

A Large-Scale Proteomics Resource of Circulating Extracellular Vesicles for Biomarker Discovery in Pancreatic Cancer

Reviewed Preprint

v2 • October 24, 2024

Revised by authors

Reviewed Preprint

v1 • July 3, 2023

Bruno Bockorny, Lakshmi Muthuswamy, Ling Huang, Marco Hadisurya, Christine Maria Lim, Leo L Tsai, Ritu R Gill, Jesse L Wei, Andrea J Bullock, Joseph E Grossman, Robert J Besaw, Supraja Narasimhan, W Andy Tao, Sofia Perea, Mandeep S Sawhney, Steven D Freedman, Manuel Hidalgo, Anton Iliuk , Senthil K Muthuswamy 

Division of Medical Oncology, Beth Israel Deaconess Medical Center, Boston, MA, USA • Harvard Medical School, Boston, MA, USA • Blueprint Medicines, Cambridge, MA, USA • Henry Ford Cancer Institute, Detroit, MI, USA • Department of Biochemistry, Purdue University, West Lafayette, IN, USA • Nanyang Technological University, Singapore • Department of Radiology, Beth Israel Deaconess Medical Center, Boston, MA, USA • Deciphera Pharmaceuticals, Waltham, MA, USA • Division of Gastroenterology, Beth Israel Deaconess Medical Center, Boston, MA, USA • Division of Hematology-Oncology, Weill Cornell Medical College, New York, NY, USA • New York-Presbyterian Hospital, New York, NY, USA • Tymora Analytical Operations, West Lafayette, IN, USA • National Cancer Institute, NIH, Bethesda, MD, USA

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

eLife Assessment

The authors analyze a comprehensive cohort of human plasma samples to identify an extracellular vesicles protein signature for early diagnosis of pancreatic cancer. The application of liquid biopsies is **valuable**, and the work addresses a key clinical problem as pancreas cancer is often diagnosed in late stages. The strength of evidence is **solid**. Altogether, this work supports the potential use of extracellular vesicles in clinical settings, with promising value to scientists and clinicians.

<https://doi.org/10.7554/eLife.87369.2.sa3>

Abstract

Pancreatic cancer has the worst prognosis of all common tumors. Earlier cancer diagnosis could increase survival rates and better assessment of metastatic disease could improve patient care. As such, there is an urgent need to develop biomarkers to diagnose this deadly malignancy. Analyzing circulating extracellular vesicles (cEVs) using ‘liquid biopsies’ offers an attractive approach to diagnose and monitor disease status. However, it is important to differentiate EV-associated proteins enriched in patients with pancreatic ductal adenocarcinoma (PDAC) from those with benign pancreatic diseases such as chronic pancreatitis and intraductal papillary mucinous neoplasm (IPMN). To meet this need, we combined the novel EVtrap method for highly efficient isolation of EVs from plasma and conducted proteomics analysis of samples from 124 individuals, including patients with

PDAC, benign pancreatic diseases and controls. On average, 912 EV proteins were identified per 100µL of plasma. EVs containing high levels of PDCD6IP, SERPINA12 and RUVBL2 were associated with PDAC compared to the benign diseases in both discovery and validation cohorts. EVs with PSMB4, RUVBL2 and ANKAR were associated with metastasis, and those with CRP, RALB and CD55 correlated with poor clinical prognosis. Finally, we validated a 7-EV protein PDAC signature against a background of benign pancreatic diseases that yielded an 89% prediction accuracy for the diagnosis of PDAC. To our knowledge, our study represents the largest proteomics profiling of circulating EVs ever conducted in pancreatic cancer and provides a valuable open-source atlas to the scientific community with a comprehensive catalogue of novel cEVs that may assist in the development of biomarkers and improve the outcomes of patients with PDAC.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) has the worst prognosis of all common tumors, with a 5-year survival of 10%¹. With rising incidence, it is expected that PDAC will become the second leading cause of cancer-related deaths by 2030². A critical factor for this dismal development is the late diagnosis, with less than 20% of patients presenting with a potentially resectable and curable tumor³⁻⁵. Earlier cancer diagnosis could increase the survival rates by an estimated 5-fold, and more reliable and real-time assessment of treatment effects in patients with cancer could improve quality of life and reduce healthcare costs^{6,7}. Unfortunately, there are no credentialed serologic biomarkers with high enough performance to assist in the early detection of PDAC. The best-established biomarker for PDAC, carbohydrate antigen 19-9 (CA19-9), is fraught with poor sensitivity and specificity and is only used for monitoring disease on treatment or after surgical resection^{8,9}.

Extracellular vesicles (EVs), including exosomes and microvesicles, are nanosized particles released by most cell types and can be detected in the circulation¹⁰. EVs play important roles in transmission of oncogenic and inflammatory signals¹¹, communications between cells and their microenvironment¹². In addition, exoDNA, exoRNA and protein profiles highly reflect parental cells, therefore offering an attractive strategy for diagnosing cancers non-invasively by analyzing EVs in the circulation^{11,13}. Previous studies employed EVs to discover biomarkers for PDAC¹³⁻¹⁶, however those discovery proteomics experiments were carried out using cell lines or tumor tissue, which are not representative of the heterogeneity of human PDAC and are unable to recapitulate the systemic responses to cancer¹⁴⁻¹⁶. In addition, the EV biomarkers discovered in those studies have been compared only against healthy controls¹⁴⁻¹⁶. It is unclear how they would perform in subjects with underlying benign diseases of the pancreas, which is highly desirable from the clinical standpoint as many patients with PDAC have underlying chronic pancreatitis and cysts.

To meet this need, we conducted a large EV proteomics study from peripheral blood across a range of patients with pancreatic cancer, benign pancreatic diseases such as chronic pancreatitis and intraductal papillary mucinous neoplasm (IPMN), and healthy controls. Circulating EV (cEV) proteins detected included those involved in metabolism and immune regulation, in addition to proteins involved in protein binding, exocytosis, endocytosis and regulation of cellular protein localization that have been identified in previous studies^{17,18}. We subsequently discovered multiple biomarker candidates for cancer diagnosis and verified several of them in an independent cohort of patients with the potential to aid in diagnosing pancreatic cancer. In addition, we identified a set of cEV proteins associated with metastasis which could provide a valuable resource for future biomarker studies.

Results

Proteomics Characterization of Circulating EVs

In this study, we sought to identify proteins in extracellular vesicles in the blood that may be used as biomarkers for the diagnosis and prognosis of pancreatic cancer. With the approval of our institutional review board, we enrolled a total of 124 patients to the discovery cohort of this biomarker study (**Methods and Supplementary Table 1**). Subjects in the pancreatic cancer group (N=93) had a mean age of 66.5 years (range, 37-91), and 48.4% were female. All subjects had biopsy-proven disease. Thirty subjects had early-stage disease (stages I-II) and 63 had advanced disease (stages III-IV). Patients with benign pancreatic diseases included chronic pancreatitis (N=12) with a mean age of 57.5 years (range 37-78) and with 50% females, whereas IPMN included individuals with main duct and side branch IPMNs (N=8) with a mean age of 68.2 years (range 50-89) and with 87.5% being females. Subjects in the healthy control group (N=11) had a mean age of 53.4 years (range, 31-83) with 54.5% females (**Supplementary Table 1**).

We employed the novel EVtrap method (Extracellular Vesicles Total Recovery And Purification) to capture EVs from plasma samples and overcome the traditional laborious techniques for EV isolation, which are not scalable for large clinical studies. As described in recent reports, EVtrap is a magnetic bead-based isolation method that enables highly efficient capture of EVs from biofluids, confirmed by multiple common EV markers^{19,23}. Previous analyses using electron microscopy and nanoparticle tracking also confirmed that the vast majority of particles isolated by EVtrap had diameters between 100-200 nm, consistent with exosomes¹⁹. In addition, EVtrap isolates demonstrates higher abundance of CD9, a common exosome marker, as compared to isolates from other traditional EV isolation methods such as size exclusion chromatography and ultracentrifugation¹⁹. Over 95% recovery yield can be achieved by EVtrap with less contamination from soluble proteins, a significant improvement over current commercially available methods as well as ultracentrifugation^{19,21,24}.

Following EV isolation, samples were digested in-solution and analyzed by liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) on a high-resolution mass spectrometer (Q-Exactive HF-X). The workflow for cEVs isolation and enrichment and subsequent cEV mass spectrometry analysis is illustrated in **Figure 1A**.

First, to confirm that EVtrap can efficiently isolate extracellular vesicles from plasma, a test plasma sample was processed to remove platelets and other large particles and enriched for EVs using EVtrap beads (see **methods** for details). Transmission electron microscopy (TEM) analysis of the EV pellet showed cup-shaped extracellular vesicles (exosomes and microvesicles) (**Supplementary Figure 1A**), and nanoparticle tracking analysis (NTA) using ZetaView instrument (Particle Metrix) demonstrated that the isolated EVs were in the 100–200 nm diameter range, with a mean diameter of 152 nm (**Supplementary Figure 1B**). Second, to assess the technical reproducibility of the EV proteomics approach, the test plasma sample was processed in six replicates and Pearson correlation analysis revealed a very high correlation ($r^2 > 0.97$) between replicates (**Supplementary Figure 2A, Supplementary Table 2**). These results provided the confidence to proceed with the analysis of our discovery set of plasma from 124 subjects. In this cohort, we identified 1,708 unique proteins (**Supplementary Table 3**). The number of unique EV proteins detected per 100μL of plasma sample varied from 817 to 1,128, with an average of 912 unique proteins per sample (**Supplementary Figure 2B**). We did not observe differences between non-tumor and tumor samples regarding the overall number of EV proteins identified. Within the PDAC group, we did not observe significant differences in the average number of EV proteins detected for different disease stages. Collectively, these data demonstrate high reproducibility of EV isolation and robust label-free MS quantification of cEVs.

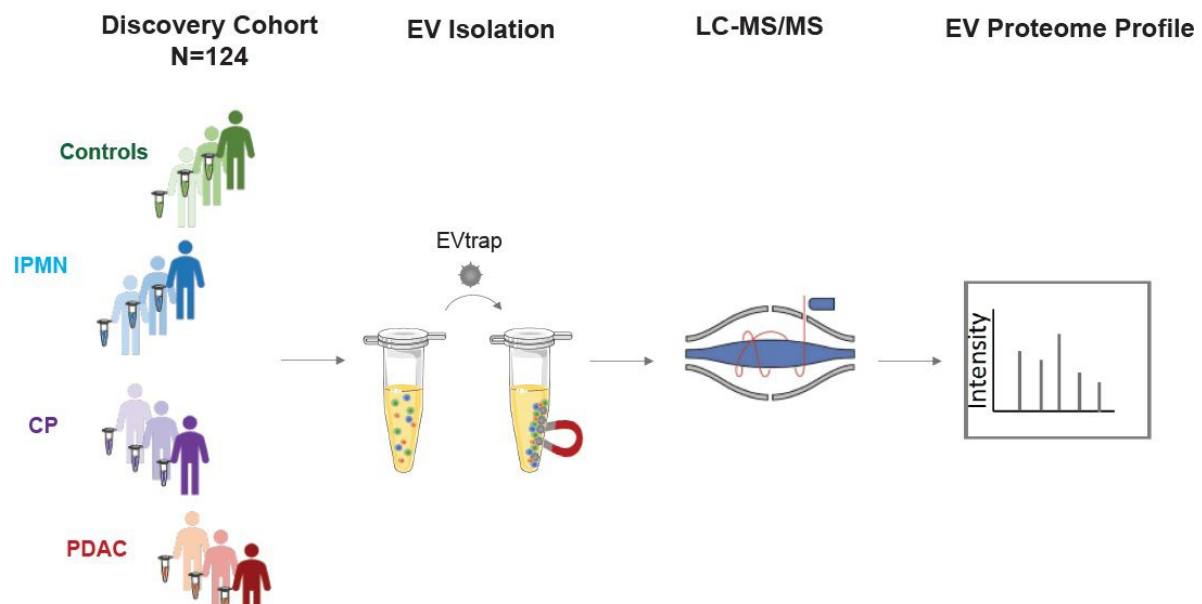


Figure 1.

Study Design

The discovery cohort was comprised of 124 individuals, including pancreatic ductal adenocarcinoma (PDAC, N=93), chronic pancreatitis (CP, N=12), intraductal papillary mucinous neoplasm (IPMN, N=8) and healthy controls (N=11). Plasma samples were processed for EV isolation using EVtrap and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

Diseases of the Pancreas Express Distinct Circulating EV Proteome Compared to Controls

Next, we aimed at identifying specific cEV proteins associated with clinical parameters with the potential to serve as diagnostic biomarkers. We first compared the proteomics profile of individuals with underlying pancreatic diseases (PDAC, chronic pancreatitis and IPMN) against healthy controls. We selected EV proteins expressed in at least 50% of subjects in the disease group with a fold change of expression ≥ 2 or ≤ 2 compared to controls and p-value ≤ 0.01 after adjusting for multiple testing. A total of 207 proteins were identified that met the criteria, with the largest number of differentially expressed markers in PDAC (176), followed by chronic pancreatitis (55) and IPMN (3) (**Supplementary Table 4**). Principal component analysis (PCA) of these markers showed control samples as a tight cluster segregated away from PDAC samples but closer to IPMN and chronic pancreatitis patients (**Figure 2A** [↗](#)).

Circulating EV Proteome Discriminates Pancreatic Cancer from Benign Pancreatic Diseases

To further assess the potential of cEV proteins for cancer detection, we compared proteomic profiles of cEVs between patients with PDAC with those with underlying benign diseases of the pancreas (chronic pancreatitis and IPMN). We identified 182 differentially expressed proteins in malignant cases (92 over-expressed and 90 with reduced expression) (**Supplementary Table 5**). Several of those markers had remarkable overexpression in PDAC (greater than 10-fold), including PDCD6IP, SERPINA12, RUVBL2, among others, as shown in the volcano plot (**Figure 2B** [↗](#)). Unsupervised clustering showed a clear separation between PDAC and benign pancreatic diseases. Individuals with IPMN were more closely related to controls, whereas chronic pancreatitis cases were more related to PDAC (**Figure 2C** [↗](#)). In addition, the PDAC cohort was separated into two subgroups: the first, enriched for early-stage tumors and more closely related to the other pancreatic diseases (chronic pancreatitis and IPMN); the second, enriched for advanced and metastatic cases with expression profiles further apart from early-stage cancer and pancreatic diseases (**Figure 2C** [↗](#)). We further noticed that some proteins such as PDCD6IP, SERPINA12, KRT20 showed statistically significant population-wise enrichment in pancreatic cancer compared to benign pancreatic diseases (**Figure 2D** [↗](#), **Supplementary Figure 3**). Together, these data indicate the existence of EV markers that can separate controls, benign and malignant pancreatic diseases, as well as proteins that separate early versus late-stage PDAC, suggesting their potential to serve as diagnostic biomarkers.

Functional and Systems Biology of cEV Proteome

To gain molecular insight into the functions of the 182 proteins differentially expressed in pancreatic cancer as compared to benign pancreatic diseases, we conducted pathway analysis using the Gene Ontology (GO) and REACTOME databases (**Supplementary Table 5**). We identified protein modules in protein localization, biomolecule binding/docking, peptidase activities among changes enriched in PDAC compared to benign diseases (**Supplementary Figure 4**). Interestingly, KRT20 (keratin 20), a gastrointestinal epithelia-associated keratin, was increased in PDAC patient EVs, while keratins associated basal cells, KRT4, KRT15, and KRT3, were reduced. KRT20 overexpression is frequently found in pancreatic tumor tissues and correlates with poor prognosis²⁵ [↗](#), suggesting a biological basis for their high levels in the cEVs of PDAC patients.

Interestingly, proteins associated with immunological functions showed complex regulation with increased representation of leukocyte mediated immunity (GO:0002443), leukocyte degranulation (GO:0043299), myeloid leukocyte activation (GO:0002274), and decrease in Fc receptor signaling (GO: 0038093), regulation of complement activation (GO:0030449), and immune effector process

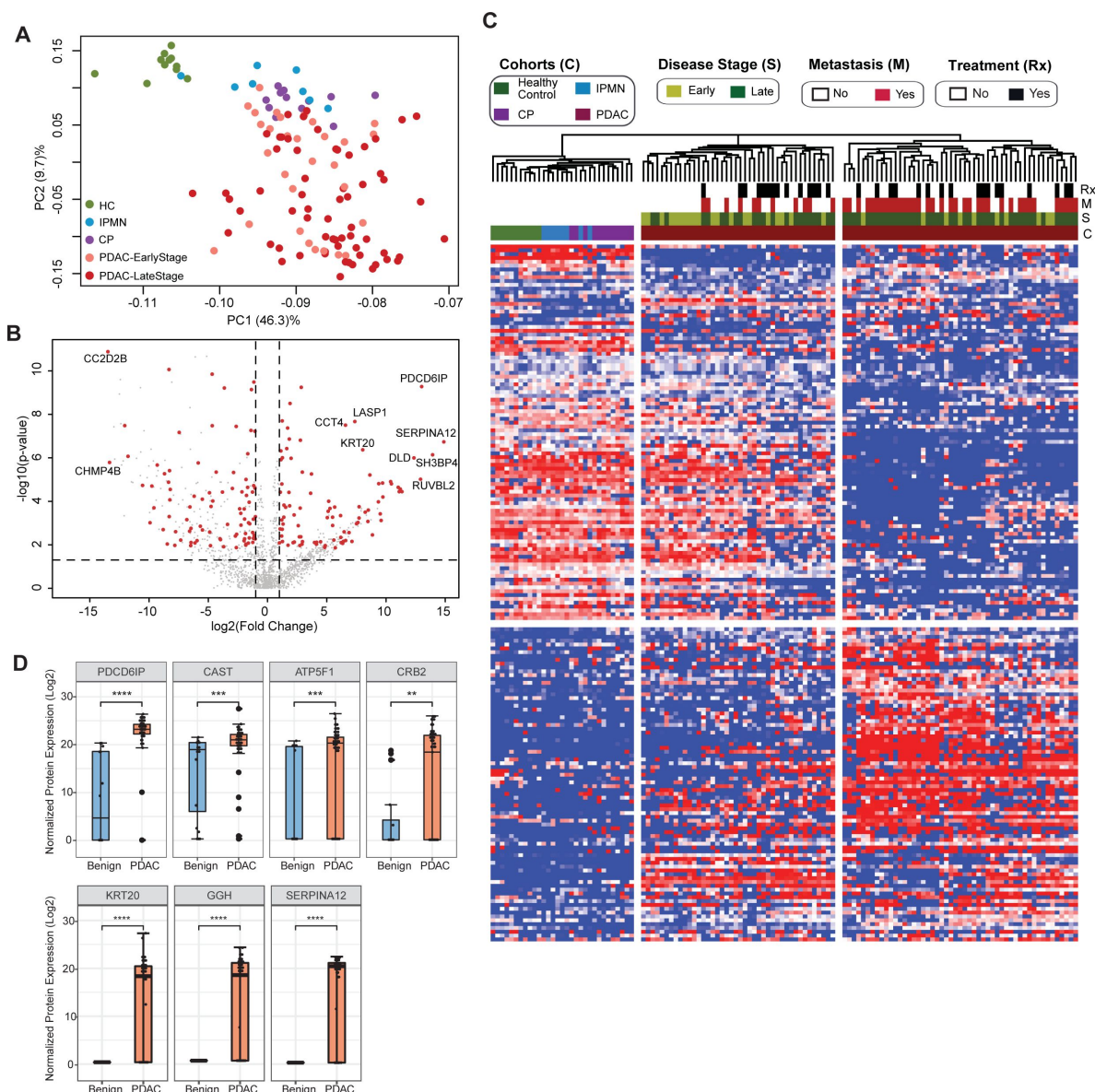


Figure 2.

Identification of cEV Proteins Differentially Expressed in Disease Groups

(A) Principal component analysis of cEV proteins differentially expressed in the plasma of patients with pancreatic diseases compared to controls. Each dot indicates one individual enrolled in the study: green, controls; blue, patients with intraductal papillary mucinous neoplasm (IPMN); purple, patients with chronic pancreatitis (CP); salmon, early stage (stages I and II) pancreatic ductal adenocarcinoma (PDAC); red, late stage (stages III and IV) PDAC.

(B) Volcano plot of circulating EV proteins enriched in the plasma of patients with PDAC versus benign pancreatic diseases. X-axis, log base 2 of fold changes; Y-axis, negative of the log base 10 of p values

(C) Heatmap of cEV proteins differentially expressed in the plasma of patients with pancreatic diseases compared to controls. Designations of clinical parameters were indicated at the top of the heatmap.

(D) Expression of enriched cEV proteins in patients with PDAC versus benign pancreatic diseases. Each dot indicates the target protein signal from one patient. Y-axis, normalized log base 2 of protein signals detected by mass spectrometry; Error bars, min and max values; lines in boxes, median values. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

(GO:0002252) (**Supplementary Table 5**). These data suggest that direct profiling of cEVs from patient plasma provided unique insights into systemic changes in immune biology during pancreatic cancer development, which is lacked in analysis restricted to tissue or cell models.

Circulating EV Proteomics Reveal Markers Associated with Metastasis and Worse Prognosis

We then investigated whether cEV proteins can assist in the distinction of metastatic versus non-metastatic pancreatic cancer. We compared the cEV proteome profiles of individuals with metastatic cancer to those without metastasis and identified 85 proteins differentially expressed between the two groups (**Supplementary Table 6**). Supervised clustering between metastatic and non-metastatic diseases showed a clear separation with two distinct expression patterns (**Figure 3A** [↗](#)). In particular, PSMB4, RUVBL2 and ANKAR (**Figure 3B** [↗](#)) EV protein levels were increased in patients with metastatic disease, whereas RAP2B, SERPINA12 and IGLV4-69 abundance levels were decreased in the cEVs of patients with metastasis (**Figure 3C** [↗](#)). Together, these findings suggest the presence of a core set of cEV proteins with the potential to distinguish early versus metastatic pancreatic cancer.

We further analyzed whether the expression of certain cEV proteins had prognostic relevance in our cohort. We first classified individuals with PDAC as having low or high expression of any given markers based on each marker's first and third quartile. Survival was estimated by the Kaplan Meier method. We identified that the cEV expression of RALB, CRP, and CD55 had a significant correlation with overall survival, with a trend for PDCD6IP (**Figure 3D** [↗](#)).

Validation of cEV Markers Using Parallel Reaction Monitoring and Identification an EV Protein Signature for Pancreatic Cancer Diagnosis

Because pancreatic cancer is extremely heterogeneous, the chance of identifying a single biomarker with sufficient diagnostic performance is likely low. Instead, the identification of a panel of candidate markers may have enhanced diagnostic performance.

To identify a signature that shows the most discriminatory power between 'benign diseases' and 'PDAC,' we employed a binary classification approach using Support Vector Machines (SVM). Classification models, built based on a large number of proteins, contain irrelevant markers that can reduce the predictive accuracy. Hence, we implemented a consensus feature selection method based on two algorithms: one using recursive feature elimination (RFE) algorithm (SVM-RFE)²⁶ [↗](#) and second, RFE combined with a non-parametric Wilcoxon rank test (sigFeature)²⁷ [↗](#). The top 16 markers were selected whose classification performance can be tested in the independent validation cohort (**Supplementary Table 7**). A summary of selection process is shown in **Supplementary Figure 5**. The classification performance of these 16 markers, individual and in all combinations, were tested using 80% training data and evaluated in the remaining 20% test data. The quality of training was assessed using five repetitions of 10-fold cross-validation. The optimal kernel parameters were estimated by tuning over a wide range of values. Receiver operating characteristic (ROC) analysis was used as the metric to assess the performance of the classifier model. We found a set of 7-EV protein signature comprised of RUVBL2, PDCD6IP, ATP5F1, DLD, KRT20, CCT4, and SERPINA12, that gave 100% accuracy when tested in the discovery cohort (**Supplementary Table 7, Supplementary Figure 6**). Recurrence of these putative markers in our dataset varied from 55% to 97%.

The model was further validated on an independent validation cohort whose proteome was obtained using an alternate technology, parallel reaction monitoring (PRM) mass spectrometry. The markers chosen for validation included 16 markers selected for SVM classification model and an additional 9 markers to result in top 25 markers that are significantly differentially expressed

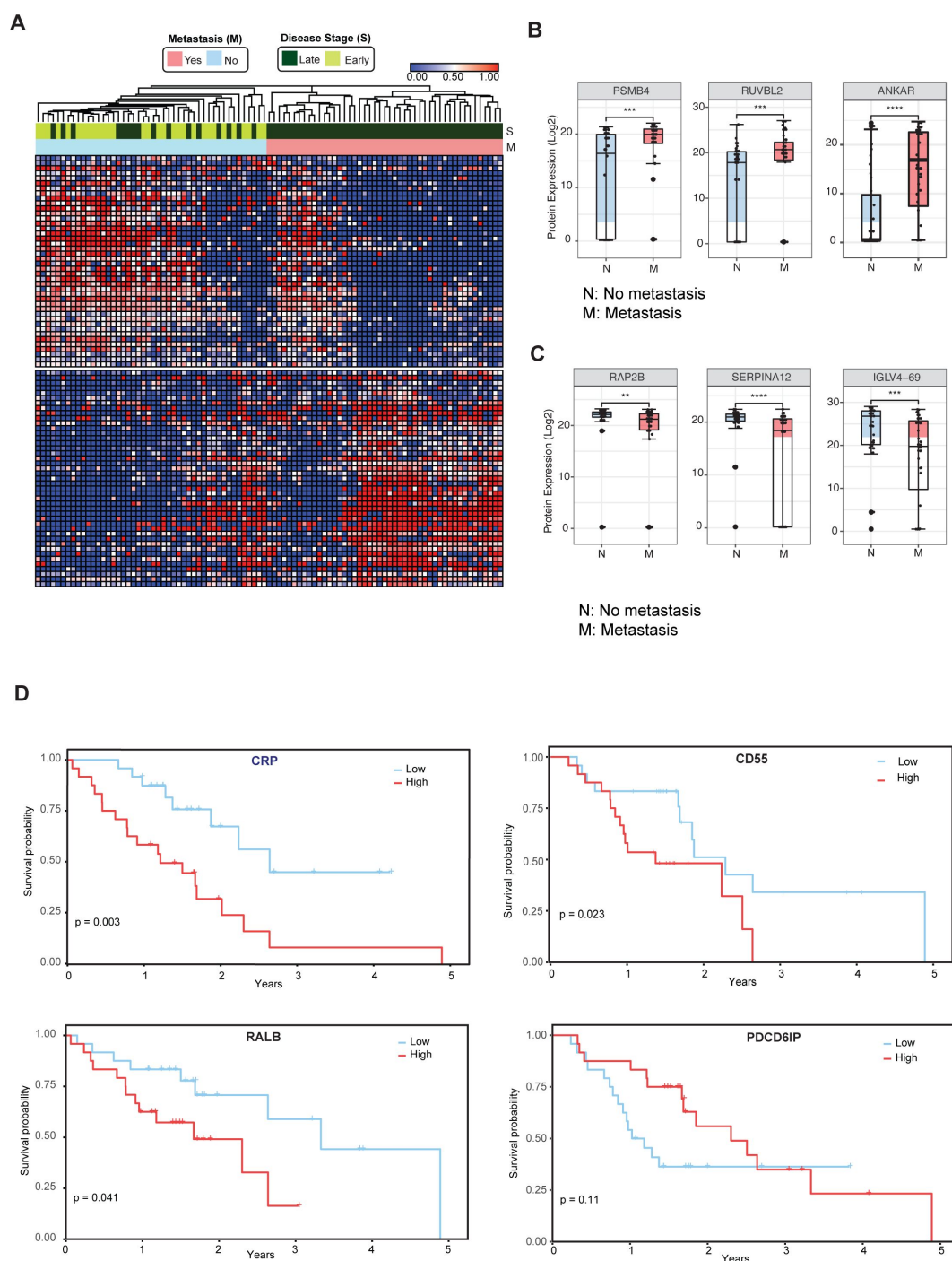


Figure 3.

Circulating EV Proteomics Reveal Markers Associated with Metastasis and Worse Prognosis.

(A) Heatmap showing EV proteins differentially expressed in the plasma of metastatic versus non-metastatic PDAC. Designations of clinical parameters are indicated at the top of the heatmap.

(B) Expression patterns of cEV proteins associated with metastatic disease. Y-axis, normalized log base 2 of protein signals detected by mass spectrometry; N, non-metastatic PDAC group; M, metastatic PDAC group. Each dot indicated the target protein signal from one patient. Error bars, min and max values; lines in boxes, median values. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

(C) As is **(B)**, except for cEV markers with increased expression in non-metastatic PDAC.

(D) Correlation of cEV marker expression with survival. Kaplan–Meier curves and log-rank test P values of representative survival cEV markers quantified in the discovery cohort.

in the discovery cohort with a fold change increase in PDAC ≥ 5.5 and p-value ≤ 0.01 (**Methods, Supplementary Table 8**). The independent validation cohort consisted of 36 new subjects (24 with PDAC, 6 with chronic pancreatitis, and 6 with IPMN) (**Supplementary Table 9**). A total of 10 proteins, including all 7 signature proteins, showed a significant difference ($p < 0.05$) in patients with PDAC as compared to benign pancreatic diseases (**Figure 4A**). The performance of individual validated markers according to the specific underlying disease in the validation cohort is presented in **Supplementary Figure 7**. The performance of 7-EV protein signature was further tested using SVM model, in our independent validation cohort, yielding an 89% prediction accuracy (**Figure 4B**, **Supplementary Figure 8**). As expected, we observed that no single marker achieved sufficiently high sensitivity and specificity as the combined model for the diagnosis of pancreatic cancer.

Discussion

Extracellular vesicles hold a great promise as a source of potential biomarkers, making them attractive candidates for liquid biopsy tests. Previously, we reported that organoid cultures of pancreatic cancer could serve as models to discover tissue-derived EV proteins with high specificity for PDAC, as opposed to chronic pancreatitis and other benign gastrointestinal diseases²⁸. A shortcoming of these tissue-based studies is the inability to discover markers associated to the systemic responses to cancer. Here, we performed a large-scale, comprehensive analysis of circulating EV proteomes directly from plasma samples of 124 patients, with subsequent validation in a separate cohort of 36 patients. To our knowledge, this represents the largest proteomics profiling dataset of circulating EVs conducted in pancreatic cancer to date. In this study, we identified and validated new EV markers from plasma that distinguish patients with pancreatic cancer from subjects with benign pancreatic diseases. Furthermore, we discovered several cEV proteins associated with metastatic disease and poor prognosis. In contrast to the prior studies of experimental cell models or tissues extracts that were examined only against healthy subjects^{14,16}, we report the identification of EV proteins in plasma of patients with pancreatic cancer compared to patients with underlying pancreatic diseases, which is clinically relevant as many patients with pancreatic cancer have underlying chronic inflammation and premalignant cystic lesions.

In addition, our study demonstrated the feasibility of using the novel EVtrap method^{19,20} for discovery of hundreds of EV proteins directly from a small volume (100uL) of plasma samples. This methodological advance can be adopted for biomarker discovery in other cancer types. Other workflows traditionally employed for EV isolation from blood samples require laborious techniques including lengthy ultracentrifugation steps which are unsuitable for large scale studies²⁹.

We identified several EV proteins as significantly associated with metastasis or survival. For instance, PSMB4 and RUVBL2 levels were increased in cEVs of patients with metastatic PDAC. Notably, PSMB4 (proteasome subunit beta type-4), a protein of the ubiquitin-proteasome degradation pathway, has been identified as the first proteasomal subunit with oncogenic properties and associated to poor prognosis in several tumors including melanoma, breast and ovarian cancers³⁰⁻³³. As expected, the EV proteomic profiles of PDAC patients exhibited significant heterogeneity. While the above-mentioned markers exhibited strong association with disease states at population levels, their abundances in individual patients varied significantly. Those observations highlight the need to develop multi-protein panels for pancreatic cancer diagnosis and prognosis.

We also discovered RALB, CRP and CD55 expression on EVs to have a significant correlation with poor survival, while PDCD6IP expression was associated with improved outcomes. Interestingly, PDCD6IP (programmed cell death 6-interacting protein), was also identified as a PDAC-enriched

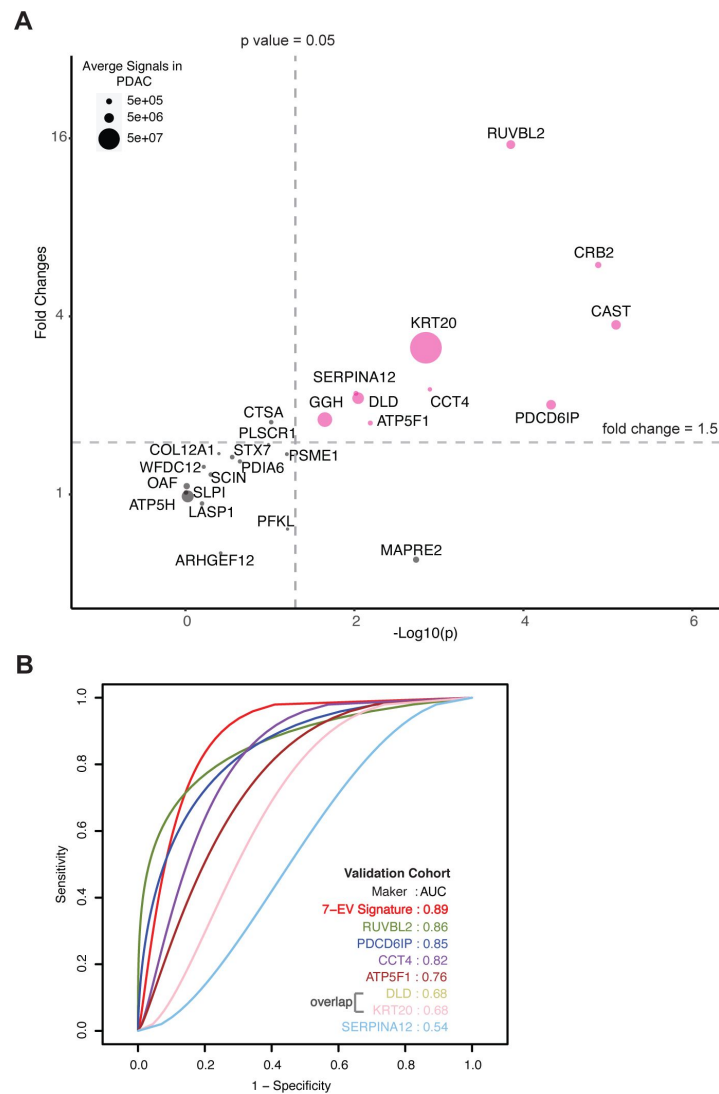


Figure 4.

Validation of cEV Markers and Identification of 7-EV Protein Signature for PDAC Diagnosis.

(A) Differences of cEV protein abundances between patients with PDAC (n=24) and benign pancreatic diseases (chronic pancreatitis and IPMN) (n=12). x axis, minus log p values of protein abundance differences between PDAC and benign groups; y axis, average fold changes of proteins in PDAC group compared to benign group. Size of bubbles indicate average protein abundances in PDAC group. Pink color, proteins that had at least two-fold enrichment in PDAC group ($p < 0.05$).

(B) ROC curves were calculated for individual cEV markers as well as for the 7-EV protein PDAC signature combination to determine optimum diagnostic performance. AUC, area under the curve.

protein in the tissue-based proteomics studies from Le Large *et al.*³⁴ and Hoshino *et al.*¹⁸. In line with our findings, its tissue expression in liver metastasis of pancreatic cancer has been found to also correlate with improved prognosis in patients with PDAC in the study of Law *et al.*³⁵. Collectively, these data suggest that some tissue-specific proteins can be isolated from circulating EVs and their quantifiable levels in the blood may have the potential to serve as diagnostic or prognostic biomarkers in pancreatic cancer.

In our validation studies, all seven putative markers identified from the model were significantly enriched in the plasma of PDAC patients. Based on the top seven markers, we derived a 7-EV protein panel that yielded an 89% prediction accuracy for diagnosing pancreatic cancer. A recent modeling study showed that a new diagnostic assay for PDAC would have to perform with a minimum sensitivity of 88% and a specificity of 85% to reduce healthcare expenditure and prolong survival⁶. Serum CA19-9, the best-established blood test for PDAC, has a pooled sensitivity of 75.4% and a specificity of 77.6%³⁶. It commonly rises late in the disease and may be elevated in nonmalignant conditions such as biliary obstruction and pancreatitis, making it unsuitable as a diagnostic biomarker for PDAC³⁷. As such, our 7-EV protein signature with 89% prediction accuracy serves as a proof-of-concept and has the potential to facilitate the further development of biomarker tests for pancreatic cancer. We anticipate that for clinical use application, an even higher diagnostic performance is needed. Future studies are warranted to investigate if combining our validated cEV proteins with other biomarkers such as cell free DNA, serum proteins or metabolites, as a multi-analyte biomarker assay, would yield higher accuracy in diagnosing pancreatic cancer.

While our work involved a large cohort with 160 patients, the single-center nature is an inherent limitation of our study. Also, it would be ideal to perform validation with a larger cohort of controls to achieve greater statistical power. To balance this limitation, we increased the rigor of our validation by selecting controls with underlying benign diseases of pancreas as opposed to healthy volunteers, and an alternate quantitative technology for measuring protein abundance (Parallel Reaction Monitoring Mass Spectrometry) instead of MS-LC. While this approach increased the generalizability, it marginally reduced model prediction. Thus, the performance of our 7-EV protein PDAC panel should be cross validated in larger and multicenter populations. In addition, in this work we only used EVtrap as EV isolation method and mass spectrometry for protein quantification, and it is possible that there was some degree of heterogeneity in the extracellular vesicles analyzed. The clinical impact of biomarkers identified in our study will need to be cross validated using other methods.

With no major treatment breakthrough for pancreatic cancer in the last decade, every effort should be made to diagnose this deadly cancer at earlier stages and to discover new proteins involved in tumorigenesis. Our study provides a valuable open resource to the scientific community with a comprehensive catalog of novel proteins packaged inside circulating EVs that may assist in the development of novel biomarkers and improve the outcomes of patients with pancreatic cancer.

Methods

Study Design and Patient Demographics

We conducted this study at Beth Israel Deaconess Medical Center with the approval of the Harvard Cancer Center Institutional Review Board. All subjects provided written informed consent. Clinical data and blood samples were prospectively collected from 2017 to 2019 from patients with pancreatic cancer, chronic pancreatitis, intraductal papillary mucinous neoplasms (IPMN), and age-matched controls. A total of 124 patients, including PDAC (N=93), chronic pancreatitis of different etiologies (N=12), IPMN (N=8), and controls (N=11), were included in the discovery

cohort. PDAC diagnosis was established by histology or cytology, and staging was performed according to the American Joint Committee on Cancer guidelines (8th Edition 2016)³⁸ (Supplementary Table 1). For the independent validation cohort, a total of 36 patients were enrolled, including PDAC (N=24), IPMN (N=6), and chronic pancreatitis (N=6) (Supplementary Table 9).

Plasma Sample Collection and Processing

All blood samples were collected and processed following the same standard operating procedure optimized for EV analysis and included the following steps: (i) whole blood was collected into one 10ml yellow-top tube containing acid citrate dextrose; (ii) blood was mixed by gently inverting the tube five times; (iii) vacutainer tubes were stored upright at room temperature (RT); (iv) samples were centrifuged at 1,300g for 15 min in RT; (v) plasma was removed from the top carefully avoiding cell pellet; (vi) repeat centrifugation of plasma at 2,500g for 15 min in RT; (vii) again, plasma was removed from the top carefully avoiding cell pellet; (viii) third centrifugation at 2,500g for 15 min in RT, then samples were aliquoted to be stored at -80°C.

Extracellular Vesicle Isolation from Plasma

We employed EVtrap for EV isolation from plasma samples¹⁹. EVtrap beads were provided by Tymora Analytical (West Lafayette, IN) as a suspension in water and were used as previously described in more details^{19,20}. Briefly, 100 µL plasma samples were diluted 20 times in the diluent buffer, the EVtrap beads were added to the samples in a 1:2 v/v ratio, and the samples were incubated by end-over-end rotation for 30 min according to the manufacturer's instructions. After supernatant removal using a magnetic separator rack, the beads were washed with PBS, and the EVs were eluted by a 10 min incubation with 200 mM triethylamine (TEA, Millipore-Sigma). The samples were fully dried in a vacuum centrifuge.

Preparation of EV samples

The isolated and dried EV samples were lysed to extract proteins using the phase-transfer surfactant (PTS) aided procedure. The proteins were reduced and alkylated by incubation in 10 mM TCEP and 40 mM CAA for 10 min at 95°C. The samples were diluted fivefold with 50 mM triethylammonium bicarbonate and digested with Lys-C (Wako) at 1:100 (wt/wt) enzyme-to-protein ratio for 3 h at 37°C. Trypsin was added to a final 1:50 (wt/wt) enzyme-to-protein ratio for overnight digestion at 37°C. To remove the PTS surfactants from the samples, the samples were acidified with trifluoroacetic acid (TFA) to a final concentration of 1% TFA, and ethyl acetate solution was added at a 1:1 ratio. The mixture was vortexed for 2 min and then centrifuged at 16,000 × g for 2 min to obtain aqueous and organic phases. The organic phase (top layer) was removed, and the aqueous phase was collected. This step was repeated once more. The samples were dried in a vacuum centrifuge and desalted using Top-Tip C18 tips (Glygen) according to the manufacturer's instructions. The samples were dried completely in a vacuum centrifuge and stored at -80°C.

LC-MS Analysis of Plasma EV Proteome

Approximate 1 µg of each dried peptide sample was dissolved in 10.5 µL of 0.05% trifluoroacetic acid with 3% (vol/vol) acetonitrile containing spiked-in indexed Retention Time Standard containing 11 artificially synthetic peptides (Biognosys). The spiked-in 11-peptides standard mixture was used to account for any variation in retention times and to normalize abundance levels among samples. 10 µL of each sample was injected into an Ultimate 3000 nano UHPLC system (Thermo Fisher Scientific). Peptides were captured on a 2-cm Acclaim PepMap trap column and separated on a heated 50-cm Acclaim PepMap column (Thermo Fisher Scientific) containing C18 resin. The mobile phase buffer consisted of 0.1% formic acid in ultrapure water (buffer A) with an eluting buffer of 0.1% formic acid in 80% (vol/vol) acetonitrile (buffer B) run with a linear 60-min gradient of 6–30% buffer B at a flow rate of 300 nL/min. The UHPLC was coupled online

with a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode, in which a full-scan MS (from m/z 375 to 1,500 with the resolution of 60,000) was followed by MS/MS of the 15 most intense ions (30,000 resolution; normalized collision energy - 28%; automatic gain control target (AGC) - $2E4$, maximum injection time - 200 ms; 60sec exclusion).

EV Proteome Data Processing

The raw files were searched directly against the human Swiss-Prot database with no redundant entries using Byonic (Protein Metrics) and Sequest search engines loaded into Proteome Discoverer 2.3 software (Thermo Fisher Scientific). MS1 precursor mass tolerance was set at 10 ppm, and MS2 tolerance was set at 20ppm. Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da) and variable modifications of oxidation (+15.9949 Da) on methionine residues and acetylation (+42.011 Da) at the N terminus of proteins. The search was performed with full trypsin/P digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. The false-discovery rates of proteins and peptides were set at 0.01. All protein and peptide identifications were grouped, and any redundant entries were removed. Only unique peptides and unique master proteins were reported.

All data were quantified using the label-free quantitation node of Precursor Ions Quantifier through the Proteome Discoverer v2.3 (Thermo Fisher Scientific). For the quantification of proteomic data, the intensities of peptides were extracted with initial precursor mass tolerance set at 10 ppm, a minimum number of isotope peaks as 2, maximum ΔRT of isotope pattern multiplets – 0.2 min, PSM confidence FDR of 0.01, with hypothesis test of ANOVA, maximum RT shift of 5 min, pairwise ratio-based ratio calculation, and 100 as the maximum allowed fold change. The abundance levels of all peptides and proteins were normalized to the spiked-in internal iRT standard. For calculations of fold-change between the groups of proteins, total protein abundance values were added together, and the ratios of these sums were used to compare proteins within different samples.

The abundances of EV proteins were normalized using indexed retention time (iRT) in Proteome Discoverer (ThermoFisher Scientific). Abundances were categorized into four different categories: Control, Chronic Pancreatitis, IPMN, and PDAC. Protein abundances were then log2 transformed and quantile normalized for further analysis.

A non-parametric Wilcox Rank Sum test was performed to test the null hypothesis that the distributions of two groups of the patient population are the same, and the fold change and p-values for each protein were estimated for the following comparisons: IPMN vs. Control, CP vs. Control, PDAC vs. Control, Benign Pancreatic Diseases (CP, IPMN) vs. PDAC. Multiple testing correction was done using Benjamini-Hochberg method to control for the false discovery rate³⁹. Volcano plots were created using those p values and fold change. Heatmaps visualization and clustering of statistically significant proteins, with adjusted p-value ≤ 0.05 and absolute fold change ≥ 2 , were created in R using the pheatmap package. Euclidean distance and average cluster method were used. The values were row-scaled for normalization. Both rows and columns were allowed to cluster.

Pathways Enrichment and Protein Network Analysis

Pathway enrichment analysis was performed on statistically significant genes using g:Profiler⁴⁰, a web-based tool that searches for pathways whose genes are significantly enriched in our dataset compared to a collection of genes representing Gene Ontology (GO) terms and Reactome pathways. We further used EnrichmentMap⁴¹, a Cytoscape, v3.8.2⁴² application to create a visual network of connected pathways that helps to identify relevant pathways and theme⁴³. A

Protein-Protein interaction network was generated using a stringApp, a Cytoscape app. This application allows to import STRING networks into Cytoscape and enables to perform complex network analysis and visualization of networks⁴⁴.

Parallel Reaction Monitoring and Data Analysis

Parallel reaction monitoring mass spectrometry (PRM-MS) was employed for validation experiments. Twenty-five cEV markers were selected for validation based on fold change increase ≥ 5.5 , p-value ≤ 0.01 , and technical aspects (number of unique peptides and coverage) (Supplementary Table 8). Thirty-six plasma samples from a new cohort were used for the validation (24 PDAC, 6 IPMN and 6 chronic pancreatitis samples). The EVs were isolated from plasma and the proteins processed as described before. Peptide samples were dissolved in 10.8 μL 0.05% TFA & 2% ACN, and 10 μL injected into the UHPLC coupled with a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific). The mobile phase buffer consisted of 0.1% formic acid in HPLC grade water (buffer A) with an eluting buffer containing 0.1% formic acid in 80% (vol/vol) acetonitrile (buffer B) run with a linear 60-min gradient of 5–35% buffer B at a flow rate of 300 nL/min. Each sample was analyzed under PRM with an isolation width of ± 0.8 Th. In these PRM experiments, an MS2 level at 30,000 resolution relative to m/z 200 (AGC target 2E5, 200 ms maximum injection time) was run as triggered by a scheduled inclusion list. Higher-energy collisional dissociation was used with 28 eV normalized collision energy. PRM data were manually curated within Skyline-daily (64-bit) 20.2.1.404 (32d27b598)⁴⁵.

Identification of EV signature for pancreatic cancer diagnosis

To identify a biomarker signature demonstrating the highest discriminatory power between ‘benign’ and ‘PDAC’ diseases, we adopted a binary classification approach utilizing Support Vector Machines (SVM). Recognizing that classification models built on an extensive array of proteins may incorporate irrelevant markers, which can diminish the predictive accuracy, we started with a list of significantly differentially expressed set of 91 proteins between ‘benign’ and ‘PDAC’ patients and further employed a consensus feature selection strategy using two algorithms, ‘Recursive Feature Elimination’ (SVM-RFE), and ‘Integrated RFE with a non-parametric Wilcoxon rank test (sigFeature)’. Subsequently, we selected the top 16 markers, the classification performance of which was subjected to testing in an independent validation cohort. The classification performance evaluation of these markers, both individually and in various combinations, involved a rigorous assessment utilizing 80% of the data for training and the remaining 20% for internal-validation. To ensure the quality of the training process, we employed five repetitions of a 10-fold cross-validation approach. The optimal kernel parameters were determined through tuning across a broad range of values. Receiver operating characteristic (ROC) analysis served as the metric to gauge the performance of the classifier model. All algorithms for identifying the EV signature predictive of pancreatic cancer diagnosis were implemented in R. We used Support Vector Machine (SVM) using CRAN package, e107⁴⁶. Ranking of genes was achieved using packages ‘sigFeature’ and ‘SVM-RFE’. An R package, ‘pROC’⁴⁷ was used to build a receiver operating characteristic curve (ROC) and to calculate the area under the curve (AUC).

Survival Analysis

The prognostic value of every protein was estimated by dividing patients into two groups: group 1, patients with expression below the 25th percentile, and group 2, patients with expression values greater than 75th percentile. The Kaplan-Meier estimator was used to estimate the survival function associating survival with EV protein expression, and the log-rank test was used to compare survival curves of two groups. ‘survival’ R package was used for the analysis.

Statistical Analysis

All statistical analyses were performed using the statistical software R. Statistical significance was calculated by two-tailed Student's t-test or Wilcoxon rank-sum test unless specified otherwise in the figure legend. Data are expressed as mean $\pm$ SEM. A p-value < 0.05 in biological experiments or FDR < 0.05 after multiple comparison corrections in proteomics data analysis was considered statistically significant.

Data Availability

All data presented all be made available upon request

Acknowledgements

We thank the patients and their families for their participation in this study. B.B. was supported in part through UM1 (CA186709-06). S.D.F. was supported in part through the Barbara Janson and Arthur Hilsinger Pancreatology Fellowship. Institutional startup funds and UO1 (CA224193) to S.K.M., and seed grant from Hirschberg Foundation for Pancreatic Cancer Research to L.H. We thank members of the Muthuswamy laboratory for their critical input throughout the development of this project.

Competing Interest Statement

B.B. reports Research Funding: Agenus Inc, NanoView Bioscience; Travel Expenses: Erytech Pharma; Advisory Board and Consulting: Blueprint Medicines, Bioline Rx. A.B. reports Advisory/Consulting with Exelixis and Geistlich Pharma. A.I. and W.A.T. are principals at Tymora Analytical Operations, which developed EVtrap beads. M.H. reports grants from Weill Cornell Medicine; personal fees from Agenus, BMS, Oncomatrix, Khar, Inxmed, Genchem, Cantargia, and Fibrogen outside the submitted work; patent for Method for ACT in PDAC licensed to Peaches SL; and Director of BMS. S.K.M. owns stocks and is a member of the Scientific Advisory Board of KAHR Medical. The remaining authors declare no related competing interests.

References

1. Siegel RL (2020) **Miller KD** *CA Cancer J Clin* **70**:7–30
2. Rahib L, Smith BD, Aizenberg R, et al (2014) **Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States** *Cancer Res* **74**:2913–21
3. Kamisawa T, Wood LD, Itoi T, et al (2016) **Pancreatic cancer** *Lancet* **388**:73–85
4. Rahib L, Fleshman JM, Matrisian LM, et al (2016) **Evaluation of Pancreatic Cancer Clinical Trials and Benchmarks for Clinically Meaningful Future Trials: A Systematic Review** *JAMA Oncol* **2**:1209–16
5. Bockorny B, Grossman JE (2022) **Hidalgo M: Facts and Hopes in Immunotherapy of Pancreatic Cancer** *Clin Cancer Res* **28**:4606–4617
6. Ghatnekar O, Andersson R, Svensson M, et al (2013) **Modelling the benefits of early diagnosis of pancreatic cancer using a biomarker signature** *Int J Cancer* **133**:2392–7
7. Matsuno S, Egawa S, Fukuyama S, et al (2004) **Pancreatic Cancer Registry in Japan: 20 years of experience** *Pancreas* **28**:219–30
8. Locker GY, Hamilton S, Harris J, et al (2006) **ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer** *J Clin Oncol* **24**:5313–27
9. Galli C, Basso D (2013) **Plebani M: CA 19-9: handle with care** *Clin Chem Lab Med* **51**:1369–83
10. Chen IH, Xue L, Hsu CC, et al (2017) **Phosphoproteins in extracellular vesicles as candidate markers for breast cancer** *Proc Natl Acad Sci U S A* **114**:3175–3180
11. Costa-Silva B, Aiello NM, Ocean AJ, et al (2015) **Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver** *Nat Cell Biol* **17**:816–26
12. van Niel G, Carter DRF, Clayton A, et al (2022) **Challenges and directions in studying cell-cell communication by extracellular vesicles** *Nat Rev Mol Cell Biol* **23**:369–382
13. Melo SA, Luecke LB, Kahlert C, et al (2015) **Glypican-1 identifies cancer exosomes and detects early pancreatic cancer** *Nature* **523**:177–82
14. Madhavan B, Yue S, Galli U, et al (2015) **Combined evaluation of a panel of protein and miRNA serum-exosome biomarkers for pancreatic cancer diagnosis increases sensitivity and specificity** *Int J Cancer* **136**:2616–27
15. Yang KS, Im H, Hong S, et al (2017) **Multiparametric plasma EV profiling facilitates diagnosis of pancreatic malignancy** *Sci Transl Med* **9**
16. Castillo J, Bernard V, San Lucas FA, et al (2018) **Surfaceome profiling enables isolation of cancer-specific exosomal cargo in liquid biopsies from pancreatic cancer patients** *Ann Oncol* **29**:223–229

17. Fahrman JF, Mao X, Irajizad E, et al (2020) **Plasma-Derived Extracellular Vesicles Convey Protein Signatures that Reflect Pathophysiology in Lung and Pancreatic Adenocarcinomas** *Cancers (Basel)* **12**
18. Hoshino A, Kim HS, Bojmar L, et al (2020) **Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers** *Cell* **182**:1044–1061
19. Iliuk A, Wu X, Li L, et al (2020) **Plasma-Derived Extracellular Vesicle Phosphoproteomics through Chemical Affinity Purification** *J Proteome Res* **19**:2563–2574
20. Wu X, Li L, Iliuk A, et al (2018) **Highly Efficient Phosphoproteome Capture and Analysis from Urinary Extracellular Vesicles** *J Proteome Res* **17**:3308–3316
21. Lopez YO Nunez, Iliuk A, Petrilli AM, et al (2022) **Proteomics and Phosphoproteomics of Circulating Extracellular Vesicles Provide New Insights into Diabetes Pathobiology** *Int J Mol Sci* **23**
22. Hinzman CP, Jayatilake M, Bansal S, et al (2022) **An optimized method for the isolation of urinary extracellular vesicles for molecular phenotyping: detection of biomarkers for radiation exposure** *J Transl Med* **20**
23. Hinzman CP, Singh B, Bansal S, et al (2022) **A multi-omics approach identifies pancreatic cancer cell extracellular vesicles as mediators of the unfolded protein response in normal pancreatic epithelial cells** *J Extracell Vesicles* **11**
24. Shuen TWH, Alunni-Fabbroni M, Ocal E, et al (2022) **Extracellular Vesicles May Predict Response to Radioembolization and Sorafenib Treatment in Advanced Hepatocellular Carcinoma: An Exploratory Analysis from the SORAMIC Trial** *Clin Cancer Res* **28**:3890–3901
25. Schmitz-Winnenthal FH, Volk C, Helmke B, et al (2006) **Expression of cytokeratin-20 in pancreatic cancer: an indicator of poor outcome after R0 resection** *Surgery* **139**:104–8
26. Guyon I, Weston J, Barnhill S, et al (2002) **Gene Selection for Cancer Classification using Support Vector Machines** *Machine Learning* **46**:389–422
27. Das P, Roychowdhury A, Das S, et al (2020) **sigFeature: Novel Significant Feature Selection Method for Classification of Gene Expression Data Using Support Vector Machine and t Statistic** *Front Genet* **11**
28. Huang L, Bockorny B, Paul I, et al (2020) **PDX-derived organoids model in vivo drug response and secrete biomarkers** *JCI Insight*
29. LeBleu VS (2020) **Kalluri R: Exosomes as a Multicomponent Biomarker Platform in Cancer** *Trends Cancer* **6**:767–774
30. Liu R, Lu S, Deng Y, et al (2016) **PSMB4 expression associates with epithelial ovarian cancer growth and poor prognosis** *Arch Gynecol Obstet* **293**:1297–307
31. Zhang X, Lin D, Lin Y, et al (2017) **Proteasome beta-4 subunit contributes to the development of melanoma and is regulated by miR-148b** *Tumour Biol* **39**
32. Zheng P, Guo H, Li G, et al (2015) **PSMB4 promotes multiple myeloma cell growth by activating NF-kappaB-miR-21 signaling** *Biochem Biophys Res Commun* **458**

33. Lee GY, Haverty PM, Li L, et al (2014) **Comparative oncogenomics identifies PSMB4 and SHMT2 as potential cancer driver genes** *Cancer Res* **74**:3114–26
34. Le Large TY, Mantini G, Meijer LL, et al (2020) **Microdissected pancreatic cancer proteomes reveal tumor heterogeneity and therapeutic targets** *JCI Insight* **5**
35. Law HC, Lagundzin D, Clement EJ, et al (2020) **The Proteomic Landscape of Pancreatic Ductal Adenocarcinoma Liver Metastases Identifies Molecular Subtypes and Associations with Clinical Response** *Clin Cancer Res* **26**:1065–1076
36. Zhang Y, Yang J, Li H, et al (2015) **Tumor markers CA19-9 CA242 and CEA in the diagnosis of pancreatic cancer: a meta-analysis.** *Int J Clin Exp Med* **8**
37. Duffy MJ, Sturgeon C, Lamerz R, et al (2010) **Tumor markers in pancreatic cancer: a European Group on Tumor Markers (EGTM) status report** *Ann Oncol* **21**:441–7
38. Amin MB, Greene FL, Edge SB, et al (2017) **The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging** *CA Cancer J Clin* **67**:93–99
39. Benjamini Y, Hochberg Y (1995) **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing** *Journal of the Royal Statistical Society: Series B (Methodological)* **57**:289–300
40. Raudvere U, Kolberg L, Kuzmin I, et al (2019) **g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)** *Nucleic Acids Res* **47**:W191–W198
41. Merico D, Isserlin R, Stueker O, et al (2010) **Enrichment map: a network-based method for gene-set enrichment visualization and interpretation** *PLoS One* **5**
42. Shannon P, Markiel A, Ozier O, et al (2003) **Cytoscape: a software environment for integrated models of biomolecular interaction networks** *Genome Res* **13**:2498–504
43. Reimand J, Isserlin R, Voisin V, et al (2019) **Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap** *Nat Protoc* **14**:482–517
44. Szklarczyk D, Gable AL, Lyon D, et al (2019) **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets** *Nucleic Acids Res* **47**:D607–D613
45. MacLean B, Tomazela DM, Shulman N, et al (2010) **Skyline: an open source document editor for creating and analyzing targeted proteomics experiments** *Bioinformatics* **26**
46. Meyer D, Dimitriadou E, Hornik K, et al (2015) **Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien**
47. Robin X, Turck N, Hainard A, et al (2011) **pROC: an open-source package for R and S+ to analyze and compare ROC curves** *BMC Bioinformatics* **12**

Editors

Reviewing Editor

Wadih Arap

Rutgers, The State University of New Jersey, Newark, United States of America

Senior Editor

Eduardo Franco

McGill University, Montreal, Canada

Reviewer #1 (Public review):

This study presents a large cohort of plasma-derived extracellular vesicle samples from 124 individuals, including patients with PDAC, benign pancreatic diseases and controls. The authors identified a panel of protein markers for the early detection of pancreatic cancer and validated in an external cohort.

<https://doi.org/10.7554/eLife.87369.2.sa2>

Reviewer #2 (Public review):

This work investigates the use of extracellular vesicles (EVs) in blood as a noninvasive 'liquid biopsy' to aid in differentiation of patients with pancreatic cancer (PDAC) from those with benign pancreatic disease and healthy controls, an important clinical question where biopsies are frequently non-diagnostic. The use of extracellular vesicles as biomarkers of disease has been gaining interest in recent history, with a variety of published methods and techniques, looking at a variety of different compositions ('the molecular cargo') of EVs particularly in cancer diagnosis (Shah R, et al, N Engl J Med 2018; 379:958-966).

This study adds to the growing body of evidence in using EVs for earlier detection of pancreatic cancer, identifying both new and known proteins of interest. Limitations in studying EVs in general include dealing with low concentrations in circulation and identifying the most relevant molecular cargo. This study provides validation of assaying EVs using the novel EVtrap method (Extracellular Vesicles Total Recovery And Purification), which the authors show to be more efficient than current standard techniques and potentially more scalable for larger clinical studies.

The strength of this study is in its numbers - the authors worked with a cohort of 124 cases, 93 of them which were PDAC samples, which considered large for an EV study (Jia, E et al. BMC Cancer 22, 573 (2022)). The benign disease group (n=20, between chronic pancreatitis and IPMNs) and healthy control groups (n=11) were relatively small, but the authors were not only able to identify candidate biomarkers for diagnosis that clearly stood out in the PDAC cohort, but also validate it in an independent cohort of 36 new subjects. Proteins they've identified as associated with pancreatic cancer over benign disease included PDCD6IP, SERPINA12 and RUVBL2. They were even able to identify a set of EV proteins associated with metastasis and poorer prognosis, which include the proteins PSMB4, RUVBL2 and ANKAR and CRP, RALB and CD55. Their 7-EV protein signature yielded an 89% prediction accuracy for the diagnosis of PDAC against a background of benign pancreatic diseases that is compelling and comparable to other studies in the literature (Jia, E. et al. BMC Cancer 22, 573 (2022)).

The limitations of this study are its containment within a single institution - further studies are warranted to apply the authors' 7-EV protein PRAC panel to multiple other cases at other institutions in a larger cohort.

Author response:

The following is the authors' response to the original reviews.

Reviewer #1 (Public Review):

In this manuscript, Bockorny, Muthuswamy, and Huang et al. performed proteomics analysis of plasma extracellular vesicles (EVs) from pancreatic ductal adenocarcinoma (PDAC) patients and patients with benign pancreatic diseases (chronic pancreatitis and intraductal papillary mucinous neoplasm, IPMN) to develop a 7-EV protein signature that predicts PDAC. Moreover, the authors identified PSMB4, RUVBL2, and ANKAR as being associated with metastasis. These studies provide important insight into alterations of EVs during PDAC progression and the data supporting predict PDAC with EV protein signatures are solid. However, there are certain concerns regarding the rigor and novelty of the data analysis and interpretation, as well as the clinical implications, as detailed below.

(1) Plasma EVs were characterized by transmission electron microscopy and nanoparticle tracking analysis to confirm their morphology and size. The authors should also include an analysis of putative EV markers (e.g., tetraspanins, syntenin, ALIX, etc.) to confirm that the analyzed particles are EVs.

We thank the reviewer for this comment. In the previous study from our co-authors who developed EVtrap method (PMID:32396726), they used electron microscopy and NTA, as well as quantification of typical EV protein markers, such as CD9, to confirm that particles isolated using EVtrap had typical characteristics of the extracellular vesicles. As such, these experiments were not replicated here. We added the following statement to the manuscript:

“Previous analyses using electron microscopy and nanoparticle tracking also confirmed that the vast majority of particles isolated by EVtrap had diameters between 100-200 nm, consistent with exosomes (PMID:32396726). In addition, EVtrap isolates demonstrates higher abundance of CD9, a common exosome marker, as compared to isolates from other traditional EV isolation methods such as size exclusion chromatography and ultracentrifugation (PMID:32396726)”

(2) The authors identified multiple over-expressed proteins in PDAC based on their foldchange and p-value; however, due to the heterogeneity of PDAC, it is necessary to show a heatmap displaying their abundance in all samples. High fold change does not necessarily indicate consistently high abundance in all PDAC samples.

We thank the reviewer for this suggestion. We have now included the heatmap in the new Supplementary Figure 3.

(3) PSMB4, RUVBL2, and ANKAR were identified as being associated with metastasis. The authors state that they intended to distinguish early and late-stage cancer samples, but it is unclear why they chose to compare metastatic and non-metastatic samples, as the non-metastatic group also includes late-stage cancer samples. This sentence should be rephrased to more accurately reflect the sample types profiled.

We thank the reviewer for pointing this out. We would like to clarify that this analyses shown in Figures 3B and 3C pertain to patients with Metastatic vs Non-Metastatic disease, not early versus late stage. We edited the text to ensure this information is clear.

(4) Non-metastatic and metastatic patients were separated based on global protein abundance. The samples within each group display significant heterogeneity, with some samples displaying similar patterns although they were classified into different groups (Figure 3A), and the samples within the same group, particularly the metastasis group, did not consistently exhibit similar patterns of protein abundance. The authors should clarify this point.

We thank the reviewer for this comment. The EV proteomic expression is anticipated not to show the exact pattern across of samples of each group. The purpose of this experiment depicted in Figure 3 heatmap is to show the enrichment for pattern of expressions, but we acknowledge that not all samples from the same group have the exact proteome pattern.

We added this statement in the discussion section:

“As expected, the EV proteomic profiles of PDAC patients exhibited significant heterogeneity. While the above mentioned markers exhibited strong association with disease states at population levels, their abundances in individual patients varied significantly. Those observations highlight the need to develop multi-protein panels for pancreatic cancer diagnosis and prognosis.”

(5) The authors performed the survival analysis on a set of EV proteins but did not specify the origin of these markers or how many markers were examined. The authors should show their abundances across different groups, such as different stages and metastasis status.

We thank the reviewer for the comments. The goal of this experiment was not to identify EV proteins that performed similarly well for diagnosing and prognostication. In Figure 3A, 3B and 3C, we identified EV proteins that had better performance for diagnosis of metastatic disease. In these experiments we made comparative analysis between patients with metastasis versus non-metastasis. In the experiment depicted in Figure 3D, the goal was to identify EV markers that had better performance is prognosticating outcomes as measured by overall survival, out of the markers identified in the previous experiments from Figure 3A. We would like to further clarify that based on our observation and others, it has become clear that EV profiles from cancer patients are highly heterogenous and we do not anticipate that a single marker will have sufficient test performance for cancer diagnosis or prognosis assessment when measured isolated. Rather, we anticipate that a panel of markers may yield better performance for diagnosis while a different combination of EV markers may have better performance for prognosis assessment.

(6) The classification model yielded a 100% accuracy, which may refer to AUC, in their discovery cohort, but it decreased to 89% in the independent cohort. This suggests that the authors have encountered overfitting issues with their model, where it performed well on the discovery cohort but did not generalize well to the independent cohort. The authors should clarify this point. The AUC score of the 7-EV signature is 0.89 and is not equivalent to prediction accuracy. In order to demonstrate prediction accuracy, the authors should show the confusion matrix of training and testing data as well as other evaluation metrics, such as accuracy, precision, and recall.

We thank the reviewer for providing these insightful comments. As you noted, the 7-biomarker signature machine learning model attained an impressive 100% accuracy within the internal Discovery Cohort, raising concerns about potential overfitting in the external validation dataset. Acknowledging the noted difference in AUROC of 0.11 in the external validation cohort, which surpasses the typical reported range of ~0.06-0.09, the model demonstrated a commendable AUROC of 0.89 in an independent patient cohort. Moreover,

the utilization of an alternate technology to measure protein abundance in the validation dataset, underscores the model's reproducibility and validity. We have provided the model metrics for both internal- and external-validation cohort. For these, please see updated Supplementary Figure 7, as well as the new Supplementary Figure 6 and Supplementary Figure 8. We also amended the discussion section to acknowledge that the validation cohort had limited sample size and proteins were measured in using a different method. Those factors likely contributed to the lower accuracy of predictions in the validation cohort. We addressed these limitations in the discussion section of the manuscript.

(7) The authors should include more details of their model and the process of selection of signatures to enhance the reproducibility and transparency of their methods.

We thank the reviewer for their valuable comments. To enhance clarity, we have incorporated additional information regarding the method employed for biomarker signature identification into the 'Methods Section' in page 23. We note that Supplementary Table 7a provides details on 'Sensitivity, Specificity, Precision, and AUC' for the 16 markers included in the external validation study. Additionally, Supplementary Table 7b presents the contingency table for 7-biomarker signature, offering insights into model accuracy for both the Internal-Discovery and External Validation cohorts.

Reviewer #2 (Public Review):

The authors intended to identify a protein signature in extracellular vesicles of serum to distinguish pancreatic ductal adenocarcinoma from benign pancreatic diseases.

A major strength of the work presented is the valuable profiling of a significant number of patient samples, with a rich cohort of patients with pancreatic cancer, benign pancreatic diseases, and healthy controls. However, despite the strong cohorts presented, the numbers of patient samples for benign pancreatic diseases as well as controls were very limited.

Also, the method used to isolate vesicles, EVTrap, recognizes double bilayers, which means that it can detect cellular debris and apoptotic bodies, which are very common in the circulation of patients that are undergoing chemotherapy. It would be important to identify the patients that are therapy naïve and the ones that are not because of this possible bias.

We thank the Reviewer for these comments. We want to point out that the experiments presented in Supplementary Figure 1 (Transmission electron microscopy images and Nanoparticle tracking analysis) confirm that the vesicles isolated with EVTrap are not cellular debris and apoptotic bodies. Rather, these structures are in the nano range expected for exosomes. This is further supported by the additional work from our co-author and collaborator describing the development of EVtrap and its performance in isolating exosomes when compared to other traditional methods such as ultracentrifugation and size exclusion chromatography (PMID:32396726).

As per the Reviewer's request, we have provided an additional heatmap figure depicting whose patients are treatment naïve to differentiate from those who have received treatment (revised Figure 2C).

Additionally, the transmission electron microscopy data reflect this heterogeneity of the samples, also with little identification of double bilayered vesicles. It would be important to identify some extracellular vesicles markers in those preparations to strengthen the quality of the samples analyzed.

We appreciate the comment from the Reviewer and acknowledge the importance of identifying exosome markers on the isolate from EVtrap. These experiments have already been done and are reported in the original paper describing the development of this method by our co-authors in a separate work. In the manuscript PMID: 30080416, our collaborators demonstrated the detection of CD9, a well-known exosome marker, using Western Blot from isolates using EVtrap or ultra-centrifugation, a traditional technique to isolate exosomes. This work showed that EVtrap yielded much higher recovery rate of exosomes with lower contamination from soluble proteins. We did not repeat these already published experiments, but we amended our manuscript to reference these results.

What is more, previously published work with this same methodology identifies around 2000 proteins per sample. It would be important to explain why in this study there seems to be a reduction in more than 50% of the amount of proteins identified in the vesicles.

We thank the Reviewer for pointing out this important detail. In the previous work in which EVtrap was developed by our co-authors, the blood samples were processed using a different protocol, with shorter centrifugation (2,500g for 10 min) (PMID: 32396726). In the current work, we employed three centrifugation steps. As detailed in the Methods section of the manuscript, blood samples were centrifuged at 1,300g for 15 min. Then plasma was removed from the top carefully avoiding cell pellet; Repeat centrifugation of plasma at 2,500g for 15 min; Again, plasma was removed from the top carefully avoiding cell pellet; Third centrifugation at 2,500g for 15 min. This more extensive centrifugation process was intended to further increase the removal of platelets, apoptotic bodies, and other large particles and aggregates. Accordingly, we anticipate that the additional centrifugation steps decreased the contamination of our isolates but may have also decreased the amount of exosome proteins, hence the lower amount of exosome proteins identified in our study as compared to the original study from our co-authors (PMID: 32396726).

One of the proteins that constantly surges on the analysis is KRT20. It would be important to proceed with the analysis by first filtering out possible contaminants of the proteomics, of which keratins are the most common ones.

We thank the Reviewer for this comment. We would like to point out that we do believe that KRT20 is, in fact, cancer related and a not a contaminant. This is supported by our results presented in this manuscript showing enrichment of KRT20 in PDAC cases, and lower expression in benign samples. If this protein was a contaminant, its expression would be found uniformly in all samples, there would be no apparent reason for different expression between malignant vs benign cases, as all samples were processed following the same procedures. In addition, increased expression of KRT20 in PDAC tissues has also been reported by others. For instance, in a study by Schmiz-Winnthal (PMID: 16364723), the authors showed that Cytokeratin 20 (KRT20) were expressed in 76% of PDAC patients and expression of KRT20 was associated with poor survival after surgical resection. Based on these observations, we believe that the KRT20 identified in our study is indeed a tumor associated EV protein rather than contamination.

Finally, none of the 7-extracellular vesicle protein signatures has been validated by other techniques, such as western blot, in extracellular vesicles isolated by other, standard, methods, such as size exclusion chromatography.

A distinct technique for protein analysis was done but not a different method of isolation of these vesicles. This would strengthen the results and the origin of the proteins.

We appreciate the Reviewer's comment. We would like to again emphasize that the goal of this manuscript was not to compare the performance of EVtrap with other traditional EV

isolation approaches such as ultracentrifugation and size exclusion chromatography. The main goal of study is to determine proteomic profiles of EVs isolated from clinical samples and provide such information to research community for further studies. As the Reviewer points out, proteins in EVs are highly heterogeneous which highlight the complexity of EV biology and interpatient heterogeneity of pancreatic cancer. We do not anticipate the development of EV-based markers for pancreatic diagnosis can be achieved by a single team, but by a community of researchers. We hope information presented in the current study will help other researchers identify additional candidates for validation in future work. Nonetheless, we edited the manuscript to discuss the limitation of not doing cross-validation of protein detection using a different method.

The conclusions that are reached do not fully meet the proposed aims of the identification of a protein signature in circulating extracellular vesicles that could improve early detection of the disease. The authors did not demonstrate the superiority of detection of these proteins in extracellular vesicles versus simply performing an ELISA, nor their superiority with respect to the current standard procedure for diagnosis.

We would like to clarify to the Reviewer that the goal of this manuscript was not to prove superiority of the EV signature biomarker in diagnosing pancreatic cancer as compared to current standard of care (SOC) practice, i.e., CT scans, endoscopic ultrasound and CA19-9. In order to prove such superiority, one would require a large, randomized phase III trial with several hundred patients. This was not the pursue of our discovery EV proteomics study and we double checked our manuscript to ensure no such claim was made. Rather, we aimed at developing a new pipeline for discovery of new EV biomarkers and we believe we were able to prove that this approach was successful in discovering a new class of biomarkers based on proteins expressed on extra-cellular vesicles that have predominant expression on patients with pancreatic cancer. Future studies should continue to advance this field with goals of improving on the current standard of care diagnostic methods.

The authors also suggest that profiling of circulating extracellular vesicles provides unique insights into systemic immune changes during pancreatic cancer development. How is this better than a regular hemogram is not clear.

We would like to clarify that the overall goal of this study is to provide patient-relevant information for the research community to further investigate biology of extracellular vesicles. For the state 'unique insights into systemic immune changes' we referred to the fact that we discovered EVs carrying proteins involved in immune responses. Previous studies have shown that EVs play important roles in cell-cell communication, discoveries from our study provide candidates for future studies on cellular mechanisms underlying immune regulation during pancreatic cancer development.

Finally, it would be important to determine how this signature compares with many others described in the literature that have the exact same aim. Why and how would this one be better?

We would like to again clarify that comparing the diagnostic performance of the EV biomarkers discovered in the study against standard of care methods (CA19-9, ctDNA, CT scan) was beyond the scope of this discovery EV proteomics work. We reviewed the manuscript to ensure that no claims were made as far as superiority against point-of-care tests available in clinic.

Reviewer #3 (Public Review):

This work investigates the use of extracellular vesicles (EVs) in blood as a noninvasive 'liquid biopsy' to aid in the differentiation of patients with pancreatic cancer (PDAC) from those with benign pancreatic disease and healthy controls, an important clinical question where biopsies are frequently non-diagnostic. The use of extracellular vesicles as biomarkers of disease has been gaining interest in recent history, with a variety of published methods and techniques, looking at a variety of different compositions ('the molecular cargo') of EVs particularly in cancer diagnosis (Shah R, et al, N Engl J Med 2018; 379:958-966).

This study adds to the growing body of evidence in using EVs for earlier detection of pancreatic cancer, identifying both new and known proteins of interest. Limitations in studying EVs, in general, include dealing with low concentrations in circulation and identifying the most relevant molecular cargo. This study provides validation of assaying EVs using the novel EVtrap method (Extracellular Vesicles Total Recovery And Purification), which the authors show to be more efficient than current standard techniques and potentially more scalable for larger clinical studies.

The strength of this study is in its numbers - the authors worked with a cohort of 124 cases, 93 of them which were PDAC samples, which are considered large for an EV study (Jia, E et al. BMC Cancer 22, 573 (2022)). The benign disease group (n=20, between chronic pancreatitis and IPMNs) and healthy control groups (n=11) were relatively small, but the authors were not only able to identify candidate biomarkers for diagnosis that clearly stood out in the PDAC cohort, but also validate it in an independent cohort of 36 new subjects.

Proteins they have identified as associated with pancreatic cancer over benign disease included PDCD6IP, SERPINA12, and RUVBL2. They were even able to identify a set of EV proteins associated with metastasis and poorer prognosis, which include the proteins PSMB4, RUVBL2 and ANKAR and CRP, RALB and CD55. Their 7-EV protein signature yielded an 89% prediction accuracy for the diagnosis of PDAC against a background of benign pancreatic diseases that is compelling and comparable to other studies in the literature (Jia, E. et al. BMC Cancer 22, 573 (2022)).

The limitations of this study are its containment within a single institution - further studies are warranted to apply the authors' 7-EV protein PRAC panel to multiple other cases at other institutions in a larger cohort.

We are very thankful to the Reviewer for the positive feedback. We are similarly optimistic that EV-based biomarkers will assist future researchers to develop better diagnostic assays for patients with pancreatic cancer, as well as other tumor types lacking accurate blood-based tests.

<https://doi.org/10.7554/eLife.87369.2.sa0>