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

Bruno Bockorny[1,2*], Lakshmi Muthuswamy[3*], Ling Huang[4*], Marco Hadisurya[5], Christine Maria Lim[6], Leo L. Tsai[2,7], Ritu R. Gill[2,7], Jesse L. Wei[2,7], Andrea J. Bullock[1,2], Joseph E. Grossman[8], Robert J. Besaw[1], Supraja Narasimhan[9], W. Andy Tao[5], Sofia Perea[1], Mandeep S. Sawhney[2,10], Steven D. Freedman[2,10], Manuel Hidalgo[11,12*], Anton Iliuk[13*†], Senthil K. Muthuswamy[14*†]

Affiliations:

1. Division of Medical Oncology, Beth Israel Deaconess Medical Center, Boston, MA, USA
2. Harvard Medical School, Boston, MA, USA
3. Blueprint Medicines, Cambridge, MA, USA
4. Henry Ford Cancer Institute, Detroit, MI, USA
5. Department of Biochemistry, Purdue University, West Lafayette, IN, USA
6. Nanyang Technological University, Singapore
7. Department of Radiology, Beth Israel Deaconess Medical Center, Boston, MA, USA
8. Agenus Inc, Lexington, MA, USA
9. Deciphera Pharmaceuticals, Waltham, MA, USA
10. Division of Gastroenterology, Beth Israel Deaconess Medical Center, Boston, MA
11. Division of Hematology-Oncology, Weill Cornell Medical College, New York, NY, USA
12. New York-Presbyterian Hospital, New York, NY, USA
13. Tymora Analytical Operations, West Lafayette, IN, USA
14. National Cancer Institute, NIH, Bethesda, MD, USA

Running Title: Large Scale EV Proteomics Resource to Discover PDAC Biomarkers

Keywords: Proteomics, exosomes, extracellular vesicles, chronic pancreatitis, pancreatic cancer

Correspondence:

Senthil K. Muthuswamy PhD

Laboratory of Cancer Biology and Genetics

LCBG, Bldg. 37, Rm4068, 37 Convent Dr. Bethesda, MD 20892 |Ph:240-858-7734

Senthil.muthuswamy@nih.gov

Anton Iliuk PhD

1201 Cumberland Ave, Suite 209, West Lafayette, IN 47906

anton.iliuk@tymora-analytical.com

* Equal contribution † Corresponding authors

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 earlier. 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

56 novel cEVs that may assist in the development of biomarkers and improve the outcomes of
57 patients with PDAC.

58 **INTRODUCTION**

59

60 Pancreatic ductal adenocarcinoma (PDAC) has the worst prognosis of all common tumors, with
61 a 5-year survival of 12%[1]. With rising incidence, it is expected that PDAC will become the
62 second leading cause of cancer-related deaths by 2030[2]. A critical factor for this dismal
63 development is the late diagnosis, with less than 20% of patients presenting with a potentially
64 resectable and curable tumor[3-5]. Earlier cancer diagnosis could increase the survival rates by an
65 estimated 5-fold, and more reliable and real-time assessment of treatment effects in patients
66 with cancer could improve quality of life and reduce healthcare costs[6,7]. Unfortunately, there
67 are no credentialed serologic biomarkers with high enough performance to assist in the early
68 detection of PDAC. The best-established biomarker for PDAC, carbohydrate antigen 19-9 (CA19-
69 9), is fraught with poor sensitivity and specificity and is only used for monitoring disease on
70 treatment or after surgical resection[8,9].

71

72 Extracellular vesicles (EVs), including exosomes and microvesicles, are nanosized particles
73 released by most cell types and can be detected in the circulation[10]. EVs play important roles in
74 the transmission of oncogenic and inflammatory signals[11], communications between cells and
75 their microenvironment[12]. In addition, exoDNA, exoRNA and protein profiles highly reflect
76 parental cells, therefore offering an attractive strategy for diagnosing cancers non-invasively by
77 analyzing EVs in the circulation[11,13]. Previous studies employed EVs to discover biomarkers for

78 PDAC[13-16], however those discovery proteomics experiments were carried out using cell lines or
79 tumor tissue, which are not representative of the heterogeneity of human PDAC and are unable
80 to recapitulate the systemic responses to cancer[14-16]. In addition, the EV biomarkers discovered
81 in those studies have been compared only against healthy controls[14-16]. It is unclear how they
82 would perform in subjects with underlying benign diseases of the pancreas, which is highly
83 desirable from the clinical standpoint as many patients with PDAC have underlying chronic
84 pancreatitis and cysts.
85
86 To meet this need, we conducted a large EV proteomics study from peripheral blood across a
87 range of patients with pancreatic cancer, benign pancreatic diseases such as chronic
88 pancreatitis and intraductal papillary mucinous neoplasm (IPMN), and healthy controls.
89 Circulating EV (cEV) proteins detected included those involved in metabolism and immune
90 regulation, in addition to proteins involved in protein binding, exocytosis, endocytosis and
91 regulation of cellular protein localization that have been identified in previous studies[17,18]. Our
92 study represents the largest EV proteomics profiling dataset of pancreatic cancer, therefore
93 providing a resource to the research community to further the understanding of EV roles in
94 pancreatic cancer biology and biomarker discovery. To demonstrate the utility of this dataset,
95 we identified multiple candidate biomarkers for cancer diagnosis and validated several in an
96 independent cohort of patients with pancreatic cancer. In addition, we identified a set of cEV
97 proteins associated with metastasis and prognosis which could provide a valuable resource for
98 future biomarker studies as well as understanding of systemic changes during pancreatic tumor
99 progression.

100

101

102 **RESULTS**

103

104 **Proteomics Characterization of Circulating EVs**

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

116

117 We employed the novel EVtrap method (Extracellular Vesicles Total Recovery And Purification)
118 to capture EVs from plasma samples and overcome the traditional laborious techniques for EV
119 isolation, which are not scalable for large clinical studies. As described in recent reports, EVtrap
120 is a magnetic bead-based isolation method that enables highly efficient capture of EVs from
121 biofluids, confirmed by multiple common EV markers[19-23]. Over 95% recovery yield can be
122 achieved by EVTRAP with less contamination from soluble proteins, a significant improvement
123 over current commercially available methods as well as ultracentrifugation[19,21,24].

124

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

129

130 First, to confirm that EVtrap can efficiently isolate extracellular vesicles from plasma, a test
131 plasma sample was processed to remove platelets and other large particles and enriched for
132 EVs using EVtrap beads (see **methods** for details). Transmission electron microscopy (TEM)
133 analysis of the EV pellet showed cup-shaped extracellular vesicles (exosomes and microvesicles)
134 (**Supplementary Figure 1A**), and nanoparticle tracking analysis (NTA) using ZetaView
135 instrument (Particle Metrix) demonstrated that the isolated EVs were in the 100−200 nm
136 diameter range, with a mean diameter of 152 nm (**Supplementary Figure 1B**). Second, to
137 assess the technical reproducibility of the EV proteomics approach, the test plasma sample was
138 processed in six replicates and Pearson correlation analysis revealed a very high correlation (r2
139 > 0.97) between replicates (**Supplementary Figure 2A, Supplementary Table 2**). These results
140 provided the confidence to proceed with the analysis of our discovery set of plasma from 124
141 subjects. In this cohort, we identified 1,708 unique proteins (**Supplementary Table 3**). The
142 number of unique EV proteins detected per 100µL of plasma sample varied from 817 to 1,128,
143 with an average of 912 unique proteins per sample (**Supplementary Figure 2B**). We did not
144 observe differences between non-tumor and tumor samples regarding the overall number of
145 EV proteins identified. Within the PDAC group, we did not observe significant differences in the

146 average number of EV proteins detected for different disease stages. Collectively, these data
147 demonstrate high reproducibility of EV isolation and robust label-free MS quantification of
148 cEVs.
149
150 **Diseases of the Pancreas Express Distinct Circulating EV Proteome Compared to Controls**
151 Next, we aimed at identifying specific cEV proteins associated with clinical parameters with the
152 potential to serve as diagnostic biomarkers. We first compared the proteomics profile of
153 individuals with underlying pancreatic diseases (PDAC, chronic pancreatitis and IPMN) against
154 healthy controls. We selected EV proteins expressed in at least 50% of subjects in the disease
155 group with a fold change of expression ≥2 or ≤2 compared to controls and p-value ≤0.01 after
156 adjusting for multiple testing. A total of 207 proteins were identified that met the criteria, with
157 the largest number of differentially expressed markers in PDAC (176), followed by chronic
158 pancreatitis (55) and IPMN (3) (**Supplementary Table 4**). Principal component analysis (PCA) of
159 these markers showed control samples as a tight cluster segregated away from PDAC samples
160 but closer to IPMN and chronic pancreatitis patients (**Figure 2A**).
161
162 **Circulating EV Proteome Discriminates Pancreatic Cancer from Benign Pancreatic Diseases**
163 To further assess the potential of cEV proteins for cancer detection, we compared proteomic
164 profiles of cEVs between patients with PDAC with those with underlying benign diseases of the
165 pancreas (chronic pancreatitis and IPMN). We identified 182 differentially expressed proteins in
166 malignant cases (92 over-expressed and 90 with reduced expression) (**Supplementary Table 5)**.
167 Several of those markers had remarkable overexpression in PDAC (greater than 10-fold),

168 including PDCD6IP, SERPINA12, RUVBL2, among others, as shown in the volcano plot (**Figure**
169 **2B**). Unsupervised clustering showed a clear separation between PDAC and benign pancreatic
170 diseases. Individuals with IPMN were more closely related to controls, whereas chronic
171 pancreatitis cases were more related to PDAC **(Figure 2C)**. In addition, the PDAC cohort was
172 separated into two subgroups: the first, enriched for early-stage tumors and more closely
173 related to the other pancreatic diseases (chronic pancreatitis and IPMN); the second, enriched
174 for advanced and metastatic cases with expression profiles further apart from early-stage
175 cancer and pancreatic diseases (**Figure 2C**). We further noticed that some proteins such as
176 PDCD6IP, SERPINA12, KRT20 showed statistically significant population-wise enrichment in
177 pancreatic cancer compared to benign pancreatic diseases (**Figure 2D**). Together, these data
178 indicate the existence of a set of EV markers that can separate controls, benign and malignant
179 pancreatic diseases, as well as proteins that separate early versus late-stage PDAC, suggesting
180 their potential to serve as diagnostic biomarkers.
181
182 **Functional and Systems Biology of cEV Proteome**
183 To gain molecular insight into the functions of the previously identified 182 proteins
184 differentially expressed in pancreatic cancer as compared to benign pancreatic diseases, we
185 conducted pathway analysis using the Gene Ontology (GO) and REACTOME databases
186 (**Supplementary Table 5**). We identified protein modules in protein localization, biomolecule
187 binding/docking, peptidase activities among changes enriched in PDAC compared to benign
188 diseases (**Supplementary Figure 3**). Interestingly, KRT20 (keratin 20), a gastrointestinal
189 epithelia-associated keratin, was increased in PDAC patient EVs, while keratins associated basal

190 cells, KRT4, KRT15, and KRT3, were reduced. KRT20 overexpression is frequently found in
191 pancreatic tumor tissues and correlates with poor prognosis[25], suggesting a biological basis for
192 their high levels in the cEVs of PDAC patients.
193
194 Interestingly, proteins associated with immunological functions showed complex regulation
195 with increased representation of leukocyte mediated immunity (GO:0002443), leukocyte
196 degranulation (GO:0043299), myeloid leukocyte activation (GO:0002274), and decrease in Fc
197 receptor signaling (GO: 0038093), regulation of complement activation (GO:0030449), and
198 immune effector process (GO:0002252) (**Supplementary Table 5**). These data suggest that
199 direct profiling of cEVs from patient plasma provided unique insights into systemic changes in
200 immune biology during pancreatic cancer development, which is lacked in analysis restricted to
201 tissue or cell models.
202
203 **Circulating EV Proteomics Reveal Markers Associated with Metastasis and Worse Prognosis**
204 We then investigated whether cEV proteins can assist in the distinction of early versus late-
205 stage pancreatic cancer. We compared the cEV proteome profiles of individuals with metastatic
206 cancer to those without metastasis and identified 85 proteins differentially expressed between
207 the two groups (**Supplementary Table 6)**. Supervised clustering between metastatic and non-
208 metastatic diseases showed a clear separation with two distinct expression patterns (**Figure**
209 **3A**). The majority of non-metastatic cases represented stages 1 and 2 diseases, with a minority
210 representing locally advanced stage 3 disease. In particular, PSMB4, RUVBL2 and ANKAR (**Figure**
211 **3B**) EV protein levels were increased, whereas RAP2B, SERPINA12 and IGLV4-69 abundance

212 levels were decreased in the cEVs of patients with metastatic disease (**Figure 3C**). Together,
213 these findings suggest the presence of a core set of cEV proteins with the potential to
214 distinguish early versus metastatic pancreatic cancer.
215
216 We further analyzed whether the expression of certain cEV proteins had prognostic relevance
217 in our cohort. We first classified individuals with PDAC as having low or high expression of any
218 given markers based on each marker's first and third quartile. Survival was estimated by the
219 Kaplan Meier method. We identified that the cEV expression of RALB, CRP, and CD55 had a
220 significant correlation with worse survival, while PDCD6IP expression had a trend for improved
221 outcomes (**Figure 3D**).
222
223 **Validation of cEV Markers Using Parallel Reaction Monitoring and Identification an EV Protein**
224 **Signature for Pancreatic Cancer Diagnosis**
225 Because pancreatic cancer is extremely heterogeneous, the chance of identifying a single
226 biomarker with sufficient diagnostic performance is likely low. Instead, the identification of a
227 panel of candidate markers may have enhanced diagnostic performance.
228
229 To identify a signature that shows the most discriminatory power between 'benign diseases'
230 and 'PDAC,' we employed a binary classification approach using Support Vector Machines
231 (SVM). Classification models, built based on a large number of proteins, contain irrelevant
232 markers that can reduce the predictive accuracy. Hence, we implemented a consensus feature
233 selection method based on two algorithms: one using recursive feature elimination (RFE)

234 algorithm (SVM-RFE)[26] and second, RFE combined with a non-parametric Wilcox rank test
235 (sigFeature)[27]. The top 16 markers were selected whose classification performance can be
236 tested in the independent validation cohort (**Supplementary Table 7**). A summary of selection
237 process is shown in **Supplementary Figure 4**. The classification performance of these 16
238 markers, individual and in all combinations, were tested using 80% training data and evaluated
239 in the remaining 20% test data. The quality of training was assessed using five repetitions of 10-
240 fold cross-validation. The optimal kernel parameters were estimated by tuning over a wide
241 range of values. Receiver operating characteristic (ROC) analysis was used as the metric to
242 assess the performance of the classifier model. We found a set of 7-EV protein signature
243 comprised of RUVBL2, PDCD6IP, ATP5F1, DLD, KRT20, CCT4, and SERPINAI2, that gave 100%
244 accuracy when tested in the discovery cohort (**Supplementary Table 7**). Recurrence of these
245 putative markers in our dataset varied from 55% to 97%.
246
247 The model was further validated on an independent validation cohort whose proteome was
248 obtained using an alternate technology, parallel reaction monitoring (PRM) mass spectrometry.
249 The markers chosen for validation included 16 markers selected for SVM classification model
250 and an additional 9 markers to result in top 25 markers that are significantly differentially
251 expressed in the discovery cohort with a fold change increase in PDAC ≥ 5.5 and p-value ≤0.01
252 (**Methods**, **Supplementary Table 8**). The independent validation cohort consisted of 36 new
253 subjects (24 with PDAC, 6 with chronic pancreatitis, and 6 with IPMN) (**Supplementary Table 9**).
254 A total of 10 proteins, including all 7 signature proteins, showed a significant difference ($p <$
255 0.05) in patients with PDAC as compared to benign pancreatic diseases (**Figure 4A**). The

256 performance of individual validated markers according to the specific underlying disease in the
257 validation cohort is presented in **Supplementary Figure 5**. The performance of 7-EV protein
258 signature was further tested using SVM model, in our independent validation cohort, yielding
259 an 89% prediction accuracy (**Figure 4B**). As expected, we observed that no single marker
260 achieved sufficiently high sensitivity and specificity as the combined model for the diagnosis of
261 pancreatic cancer.
262
263
264 **DISCUSSION**
265
266 Extracellular vesicles hold a great promise as a source of potential biomarkers, making them
267 attractive candidates for liquid biopsy tests. Previously, we reported that organoid cultures of
268 pancreatic cancer could serve as models to discover tissue-derived EV proteins with high
269 specificity for PDAC, as opposed to chronic pancreatitis and other benign gastrointestinal
270 diseases[28]. A shortcoming of these tissue-based studies is the inability to discover markers
271 associated to the systemic responses to cancer. Here, we performed a large-scale,
272 comprehensive analysis of circulating EV proteomes directly from plasma samples of 124
273 patients, with subsequent validation in a separate cohort of 36 patients. To our knowledge, this
274 represents the largest proteomics profiling dataset of circulating EVs conducted in pancreatic
275 cancer to date. Using this dataset, we identified and validated new EV markers from plasma
276 that distinguish patients with pancreatic cancer from subjects with benign pancreatic diseases.
277 Furthermore, we discovered several cEV proteins associated with metastatic disease and poor

278 prognosis. In contrast to the prior studies of experimental cell models or tissues extracts that
279 were examined only against healthy subjects[14-16], we report the identification of EV proteins in
280 plasma of patients with pancreatic cancer compared to patients with underlying pancreatic
281 diseases, which is clinically relevant as many patients with pancreatic cancer have underlying
282 chronic inflammation and premalignant cystic lesions.
283
284 In addition, our study demonstrated the feasibility of utilizing the EVtrap method[19,20] to process
285 patient samples in large scale within a reasonable time frame, allowing the discovery of
286 hundreds of EV proteins directly from a small volume (100uL) of plasma samples. This
287 methodological advance can be adopted for biomarker discovery in other cancer types. Other
288 workflows traditionally employed for EV isolation from blood samples require laborious
289 techniques including lengthy ultracentrifugation steps which are unsuitable for large scale
290 studies[29].
291
292 In this dataset, we identified several EV proteins as significantly associated with metastasis or
293 survival. For instance, PSMB4 and RUVBL2 levels were increased in cEVs of patients with
294 metastatic PDAC. Notably, PSMB4 (proteasome subunit beta type-4), a protein of the ubiquitin-
295 proteasome degradation pathway, has been identified as the first proteasomal subunit with
296 oncogenic properties and associated to poor prognosis in several tumors including melanoma,
297 breast and ovarian cancers[30-33].
298

299 We also discovered RALB, CRP and CD55 expression on EVs to have a significant correlation
300 with poor survival, while PDCD6IP expression was associated with improved outcomes.
301 Interestingly, PDCD6IP (programmed cell death 6-interacting protein), was also identified as a
302 PDAC-enriched protein in the tissue-based proteomics studies from Le Large *et.al*[34] and
303 Hoshino *et.al*[18]. In line with our findings, tissue expression of PDCD6IP in liver metastasis of
304 pancreatic cancer has been found to also correlate with improved prognosis in patients with
305 PDAC in the study of Law *et.al*[35]. Collectively, these data suggest that some tissue-specific
306 proteins can be detected in circulating EVs and their quantifiable levels in the blood may have
307 the potential to serve as diagnostic or prognostic biomarkers in pancreatic cancer.
308
309 In our validation studies, all seven putative markers identified from the model were significantly
310 enriched in the plasma of PDAC patients. Based on the top seven markers, we derived a 7-EV
311 protein panel that yielded an 89% prediction accuracy for diagnosing pancreatic cancer. A
312 recent modeling study showed that a new diagnostic assay for PDAC would have to perform
313 with a minimum sensitivity of 88% and a specificity of 85% to reduce healthcare expenditure
314 and prolong survival[6]. Serum CA19-9, the best-established blood test for PDAC, has a pooled
315 sensitivity of 75.4% and a specificity of 77.6%[36]. It commonly rises late in the disease and may
316 be elevated in nonmalignant conditions such as biliary obstruction and pancreatitis, making it
317 unsuitable as a diagnostic biomarker for PDAC[37]. As such, our 7-EV protein signature with 89%
318 prediction accuracy serves as a proof-of-concept and has the potential to facilitate the further
319 development of biomarker tests for pancreatic cancer. We anticipate that for clinical use
320 application, an even higher diagnostic performance is needed. Future studies are warranted to

321 investigate if combining our validated cEV proteins with other biomarkers such as cell free DNA,
322 serum proteins or metabolites, as a multi-analyte biomarker assay, would yield higher accuracy
323 in diagnosing pancreatic cancer.
324
325 While our work involved a large cohort with 160 patients, the single-center nature is an
326 inherent limitation of our study. Also, it would be ideal to perform validation with a larger
327 cohort of controls to achieve greater statistical power. To balance this limitation, we increased
328 the rigor of our validation by selecting controls with underlying benign diseases of pancreas as
329 opposed to healthy volunteers. Thus, the performance of our 7-EV protein PDAC panel should
330 be cross validated in larger and multicenter populations.
331
332 With no major treatment breakthrough for pancreatic cancer in the last decade, every effort
333 should be made to diagnose this deadly cancer at earlier stages and to discover new proteins
334 involved in tumorigenesis. Our study provides a valuable open resource to the scientific
335 community with a comprehensive catalog of novel proteins packaged inside circulating EVs that
336 may assist in the development of novel biomarkers and improve the outcomes of patients with
337 pancreatic cancer.

338 **METHODS**

339

340 **Study Design and Patient Demographics**

341 We conducted this study at Beth Israel Deaconess Medical Center with the approval of the
342 Harvard Cancer Center Institutional Review Board. All subjects provided written informed
343 consent. Clinical data and blood samples were prospectively collected from 2017 to 2019 from
344 patients with pancreatic cancer, chronic pancreatitis, intraductal papillary mucinous neoplasms
345 (IPMN), and age-matched controls. A total of 124 patients, including PDAC (N=93), chronic
346 pancreatitis of different etiologies (N=12), IPMN (N=8), and controls (N=11), were included in
347 the discovery cohort. PDAC diagnosis was established by histology or cytology, and staging was
348 performed according to the American Joint Committee on Cancer guidelines (8th Edition
349 2016)[38](**Supplementary Table 1**). For the independent validation cohort, a total of 36 patients
350 were enrolled, including PDAC (N=24), IPMN (N=6), and chronic pancreatitis (N=6)
351 (**Supplementary Table 9**).

352

353 **Plasma Sample Collection and Processing**

354 All blood samples were collected and processed following the same standard operating
355 procedure optimized for EV analysis and included the following steps: (i) whole blood was
356 collected into one 10ml yellow-top tube containing acid citrate dextrose; (ii) blood was mixed
357 by gently inverting the tube five times; (iii) vacutainer tubes were stored upright at room
358 temperature (RT); (iv) samples were centrifuged at 1,300g for 15 min in RT; (v) plasma was
359 removed from the top carefully avoiding cell pellet; (vi) repeat centrifugation of plasma at

360 2,500g for 15 min in RT; (vii) again, plasma was removed from the top carefully avoiding cell
361 pellet; (viii) third centrifugation at 2,500g for 15 min in RT, then samples were aliquoted to be
362 stored at -80^{0}C.
363
364 **Extracellular Vesicle Isolation from Plasma**
365 We employed EVtrap for EV isolation from plasma samples[19]. EVtrap beads were provided by
366 Tymora Analytical (West Lafayette, IN) as a suspension in water and were used as previously
367 described in more details[19,20]. Briefly, 100 µL plasma samples were diluted 20 times in the
368 diluent buffer, the EVtrap beads were added to the samples in a 1:2 v/v ratio, and the samples
369 were incubated by end-over-end rotation for 30 min according to the manufacturer's
370 instructions. After supernatant removal using a magnetic separator rack, the beads were
371 washed with PBS, and the EVs were eluted by a 10 min incubation with 200 mM triethylamine
372 (TEA, Millipore-Sigma). The samples were fully dried in a vacuum centrifuge.
373
374 **Preparation of EV samples**
375 The isolated and dried EV samples were lysed to extract proteins using the phase-transfer
376 surfactant (PTS) aided procedure. The proteins were reduced and alkylated by incubation in 10
377 mM TCEP and 40 mM CAA for 10 min at 95°C. The samples were diluted fivefold with 50 mM
378 triethylammonium bicarbonate and digested with Lys-C (Wako) at 1:100 (wt/wt) enzyme-to-
379 protein ratio for 3 h at 37°C. Trypsin was added to a final 1:50 (wt/wt) enzyme-to-protein ratio
380 for overnight digestion at 37°C. To remove the PTS surfactants from the samples, the samples
381 were acidified with trifluoroacetic acid (TFA) to a final concentration of 1% TFA, and ethyl

382 acetate solution was added at a 1:1 ratio. The mixture was vortexed for 2 min and then
383 centrifuged at 16,000 × g for 2 min to obtain aqueous and organic phases. The organic phase
384 (top layer) was removed, and the aqueous phase was collected. This step was repeated once
385 more. The samples were dried in a vacuum centrifuge and desalted using Top-Tip C18 tips
386 (Glygen) according to the manufacturer's instructions. The samples were dried completely in a
387 vacuum centrifuge and stored at -80°C.
388
389 **LC–MS Analysis of Plasma EV Proteome**
390 Approximate 1 µg of each dried peptide sample was dissolved in 10.5 µL of 0.05%
391 trifluoroacetic acid with 3% (vol/vol) acetonitrile containing spiked-in indexed Retention Time
392 Standard containing 11 artificially synthetic peptides (Biognosys). The spiked-in 11-peptides
393 standard mixture was used to account for any variation in retention times and to normalize
394 abundance levels among samples. 10 µL of each sample was injected into an Ultimate 3000
395 nano UHPLC system (Thermo Fisher Scientific). Peptides were captured on a 2-cm Acclaim
396 PepMap trap column and separated on a heated 50-cm Acclaim PepMap column (Thermo
397 Fisher Scientific) containing C18 resin. The mobile phase buffer consisted of 0.1% formic acid in
398 ultrapure water (buffer A) with an eluting buffer of 0.1% formic acid in 80% (vol/vol)
399 acetonitrile (buffer B) run with a linear 60-min gradient of 6–30% buffer B at a flow rate of 300
400 nL/min. The UHPLC was coupled online with a Q-Exactive HF-X mass spectrometer (Thermo
401 Fisher Scientific). The mass spectrometer was operated in the data-dependent mode, in which a
402 full-scan MS (from m/z 375 to 1,500 with the resolution of 60,000) was followed by MS/MS of

403 the 15 most intense ions (30,000 resolution; normalized collision energy - 28%; automatic gain
404 control target (AGC) - 2E4, maximum injection time - 200 ms; 60sec exclusion).
405
406 **EV Proteome Data Processing**
407 The raw files were searched directly against the human Swiss-Prot database with no redundant
408 entries using Byonic (Protein Metrics) and Sequest search engines loaded into Proteome
409 Discoverer 2.3 software (Thermo Fisher Scientific). MS1 precursor mass tolerance was set at 10
410 ppm, and MS2 tolerance was set at 20ppm. Search criteria included a static
411 carbamidomethylation of cysteines (+57.0214 Da) and variable modifications of oxidation
412 (+15.9949 Da) on methionine residues and acetylation (+42.011 Da) at the N terminus of
413 proteins. The search was performed with full trypsin/P digestion and allowed a maximum of
414 two missed cleavages on the peptides analyzed from the sequence database. The false-
415 discovery rates of proteins and peptides were set at 0.01. All protein and peptide identifications
416 were grouped, and any redundant entries were removed. Only unique peptides and unique
417 master proteins were reported.

418 All data were quantified using the label-free quantitation node of Precursor Ions Quantifier
419 through the Proteome Discoverer v2.3 (Thermo Fisher Scientific). For the quantification of
420 proteomic data, the intensities of peptides were extracted with initial precursor mass tolerance
421 set at 10 ppm, a minimum number of isotope peaks as 2, maximum ΔRT of isotope pattern
422 multiplets – 0.2 min, PSM confidence FDR of 0.01, with hypothesis test of ANOVA, maximum RT
423 shift of 5 min, pairwise ratio-based ratio calculation, and 100 as the maximum allowed fold
424 change. The abundance levels of all peptides and proteins were normalized to the spiked-in

425 internal iRT standard. For calculations of fold-change between the groups of proteins, total
426 protein abundance values were added together, and the ratios of these sums were used to
427 compare proteins within different samples.

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

432 A non-parametric Wilcox Rank Sum test was performed to test the null hypothesis that the
433 distributions of two groups of the patient population are the same, and the fold change and p-
434 values for each protein were estimated for the following comparisons: IPMN vs. Control, CP vs.
435 Control, PDAC vs. Control, Benign Pancreatic Diseases (CP, IPMN) vs. PDAC. Multiple testing
436 correction was done using Benjamini-Hochberg method to control for the false discovery rate[39].
437 Volcano plots were created using those p values and fold change. Heatmaps visualization and
438 clustering of statistically significant proteins, with adjusted p-value ≤ 0.05 and absolute fold
439 change ≥ 2, were created in R using the pheatmap package. Euclidean distance and average
440 cluster method were used. The values were row-scaled for normalization. Both rows and
441 columns were allowed to cluster.

442

443 **Pathways Enrichment and Protein Network Analysis**

444 Pathway enrichment analysis was performed on statistically significant genes using g:Profiler[40],
445 a web-based tool that searches for pathways whose genes are significantly enriched in our

446 dataset compared to a collection of genes representing Gene Ontology (GO) terms and
447 Reactome pathways. We further used EnrichmentMap[41], a Cytoscape, v3.8.2[42] application to
448 create a visual network of connected pathways that helps to identify relevant pathways and
449 theme[43]. A Protein-Protein interaction network was generated using a stringApp, a Cytoscape
450 app. This application allows to import STRING networks into Cytoscape and enables to perform
451 complex network analysis and visualization of networks[44].

452

453 **Parallel Reaction Monitoring and Data Analysis**

454 Parallel reaction monitoring mass spectrometry (PRM-MS) was employed for validation
455 experiments. Twenty-five cEV markers were selected for validation based on fold change
456 increase ≥5.5, p-value ≤0.01, and technical aspects (number of unique peptides and coverage)
457 (**Supplementary Table 8**). Thirty-six plasma samples from a new cohort were used for the
458 validation (24 PDAC, 6 IPMN and 6 chronic pancreatitis samples). The EVs were isolated from
459 plasma and the proteins processed as described before. Peptide samples were dissolved in 10.8
460 µL 0.05% TFA & 2% ACN, and 10 µL injected into the UHPLC coupled with a Q-Exactive HF-X
461 mass spectrometer (Thermo Fisher Scientific). The mobile phase buffer consisted of 0.1% formic
462 acid in HPLC grade water (buffer A) with an eluting buffer containing 0.1% formic acid in 80%
463 (vol/vol) acetonitrile (buffer B) run with a linear 60-min gradient of 5–35% buffer B at a flow
464 rate of 300 nL/min. Each sample was analyzed under PRM with an isolation width of ±0.8 Th. In
465 these PRM experiments, an MS2 level at 30,000 resolution relative to m/z 200 (AGC target 2E5,
466 200 ms maximum injection time) was run as triggered by a scheduled inclusion list. Higher-

467 energy collisional dissociation was used with 28 eV normalized collision energy. PRM data were
468 manually curated within Skyline-daily (64-bit) 20.2.1.404 (32d27b598)[45].

469

470 **Identification of EV signature for pancreatic cancer diagnosis**

471 All algorithms for identifying the EV signature predictive of pancreatic cancer diagnosis were
472 implemented in R. We used Support Vector Machine (SVM) using CRAN package, e107[46].
473 Ranking of genes was achieved using packages 'sigFeature' and 'SVM-RFE'. An R package,
474 'pROC'[47] was used to build a receiver operating characteristic curve (ROC) and to calculate the
475 area under the curve (AUC).

476

477 **Survival Analysis**

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

484

485 **Statistical Analysis**

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

491
492

Acknowledgments

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.

514 **REFERENCES**
515
516
517 1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2020. CA Cancer J Clin 70:7-30,
518 2020
519 2. Rahib L, Smith BD, Aizenberg R, et al: Projecting cancer incidence and deaths to
520 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States.
521 Cancer Res 74:2913-21, 2014
522 3. Kamisawa T, Wood LD, Itoi T, et al: Pancreatic cancer. Lancet 388:73-85, 2016
523 4. Rahib L, Fleshman JM, Matrisian LM, et al: Evaluation of Pancreatic Cancer
524 Clinical Trials and Benchmarks for Clinically Meaningful Future Trials: A Systematic Review.
525 JAMA Oncol 2:1209-16, 2016
526 5. Bockorny B, Grossman JE, Hidalgo M: Facts and Hopes in Immunotherapy of
527 Pancreatic Cancer. Clin Cancer Res 28:4606-4617, 2022
528 6. Ghatnekar O, Andersson R, Svensson M, et al: Modelling the benefits of early
529 diagnosis of pancreatic cancer using a biomarker signature. Int J Cancer 133:2392-7, 2013
530 7. Matsuno S, Egawa S, Fukuyama S, et al: Pancreatic Cancer Registry in Japan: 20
531 years of experience. Pancreas 28:219-30, 2004
532 8. Locker GY, Hamilton S, Harris J, et al: ASCO 2006 update of recommendations for
533 the use of tumor markers in gastrointestinal cancer. J Clin Oncol 24:5313-27, 2006
534 9. Galli C, Basso D, Plebani M: CA 19-9: handle with care. Clin Chem Lab Med
535 51:1369-83, 2013
536 10. Chen IH, Xue L, Hsu CC, et al: Phosphoproteins in extracellular vesicles as
537 candidate markers for breast cancer. Proc Natl Acad Sci U S A 114:3175-3180, 2017
538 11. Costa-Silva B, Aiello NM, Ocean AJ, et al: Pancreatic cancer exosomes initiate
539 pre-metastatic niche formation in the liver. Nat Cell Biol 17:816-26, 2015
540 12. van Niel G, Carter DRF, Clayton A, et al: Challenges and directions in studying
541 cell-cell communication by extracellular vesicles. Nat Rev Mol Cell Biol 23:369-382, 2022
542 13. Melo SA, Luecke LB, Kahlert C, et al: Glypican-1 identifies cancer exosomes and
543 detects early pancreatic cancer. Nature 523:177-82, 2015
544 14. Madhavan B, Yue S, Galli U, et al: Combined evaluation of a panel of protein and
545 miRNA serum-exosome biomarkers for pancreatic cancer diagnosis increases sensitivity and
546 specificity. Int J Cancer 136:2616-27, 2015
547 15. Yang KS, Im H, Hong S, et al: Multiparametric plasma EV profiling facilitates
548 diagnosis of pancreatic malignancy. Sci Transl Med 9, 2017
549 16. Castillo J, Bernard V, San Lucas FA, et al: Surfaceome profiling enables isolation
550 of cancer-specific exosomal cargo in liquid biopsies from pancreatic cancer patients. Ann Oncol
551 29:223-229, 2018
552 17. Fahrmann JF, Mao X, Irajizad E, et al: Plasma-Derived Extracellular Vesicles
553 Convey Protein Signatures that Reflect Pathophysiology in Lung and Pancreatic
554 Adenocarcinomas. Cancers (Basel) 12, 2020
555 18. Hoshino A, Kim HS, Bojmar L, et al: Extracellular Vesicle and Particle Biomarkers
556 Define Multiple Human Cancers. Cell 182:1044-1061 e18, 2020

557 19. Iliuk A, Wu X, Li L, et al: Plasma-Derived Extracellular Vesicle Phosphoproteomics
558 through Chemical Affinity Purification. J Proteome Res 19:2563-2574, 2020
559 20. Wu X, Li L, Iliuk A, et al: Highly Efficient Phosphoproteome Capture and Analysis
560 from Urinary Extracellular Vesicles. J Proteome Res 17:3308-3316, 2018
561 21. Nunez Lopez YO, Iliuk A, Petrilli AM, et al: Proteomics and Phosphoproteomics of
562 Circulating Extracellular Vesicles Provide New Insights into Diabetes Pathobiology. Int J Mol Sci
563 23, 2022
564 22. Hinzman CP, Jayatilake M, Bansal S, et al: An optimized method for the isolation
565 of urinary extracellular vesicles for molecular phenotyping: detection of biomarkers for
566 radiation exposure. J Transl Med 20:199, 2022
567 23. Hinzman CP, Singh B, Bansal S, et al: A multi-omics approach identifies pancreatic
568 cancer cell extracellular vesicles as mediators of the unfolded protein response in normal
569 pancreatic epithelial cells. J Extracell Vesicles 11:e12232, 2022
570 24. Shuen TWH, Alunni-Fabbroni M, Ocal E, et al: Extracellular Vesicles May Predict
571 Response to Radioembolization and Sorafenib Treatment in Advanced Hepatocellular
572 Carcinoma: An Exploratory Analysis from the SORAMIC Trial. Clin Cancer Res 28:3890-3901,
573 2022
574 25. Schmitz-Winnenthal FH, Volk C, Helmke B, et al: Expression of cytokeratin-20 in
575 pancreatic cancer: an indicator of poor outcome after R0 resection. Surgery 139:104-8, 2006
576 26. Guyon I, Weston J, Barnhill S, et al: Gene Selection for Cancer Classification using
577 Support Vector Machines. Machine Learning 46:389-422, 2002
578 27. Das P, Roychowdhury A, Das S, et al: sigFeature: Novel Significant Feature
579 Selection Method for Classification of Gene Expression Data Using Support Vector Machine and
580 t Statistic. Front Genet 11:247, 2020
581 28. Huang L, Bockorny B, Paul I, et al: PDX-derived organoids model in vivo drug
582 response and secrete biomarkers. JCI Insight, 2020
583 29. LeBleu VS, Kalluri R: Exosomes as a Multicomponent Biomarker Platform in
584 Cancer. Trends Cancer 6:767-774, 2020
585 30. Liu R, Lu S, Deng Y, et al: PSMB4 expression associates with epithelial ovarian
586 cancer growth and poor prognosis. Arch Gynecol Obstet 293:1297-307, 2016
587 31. Zhang X, Lin D, Lin Y, et al: Proteasome beta-4 subunit contributes to the
588 development of melanoma and is regulated by miR-148b. Tumour Biol 39:1010428317705767,
589 2017
590 32. Zheng P, Guo H, Li G, et al: PSMB4 promotes multiple myeloma cell growth by
591 activating NF-kappaB-miR-21 signaling. Biochem Biophys Res Commun 458:328-33, 2015
592 33. Lee GY, Haverty PM, Li L, et al: Comparative oncogenomics identifies PSMB4 and
593 SHMT2 as potential cancer driver genes. Cancer Res 74:3114-26, 2014
594 34. Le Large TY, Mantini G, Meijer LL, et al: Microdissected pancreatic cancer
595 proteomes reveal tumor heterogeneity and therapeutic targets. JCI Insight 5, 2020
596 35. Law HC, Lagundzin D, Clement EJ, et al: The Proteomic Landscape of Pancreatic
597 Ductal Adenocarcinoma Liver Metastases Identifies Molecular Subtypes and Associations with
598 Clinical Response. Clin Cancer Res 26:1065-1076, 2020
599 36. Zhang Y, Yang J, Li H, et al: Tumor markers CA19-9, CA242 and CEA in the
600 diagnosis of pancreatic cancer: a meta-analysis. Int J Clin Exp Med 8:11683-91, 2015

601 37. Duffy MJ, Sturgeon C, Lamerz R, et al: Tumor markers in pancreatic cancer: a
602 European Group on Tumor Markers (EGTM) status report. Ann Oncol 21:441-7, 2010
603 38. Amin MB, Greene FL, Edge SB, et al: The Eighth Edition AJCC Cancer Staging
604 Manual: Continuing to build a bridge from a population-based to a more "personalized"
605 approach to cancer staging. CA Cancer J Clin 67:93-99, 2017
606 39. Benjamini Y, Hochberg Y: Controlling the False Discovery Rate: A Practical and
607 Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B
608 (Methodological) 57:289-300, 1995
609 40. Raudvere U, Kolberg L, Kuzmin I, et al: g:Profiler: a web server for functional
610 enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 47:W191-
611 W198, 2019
612 41. Merico D, Isserlin R, Stueker O, et al: Enrichment map: a network-based method
613 for gene-set enrichment visualization and interpretation. PLoS One 5:e13984, 2010
614 42. Shannon P, Markiel A, Ozier O, et al: Cytoscape: a software environment for
615 integrated models of biomolecular interaction networks. Genome Res 13:2498-504, 2003
616 43. Reimand J, Isserlin R, Voisin V, et al: Pathway enrichment analysis and
617 visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc
618 14:482-517, 2019
619 44. Szklarczyk D, Gable AL, Lyon D, et al: STRING v11: protein-protein association
620 networks with increased coverage, supporting functional discovery in genome-wide
621 experimental datasets. Nucleic Acids Res 47:D607-D613, 2019
622 45. MacLean B, Tomazela DM, Shulman N, et al: Skyline: an open source document
623 editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26:966-8,
624 2010
625 46. Meyer D, Dimitriadou E, Hornik K, et al: Misc Functions of the Department of
626 Statistics, ProbabilityTheory Group (Formerly: E1071), TU Wien, 2015
627 47. Robin X, Turck N, Hainard A, et al: pROC: an open-source package for R and S+ to
628 analyze and compare ROC curves. BMC Bioinformatics 12:77, 2011
629
630
631
632
633
634

635 **FIGURES**
636
637

638 **Figure 1. Study Design**

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

644

645 **Figure 2. Identification of cEV Proteins Differentially Expressed in Disease Groups**

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

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

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

657 **(D)** Expression of enriched cEV proteins in patients with PDAC versus benign pancreatic
658 diseases. Each dot indicates the target protein signal from one patient. Y-axis, normalized log
659 base 2 of protein signals detected by mass spectrometry; Error bars, min and max values; lines
660 in boxes, median values. ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.
661
662 **Figure 3. Circulating EV Proteomics Reveal Markers Associated with Metastasis and Worse**
663 **Prognosis.**
664 **(A)** Heatmap showing EV proteins differentially expressed in the plasma of metastatic versus
665 non-metastatic PDAC. Designations of clinical parameters are indicated at the top of the
666 heatmap.
667 **(B)** Expression patterns of cEV proteins associated with metastatic disease. Y-axis, normalized
668 log base 2 of protein signals detected by mass spectrometry; N, non-metastatic PDAC group; M,
669 metastatic PDAC group. Each dot indicated the target protein signal from one patient. Error
670 bars, min and max values; lines in boxes, median values. ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq$
671 0.0001.
672 **(C)** As is **(B)**, except for cEV markers with increased expression in non-metastatic PDAC.
673 **(D)** Correlation of cEV marker expression with survival. Kaplan–Meier curves and log-rank test P
674 values of representative survival cEV markers quantified in the discovery cohort.
675
676 **Figure 4. Validation of cEV Markers and Identification of 7-EV Protein Signature for PDAC**
677 **Diagnosis.**

678 **(A)** Differences of cEV protein abundances between patients with PDAC (n=24) and benign
679 pancreatic diseases (chronic pancreatitis and IPMN) (n=12). x axis, minus log *p* values of protein
680 abundance differences between PDAC and benign groups; y axis, average fold changes of
681 proteins in PDAC group compared to benign group. Size of bubbles indicate average protein
682 abundances in PDAC group. Pink color, proteins that had at least two-fold enrichment in PDAC
683 group ($p<0.05$).
684 **(B)** ROC curves were calculated for individual cEV markers as well as for the 7-EV protein PDAC
685 signature combination to determine optimum diagnostic performance. AUC, area under the
686 curve.
687
688 **SUPPLEMENTARY MATERIAL**
689
690 **Supplementary Table 1**. Baseline characteristics of patients enrolled on the discovery cohort.
691
692 **Supplementary Table 2.** Plasma EV analysis reproducibility
693
694 **Supplementary Table 3.** LC-MS results of EV analysis of plasma from patients with PDAC (PA),
695 IPMN, Chronic Pancreatitis (CP) and Control individuals.
696
697 **Supplementary Table 4.** List of EV proteins that met the eligibility criteria for principal
698 component analysis.
699

700 **Supplementary Table 5**. List of 182 proteins differentially expressed in PDAC compared to
701 benign diseases.
702
703 **Supplementary Table 6**. List of EV proteins that are significantly altered in patients with
704 metastatic versus non-metastatic diseases.
705
706 **Supplementary Table 7.** Support Vector Machine Prediction of 16 individual genes in Discovery
707 Test and Validation Cohorts
708
709 **Supplementary Table 8**. List of 25 cEV proteins that met the eligibility criteria for validation
710 studies.
711
712 **Supplementary Table 9**. Baseline characteristics of patients enrolled in the validation cohort.
713
714 **Supplementary Figure 1. EVtrap isolation of extracellular vesicles**
715 **(A)** Transmission electron microscopy (TEM) images collected of a single EV and multiple EVs
716 captured from plasma by EVtrap. TEM imaging of EVs was carried out on a HITACHI H-8100
717 electron microscope (Hitachi, Tokyo, Japan) with an accelerating applied potential of 200 kV.
718 **(B)** Nanoparticle tracking analysis (NTA) of EVs after elution off EVtrap beads. NTA was carried
719 out using ZetaView instrument (Particle Metrix) after calibration with 100 nm polystyrene
720 particles.
721

722 **Supplementary Figure 2. EV proteomics analytical performance**

723 **(A)** Reproducibility of the method. A test plasma sample was processed in six replicates and
724 performed a Pearson correlation analysis that revealed a very high correlation between
725 replicates.

726 **(B)** Number of quantified EV proteins per sample according to different patient cohort.

727

728 **Supplementary Figure 3**. **Network Analyses of cEV Proteins Differentially Expressed in PDAC**
729 **Compared to Benign Pancreatic Diseases.**

730 **(A)** Functional association of proteins identified by STRING database. Red, cEV proteins
731 enriched in PDAC patients as compared to benign pancreatic diseases. Green, cEV proteins
732 decreased in patients with PDAC as compared to benign pancreatic diseases. Red, cEV proteins
733 increased in PDAC as compared to benign pancreatic diseases. Thickness of lines indicate
734 confidence of association.

735 **(B, C)** Clustering of cEV protein pathways enriched **(B)** or downregulated **(C)** in PDAC cohorts.
736 Pathways were identified using Gene Ontology database and REACTOME database.

737

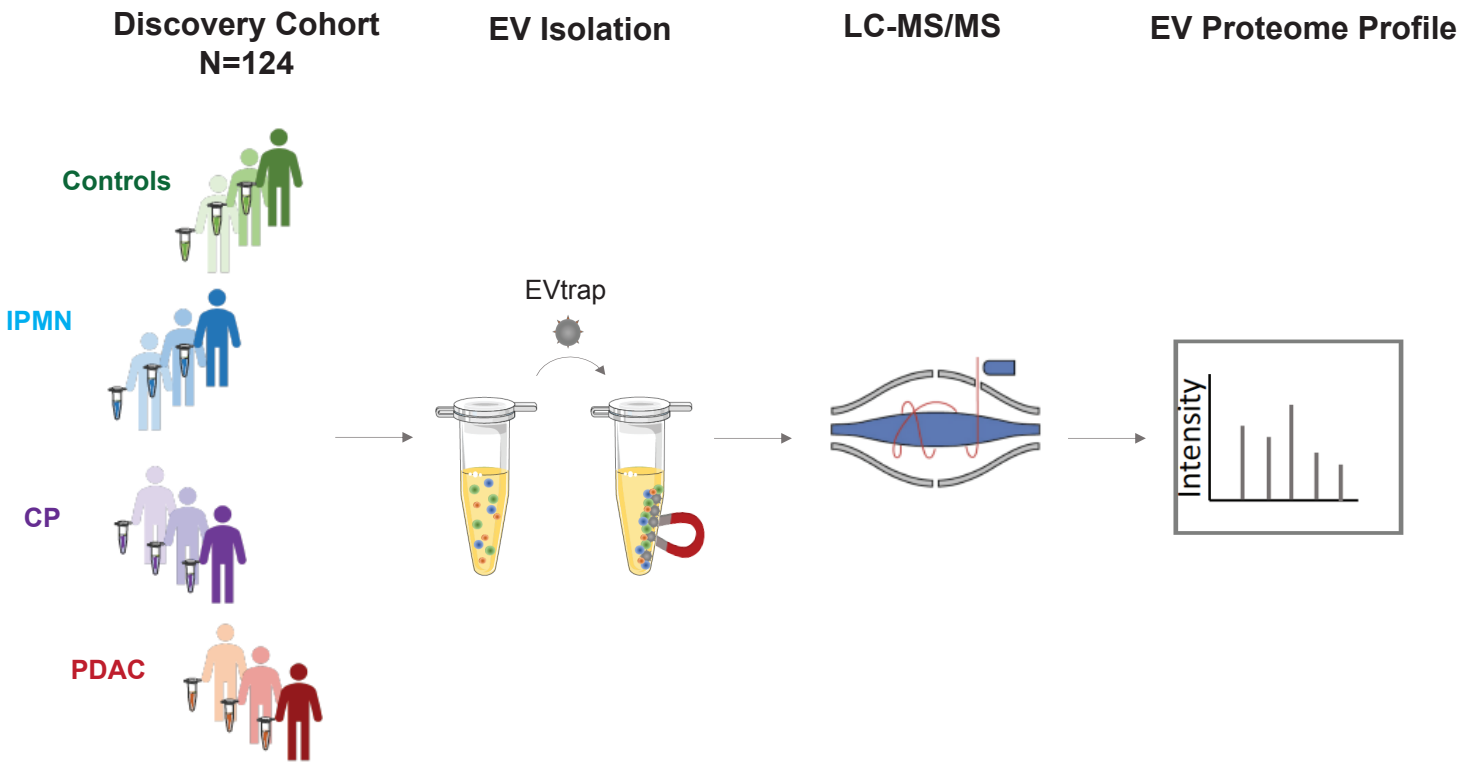
738 **Supplementary Figure 4. Summary of selection process to develop EV signature for pancreatic**
739 **cancer diagnosis.**

740

741 **Supplementary Figure 5. Validation of individual cEV proteins in an independent cohort of**
742 **patients.**

743 Expression of biomarker candidates detected by Parallel Reaction Monitoring (PRM) analyses. A
744 total of 25 cEV proteins with significant overexpression in PDAC in the discovery cohorts were
745 quantified by PRM in a separate validation cohort of patients.
746
747
748
749
750
751

Figure 1


Discovery Cohort
N=124
EV Isolation
LC-MS/MS
EV Proteome Profile
Controls
IPMN
CP
PDAC
EVtrap
Intensity

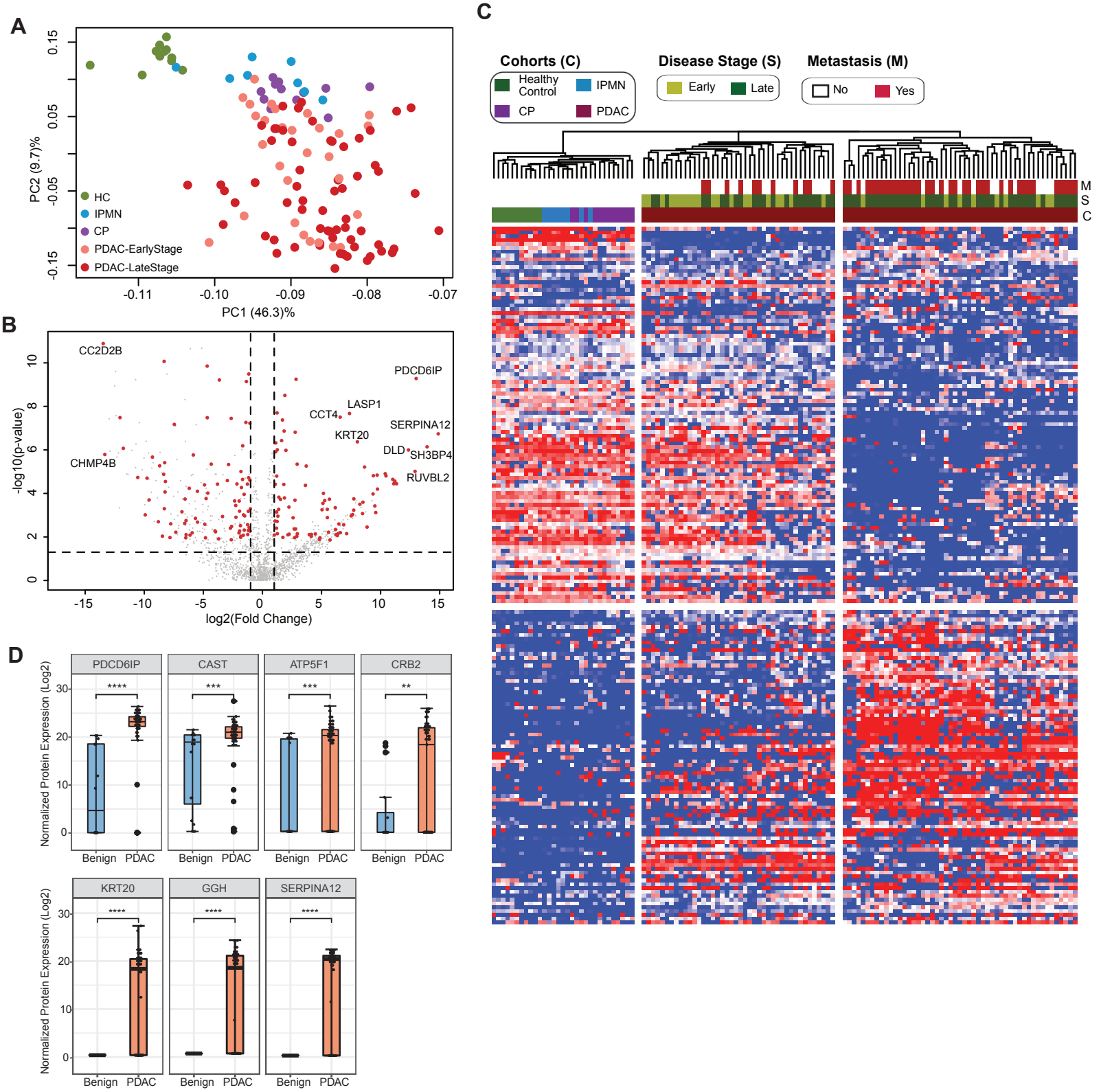

A
PC2 (9.7)%
PC1 (46.3)%
HC
IPMN
CP
PDAC-EarlyStage
PDAC-LateStage
B
-log10(p-value)
log2(Fold Change)
CC2D2B
CHMP4B
PDCD6IP
LASP1
CCT4
SERPINA12
KRT20
DLD
SH3BP4
RUVBL2
C
Cohorts (C)
Healthy Control
IPMN
CP
PDAC
Disease Stage (S)
Early
Late
Metastasis (M)
No
Yes
M
S
C
D
PDCD6IP
CAST
ATP5F1
CRB2
KRT20
GGH
SERPINA12
Normalized Protein Expression (Log2)
Benign
PDAC

Figure 2

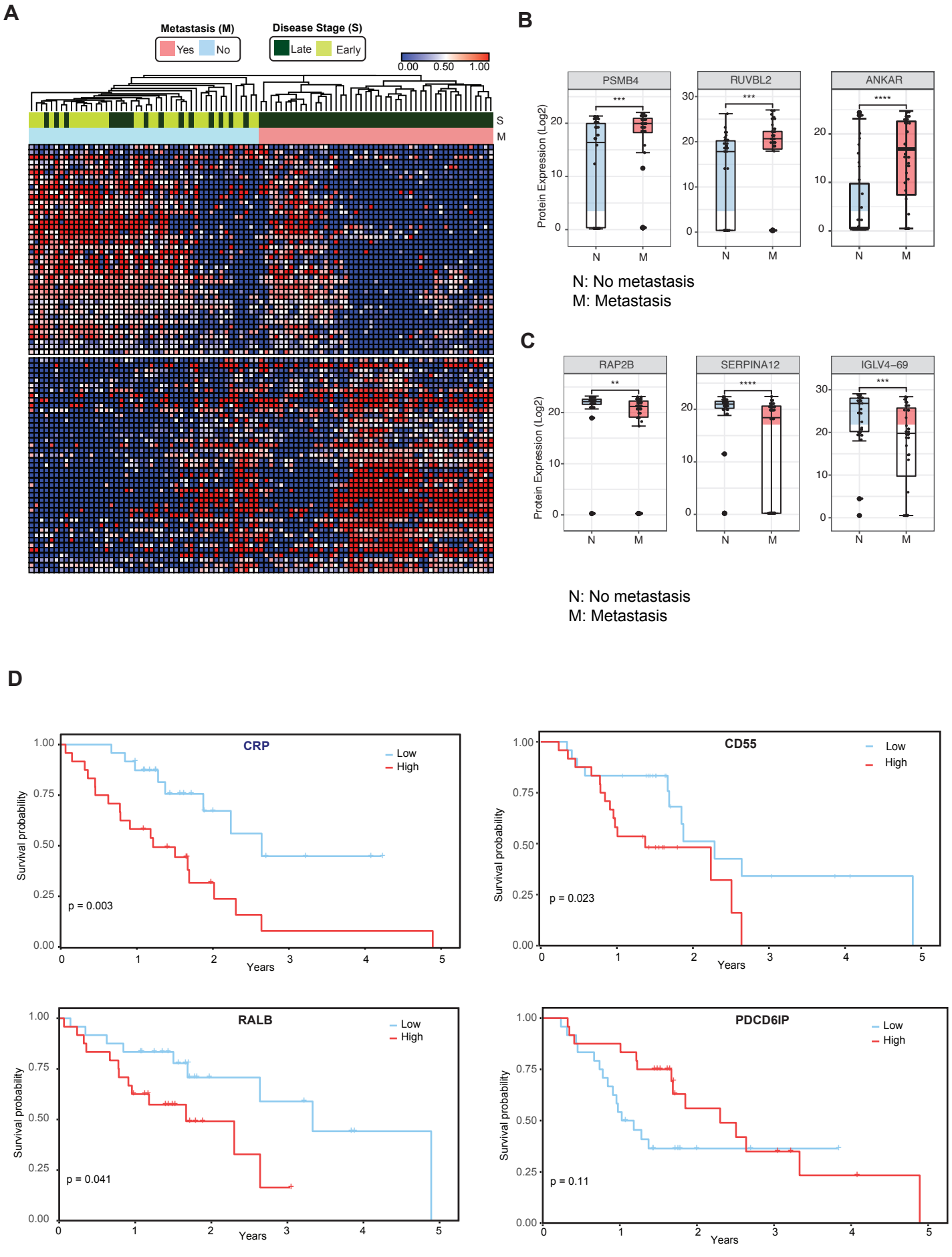
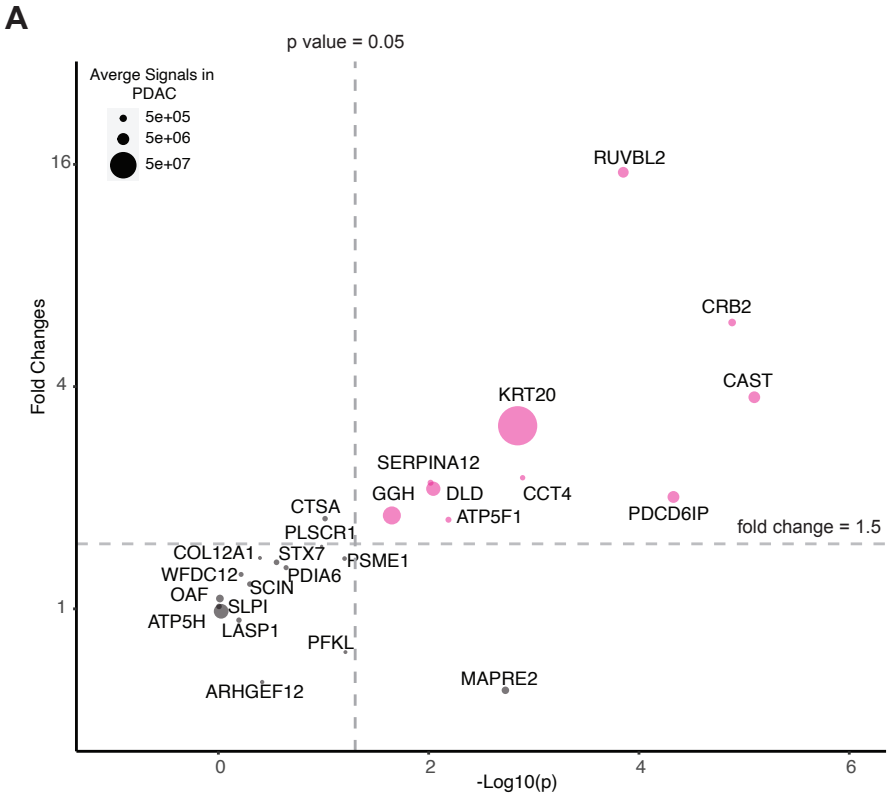

A
Metastasis (M)
Yes
No
Disease Stage (S)
Late
Early
0.00
0.50
1.00
S
M
B
PSMB4
RUVBL2
ANKAR
Protein Expression (Log2)
N
M
N: No metastasis
M: Metastasis
C
RAP2B
SERPINA12
IGLV4−69
Protein Expression (Log2)
N
M
N: No metastasis
M: Metastasis
D
CRP
Low
High
p = 0.003
Survival probability
Years
CD55
Low
High
p = 0.023
Survival probability
Years
RALB
Low
High
p = 0.041
Survival probability
Years
PDCD6IP
Low
High
p = 0.11
Survival probability
Years

Figure 3


A
p value = 0.05
Averge Signals in PDAC
5e+05
5e+06
5e+07
Fold Changes
16
4
1
RUVBL2
CRB2
CAST
KRT20
SERPINA12
GGH
DLD
CCT4
ATP5F1
PDCD6IP
CTSA
PLSCR1
fold change = 1.5
COL12A1
STX7
PSME1
WFDC12
PDIA6
SCIN
OAF
SLPI
ATP5H
LASP1
PFKL
MAPRE2
ARHGEF12
0
2
4
6
-Log10(p)

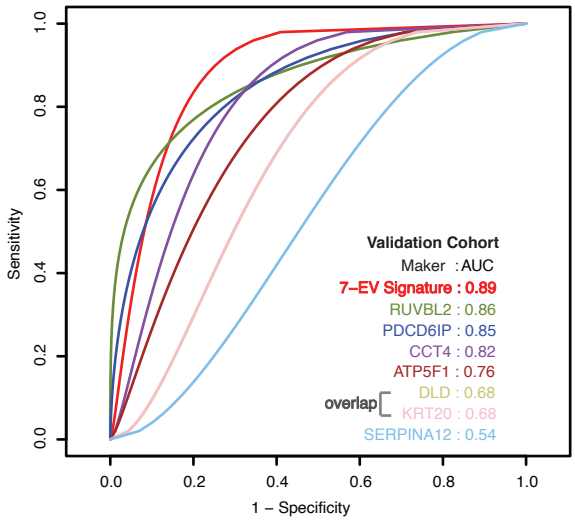
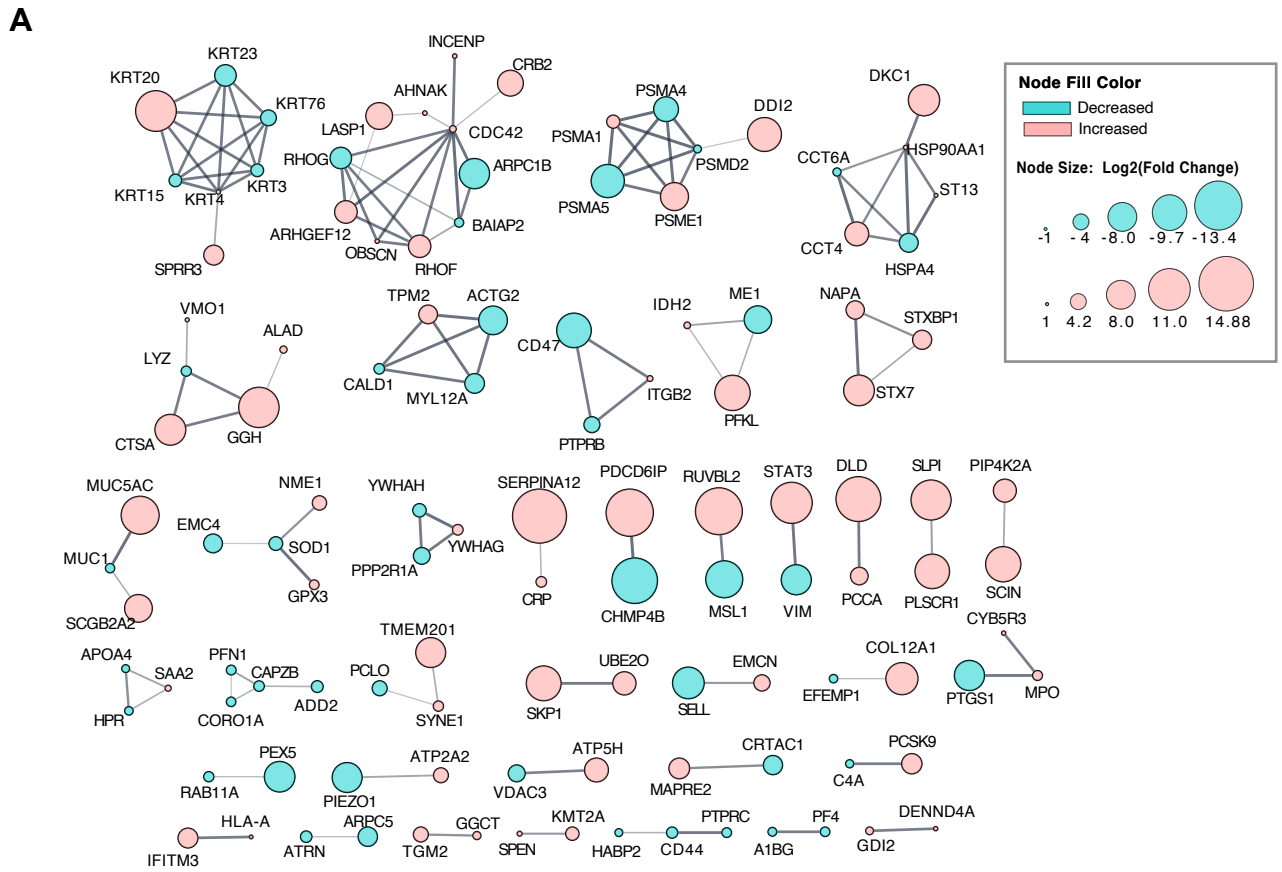
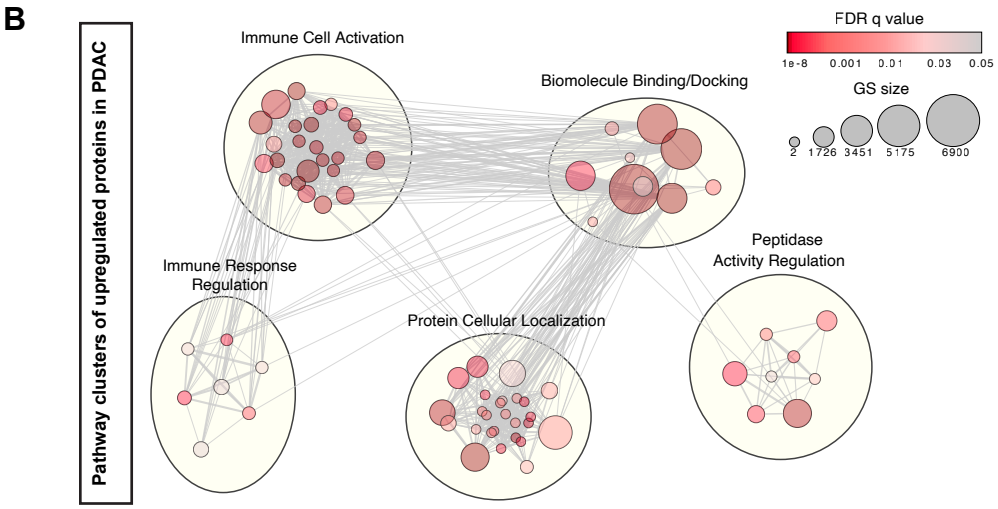
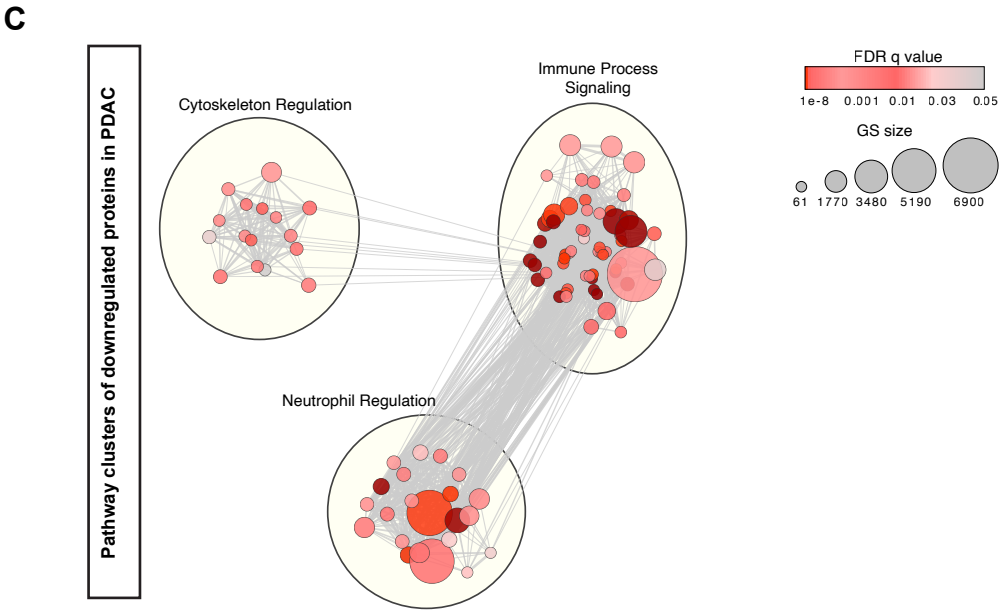

B
Sensitivity
1 – Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Validation Cohort
Maker : AUC
7–EV Signature : 0.89
RUVBL2 : 0.86
PDCD6IP : 0.85
CCT4 : 0.82
ATP5F1 : 0.76
overlap
DLD : 0.68
KRT20 : 0.68
SERPINA12 : 0.54

Figure 4







Suppl. Figure 3