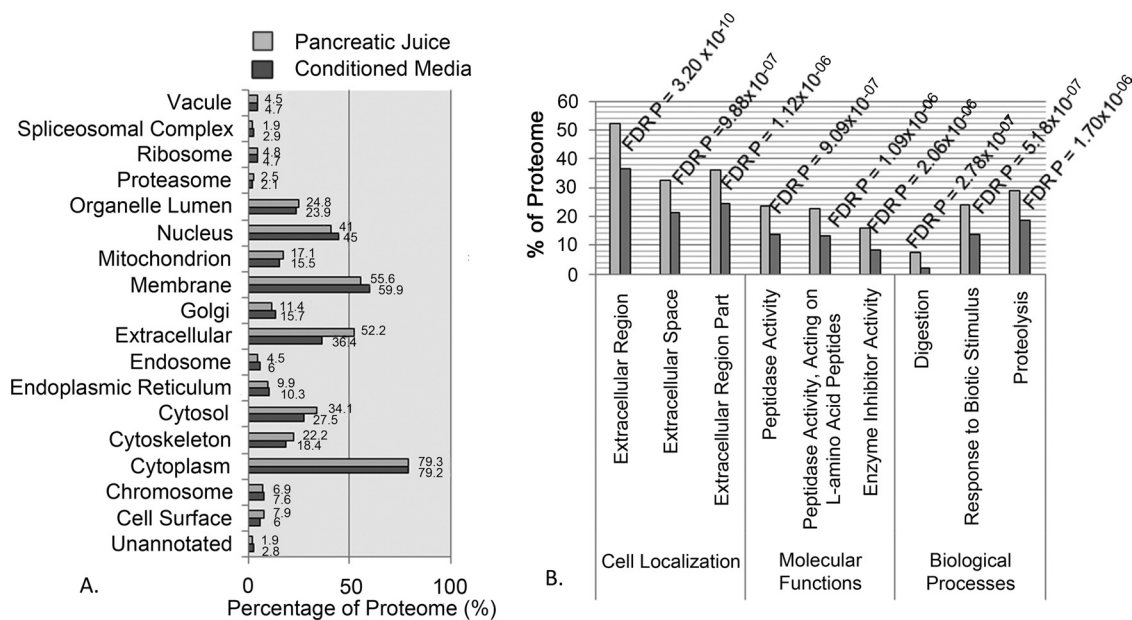
<sup>c</sup> Average emPAI values (semiquantitative measure of protein abundance) were obtained from the triplicate analysis of each cell line for each protein and then an average was calculated for each protein if the protein was identified in multiple cell lines. Subsequently an average (and standard deviation) for each group of proteins common to the stated number of cell lines was calculated and reported. The proteins common to multiple cell lines had a higher average emPAI than those common to a small number of cell lines or unique to one cell line indicating proteins common to a larger number of cell lines were on average of higher abundance.

<sup>d</sup> Indicates the percentage of proteins common to the multiple cell lines that were also identified in the pancreatic juice proteome.

sult, the categories for each are nonexclusive and the sum of the percentages can be  $>100\%$ .

The top three molecular functions for the cell lines and the pancreatic juice were the same: protein binding ( $\sim 80.2\%$ ,  $79.9\%$ ), catalytic activity ( $69.4\%$ ,  $70.2\%$ ), and metal ion binding ( $45.7\%$ ,  $49.7\%$ ), respectively. Both fluids also shared the top two biological processes—metabolic process ( $81.2\%$ ,  $83.5\%$ ) and regulation of biological process ( $61.9\%$ ,  $63.1\%$ ), respectively (supplemental Fig. S3A, B). In a comparison between the cell line and pancreatic juice proteomes as a whole, several molecular functions related to enzyme activity and biological processes related to digestion and proteolysis were found overrepresented in the pancreatic juice proteome (Fig. 3B). The only GO category overrepresented in the cell line proteome was the biological process “biopolymer metabolic process” (GO:0043283;  $p = 4.053 \times 10^{-7}$ ; FDR  $p = 2.524 \times 10^{-3}$ ). No GO terms were over- or under-represented in a comparison between the cancer cell lines and HPDE. Specifics for localization, function, and process for each protein are provided in supplemental Table S6.

**Hierarchical Clustering**—One of the difficulties in dealing with large data sets is visualizing the proteomes as a whole and identifying subsets of proteins that might be of importance within certain biological contexts. In an initial attempt to mine and explore the CM and pancreatic juice proteomes, unsupervised two-way hierarchical clustering analysis was performed using average emPAI values of the three replicates



**FIG. 3. Cellular localization and comparison of GO categories between cell line conditioned media and pancreatic juice proteomes.** Cellular localization of proteins annotated using gene ontology (GO) consortium annotations (A). Depicted in dark gray is the percentage of proteins from the cell line CM analysis and light gray is percentage of proteins from the pancreatic juice analysis for each cell localization category. Proteins can contain multiple GO annotations resulting in a sum of percentages >100%. Top three significantly overrepresented GO categories in pancreatic juice proteins in comparison to the cell line conditioned media proteome for cellular localization, molecular function, and biological process are also depicted (B). FDR (false discovery rate)  $p$  values are based on application of a hypergeometric test corrected for a FDR of 1% as obtained through Protein Center Software.

of each sample, normalized through Z-scores. Through these means, proteins were clustered based on abundance within each sample. The concentrations of two proteins (KLK6 and KLK10) were assessed in the CM through ELISA to determine if Z-scores of emPAI values are a suitable indicator of protein abundance. Good correlation ( $r^2 = 0.74$ ) was seen between Z-scores of ELISA concentrations and Z-scores of emPAIs (Fig. 4A) (supplemental Table S8). In addition, the lowest ELISA concentration measured was  $0.80 \mu\text{g/L}$  for KLK10 in CAPAN1 CM, which indicates the sensitivity of our mass spectrometry analysis in general, to be at least in the low  $\mu\text{g/L}$  range for the CM analysis.

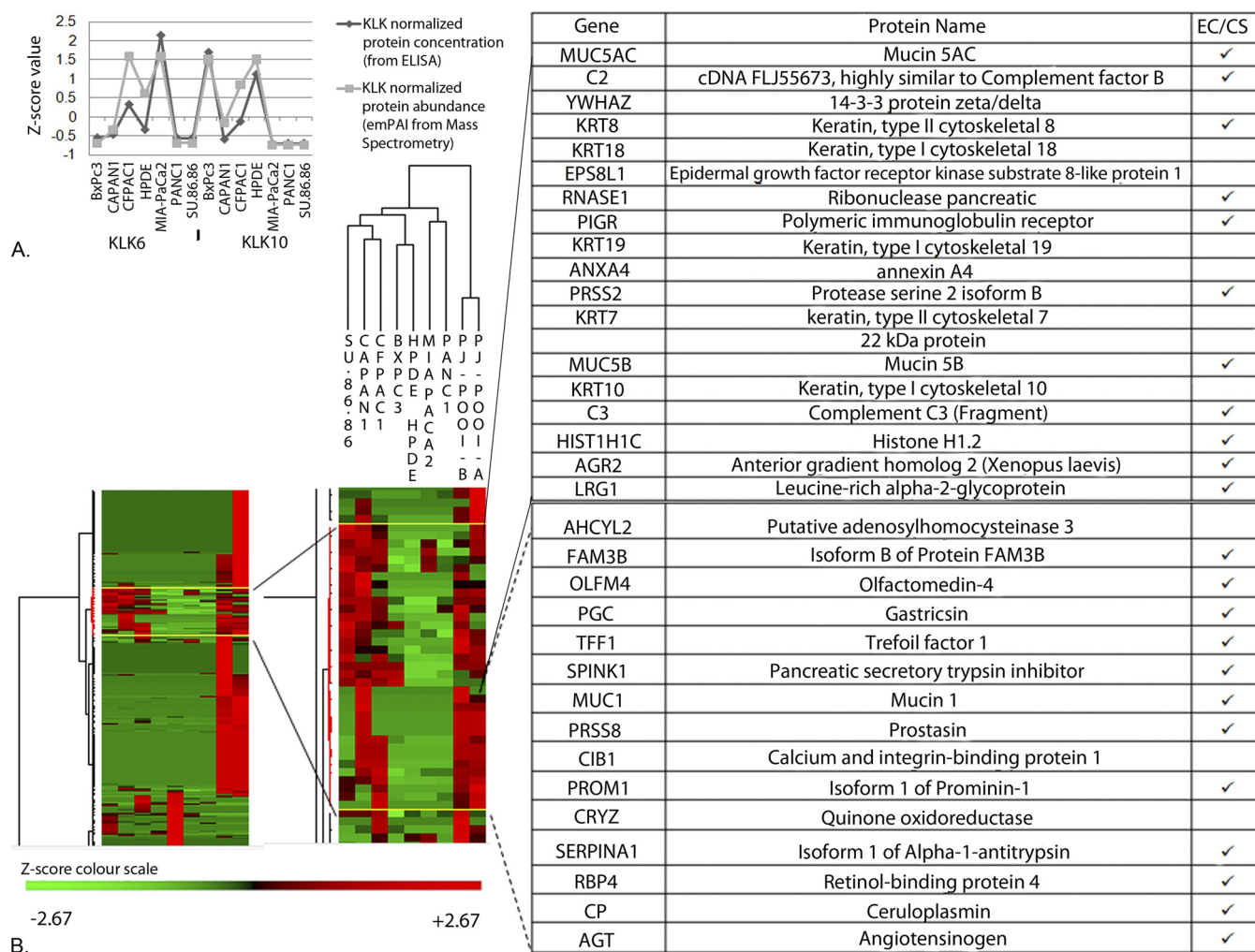
Hierarchical clustering analysis was performed on the entire data set of 3479 proteins and based on normalized emPAI values, the pancreatic juice samples were distinctly clustered separately from the cell lines, and within the cell lines, the three derived from metastatic sites (SU.86.86, CFPAC1, and CAPAN1) were clustered together. MIA-PaCa2, PANC1, and BxPc3 are cell lines derived from the primary tumor site of three patients (43). The MIA-PaCa2 and PANC1 proteomes were clustered together, as were the BxPc3 and HPDE cell lines. Heat-map visualization facilitated a first exploration of the data set and the identification of several regions or protein clusters of interest. Among them, two clusters containing 34 proteins were shown to be highly expressed in multiple cancer cell lines and the pancreatic juice samples, all with minimal expression in HPDE (Fig. 4B) (supplemental Table S9). This included proteins such as MUC1 (Mucin-1) (8) and

RNASE1 (pancreatic ribonuclease) (60) which have been shown to be elevated in pancreatic cancer and studied previously as pancreatic cancer biomarkers in serum. This prompted us to further examine proteins that are differentially expressed between the cancer cell lines and the HPDE cell line. As an additional note, from Fig. 4B, the pancreatic juice protein expression profile appears more closely associated with that of the metastatic cell lines. In general, a two-sample test for equality of proportions with continuity correction showed that there was a closer association of the pancreatic juice proteome with the combined metastatic cell line proteome than with the combined primary cell line proteome ( $p$  value =  $8.655 \times 10^{-05}$ ).

**Differential Expression of Proteins in Cancer versus Normal Cell Lines**—Normalized spectral counts of the cancer cell lines were compared with those of the normal HPDE cell line as described in the *Experimental Procedures* section. The Pearson correlation coefficient was evaluated for all pairs of the 21 replicates from the seven cell lines using normalized spectral count values (supplemental Table S10). With the exception of replicate 2 from CFPAC1, which showed 0.727 and 0.851 correlation with CFPAC1 replicates 1 and 3, good correlation (Pearson  $r$  ranging from 0.944 to 0.993) was seen among replicates of each cell line (including CFPAC1 replicates 1 and 3) indicating good reproducibility (supplemental Table S10).

Analysis of variance (ANOVA) identified 1293 proteins (each with a minimum number of 10 spectra in at least one cell line),





**FIG. 4. Hierarchical clustering analysis in the seven cell lines and two pancreatic juice pools based on normalized emPAI values for the 3479 total nonredundant proteins identified.** Comparison of z-scores between emPAI values and ELISA concentrations for two control proteins, KLK6 and KLK10 (A). Good correlation was seen ( $R^2 = 0.74$ ). Clustering analysis using Pearson and Ward's minimum variance methods for distance and aggregation (B) depicting the seven cell lines and two pancreatic juice pools on the x axis and proteins on the y axis. Shown is a segment of the resulting dendrogram depicting two clusters of proteins found highly expressed in the cancer cell lines and pancreatic juice (low expression in the normal HPDE cell line). See also [supplemental Table S9](#) for z-score values for each protein in Fig. 4B. KLK6, kallikrein 6; KLK10, kallikrein 10; emPAI, exponentially modified protein abundance index; EC, extracellular; CS, cell surface.

with a statistically significant difference in normalized spectra among the seven cell lines ( $p < 0.05$ ). Based on the criteria described in the *Experimental Procedures*, 483 of these proteins showed  $\geq$ fivefold increase in at least one cancer cell line compared with HPDE. These proteins, along with normalized spectral count values, %CV, fold-change and other details are provided in [supplemental Table S11](#). One-hundred and nineteen proteins further demonstrated  $\geq$ fivefold increase in at least three cancer cell lines in comparison to HPDE, of which 53 proteins showed over 10-fold increase and 18 showed over a 20-fold increase in at least three cancer cell lines. Sixty-three of the 119 proteins were extracellular and cell surface-annotated and are listed in Table III. Pathway analysis show cellular movement, cancer, and cell-to-cell signaling and interaction as the top pathways assigned to this set of

proteins ([supplemental Fig. S4](#)). In addition, 17 of these proteins have been previously shown to be up-regulated in pancreatic cancer in at least four studies (61), and 10 have been shown to be elevated in pancreatic cancer serum in comparison to controls (60, 63–71) (Table III). The unstudied proteins might yield promising new candidate biomarkers for pancreatic cancer.

Many of the proteins were also identified in a comprehensive database of human plasma proteins (72) and five proteins, COPS4 (COP9 signalosome complex subunit 4), PXN (Paxillin), MYO1C (Myosin-1c), GBA (protein similar to Glucosylceramidase), and LMAN2 (Vesicular integral-membrane protein VIP36), were also identified in a recent global genomic analysis of pancreatic cancer by Jones *et al.* (73) as overexpressed in the large majority of pancreatic cancer cases stud-

TABLE III  
 Extracellular and cell surface annotated proteins with over fivefold increase in at least three pancreatic cancer cell lines compared to the HPDE cell line. One hundred and nineteen proteins were expressed over fivefold in at least three cancer cell lines in comparison to the HPDE cell line. Of these 63 were extracellular and cell surface annotated according to Genome Ontology and are listed in this table. See also Supplemental Table XI for entire list of 483 proteins expressed over fivefold in at least one cancer cell line compared to HPDE

Gene	Anova P-Value	Accession	BxPc3			MIA-PaCa2			PANC1			CAPANI			CFPAC1			SUJ.86.86			Previously Studied as a Pancreatic Cancer Serum Biomarker (reference shown)
			% CV	FC	% CV	FC	% CV	FC	% CV	FC	% CV	FC	% CV	FC	% CV	FC	% CV	FC	Asciites <sup>a</sup>	Human Plasma Proteome <sup>b</sup>	
RNASE1	1.1E-16	IP00014048																✓		(60)	
PIGR	3.3E-16	IP00004573																✓			
LOXL2; ENTPD4	7.8E-16	IP00294839																✓			
MUC5AC	1.4E-15	IP00103397	7	125	7	35												✓		(69)	
PRSS2	8.1E-15	IP00011695																✓		(65)	
MUC5B	1.2E-14	IP00902941																✓			
FCGBP	1.4E-14	IP00242956																✓			
MMP13	3.0E-14	IP00021738	8	94														✓			
VWA1	4.2E-14	IP00396383	20	6			74											✓			
CP	5.2E-14	IP00017601	12	7														✓			
C3	9.2E-14	IP00783987	12	44														✓			
SEMA3A	9.3E-14	IP00031510	47	5	5	86	25	13	19	22	16						✓		(63, 64)		
SEMA3C	1.3E-13	IP00019209			14	21	3	51									✓		(63, 70)		
SPOCK1	4.3E-13	IP00005292					26	8									✓				
MMP7	4.4E-13	IP00013400	12	13					2	26							✓			(68)	
SERPINA1	6.0E-13	IP00553177							42	40	11	481	24	9			✓				
PLAT	3.1E-12	IP00019590							22	37	11	304	11	10			✓				
NRP1	1.8E-11	IP00299594					8	21	12	130	9	40	9	49			✓				
ST14	2.5E-11	IP00001922							32	6	8	38	29	14			✓				
LYZ	2.7E-11	IP00019038							13	27	13	19	14	10			✓				
TGM2	6.8E-11	IP00294578	7	32	7	15	11	13	13	111	28	31	17	10			✓				
EPHA2	7.5E-11	IP00021267							15	6				55			✓				
MF2	1.4E-10	IP00029275			30	13			13	12	7	37	19	18			✓				
CTSH	1.9E-10	IP00297487	19	18	5	12			14	48							✓				
NAGLU	2.1E-10	IP00008787			6	24	19	10	16	33	4	10					✓				
AGR2	2.4E-10	IP00007427	25	31	3	73	25	63	13	101	19	56	9	76			✓			(62)	
CSF1	6.4E-10	IP00015881	23	7	8	10			13	29	7	26	27	14			✓				
CFB;C2	7.2E-10	IP00019591							13	13	17	9	12	6			✓				
RNASE4	7.3E-10	IP00029699			8	9	21	6	6	13	17	9	12	6			✓				
PLBD1	7.8E-10	IP00016255	42	5	14	26	21	6	21	7							✓				
COL6A1	8.2E-10	IP00291136			14	121	14	47	8	27	15	55	21	43			✓				
FUCA1	1.1E-09	IP00843910	15	43	15	7			9	20	13	35	48	5			✓				
LRG1	1.8E-09	IP00022417	14	18					19	39	11	34	19	11			✓			(97)	
LTBP3	1.9E-09	IP00073196			32	10	14	33	26	6	18	18	5	22			✓				
PLBD2	1.9E-09	IP00169285	31	7	4				29	32	17	137	11	20	22	43	✓			(66, 67)	
GDF15	2.3E-09	IP00306543			16	9	20	60	42	10	24	22	8				✓				
SIAE	6.4E-09	IP00010949	16	9	20	18	11	28									✓				
DPP7	8.4E-09	IP00296141	16	18					20	14							✓				
TGFB2	9.0E-09	IP00235354							22	35	15	34	12	16			✓				
B3GNT3	9.6E-09	IP00031983	40	8					15	19	5	11					✓				
ITGA2	2.8E-08	IP00013744							10	21	9	24	12	17			✓				
CTSS	3.4E-08	IP00299150	22	22					20	15	14	9					✓				
LTBP1	5.7E-08	IP00302679	43	7					23	9							✓				
BSG	1.1E-07	IP00019906			27	11	13	24	36	10	34	6	24	6			✓				
ACE	1.6E-07	IP00437751			14	15	22	7	24	5	36	6					✓				
LCN2	1.9E-07	IP00299547							13	28	9	31	31	51			✓			(71)	

TABLE III—continued

Gene	Anova P-Value	Accession	BxPc3		MIA-PaCa2		PANC1		CAPAN1		CFPAC1		SU.86.86		Identified in/as..		Previously Studied as a Pancreatic Cancer Serum Biomarker (reference shown)
			% CV	FC	% CV	FC	% CV	FC	% CV	FC	% CV	FC	% CV	FC	Pancreatic Juice	Ascites <sup>a</sup>	
CDH2	1.9E-07	IP100290085	14	8	19	6	28	53	4	18	17	28	18	15	✓	✓	
ITGB1	4.4E-07	IP100217563	12	38	24	8	22	34	4	15	19	15	30	12	✓	✓	✓
MSLN	5.0E-07	IP100025110	38	5	19	48	14	12	11	18	49	22	11	13	✓	✓	✓
SDCBP	6.2E-07	IP100299086	14	26	25	14	38	24	17	14	14	16	49	22	✓	✓	
CXCL1	1.8E-06	IP100013874	23	8	28	6	45	6	5	10	24	38	36	13	✓	✓	
AHSG	5.0E-06	IP100022431	17	13	7	37	14	90	14	20	24	17	38	226	✓	✓	✓
DNASE2	5.2E-06	IP100010348	19	15	17	9	29	9	12	12	43	45	12	41	✓	✓	
CXCL5	5.3E-06	IP100292936	31	5	9	12	6	6	32	22	49	8	38	9	✓	✓	
WFD2	8.1E-06	IP100291488	18	12	14	9	35	12	42	6	24	6	41	29	✓	✓	
NEU1	8.5E-06	IP100029817	20	19	6	235	27	213	34	29	12	5	47	20	✓	✓	
SERPINB9	1.2E-05	IP100032139	20	19	6	235	27	213	34	29	12	5	47	20	✓	✓	
RARRS1	1.4E-05	IP100410240	31	5	9	12	6	6	42	6	24	6	41	29	✓	✓	
PLA2G15	2.7E-05	IP100301459	18	12	14	9	35	12	42	6	24	6	41	29	✓	✓	
CTBS	2.8E-05	IP100007778	20	19	6	235	27	213	34	29	12	5	47	20	✓	✓	
HS3ST1	9.3E-05	IP100021377	20	19	6	235	27	213	34	29	12	5	47	20	✓	✓	
LFNG	1.1E-04	IP100455739	20	19	6	235	27	213	34	29	12	5	47	20	✓	✓	
ENO2	2.0E-02	IP100216171	20	19	6	235	27	213	34	29	12	5	47	20	✓	✓	

<sup>a</sup> Proteome of ascites samples from pancreatic cancer patients (Kosanam H., Makawita S., Diamandis E.P.; unpublished data).

<sup>b</sup> Identification in 12,787 protein containing plasma proteome database (72). FC, fold change between cancer cell line and HPDE; %CV, percent coefficient of variation in normalized spectral counts for triplicates of cell line; P.J, pancreatic juice.

ied (supplemental Table S11). Examination of underexpressed proteins revealed 21 proteins consistently decreased at least fivefold in all six cancer cell lines and 46 consistently decreased in five cancer cell lines in comparison to HPDE (supplemental Table S11, separate tab).

It should be noted that all proteins had to be identified with at least 10 spectra in one cell line to be included in the quantitative analysis described above and proteins with %CVs greater than 50% (for normalized spectral counts in triplicate analysis) were not included. Mean %CVs ranged from 13.2 to 21.8 for the proteins that showed over fivefold increase in the seven cell lines (supplemental Table S11). Percent CVs >50% were only accepted in comparison cell lines (i.e. comparison cell line refers to the normal HPDE cell line when overexpression between the cancer versus HPDE cell line was the focus of analysis, or each of the cancer cell lines when underexpression between the cancer versus HPDE cell lines was the focus of analysis) in which proteins were identified with <10 spectra in the comparison cell line.

*Further Prioritization of Candidates through Integration of Biofluids and Tissue Specificity*—Recent evidence suggests the integration of data from different biological fluids might yield stronger candidates for verification phases of biomarker discovery (15, 16). As such, in addition to the differential protein expression analysis of the cell line CM, we also applied a comparative, or integrative, approach based on overlap of proteins between different biological sources and cellular localization for generation of additional candidates. Of the 488 proteins common to the pancreatic juice and cell lines (Fig. 2B), 235 had been annotated to extracellular and cell surface compartments. One-hundred and nine of these proteins were also identified in the proteome of ascites fluid from three patients with pancreatic adenocarcinoma (Kosanam Kosanam H., Makawita S., Judd B., Newman A., and Diamandis E.P. Mining the malignant ascites proteome for pancreatic cancer biomarkers, unpublished data), and of these, 43 were not identified in the HPDE cell line (supplemental Table S12).

Because there might be pertinent proteins in the pancreatic juice that might not be identified in the CM and vice versa, proteins identified in either proteome that were shown to be highly expressed in or highly specific to the pancreas were also included. To examine tissue specificity, we used microarray, expressed sequence tags (ESTs), and immunohistochemistry data using TiSGeD (39), TiGER (40), Unigene (41), and the Human Protein Atlas (42) databases, respectively. Specifically, we compared our list of identified proteins to 150 proteins reported as specific to pancreas tissue using TiSGeD specificity measure >0.90 (39), 55 “pancreas-restricted” proteins from Unigene (41), 205 proteins preferentially expressed in the pancreas based on the TiGER database, and 198 proteins showing “strong” pancreatic exocrine cell staining and annotated on the Human Protein Atlas. Twenty proteins were common to at least three or more of the databases, of which two proteins, PRSS1 and SPINK1, were identified in the cell

TABLE IV

List of 15 pancreas specific proteins ( $\geq 3$  databases) identified in CM and pancreatic juice. Pancreas-specific proteins as it applies to this table indicates proteins identified in at least three out of the four databases queried. (Unigene, TiSGeD, TiGER and HPA databases) as highly/preferentially expressed in the normal pancreas. HPA, Human Protein Atlas; TiSGeD, Tissue-Specific Genes Database; TiGER, Tissue-specific and Gene Expression and Regulation

Gene	Protein Name	HPA (42)	UniGene (41)	TiGER (40)	TiSGeD (39)	Identified in		Previous shown Elevated in Pancreatic Cancer or Pancreatitis Serum/Plasma (reference shown)
						Pancreatic Juice Proteome	CM Proteome	
<b>CPA1</b>	Carboxypeptidase A1	✓		✓	✓	✓		(82)
<b>PRSS1</b>	Trypsin-1	✓	✓	✓	✓	✓	✓	(74, 75)
<b>CPA1</b>	cDNA FLJ53709, highly similar to Carboxypeptidase A1	✓	✓	✓	✓	✓		(82)
<b>CPA2</b>	Carboxypeptidase A2	✓		✓	✓	✓		(82)
<b>GP2</b>	Isoform Alpha of Pancreatic secretory granule membrane major glycoprotein GP2	✓		✓	✓	✓		(78)
<b>REG1A</b>	Lithostathine-1-alpha		✓	✓	✓	✓		(80)
<b>CTRC</b>	Chymotrypsin-C		✓	✓	✓	✓		
<b>CPB1</b>	Carboxypeptidase B		✓	✓	✓	✓		(76)
<b>GP2</b>	Isoform 1 of Pancreatic secretory granule membrane major glycoprotein GP2	✓		✓	✓	✓		(78)
<b>PNLIP</b>	Pancreatic triacylglycerol lipase		✓	✓	✓	✓		(79, 84)
<b>SYCN</b>	Syncollin		✓	✓	✓	✓		
<b>REG1B</b>	Lithostathine-1-beta		✓	✓	✓	✓		
<b>CLPS</b>	Colipase		✓	✓	✓	✓		(81)
<b>SPINK1</b>	Pancreatic secretory trypsin inhibitor	✓		✓	✓	✓	✓	(83)
<b>PLA2G1B</b>	Phospholipase A2		✓	✓	✓	✓		(77)

line CM, and 15 proteins (which included PRSS1 and SPINK1) were identified in the pancreatic juice proteome (Table IV). Twelve of these proteins have been previously shown to be elevated in serum and plasma of patients with pancreatitis or pancreatic cancer (74–84), leaving CTRC (chymotrypsin C), SYCN (syncollin), and REG1B (Lithostathine-1-beta) (Table IV).

**Candidate Verification in Plasma**—Several candidates identified through the differential protein expression, integration of multiple biological fluids, and tissue specificity analyses, as described above and listed in Tables III, IV and supplemental Table S12 and not previously shown to be increased in plasma and serum from pancreatic cancer patients were verified based on availability of enzyme-linked immunosorbent assays. Specifically, seven proteins were verified, of which five, Anterior Gradient Homolog 2 (AGR2), Olfactomedin-4 (OLFM4), Syncollin (SYCN), Collagen alpha-1(VI) chain (COL6A1), and Polymeric Immunoglobulin Receptor (PIGR) showed a significant increase in plasma from pancreatic cancer patients in our sample set (Fig. 5). Two other proteins (NUCB2, nucleobindin-2; PLAT, tissue-type plasminogen activator) did not show a significant increase in pancreatic cancer plasma (data not shown).

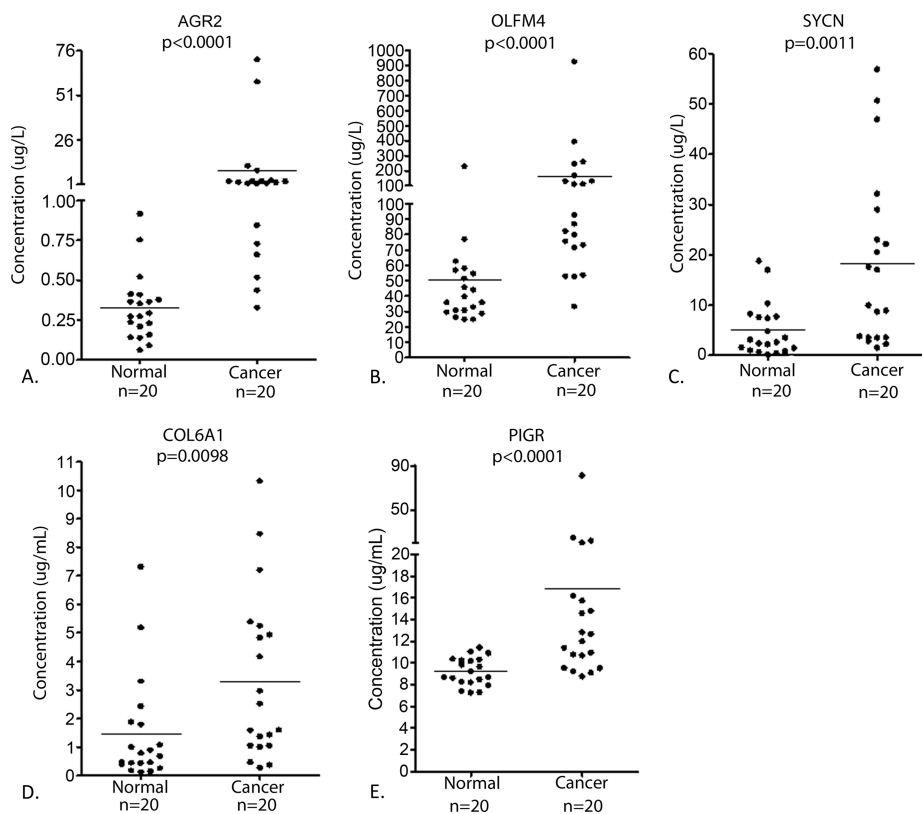
In the CM analysis, AGR2 showed over 10-fold increase in the BxPc3, CAPAN1, CFPAC1, and SU-86–86 cell lines compared with the near normal HPDE cell line (Table III). As well, AGR2 was common to the CM and pancreatic juice proteomes and was identified in the cluster of proteins highly

expressed in many cancer cell lines and pancreatic juice in comparison to HPDE (Fig. 4B). In plasma, AGR2 levels were significantly increased in pancreatic cancer patients ( $p < 0.0001$ ) in comparison to controls (Fig. 5A). Mean and median plasma levels in the pancreatic cancer patients were 8.8  $\mu\text{g/L}$  and 2.1  $\mu\text{g/L}$  whereas mean and median levels in controls were 0.33  $\mu\text{g/L}$  and 0.28  $\mu\text{g/L}$ .

OLFM4 was a protein identified based on the integrated method (supplemental Table S12), and also identified in the cluster shown in Fig. 4B. In the plasma samples, OLFM4 also showed a significant elevation ( $p < 0.0001$ ) in cancer (mean = 161  $\mu\text{g/L}$ , median = 90  $\mu\text{g/L}$ ) in comparison to controls (mean = 51  $\mu\text{g/L}$ , median = 38  $\mu\text{g/L}$ ) (Fig. 5B). SYCN was a protein identified solely in the pancreatic juice samples. It is monospecific to the pancreas based on TiGER, TiSGeD, and Unigene databases (data was unavailable in the Human Protein Atlas) (Table IV). This protein is part of the secretory granule membranes of the exocrine pancreas, and because of its tissue specificity, it was identified for verification. In plasma, SYCN also showed a significant increase in pancreatic cancer patients ( $p = 0.0011$ ; mean cancer = 18.2  $\mu\text{g/L}$ , median cancer = 13.5  $\mu\text{g/L}$ ; mean controls = 5.1  $\mu\text{g/L}$ , median controls = 2.9  $\mu\text{g/L}$ ) (Fig. 5C).

COL6A1 was expressed over 20-fold in all of the cancer cell lines except for the BxPc3 cell line in comparison to the HPDE cell line. Similarly, PIGR was expressed over 20-fold in three of the cancer cell lines (Table III). Both proteins showed a significant increase in pancreatic cancer plasma in our pre-

**FIG. 5. Verification of AGR2 (A), OLFM4 (B), SYCN (C), COL6A1 (D) and PIGR (E) in plasma from pancreatic cancer patients and healthy controls of similar age and sex.** Plasma concentrations of the proteins were measured by ELISA. Mean values are indicated by a horizontal line and  $p$  values were calculated using the Mann-Whitney  $U$  test. Number of subjects = 20 pancreatic cancer and 20 controls.



liminary analysis ( $p = 0.0098$ ; mean cancer = 3.3 mg/L, median cancer = 2.1 mg/L; mean controls = 1.5 mg/L, median controls = 0.73 mg/L for COL6A1 and  $p < 0.0001$ ; mean cancer = 16.8 mg/L, median cancer = 12.3 mg/L; mean controls = 9.2 mg/L, median controls = 8.96 mg/L for PIGR) (Fig. 5D, 5E).

At present, CA19.9 is the most widely used pancreatic cancer biomarker and CA19.9 levels were also assessed in our screening set of plasma samples. Because we used mostly pancreatic cancer patients with established disease, the AUC of CA19.9 alone in this set of samples was 0.97 (Fig. 6A). The AUCs for each one of our newly discovered five biomarkers alone ranged from 0.95 to 0.74 (Fig. 6A). Each one of our biomarkers was able to improve CA19.9 AUC to 1.00 (CA19.9 + AGR2, data not shown) or to 0.98 (Fig. 6B). The scatterplot of Fig. 6C shows that AGR2 was able to identify two patients with pancreatic cancer, which did not show an CA19.9 elevation (Fig. 6D). Fig. 6E shows that our panel of five newly discovered biomarkers is slightly more discriminatory than CA19.9 alone in this sample set.

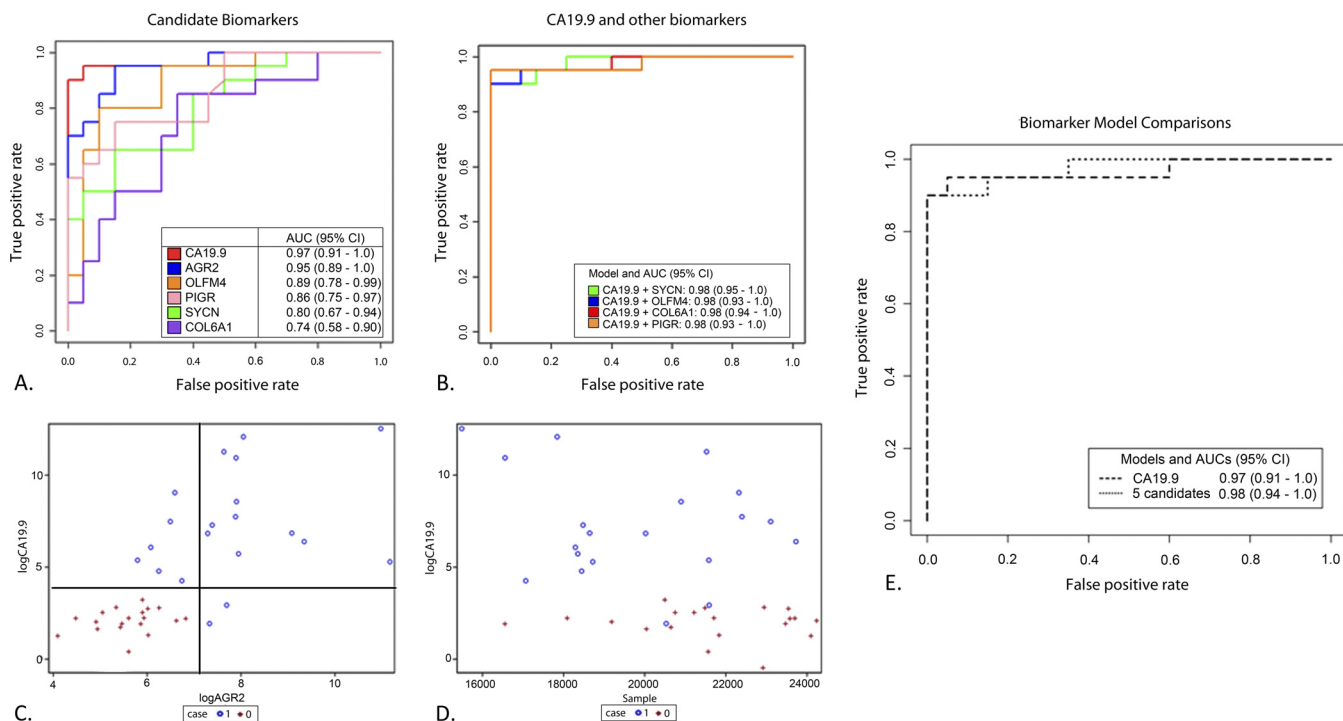
#### DISCUSSION

The proteome presented in this article, is, to our knowledge one of the largest and most comprehensive proteomes to date for pancreatic cancer-related biological fluids in a single study.

Although many notable differences exist, the genomic and transcriptional make-up of cancer cell lines have been shown

to recapitulate many salient aspects of primary tumors (43, 44, 85, 86). In addition, the identification of many known biomarkers in the conditioned media of cancer cell lines for numerous cancer sites, supports that it is a good source to mine (37, 87). Previously, our group has characterized the CM of breast, ovarian, prostate, and lung cancer-related cell lines using three to four cell lines per cancer site (27–30). Using an LTQ mass spectrometer, 1139, 1830, 2124, and 2039 proteins were identified with at least one peptide in the breast (28), lung (29), prostate (30), and ovarian cancer (27) analyses, respectively. Given the vast heterogeneity of cancer, from our previous work we concluded that a larger number of cell lines per cancer site, as well as study of proximal biological fluids from patients, might provide a more complete picture of disease heterogeneity and the tumor-host interface, thereby facilitating the identification of stronger candidates for verification.

In the present study, we applied such an approach to pancreatic cancer. By using 2D LC-MS/MS, we characterized the proteomes of conditioned media from six pancreatic cancer cell lines, one near normal pancreatic ductal epithelial cell line and six pancreatic juice samples in two pools. All experiments were performed in triplicate and multiple search engines (MASCOT and X!Tandem), which employ different search algorithms, were used for protein identification. Previously, it has been reported that use of multiple search engines results in increased confidence of the proteins identified (88,



**FIG. 6. Receiver Operating Characteristic curve analysis for CA19.9 and candidates and scatter plot of CA19.9 and AGR2.** AUC (area under curve) is given at 95% confidence intervals. AUC of CA19.9, AGR2, OLFM4, SYCN, COL6A1, and PIGR is depicted individually (A). CA19.9 performs best individually in this sample set of 20 cancer and 20 controls (AUC of 0.97). The combination of each candidate with CA19.9 shows improved AUC to CA19.9 alone (B). The combination of AGR2 and CA19.9 produced a complete separation of cases from controls in this sample set and was not modeled. Instead, the combination of AGR2 and CA19.9 is depicted in the log-transformed scatterplot showing complete separation of cases (blue circles) from controls (red crosses) (C) in comparison to the separation of cases from controls for CA19.9 alone (D). Note that two cancer samples which had been missed by CA19.9 had elevated levels of AGR2. The combination of AGR2, OLFM4, SYCN, COL6A1, and PIGR in panel shows an improvement to the AUC of CA19.9 alone (E).

89). In addition, only proteins identified with multiple peptides ( $\geq 2$  peptides) were used in the analysis. With such strategies, we identified 3324 nonredundant proteins in the CM of the seven cell lines and 648 proteins in the pancreatic juice. In total, 3479 nonredundant proteins were identified.

In the first part of the study, an increase in protein yield of  $\sim 50\%$  was achieved by applying a prefractionation strategy that was tailored to the SCX elution profile. SCX was the first dimension of fractionation in our multidimensional approach. Different modes of fractionation from isoelectric focusing (IEF) to SDS-PAGE fractionation and SCX have been previously compared, with different studies reporting different methods as the most effective, when coupled with MS analysis (90–93). Fractionation of complex samples before MS analysis minimizes sample complexity and allows deeper mining into the proteome, thereby achieving increased coverage of proteins. A corollary of increased fractionation is decreased throughput. In the present study, reduced gradient times during the second dimension of separation (reverse-phase) helped to keep any increase in analysis time to a minimum.

Not all proteins identified in shotgun proteomics-driven discovery approaches will be suitable as serological biomarkers, and one of the challenges in the field is in the selection of the most promising candidates for further investigation. In the

present study, we used three strategies: (1) semiquantitative analysis through label-free protein quantification between the cancer and normal cell lines, (2) integrative and comparative analysis of the multiple biofluids, and (3) tissue specificity analysis. Label-free approaches typically employ chromatographic ion intensity-based methods or spectral count-based means to obtain relative quantification of proteins among LC-MS/MS run samples (50, 52). Further approximations of absolute protein abundance can be obtained through reported indices such as emPAI and absolute protein expression (APEX) (94, 95). Normalized spectral counts have been reported previously to be reliable indicators of protein abundance in studies comparing different label-free methods, and strong correlation between spectral counts and protein abundance have been shown (51). When restricting analysis to proteins identified with five or more spectra, results comparable to label-based approaches are obtainable (96). In the present study, this method was used for relative quantification between the cancer cell lines and the HPDE cell line.

Using the criteria outlined in the *Experimental Procedures*, 119 proteins were found to be consistently expressed over fivefold in at least three cancer cell lines (compared with the HPDE cell line), 63 of which were extracellular or cell surface annotated (Table III). Included in this list were many proteins

previously shown to be up-regulated in pancreatic cancer. For instance, the protein GDF15, also known as macrophage inhibitory cytokine 1 (MIC1), showed >10-fold increase in the CAPAN1, CFPAC1, PANC1, and SU.86.86 cell lines. Increased GDF15 mRNA and protein levels have been shown previously in pancreatic cancer tissue in comparison to adjacent normal controls (66) and evaluation of this protein in serum has also shown it to have diagnostic potential (67). Similarly, neutrophil gelatinase-associated lipocalin (LCN2) (71), matrix metalloproteinase 7 (MMP7) (68), complement component 3 (C3) (63, 70), and leucine-rich alpha-2-glycoprotein (LRG1) (97) have been reported to be elevated in serum of pancreatic cancer patients, whereas mesothelin (MSLN) (98), tissue-type plasminogen activator (PLAT) (99), C-X-C motif chemokine 5 (CXCL5) (100) and other proteins highlighted in Table III have been shown to be up-regulated in pancreatic cancer or pancreatic neoplasia at the level of tissue and/or mRNA. In addition, other proteins, such as Lysyl oxidase-like 2 (LOXL2), have been associated with pancreatic cancer pathology and tumorigenicity. For instance, when inhibited, LOXL2 has been shown to result in reduced viability of pancreatic cancer cells and their increased sensitivity to chemotherapy (101).

Identification of these proteins provides some credence to our label-free discovery approach; however proteomic comparisons between nonmalignant and malignant biological sources are limited by the possibility that the observed differences might be because of many factors, not solely because of differences in tumorigenic potential alone. This was demonstrated as two of the seven proteins verified did not show a significant increase in plasma from pancreatic cancer patients. These proteins, PLAT and NUCB2, were expressed over fivefold in three and one of the pancreatic cancer cell lines, respectively; however this failed to translate into our plasma analysis (individual value plots for these two proteins are not shown; fold change in CM is available in [supplemental Table S11](#)).

We further investigated three other proteins, AGR2, PIGR, and COL6A1, which showed over fivefold increase in four, three, and five cancer cell lines, respectively (Table III). Our analysis of these proteins in human plasma also showed a significant increase in pancreatic cancer patients. Except for AGR2, to the best of our knowledge, neither of these proteins has previously been studied in sera/plasma of pancreatic cancer patients. AGR2 is an ortholog of the *Xenopus laevis* protein XAG-2, which is a protein shown to play a role in ectodermal patterning (102). The function of AGR2 in normal human states is largely unknown; however in human cancers, AGR2 has been associated with several cancer types (103–105) and recently, increased AGR2 levels were reported in pancreatic juice (62). In this latter study, Chen *et al.*, used quantitative proteomics to profile pancreatic juice samples from pancreatic intraepithelial neoplasia (PanIN) patients in comparison to controls, and AGR2 was one of the proteins

with an over twofold increase in PanIN-stage III. Although Chen *et al.*, found diagnostic relevance for AGR2 in pancreatic juice, their analysis in six paired serum and pancreatic juice samples from PanIN patients found no correlation between serum and pancreatic juice AGR2 levels. Further analysis by this group in serum of nine pancreatic cancer and nine cancer-free controls showed no significant difference in AGR2 levels as well (62). Despite this, given that AGR2 was highly elevated in the majority of cancer cell line CM (based on spectral counting in this study) as well as its identification in pancreatic juice, we tested its levels in our screening set of plasma samples by ELISA and found a significant elevation in AGR2 levels in pancreatic cancer plasma *versus* controls (Fig 5A). AGR2 has been previously shown to play a role in invasion and metastasis (103, 106, 107), and it might be that elevated levels of this protein occur in blood in the later stages of pancreatic cancer; however our initial results warrant further evaluation of this protein in plasma and sera in larger sample sets.

PIGR has been shown previously through MRM to be increased in endometrial cancer tissue homogenates (108); however it has not been studied in clinical samples from many other cancer sites. In the present study we demonstrate its significant increase in pancreatic cancer plasma. COL6A1 is an important component of microfibrillar network formation, associating closely with basement membranes in many tissues. It is an extracellular matrix protein also found in stromal tissue (109). Mutations in this gene play a role in muscular disorders and differential COL6A1 gene expression has been associated with astrocytomas (110, 111); however it has not been studied in pancreatic cancer and was found to be significantly increased in our preliminary assessment in plasma. Taken together, the increased levels of these proteins in pancreatic cancer plasma demonstrate the utility of our label-free differential protein quantification approach to identify proteins relevant for study as potential serological biomarkers of pancreatic cancer.

The identification of cancer-derived protein alterations through integration of different biological sources is another area of interest in cancer proteomics and the integrative mining of multiple biological fluids might result in the identification of relevant candidates (15). For instance, in a recent analysis done in our laboratory (16), which compared the proteins and genes identified in six publications chosen arbitrarily to represent various biological sources and both proteomic and genomic data pertaining to ovarian cancer (two cell line CM studies, two ascites, one tissue proteomics study, and one microarray study), no proteins were found common to all six; however two proteins were found common to four of the studies. The proteins identified were WAP four-disulfide core domain protein two precursor (HE4) and GRN (granulin). Both have been implicated in ovarian cancer and WAP four-disulfide core domain protein two precursor is a recently FDA-approved ovarian cancer biomarker (112). In this respect, we

also looked at proteins common to the cancer CM and pancreatic juice for identification of further candidates. These proteins were also compared with a pancreatic cancer ascites proteome (Kosanam H., Makawita S., Judd B., Newman A., Diamandis E.P. Mining the malignant ascites proteome for pancreatic cancer biomarkers. (unpublished data)) for additional filtering. Most, if not all, current biomarkers, such as prostate specific antigen for prostate cancer, CA125 for ovarian cancer, hCG for testicular cancer, etc. are secreted and shed proteins. Consequently, special focus was given to extracellular and cell surface proteins, as they have the highest likelihood of entering into the circulation (14, 113). Focus was also given to proteins highly or specifically expressed in the pancreas. If a protein is only expressed in one tissue in healthy individuals, that tissue is likely the only contributor to endogenous serum levels of that protein. As such, increasing serum contributions of such a protein by a growing tumor might be more easily detected. Furthermore, many good currently used cancer biomarkers, such as prostate specific antigen, are highly specific to one tissue (38).

Interestingly, the great majority of pancreas-specific proteins (as denoted by several databases) were unique to the pancreatic juice and were not identified in the cell lines (Table IV). Similarly, the KEGG pancreatic secretion pathway was overrepresented in the pancreatic juice proteome (supplemental Fig. S2). In the exocrine pancreas, acinar cells are responsible for secretion of enzymes (zymogens) whereas ductal cells secrete primarily an alkaline fluid (114, 115). Although the majority of pancreatic cancers are ductal adenocarcinomas with pancreatic ductal cell-like properties, the cell of origin of these cancers is still unclear (116, 117). Previously, it has been shown that acinar cells, once having undergone a transformation to ductlike cells, show a reduced secretion of zymogens (117). The lack of pancreas specific proteins (enzymes, zymogens, etc.) in the cell line CM might likely reflect the ductal-like nature of the cell lines, whereas the presence of such proteins in the pancreatic juice might be reflective its acinar cell contributions.

Among the proteins common to all three biological fluids (supplemental Table S12), were several proteins shown previously to be increased in the serum of pancreatic cancer patients and studied as pancreatic cancer biomarkers, such as MUC1 (mucin 1) and CEACAM5 (CEA) (6–8). Two proteins identified through integration of the biological fluids and tissue specificity analysis, not previously assessed in the serum/plasma of pancreatic cancer patients, OLFM4 and SYCN, were verified in this study. OLFM4 has been shown to promote proliferation in the PANC1 cell line by Kobayashi *et al.* (118). Its mRNA levels were shown to be elevated in five cancerous, *versus* noncancerous pancreatic tissue samples in the same study, and OLFM4 serum protein levels have shown potential diagnostic utility for gastric cancer (119). In this study, OLFM4 showed over fivefold expression in the CA-PAN1 cell line in comparison to the HPDE cell line. It was also

identified by us in the pancreatic juice and ascites and our preliminary assessment shows that it is significantly increased in plasma from pancreatic cancer patients (Fig 5B). Syncollin is a zymogen granule protein specific to the pancreas and is believed to play a role in the concentration and/or efficient maturation of zymogens (120). Syncollin has been previously identified through mass spectrometry in human pancreatic juice and shown to be increased in the quantitative proteomic analysis of plasma from a murine pancreatic cancer model (22, 121); however little is known about the role of this pancreas-specific protein in pancreatic cancer and other pathologies. Our data show that it is significantly elevated in human pancreatic cancer plasma through ELISA (Fig. 5C).

The growing consensus in this field is toward the development of panels of biomarkers, as the combined assessment of multiple molecules can result in increased sensitivity and specificity, in comparison to the assessment of molecules individually. In the present study, this was demonstrated preliminarily to be true, as the combined assessment of AGR2, OLFM4, PIGR, SYCN, and COL6A1 showed improved AUC, compared with CA19.9 alone. CA19.9 has reported sensitivity and specificity values between 70 and 90% (median ~79%) and 68 and 91% (median ~82%), respectively, for detection of pancreatic cancer (note: sensitivity decreases to ~55% in early-stage disease and CA19.9 is often undetectable in many asymptomatic individuals; specificity decreases with benign disease) (3). In the present study, CA19.9 showed a very high AUC (0.97), likely because the cancer plasma samples used were from patients with established (primarily late-stage) pancreatic ductal adenocarcinoma. We used such samples as the goal in the present study was to determine the utility of our approach to identify proteins increased in serum/plasma of pancreatic cancer patients and to preliminarily examine if the candidates are informative. Our marker panel requires further validation with samples that have low or normal CA19.9 values and include patients with early-stage disease, as well as with benign abdominal pathologies. Of considerable note was the ability of one of our proteins (AGR2) to supplement CA19.9 and raise its sensitivity, in this small series, from approx. 90% to 100%, at 100% specificity (Fig. 6C, 6D).

Three of the five proteins increased in pancreatic cancer plasma (AGR2, PIGR, and OLFM4) were also identified in relevant clusters through hierarchical clustering analysis. emPAI is another means of label-free protein quantification (94), and the identification of these three proteins through emPAI-based quantification, and several other proteins, that were also identified through spectral counting, is not unexpected. It aids in further corroborating our results. Recently, Wu *et al.* (32), used emPAI values of proteins normalized through z-scores for pathway-based biomarker discovery as part of their study of 23 human cancer cell lines. In the present study, we used normalized emPAI values of proteins to gain a preliminary understanding of the data set through hierarchical clustering analysis. The six cancer cell lines chosen for analysis in



the study are well characterized and already highly studied. They contain many of the major genetic aberrations present in pancreatic cancer such as mutations in *Kras*, *SMAD4*, *CD16*, and *TP53* (43, 44). Interestingly, the cancer cell lines derived from metastatic sites (SU-86-86, CFPAC1, and CAPAN1) were clustered together, whereas MIA-PaCa2 and PANC1, which are cell lines derived from a primary tumor site were clustered together, as were the BxPc3 and HPDE cell lines. The pancreatic juice proteome was more closely associated with the metastatic cell lines than the primary cell lines. This might be because the patients had progressed to late stage (metastatic) disease at the time the pancreatic juice was obtained. BxPc3 is a cancer cell line derived from a primary tumor site and HPDE is a widely used surrogate for normal pancreatic ductal epithelial cells. Incidentally these two cell lines were the only ones with wild-type *Kras* expression (43), a gene that is mutated in the vast majority (>90%) of pancreatic cancers; however firm conclusions cannot be drawn regarding the clustering without further investigation. Nonetheless, identification of three of the five proteins verified in this study render the proteins identified in relevant clusters through normalized emPAI values a potentially viable means for the generation of biologically relevant leads.

Pancreatic cancer bodes one of the lowest five-year survival rates (<5%) of all cancer types (1). This is largely associated with the existence of locally advanced or metastatic disease in the majority of patients at the time of diagnosis. Genomic sequencing studies reveal that a broad window might exist for the detection of pancreatic cancer between the initial stages of tumor development and dissemination to secondary sites (122). Here, we present the proteomic analysis of pancreatic cancer-related cell lines and pancreatic juice for the identification of novel diagnostic leads. Label-free protein quantification methods revealed a group of proteins differentially expressed in pancreatic cancer. Included within this group were numerous proteins previously studied as pancreatic cancer biomarkers and associated with pancreatic cancer pathology. Further candidates were generated through integrative analysis of multiple biological fluids and tissue specificity analysis. Through a preliminary assessment, five proteins (AGR2, OLFM4, PIGR, SYCN, and COL6A1) were shown to be significantly increased in plasma from pancreatic cancer patients. Appropriate validation of potential biomarkers requires the use of clearly defined clinical specimen, appropriate controls and a large number of samples (preclinical, early and late-stage, benign disease, healthy controls) (123). Our preliminary assessment warrants further validation of these five proteins in larger cohorts of samples (early and late-stage pancreatic cancer, benign disease and healthy controls) as well as consideration of these proteins in the development of biomarker panels for pancreatic cancer.

The current state of cancer proteomics boasts a large number of discovery studies resulting in the generation of many potential diagnostic and therapeutic leads; however due in

part to a lack of parallel high-throughput or multiplexed quantitative technologies, subsequent verification and validation of these leads is lagging. In this regard, the proteomic data-set presented in this study, when combined with existing repositories or compendiums (61, 124), might also aid in further prioritizing candidates for future diagnostic and therapeutic applications.

**Acknowledgments**—We thank Dr. Irvin Bromberg and Dr. Adrian Pasulescu for development of in-house software used for data analysis. We also thank Jana Lee at Proteome Software, Maria Pavlou, George Karagiannis, Dr. Vathany Kulasingam, and Dr. Ivan Blasutig for helpful discussions, Elissa Brown and Apostolos Dimitromanolakis for statistical assistance and Antoninus Soosaipillai for technical assistance with ELISAs.

§ This article contains [supplemental Figs. S1 to S4 and Tables S1 to S12](#).

§§ To whom correspondence should be addressed: Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 6th Floor, Room 6-201, Box 32, 60 Murray Street, Toronto, Ontario, Canada M5T 3L9. Tel.: (416) 586-8443; Fax: (416) 619-5521; E-mail: ediamandis@mtsinai.on.ca.

#### REFERENCES

- Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., and Thun, M. J. (2009) Cancer statistics, 2009. *CA Cancer J. Clin.* **59**, 225–249
- Maitra, A., and Hruban, R. H. (2008) Pancreatic cancer. *Annu. Rev. Pathol.* **3**, 157–188
- Goonetilleke, K. S., and Siriwardena, A. K. (2007) Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur. J. Surg. Oncol.* **33**, 266–270
- Magnani, J. L., Steplewski, Z., Koprowski, H., and Ginsburg, V. (1983) Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin. *Cancer Res.* **43**, 5489–5492
- Marrelli, D., Caruso, S., Pedrazzani, C., Neri, A., Fernandes, E., Marini, M., Pinto, E., and Roviello, F. (2009) CA19-9 serum levels in obstructive jaundice: clinical value in benign and malignant conditions. *Am. J. Surg.* **198**, 333–339
- Nazli, O., Bozdogan, A. D., Tansug, T., Kir, R., and Kaymak, E. (2000) The diagnostic importance of CEA and CA 19-9 for the early diagnosis of pancreatic carcinoma. *Hepatogastroenterology* **47**, 1750–1752
- Tsavaris, N., Kosmas, C., Papadoniou, N., Kopteridis, P., Tsigritis, K., Dokou, A., Sarantonis, J., Skopelitis, H., Tzivras, M., Gennatas, K., Polyzos, A., Papastratis, G., Karatzas, G., and Papalambros, A. (2009) CEA and CA-19.9 serum tumor markers as prognostic factors in patients with locally advanced (unresectable) or metastatic pancreatic adenocarcinoma: a retrospective analysis. *J. Chemother.* **21**, 673–680
- Gold, D. V., Modrak, D. E., Ying, Z., Cardillo, T. M., Sharkey, R. M., and Goldenberg, D. M. (2006) New MUC1 serum immunoassay differentiates pancreatic cancer from pancreatitis. *J. Clin. Oncol.* **24**, 252–258
- Ringel, J., and Löhner, M. (2003) The MUC gene family: their role in diagnosis and early detection of pancreatic cancer. *Mol. Cancer* **2**, 9
- Rückert, F., Pilarsky, C., and Grützmann, R. (2010) Serum tumor markers in pancreatic cancer—recent discoveries. *Cancers* **2**, 1107–1124
- Robin, X., Turck, N., Hainard, A., Lisacek, F., Sanchez, J. C., and Müller, M. (2009) Bioinformatics for protein biomarker panel classification: what is needed to bring biomarker panels into in vitro diagnostics? *Expert Rev. Proteomics* **6**, 675–689
- Tanase, C. P., Neagu, M., Albuлесcu, R., and Hinescu, M. E. (2010) Advances in pancreatic cancer detection. *Adv. Clin. Chem.* **51**, 145–180
- Yurkovetsky, Z. R., Linkov, F. Y., D, E. M., and Lokshin, A. E. (2006) Multiple biomarker panels for early detection of ovarian cancer. *Future Oncol.* **2**, 733–741
- Kulasingam, V., and Diamandis, E. P. (2008) Strategies for discovering novel cancer biomarkers through utilization of emerging technologies.

- Nat. Clin. Pract. Oncol.* **5**, 588–599
15. Hanash, S., and Taguchi, A. (2010) The grand challenge to decipher the cancer proteome. *Nat. Rev. Cancer* **10**, 652–660
  16. Kulasingam, V., Pavlou, M. P., and Diamandis, E. P. (2010) Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. *Nat. Rev. Cancer* **10**, 371–378
  17. Rifai, N., Gillette, M. A., and Carr, S. A. (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24**, 971–983
  18. Sedlaczek, P., Frydecka, I., Gabryś, M., Van Dalen, A., Einarsson, R., and Harlozińska, A. (2002) Comparative analysis of CA125, tissue polypeptide specific antigen, and soluble interleukin-2 receptor alpha levels in sera, cyst, and ascitic fluids from patients with ovarian carcinoma. *Cancer* **95**, 1886–1893
  19. Chen, R., Pan, S., Cooke, K., Moyes, K. W., Bronner, M. P., Goodlett, D. R., Aebersold, R., and Brentnall, T. A. (2007) Comparison of pancreas juice proteins from cancer versus pancreatitis using quantitative proteomic analysis. *Pancreas* **34**, 70–79
  20. Chen, R., Pan, S., Yi, E. C., Donohoe, S., Bronner, M. P., Potter, J. D., Goodlett, D. R., Aebersold, R., and Brentnall, T. A. (2006) Quantitative proteomic profiling of pancreatic cancer juice. *Proteomics* **6**, 3871–3879
  21. Farina, A., Dumonceau, J. M., Frossard, J. L., Hadengue, A., Hochstrasser, D. F., and Lescuyer, P. (2009) Proteomic analysis of human bile from malignant biliary stenosis induced by pancreatic cancer. *J. Proteome Res.* **8**, 159–169
  22. Gronborg, M., Bunkenborg, J., Kristiansen, T. Z., Jensen, O. N., Yeo, C. J., Hruban, R. H., Maitra, A., Goggins, M. G., and Pandey, A. (2004) Comprehensive proteomic analysis of human pancreatic juice. *J. Proteome Res.* **3**, 1042–1055
  23. Ke, E., Patel, B. B., Liu, T., Li, X. M., Haluszka, O., Hoffman, J. P., Ehya, H., Young, N. A., Watson, J. C., Weinberg, D. S., Nguyen, M. T., Cohen, S. J., Meropol, N. J., Litwin, S., Tokar, J. L., and Yeung, A. T. (2009) Proteomic analyses of pancreatic cyst fluids. *Pancreas* **38**, e33–42
  24. Rosty, C., Christa, L., Kuzdzal, S., Baldwin, W. M., Zahurak, M. L., Carnot, F., Chan, D. W., Canto, M., Lillemoe, K. D., Cameron, J. L., Yeo, C. J., Hruban, R. H., and Goggins, M. (2002) Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein 1 as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. *Cancer Res.* **62**, 1868–1875
  25. Tian, M., Cui, Y. Z., Song, G. H., Zong, M. J., Zhou, X. Y., Chen, Y., and Han, J. X. (2008) Proteomic analysis identifies MMP-9, DJ-1 and A1BG as overexpressed proteins in pancreatic juice from pancreatic ductal adenocarcinoma patients. *BMC Cancer* **8**, 241
  26. Zhou, L., Lu, Z., Yang, A., Deng, R., Mai, C., Sang, X., Faber, K. N., and Lu, X. (2007) Comparative proteomic analysis of human pancreatic juice: methodological study. *Proteomics* **7**, 1345–1355
  27. Gunawardana, C. G., Kuk, C., Smith, C. R., Batruch, I., Soosaipillai, A., and Diamandis, E. P. (2009) Comprehensive analysis of conditioned media from ovarian cancer cell lines identifies novel candidate markers of epithelial ovarian cancer. *J. Proteome Res.* **8**, 4705–4713
  28. Kulasingam, V., and Diamandis, E. P. (2007) Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. *Mol. Cell Proteomics* **6**, 1997–2011
  29. Planque, C., Kulasingam, V., Smith, C. R., Reckamp, K., Goodglick, L., and Diamandis, E. P. (2009) Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines. *Mol. Cell Proteomics* **8**, 2746–2758
  30. Sardana, G., Jung, K., Stephan, C., and Diamandis, E. P. (2008) Proteomic analysis of conditioned media from the PC3, LNCaP, and 22Rv1 prostate cancer cell lines: discovery and validation of candidate prostate cancer biomarkers. *J. Proteome Res.* **7**, 3329–3338
  31. Feng, X. P., Yi, H., Li, M. Y., Li, X. H., Yi, B., Zhang, P. F., Li, C., Peng, F., Tang, C. E., Li, J. L., Chen, Z. C., and Xiao, Z. Q. (2010) Identification of biomarkers for predicting nasopharyngeal carcinoma response to radiotherapy by proteomics. *Cancer Res.* **70**, 3450–3462
  32. Wu, C. C., Hsu, C. W., Chen, C. D., Yu, C. J., Chang, K. P., Tai, D. I., Liu, H. P., Su, W. H., Chang, Y. S., and Yu, J. S. (2010) Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Mol. Cell Proteomics* **9**, 1100–1117
  33. Xue, H., Lü, B., Zhang, J., Wu, M., Huang, Q., Wu, Q., Sheng, H., Wu, D., Hu, J., and Lai, M. (2010) Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. *J. Proteome Res.* **9**, 545–555
  34. Grønborg, M., Kristiansen, T. Z., Iwahori, A., Chang, R., Reddy, R., Sato, N., Molina, H., Jensen, O. N., Hruban, R. H., Goggins, M. G., Maitra, A., and Pandey, A. (2006) Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Mol. Cell Proteomics* **5**, 157–171
  35. Mauri, P., Scarpa, A., Nascimbeni, A. C., Benazzi, L., Parmagnani, E., Mafficini, A., Della Peruta, M., Bassi, C., Miyazaki, K., and Sorio, C. (2005) Identification of proteins released by pancreatic cancer cells by multidimensional protein identification technology: a strategy for identification of novel cancer markers. *FASEB J.* **19**, 1125–1127
  36. Schiarea, S., Solinas, G., Allavena, P., Scigliuolo, G. M., Bagnati, R., Fanelli, R., and Chiabrando, C. (2010) Secretome analysis of multiple pancreatic cancer cell lines reveals perturbations of key functional networks. *J. Proteome Res.* **9**, 4376–4392
  37. Kulasingam, V., and Diamandis, E. P. (2008) Tissue culture-based breast cancer biomarker discovery platform. *Int. J. Cancer* **123**, 2007–2012
  38. Björler, E., Lindskog, C., Oksvold, P., Linné, J., Kampf, C., Hober, S., Uhlén, M., and Ponten, F. (2008) A web-based tool for in silico biomarker discovery based on tissue-specific protein profiles in normal and cancer tissues. *Mol. Cell Proteomics* **7**, 825–844
  39. Xiao, S. J., Zhang, C., Zou, Q., and Ji, Z. L. (2010) TiSGeD: a database for tissue-specific genes. *Bioinformatics* **26**, 1273–1275
  40. Liu, X., Yu, X., Zack, D. J., Zhu, H., and Qian, J. (2008) TIGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics* **9**, 271
  41. Pontius J. U., Wagner L., and Schuler G. D. (2003) UniGene: a unified view of the transcriptome. The NCBI Handbook, pp. 21.21–21.12, National Center for Biotechnology Information, Bethesda, MD.
  42. Pontén, F., Jirström, K., and Uhlen, M. (2008) The Human Protein Atlas—a tool for pathology. *J. Pathol.* **216**, 387–393
  43. Deer, E. L., González-Hernández, J., Coursens, J. D., Shea, J. E., Ngatia, J., Scaife, C. L., Firpo, M. A., and Mulvihill, S. J. (2010) Phenotype and genotype of pancreatic cancer cell lines. *Pancreas* **39**, 425–435
  44. Sipos, B., Möser, S., Kalthoff, H., Török, V., Löhr, M., and Klöppel, G. (2003) A comprehensive characterization of pancreatic ductal carcinoma cell lines: towards the establishment of an in vitro research platform. *Virchows Arch.* **442**, 444–452
  45. Furukawa, T., Duguid, W. P., Rosenberg, L., Viallet, J., Galloway, D. A., and Tsao, M. S. (1996) Long-term culture and immortalization of epithelial cells from normal adult human pancreatic ducts transfected by the E6E7 gene of human papilloma virus 16. *Am. J. Pathol.* **148**, 1763–1770
  46. Sedmak, J. J., and Grossberg, S. E. (1977) A rapid, sensitive, and versatile assay for protein using Coomassie brilliant blue G250. *Analyt. Biochem.* **79**, 544–552
  47. Itzhaki, R. F., and Gill, D. M. (1964) A Micro-Biuret Method for Estimating Proteins. *Analyt. Biochem.* **9**, 401–410
  48. Caraux, G., and Pinloche, S. (2005) PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* **21**, 1280–1281
  49. Meunier, B., Dumas, E., Picot, I., Béchet, D., Hébraud, M., and Hocquette, J. F. (2007) Assessment of hierarchical clustering methodologies for proteomic data mining. *J. Proteome Res.* **6**, 358–366
  50. Bachi, A., and Bonaldi, T. (2008) Quantitative proteomics as a new piece of the systems biology puzzle. *J. Proteomics* **71**, 357–367
  51. Liu, H., Sadygov, R. G., and Yates, J. R., 3rd (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201
  52. Zhu, W., Smith, J. W., and Huang, C. M. Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* 2010: 840518, 2010
  53. Luo, L. Y., Soosaipillai, A., Grass, L., and Diamandis, E. P. (2006) Characterization of human kallikreins 6 and 10 in ascites fluid from ovarian cancer patients. *Tumour Biol.* **27**, 227–234
  54. Shaw, J. L., and Diamandis, E. P. (2007) Distribution of 15 human kallikreins in tissues and biological fluids. *Clin. Chem.* **53**, 1423–1432
  55. Choi, H., and Nesvizhskii, A. I. (2008) False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J. Proteome Res.* **7**, 47–50

56. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214
57. Higdon, R., Hogan, J. M., Van Belle, G., and Kolker, E. (2005) Randomized sequence databases for tandem mass spectrometry peptide and protein identification. *Omics* **9**, 364–379
58. Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* **28**, 27–30
59. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucl. Acids Res.* **38**, D355–360
60. Reddi, K. K., and Holland, J. F. (1976) Elevated serum ribonuclease in patients with pancreatic cancer. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 2308–2310
61. Harsha, H. C., Kandasamy, K., Ranganathan, P., Rani, S., Ramabadrans, S., Gollapudi, S., Balakrishnan, L., Dwivedi, S. B., Telkicherla, D., Selvan, L. D., Goel, R., Mathivanan, S., Marimuthu, A., Kashyap, M., Vizza, R. F., Mayer, R. J., Decaprio, J. A., Srivastava, S., Hanash, S. M., Hruban, R. H., and Pandey, A. (2009) A compendium of potential biomarkers of pancreatic cancer. *PLoS Med.* **6**, e1000046
62. Chen, R., Pan, S., Duan, X., Nelson, B. H., Sahota, R. A., de Rham, S., Kozarek, R. A., McIntosh, M., and Brentnall, T. A. (2010) Elevated level of anterior gradient-2 in pancreatic juice from patients with premalignant pancreatic neoplasia. *Mol. Cancer* **9**, 149
63. Hanas, J. S., Hocker, J. R., Cheung, J. Y., Larabee, J. L., Lerner, M. R., Lightfoot, S. A., Morgan, D. L., Denson, K. D., Prejeant, K. C., Gusev, Y., Smith, B. J., Hanas, R. J., Postier, R. G., and Brackett, D. J. (2008) Biomarker identification in human pancreatic cancer sera. *Pancreas* **36**, 61–69
64. Irigoyen Oyarzabal, A. M., Amiguet Garcia, J. A., López Vivanco, G., Genollá Subirats, J., Muñoz Villafranca, M. C., Ojembarrena Martínez, E., and Liso Irurzun, P. (2003) [Tumoral markers and acute-phase reactants in the diagnosis of pancreatic cancer]. *Gastroenterol. Hepatol.* **26**, 624–629
65. Itonen, O., Koivunen, E., Hurme, M., Alfthan, H., Schröder, T., and Stenman, U. H. (1990) Time-resolved immunofluorometric assays for trypsinogen-1 and 2 in serum reveal preferential elevation of trypsinogen-2 in pancreatitis. *J. Lab. Clin. Med.* **115**, 712–718
66. Koopmann, J., Buckhaults, P., Brown, D. A., Zahurak, M. L., Sato, N., Fukushima, N., Sokoll, L. J., Chan, D. W., Yeo, C. J., Hruban, R. H., Breit, S. N., Kinzler, K. W., Vogelstein, B., and Goggins, M. (2004) Serum macrophage inhibitory cytokine 1 as a marker of pancreatic and other periampullary cancers. *Clin. Cancer Res.* **10**, 2386–2392
67. Koopmann, J., Rosenzweig, C. N., Zhang, Z., Canto, M. I., Brown, D. A., Hunter, M., Yeo, C., Chan, D. W., Breit, S. N., and Goggins, M. (2006) Serum markers in patients with resectable pancreatic adenocarcinoma: macrophage inhibitory cytokine 1 versus CA19–9. *Clin. Cancer Res.* **12**, 442–446
68. Kuhlmann, K. F., van Till, J. W., Boermeester, M. A., de Reuver, P. R., Zvetanova, I. D., Offerhaus, G. J., Ten Kate, F. J., Busch, O. R., van Gulik, T. M., Gouma, D. J., and Crawford, H. C. (2007) Evaluation of matrix metalloproteinase 7 in plasma and pancreatic juice as a biomarker for pancreatic cancer. *Cancer Epidemiol. Biomarkers Prev.* **16**, 886–891
69. Maker, A. V., Katabi, N., Gonen, M., DeMatteo, R. P., D'Angelica, M. I., Fong, Y., Jamagin, W. R., Brennan, M. F., and Allen, P. J. (2011) Pancreatic cyst fluid and serum mucin levels predict dysplasia in intra-ductal papillary mucinous neoplasms of the pancreas. *Ann. Surg. Oncol.* **18**, 199–206
70. Märten, A., Bächler, M. W., Werft, W., Wente, M. N., Kirschfink, M., and Schmidt, J. (2010) Soluble iC3b as an early marker for pancreatic adenocarcinoma is superior to CA19.9 and radiology. *J. Immunother.* **33**, 219–224
71. Moniaux, N., Chakraborty, S., Yalniz, M., Gonzalez, J., Shostrom, V. K., Standop, J., Lele, S. M., Ouellette, M., Pour, P. M., Sasson, A. R., Brand, R. E., Hollingsworth, M. A., Jain, M., and Batra, S. K. (2008) Early diagnosis of pancreatic cancer: neutrophil gelatinase-associated lipocalin as a marker of pancreatic intraepithelial neoplasia. *Br. J. Cancer* **98**, 1540–1547
72. Saha, S., Harrison, S. H., Shen, C., Tang, H., Radivojac, P., Arnold, R. J., Zhang, X., and Chen, J. Y. (2008) HIP2: an online database of human plasma proteins from healthy individuals. *BMC Med. Genomics* **1**, 12
73. Jones, S., Zhang, X., Parsons, D. W., Lin, J. C., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S. M., Fu, B., Lin, M. T., Calhoun, E. S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., Smith, D. R., Hidalgo, M., Leach, S. D., Klein, A. P., Jaffee, E. M., Goggins, M., Maitra, A., Iacobuzio-Donahue, C., Eshleman, J. R., Kern, S. E., Hruban, R. H., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V. E., and Kinzler, K. W. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806
74. Adrian, T. E., Besterman, H. S., Mallinson, C. N., Pera, A., Redshaw, M. R., Wood, T. P., and Bloom, S. R. (1979) Plasma trypsin in chronic pancreatitis and pancreatic adenocarcinoma. *Clin. Chim. Acta* **97**, 205–212
75. Artigas, J. M., Garcia, M. E., Faure, M. R., and Gimeno, A. M. (1981) Serum trypsin levels in acute pancreatic and non-pancreatic abdominal conditions. *Postgrad. Med. J.* **57**, 219–222
76. Borgström, A., and Regné, S. (2005) Active carboxypeptidase B is present in free form in serum from patients with acute pancreatitis. *Pancreatology* **5**, 530–536
77. Funakoshi, A., Yamada, Y., Ito, T., Ishikawa, H., Yokota, M., Shinozaki, H., Wakasugi, H., Misaki, A., and Kono, M. (1991) Clinical usefulness of serum phospholipase A2 determination in patients with pancreatic diseases. *Pancreas* **6**, 588–594
78. Hao, Y., Wang, J., Feng, N., and Lowe, A. W. (2004) Determination of plasma glycoprotein 2 levels in patients with pancreatic disease. *Arch. Pathol. Lab. Med.* **128**, 668–674
79. Hayakawa, T., Kondo, T., Shibata, T., Kitagawa, M., Ono, H., Sakai, Y., and Kiriya, S. (1989) Enzyme immunoassay for serum pancreatic lipase in the diagnosis of pancreatic diseases. *Gastroenterol. Jpn.* **24**, 556–560
80. Hayakawa, T., Kondo, T., Shibata, T., Kitagawa, M., Sakai, Y., Sobajima, H., Tanikawa, M., Nakae, Y., Hayakawa, S., Katsuzaki, T., et al. (1993) Serum pancreatic stone protein in pancreatic diseases. *Int. J. Pancreatol.* **13**, 97–103
81. Junge, W., and Leybold, K. (1982) Detection of colipase in serum and urine of pancreatitis patients. *Clin. Chim. Acta* **123**, 293–302
82. Matsugi, S., Hamada, T., Shioi, N., Tanaka, T., Kumada, T., and Satomura, S. (2007) Serum carboxypeptidase A activity as a biomarker for early-stage pancreatic carcinoma. *Clin. Chim. Acta* **378**, 147–153
83. Pasanen, P. A., Eskelinen, M., Partanen, K., Pikkariainen, P., Penttilä, I., and Alhava, E. (1994) Tumor-associated trypsin inhibitor in the diagnosis of pancreatic carcinoma. *J. Cancer Res. Clin. Oncol.* **120**, 494–497
84. Smith, R. C., Southwell-Keely, J., and Chesher, D. (2005) Should serum pancreatic lipase replace serum amylase as a biomarker of acute pancreatitis? *ANZ J. Surg.* **75**, 399–404
85. Wistuba, II, Behrens, C., Milchgrub, S., Syed, S., Ahmadian, M., Virmani, A. K., Kurvari, V., Cunningham, T. H., Ashfaq, R., Minna, J. D., and Gazdar, A. F. (1998) Comparison of features of human breast cancer cell lines and their corresponding tumors. *Clin. Cancer Res.* **4**, 2931–2938
86. Wistuba, II, Bryant, D., Behrens, C., Milchgrub, S., Virmani, A. K., Ashfaq, R., Minna, J. D., and Gazdar, A. F. (1999) Comparison of features of human lung cancer cell lines and their corresponding tumors. *Clin. Cancer Res.* **5**, 991–1000
87. Sobel, R. E., and Sadar, M. D. (2005) Cell lines used in prostate cancer research: a compendium of old and new lines—part 1. *J. Urol.* **173**, 342–359
88. Doman, B., and Aebersold, R. (2006) Challenges and opportunities in proteomics data analysis. *Mol. Cell Proteomics* **5**, 1921–1926
89. Kapp, E. A., Schütz, F., Connolly, L. M., Chakel, J. A., Meza, J. E., Miller, C. A., Fenyó, D., Eng, J. K., Adkins, J. N., Omenn, G. S., and Simpson, R. J. (2005) An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics* **5**, 3475–3490
90. Barnea, E., Sorkin, R., Ziv, T., Beer, I., and Admon, A. (2005) Evaluation of pre-fractionation methods as a preparatory step for multidimensional based chromatography of serum proteins. *Proteomics* **5**, 3367–3375
91. Das, S., Bosley, A. D., Ye, X., Chan, K. C., Chu, I., Green, J. E., Issaq, H. J., Veenstra, T. D., and Andresson, T. (2010) Comparison of Strong Cation Exchange and SDS-PAGE Fractionation for Analysis of Multiprotein Complexes. *J. Proteome Res.* **9**, 6696–6704

92. Fang, Y., Robinson, D. P., and Foster, L. J. (2010) Quantitative analysis of proteome coverage and recovery rates for upstream fractionation methods in proteomics. *J. Proteome Res.* **9**, 1902–1912
93. Slebos, R. J., Brock, J. W., Winters, N. F., Stuart, S. R., Martinez, M. A., Li, M., Chambers, M. C., Zimmerman, L. J., Ham, A. J., Tabb, D. L., and Liebler, D. C. (2008) Evaluation of strong cation exchange versus isoelectric focusing of peptides for multidimensional liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **7**, 5286–5294
94. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (em-PAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell Proteomics* **4**, 1265–1272
95. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25**, 117–124
96. Collier, T. S., Sarkar, P., Franck, W. L., Rao, B. M., Dean, R. A., and Muddiman, D. C. (2010) Direct Comparison of Stable Isotope Labeling by Amino Acids in Cell Culture and Spectral Counting for Quantitative Proteomics. *Anal. Chem.* **82**, 8696–8702
97. Kakisaka, T., Kondo, T., Okano, T., Fujii, K., Honda, K., Endo, M., Tsuchida, A., Aoki, T., Itoi, T., Moriyasu, F., Yamada, T., Kato, H., Nishimura, T., Todo, S., and Hirohashi, S. (2007) Plasma proteomics of pancreatic cancer patients by multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis (2D-DIGE): up-regulation of leucine-rich alpha-2-glycoprotein in pancreatic cancer. *J. Chromatogr. B* **852**, 257–267
98. Inami, K., Kajino, K., Abe, M., Hagiwara, Y., Maeda, M., Suyama, M., Watanabe, S., and Hino, O. (2008) Secretion of N-ERC/mesothelin and expression of C-ERC/mesothelin in human pancreatic ductal carcinoma. *Oncol. Rep.* **20**, 1375–1380
99. Paciucci, R., Torà, M., Diaz, V. M., and Real, F. X. (1998) The plasminogen activator system in pancreas cancer: role of t-PA in the invasive potential in vitro. *Oncogene* **16**, 625–633
100. Frick, V. O., Rubie, C., Wagner, M., Graeber, S., Grimm, H., Kopp, B., Rau, B. M., and Schilling, M. K. (2008) Enhanced ENA-78 and IL-8 expression in patients with malignant pancreatic diseases. *Pancreatology* **8**, 488–497
101. Rückert, F., Joensson, P., Saeger, H. D., Grützmann, R., and Pilarsky, C. (2010) Functional analysis of LOXL2 in pancreatic carcinoma. *Int. J. Colorectal Dis.* **25**, 303–311
102. Aberger, F., Weidinger, G., Grunz, H., and Richter, K. (1998) Anterior specification of embryonic ectoderm: the role of the *Xenopus* cement gland-specific gene XAG-2. *Mech. Dev.* **72**, 115–130
103. Barraclough, D. L., Platt-Higgins, A., de Silva Rudland, S., Barraclough, R., Winstanley, J., West, C. R., and Rudland, P. S. (2009) The metastasis-associated anterior gradient 2 protein is correlated with poor survival of breast cancer patients. *Am. J. Pathol.* **175**, 1848–1857
104. Fritzsche, F. R., Dahl, E., Dankof, A., Burkhardt, M., Pahl, S., Petersen, I., Dietel, M., and Kristiansen, G. (2007) Expression of AGR2 in non small cell lung cancer. *Histol. Histopathol.* **22**, 703–708
105. Zhang, Y., Forootan, S. S., Liu, D., Barraclough, R., Foster, C. S., Rudland, P. S., and Ke, Y. (2007) Increased expression of anterior gradient-2 is significantly associated with poor survival of prostate cancer patients. *Prostate Cancer Prostatic Dis.* **10**, 293–300
106. Ramachandran, V., Arumugam, T., Wang, H., and Logsdon, C. D. (2008) Anterior gradient 2 is expressed and secreted during the development of pancreatic cancer and promotes cancer cell survival. *Cancer Res.* **68**, 7811–7818
107. Zhang, Y., Ali, T. Z., Zhou, H., D'Souza, D. R., Lu, Y., Jaffe, J., Liu, Z., Passaniti, A., and Hamburger, A. W. (2010) ErbB3 binding protein 1 represses metastasis-promoting gene anterior gradient protein 2 in prostate cancer. *Cancer Res.* **70**, 240–248
108. DeSouza, L. V., Romaschin, A. D., Colgan, T. J., and Siu, K. W. (2009) Absolute quantification of potential cancer markers in clinical tissue homogenates using multiple reaction monitoring on a hybrid triple quadrupole/linear ion trap tandem mass spectrometer. *Anal. Chem.* **81**, 3462–3470
109. Hessle, H., and Engvall, E. (1984) Type VI collagen. Studies on its localization, structure, and biosynthetic form with monoclonal antibodies. *J. Biol. Chem.* **259**, 3955–3961
110. Fujita, A., Sato, J. R., Festa, F., Gomes, L. R., Oba-Shinjo, S. M., Marie, S. K., Ferreira, C. E., and Sogayar, M. C. (2008) Identification of COL6A1 as a differentially expressed gene in human astrocytomas. *Genet. Mol. Res.* **7**, 371–378
111. Rovira, M., Delaspre, F., Massumi, M., Serra, S. A., Valverde, M. A., Lloreta, J., Dufresne, M., Payre, B., Konieczny, S. F., Savatier, P., Real, F. X., and Skoudy, A. (2008) Murine embryonic stem cell-derived pancreatic acinar cells recapitulate features of early pancreatic differentiation. *Gastroenterology* **135**, 1301–1310, 1310. e1–5
112. Li, J., Dowdy, S., Tipton, T., Podratz, K., Lu, W. G., Xie, X., and Jiang, S. W. (2009) HE4 as a biomarker for ovarian and endometrial cancer management. *Expert Rev. Mol. Diagn.* **9**, 555–566
113. Welsh, J. B., Sapinoso, L. M., Kern, S. G., Brown, D. A., Liu, T., Bauskin, A. R., Ward, R. L., Hawkins, N. J., Quinn, D. I., Russell, P. J., Sutherland, R. L., Breit, S. N., Moskaluk, C. A., Frierson, H. F., Jr., and Hampton, G. M. (2003) Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3410–3415
114. Grapin-Botton, A. (2005) Ductal cells of the pancreas. *Int. J. Biochem. Cell Biol.* **37**, 504–510
115. Rovira, M., Delaspre, F., Massumi, M., Serra, S. A., Valverde, M. A., Lloreta, J., Dufresne, M., Payre, B., Konieczny, S. F., Savatier, P., Real, F. X., and Skoudy, A. (2008) Murine embryonic stem cell-derived pancreatic acinar cells recapitulate features of early pancreatic differentiation. *Gastroenterology* **135**, 1301–1310, e1–5
116. Schmid, R. M. (2002) Acinar-to-ductal metaplasia in pancreatic cancer development. *J. Clin. Invest.* **109**, 1403–1404
117. Schmid, R. M., Klöppel, G., Adler, G., and Wagner, M. (1999) Acinar-ductal-carcinoma sequence in transforming growth factor-alpha transgenic mice. *Ann. N Y Acad. Sci.* **880**, 219–230
118. Kobayashi, D., Koshida, S., Moriai, R., Tsuji, N., and Watanabe, N. (2007) Olfactomedin 4 promotes S-phase transition in proliferation of pancreatic cancer cells. *Cancer Sci.* **98**, 334–340
119. Oue, N., Sentani, K., Noguchi, T., Ohara, S., Sakamoto, N., Hayashi, T., Anami, K., Motoshita, J., Ito, M., Tanaka, S., Yoshida, K., and Yasui, W. (2009) Serum olfactomedin 4 (GW112, hGC-1) in combination with Reg IV is a highly sensitive biomarker for gastric cancer patients. *Int. J. Cancer* **125**, 2383–2392
120. Antonin, W., Wagner, M., Riedel, D., Brose, N., and Jahn, R. (2002) Loss of the zymogen granule protein syncoilin affects pancreatic protein synthesis and transport but not secretion. *Mol. Cell. Biol.* **22**, 1545–1554
121. Faca, V. M., Song, K. S., Wang, H., Zhang, Q., Krasnoselsky, A. L., Newcomb, L. F., Plentz, R. R., Gurumurthy, S., Redston, M. S., Pitteri, S. J., Pereira-Faca, S. R., Ireton, R. C., Katayama, H., Glukhova, V., Phanstiel, D., Brenner, D. E., Anderson, M. A., Misek, D., Scholler, N., Urban, N. D., Barnett, M. J., Edelstein, C., Goodman, G. E., Thornquist, M. D., McIntosh, M. W., DePinho, R. A., Bardeesy, N., and Hanash, S. M. (2008) A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS Med.* **5**, e123
122. Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R. H., Eshleman, J. R., Nowak, M. A., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., and Iacobuzio-Donahue, C. A. (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117
123. Mischak, H., Allmaier, G., Apweiler, R., Attwood, T., Baumann, M., Benigni, A., Bennett, S. E., Bischoff, R., Bongcam-Rudloff, E., Capasso, G., Coon, J. J., D'Haese, P., Dominiczak, A. F., Dakna, M., Dihazi, H., Ehrlich, J. H., Fernandez-Llama, P., Fliser, D., Frokiaer, J., Garin, J., Girolami, M., Hancock, W. S., Haubitz, M., Hochstrasser, D., Holman, R. R., Ioannidis, J. P., Jankowski, J., Julian, B. A., Klein, J. B., Kolch, W., Luider, T., Massy, Z., Mattes, W. B., Molina, F., Monsarrat, B., Novak, J., Peter, K., Rossing, P., Sánchez-Carbayo, M., Schanstra, J. P., Semmes, O. J., Spasovski, G., Theodorescu, D., Thongboonkerd, V., Vanholder, R., Veestra, T. D., Weissinger, E., Yamamoto, T., and Vlahou, A. (2010) Recommendations for biomarker identification and qualification in clinical proteomics. *Sci. Transl. Med.* **2**, 46ps42
124. Cutts, R. J., Gadaleta, E., Hahn, S. A., Crnogorac-Jurcovic, T., Lemoine, N. R., and Chelala, C. (2010) The Pancreatic Expression database: 2011 update. *Nucleic Acids Res.* **39**, 01023–1020