TUMOR MARKERS AND SIGNATURES

IJC
International Journal of Cancer

Plasma protein biomarkers for early detection of pancreatic ductal adenocarcinoma

Jingru Yu[1] | Alexander Ploner[1] | Maximilian Kordes[2,3] | Matthias Löhr[2,3] | Magnus Nilsson[2,3] | Maria Evangelina Lopez de Maturana[4] | Lidia Estudillo[4] | Harald Renz[5] | Alfredo Carrato[6] | Xavier Molero[7] | Francisco X Real[8] | Núria Malats[4] | Weimin Ye[1,9]

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[2]Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institutet, Stockholm, Sweden

[3]Department of Upper Abdominal Diseases, Karolinska University Hospital, Stockholm, Sweden

[4]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), and CIBERONC, Madrid, Spain

[5]Institute of Laboratory Medicine and Pathobiochemistry, Member of the German Center for Lung Research (DZL) and the Universities of Giessen and Marburg Lung Center (UGMLC), Philipps University of Marburg, Marburg, Germany

[6]Department of Medical Oncology, Hospital Ramón y Cajal, Madrid, Spain

[7]Exocrine Pancreas Research Unit, Hospital Universitari Vall d'Hebron—Institut de Recerca, Autonomous University of Barcelona, and CIBEREHD, Barcelona, Spain

[8]Epithelial Carcinogenesis Group, Spanish National Cancer Research Center (CNIO), and CIBERONC, Madrid, Spain

[9]Department of Epidemiology and Health Statistics & Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Fujian Medical University, Fuzhou, China

Correspondence
Weimin Ye, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, 171 77 Stockholm, Sweden.
Email: weimin.ye@ki.se

Funding information
Cancerfonden, Grant/Award Number: 2019-0264; China Scholarship Council; EU-6FP Integrated Project, Grant/Award Number: 018771-MOLDIAG-PACA; Instituto de Salud Carlos III, Grant/Award Numbers: PI11/01542, PI09/02102, PI12/01635, PI12/00815

Abstract

Pancreatic ductal adenocarcinoma (PDAC) has a poor prognosis, mainly due to late diagnosis at advanced tumor stages. In this study, we aimed to identify plasma protein biomarkers for early detection of PDAC. Totally, 135 PDAC patients (early PDAC, Stage I/II, n = 71; advanced PDAC, Stage III/IV, n = 64), 13 benign lesions/chronic pancreatitis patients and 72 healthy individuals, with corresponding plasma samples from a case-control study in Sweden were included. A proximity extension assay was used to detect 92 cancer-related proteins, and an enzyme-linked immuno-sorbent assay/electrochemiluminescence immunoassay was used to detect CA19-9. Predictive features were selected from these 93 candidate proteins and three covariates in the Swedish participants, and then validated in Spanish participants, including 37 early PDAC patients, 38 advanced PDAC patients, 19 chronic pancreatitis patients and 36 healthy controls. A panel of eight proteins discriminating early

PDAC from healthy individuals was identified, and the cross-validated area under the curves (AUCs) were 0.85 (95% confidence interval, 95% CI, 0.78-0.91) and 0.81 (95% CI, 0.70-0.92) in the Swedish and Spanish participants, respectively. Another eight-protein panel was predictive for classifying advanced PDAC from healthy controls in two populations, with cross-validated AUCs of 0.89 (95% CI, 0.83-0.95) and 0.90 (95% CI, 0.83-0.98), respectively. In conclusion, eight protein biomarkers were identified and externally validated, potentially allowing early detection of PDAC patients if validated in additional prospective studies.

KEYWORDS
AUC, biomarkers, early diagnosis, pancreatic ductal adenocarcinoma, proteomics

1 | INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal malignancies, with an overall 5-year survival rate of less than 8%,[1] and it could become the second leading cause of cancer-related death by 2030 if no action is taken.[2] The dismal prognosis of PDAC is largely attributable to the late diagnosis at an advanced tumor stage. Surgical resection is the only potentially curative treatment for PDAC patients, but less than 20% of PDAC patients are eligible for initial resection.[3] Evolution from primary PDAC initiation to metastasis takes about 18.5 years, and patients die on average 3 years after metastasis, suggesting that there is a wide window of opportunity for early detection and intervention.[4,5]

To date, carbohydrate antigen (CA19-9) is the only approved biomarker in clinical use for early diagnosis in symptomatic patients and treatment management in cancer patients. However, it has limitations for screening or early detection of PDAC, as some patients with a Lewis blood-group negative genotype cannot express the biomarker.[6] Also, CA19-9 is not specific to PDAC and has been shown to be elevated in patients with extra-pancreatic malignancies[7] and benign hepatic, biliary or pancreatic conditions. Some studies have postulated to use protein biomarker panels for distinguishing PDAC from healthy controls or chronic pancreatitis with higher sensitivity and specificity than CA19-9 alone by using blood, tumor tissues or urine samples.[8-10] A review summarized the multiparametric analysis for early detection of PDAC, specifically by searching for candidate biomarkers, including DNA, RNA, proteins and metabolites in different biospecimens that can be easily obtained, such as blood, saliva, urine and pancreatic juice[11]; however, none of the available biomarkers have sufficient accuracy for screening, even in high-risk populations, and all of them are in early phases of validation and not clinically applicable yet.

So, identifying candidate biomarkers that can be detected by sensitive, specific, cost-effective and practical measurement methods in early-stage PDAC patients is still an unsolved problem. To meet this challenge, we analyzed plasma samples using the proximity extension assay (PEA), which enables high-throughput, multiplex immunoassays that measure multiple proteins simultaneously on smaller volume of samples than general methods. This technology uses fixed panels of protein biomarkers that have been suggested as generic biomarker

What's new?

Pancreatic ductal adenocarcinoma (PDAC) has a poor prognosis, mainly due to lack of symptoms and late diagnosis. Currently, only a single, unsatisfactory biomarker (CA19-9), is approved for earlier diagnosis. Thus there is an urgent need for more practical, cost-effective tests. In this study, the authors used proximity extension assay (PEA) to develop a panel of eight plasma-protein biomarkers that can distinguish early PDAC from healthy controls. Although further validation studies are needed, these results support the potential of targeted proteomics for early detection of PDAC.

candidates for different diseases, such as colorectal cancer,[12] breast cancer,[13] heart disease and diabetes.[14] Our aim was to investigate whether the selected protein biomarkers could discriminate PDAC, especially in early stage, from healthy control individuals as well as chronic pancreatitis patients.

2 | MATERIALS AND METHODS

2.1 | Study design and sample collection

We collected plasma samples from 220 participants enrolled in a Swedish case-control study, which has been described previously.[15] Briefly, we used materials collected in a population-based case-control study of pancreatic cancer performed in Stockholm, Sweden. Patients undergoing surgical resection from 2007 to 2012 at Karolinska University Hospital were recruited, and corresponding control participants matched by sex and age were randomly identified from the general population of Stockholm County. In this study, 135 PDAC patients who had histopathological diagnosis, TNM staging information and sufficient plasma samples were included: early PDAC (Stage I/II, n = 71) and advanced PDAC (Stage III/IV, n = 64). PDAC staging was performed according to the American Joint

Commission on Cancer (AJCC) guidelines, seventh edition.[16] Another 13 patients with benign lesions or chronic pancreatitis were reidentified by histopathological diagnosis, since they were collected by clinical diagnosis with possibility of misclassification. A total of 72 healthy controls were enrolled by stratified randomization, who were frequency matched by sex and age (± 5 years) with early PDAC patients.

We also collected plasma samples from 130 participants in Spain as the external validation set, including early PDAC (Stage I/II, n = 37), advanced PDAC (Stage III/IV, n = 38), chronic pancreatitis (n = 19) and healthy controls (n = 36). Those participants were randomly selected from the European Study into Digestive Illnesses and Genetics (PanGenEU), a mostly hospital-based case-control study of pancreatic cancer conducted in six European countries (Spain, Germany, Ireland, the United Kingdom, Italy and Sweden).[17] Cases diagnosed or suspected of having pancreatic cancer and the corresponding controls

matched by region, sex and age (± 10 years) were included from 2007 to 2014.

Peripheral blood samples of both study populations were collected in EDTA tubes, then plasma was isolated and stored as 250-1000 μL aliquots at −80°C. Interviews were conducted by professional interviewers with participants, concerning demographics, medical history and smoking status. Furthermore, information about cancer diagnosis date, tumor location and histopathological diagnosis was collected from clinical charts.

2.2 | Multiplex protein measurement

Plasma samples from Swedish and Spanish participants were shipped to Olink Proteomics AB (Uppsala, Sweden) on dry ice. Samples from both populations were analyzed independent of each other and protein

TABLE 1 Demographical and clinical characteristics of the Swedish and Spanish participants

Swedish participants	Early PDAC n = 71 (%)	Advanced PDAC n = 64 (%)	Benign lesions/chronic pancreatitis (n = 13, %)	Healthy controls n = 72 (%)	P value[a]
Age					
Mean ± SD	67.7 ± 8.0	66.3 ± 9.1	57.9 ± 9.4	67.9 ± 8.1	.001**
Sex					.814
Male	44 (62.0)	42 (65.6)	7 (53.8)	43 (59.7)	
Female	27 (38.0)	22 (34.4)	6 (46.2)	29 (40.3)	
Smoking status					.813
Never	23 (32.4)	18 (28.1)	5 (38.4)	25 (34.7)	
Ex-smoker	31 (43.7)	29 (45.3)	4 (30.8)	34 (47.2)	
Current	17 (23.9)	17 (26.6)	4 (30.8)	13 (18.1)	
Diabetes					<.001***
No	37 (52.1)	33 (51.6)	8 (61.5)	69 (95.8)	
Yes	14 (19.7)	9 (14.1)	0 (0.0)	3 (4.2)	
Unknown	20 (28.2)	22 (34.4)	5 (38.5)	0 (0.0)	
Spanish participants	**Early PDAC, n = 37 (%)**	**Advanced PDAC, n = 38 (%)**	**Chronic pancreatitis (n = 19, %)**	**Healthy controls, n = 36 (%)**	**P value[a]**
Age					
Mean ± SD	61.9 ± 12.9	63.9 ± 14.4	63.5 ± 9.9	70.7 ± 12.3	.023*
Sex					.213
Male	22 (59.5)	17 (44.7)	14 (73.7)	20 (55.6)	
Female	15 (40.5)	21 (55.3)	5 (26.3)	16 (44.4)	
Smoking status					.009**
Never	12 (32.4)	16 (42.1)	2 (10.5)	19 (52.8)	
Ex-smoker	10 (27.0)	11 (28.9)	7 (36.8)	13 (36.1)	
Current	15 (40.5)	11 (28.9)	10 (52.6)	4 (11.1)	
Diabetes					<.001***
No	30 (81.1)	31 (81.6)	6 (31.6)	32 (88.9)	
Yes	7 (18.9)	7 (18.4)	13 (68.4)	4 (11.1)	

[a]For continuous variable, based on one-way analysis of variance; for categorical variables, based on Fisher's exact test.
*P value < .05. **P value < .01. ***P value < .001.

measurements were blinded to disease status. They were randomly placed on 96-well plates with 50 μL in each cell (Supplementary Table 1a), achieving an approximately balanced distribution of disease status across plates. A panel of 92 protein biomarkers were measured via PEA technique using the Proseek® Multiplex Oncology II 96 × 92 reagent kit (Olink Proteomics, Uppsala, Sweden), which measures multiple cancer-related human protein biomarker candidates simultaneously (listed in Supplementary Table 2). In short, a pair of oligonucleotide-labeled antibodies, Proseek probes, binds to the target protein in plasma sample; when the two probes are in close proximity, a new polymerase chain reaction (PCR) target sequence is formed by a proximity-dependent DNA polymerization event. This complex is then amplified and quantified using standard real-time PCR.

2.3 | CA19-9 measurement

We also measured CA19-9 concentration by using an enzyme-linked immunosorbent assay (ELISA, ThermoFisher Scientific) in the Swedish samples and electrochemiluminescence immunoassay (ECLIA, Roche Diagnostics, Germany) in the Spanish samples, following the manufacturers' instruction. Plasma samples were also randomly placed on plates (Supplementary Table 1b).

2.4 | Preprocessing and quality control of protein measurements

The 92 protein levels measured using PEA were log2-transformed and normalized with the plate median as the normalization factor.[18] Median value of the interplate coefficient of variation (CV) was 1.8% for detection of 92 proteins in both Swedish and Spanish samples (Supplementary Figure 1). Interplate/assay CVs of CA19-9 were 14.9% and 2.4% of the Swedish and Spanish samples, respectively. Calibration of 93 proteins between the Swedish and Spanish samples was conducted before any comparison (Supplementary Figure 2).

2.5 | Data analysis

Summaries of categorical and continuous variables are presented as frequencies with percentages and means with SDs, respectively. Fisher's exact test or one-way analysis of variance was used for testing distribution differences of categorical or continuous characteristic variables (in Table 1) between groups. Wilcoxon-Mann-Whitney test/Kruskal-Wallis test was used for comparing differences of
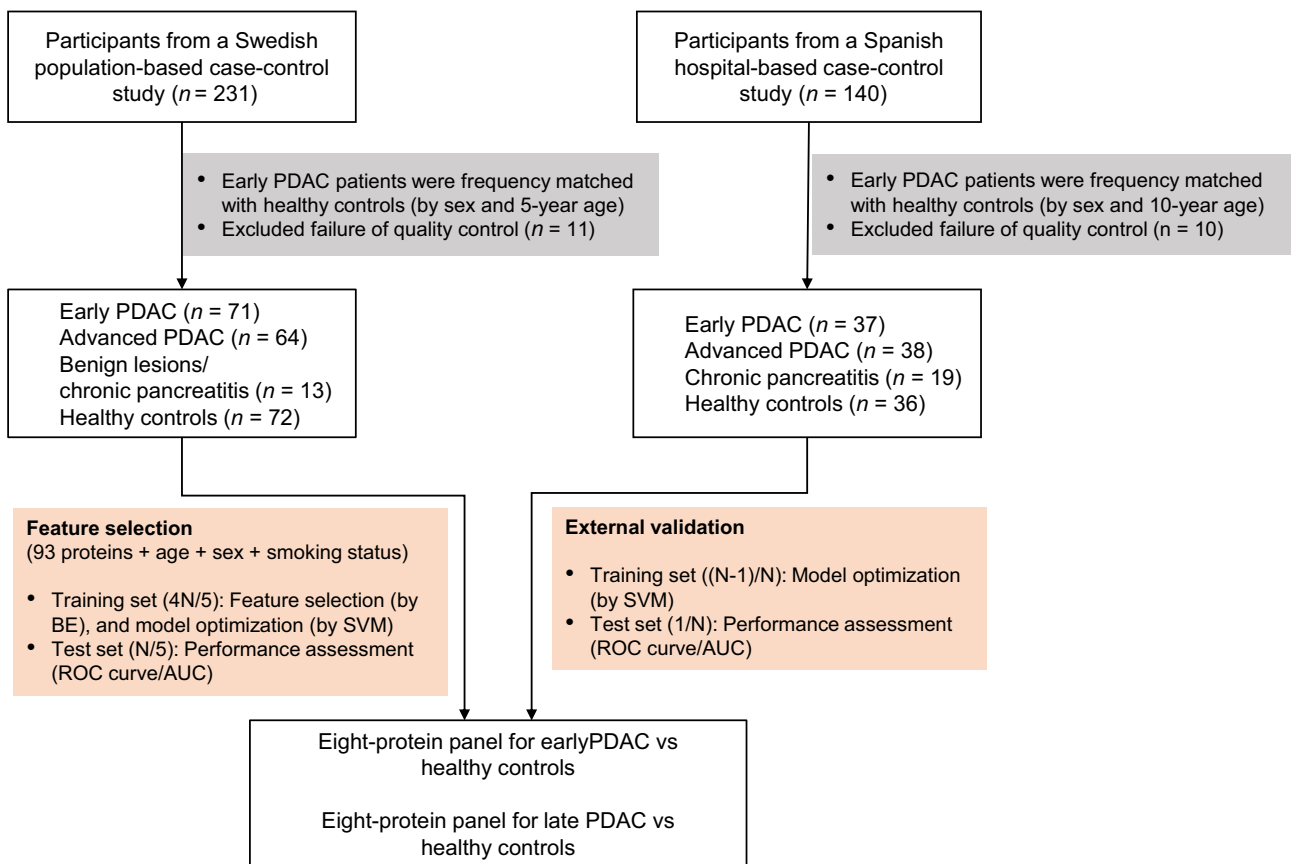

FIGURE 1 Workflow for the study. AUC, area under the curve; BE, backward elimination regression; ROC curve, receiver operating characteristic curve; SVM, support vector machine [Color figure can be viewed at wileyonlinelibrary.com]
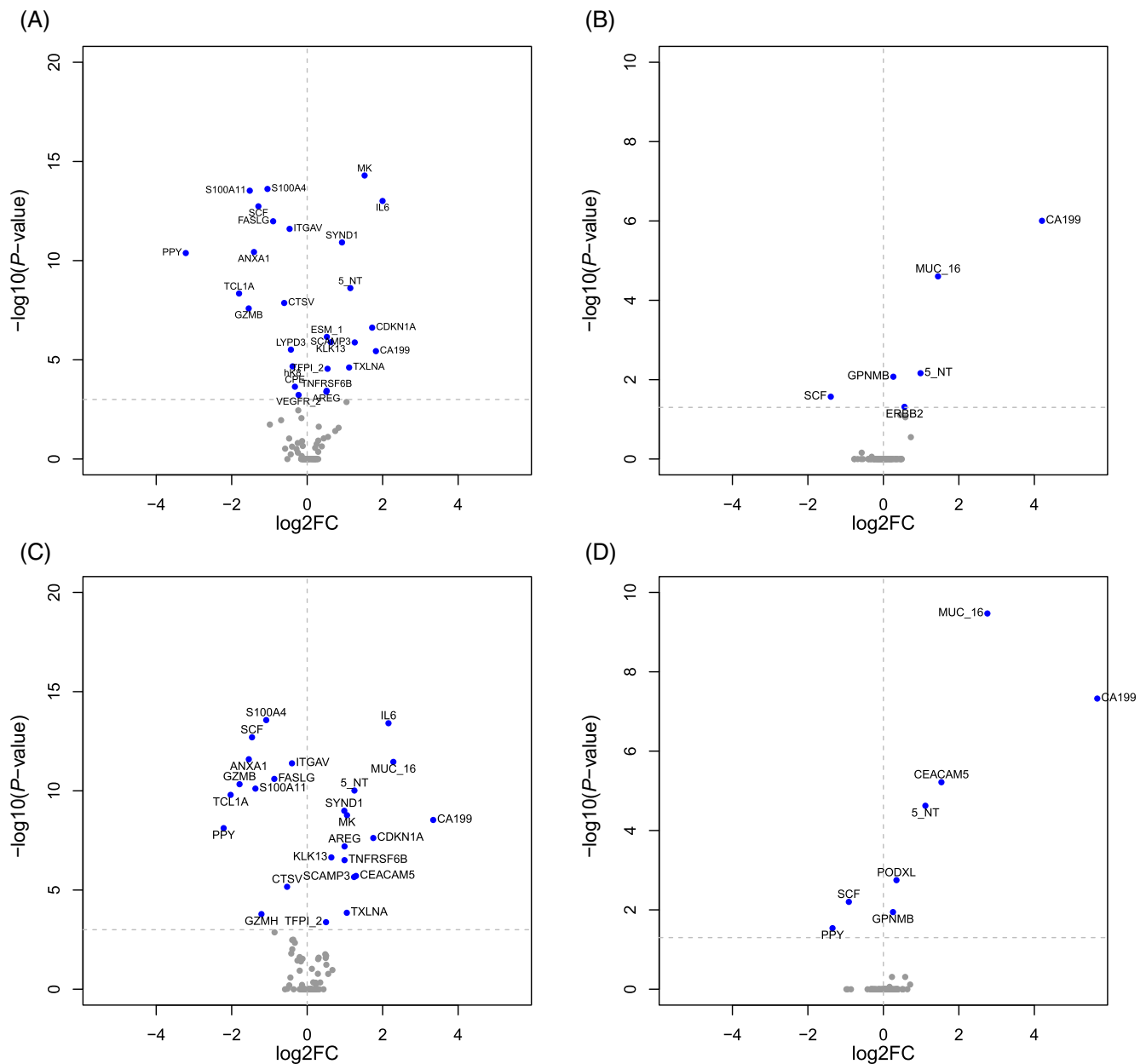
FIGURE 2 Plasma protein levels in early PDAC and advanced PDAC patients compared to healthy controls in the Swedish and Spanish participants. The volcano plots represent log2-fold change values of protein expressions in Swedish and Spanish plasma samples of early PDAC patients and healthy controls (A,B) and of advanced PDAC patients and healthy controls (C,D). Blue points indicate adjusted P value < .001 and .05 by Wilcoxon-Mann-Whitney test with Bonferroni correction in two populations, respectively; gray points indicate adjusted P value ≥ .001 and .05 in two populations, respectively; horizontal gray dashed line indicates adjusted P value = .001 and .05 in two populations, respectively; vertical gray dashed line indicates log2-fold change is zero; the statistically, significantly different proteins are labeled

protein expressions, and multiple comparisons were adjusted using Bonferroni correction, at significance levels of 0.001 and 0.05 for adjusted P-values in Swedish and Spanish samples, respectively. Protein expression was visualized using heatmaps,[19] with samples arranged together based on similar protein expression patterns using hierarchical cluster analysis with correlation distances and average linkage. Volcano plots were used to visualize marginal mean

differences in protein expression between PDAC patients and healthy control individuals.

We identified two panels of features for prediction modeling in (a) early PDAC patients vs healthy controls and (b) advanced PDAC patients vs healthy controls with a two-step method inspired by a previous study.[20] Briefly, backward elimination regression[21] was conducted to select a panel of features among 96 variables,

consisting of 93 protein biomarker candidates, age, sex and smoking status in the Swedish data. For performance assessment, to improve model robustness and avoid overfitting, participants were split into training sets and test sets (8:2) randomly by 5-fold cross-validation; in each training set, backward elimination regression was used to select features, and a support vector machine (SVM) was fit as a predictor via 5-fold cross-validation repeated three times[22]; the fit model was used in the corresponding test set to predict disease status, and the overall model performance was assessed via receiver operating characteristic (ROC) curve, and area under the curve (AUC) value with 95% confidence interval (CI) for the pooled prediction in all test sets. Furthermore, predictive performance of the panel identified from Swedish data was validated in an independent set of Spanish participants. For this external validation process, again an SVM was fit as predictor, here with leave-one-out cross-validation due to the smaller sample size, and predictive performance was assessed via ROC curve and AUC value. An identical process was conducted to identify a panel discriminating advanced PDAC from healthy controls, and ROC curve and AUC value were used to assess predictive performance.

Statistical analyses and data visualizations were performed using SAS (version 9.4; SAS Institute Inc., Cary, NC) and R (Version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria).

3 | RESULTS

3.1 | Patients overview and characteristic group comparisons

Overall, 220 Swedish and 130 Spanish samples were included for analysis as shown in the workflow (Figure 1). Among Swedish participants, we found no statistically significant differences in sex, and smoking status between early PDAC, advanced PDAC, benign lesions/chronic pancreatitis and healthy controls; however, they had different ages ($P = .001$), and diabetes condition ($P < 0.001$) with statistical significance (Table 1). Among Spanish participants, healthy controls were on average older (mean ages were 61.9, 63.9, 63.5 and 70.7 in early PDAC, advanced PDAC, chronic pancreatitis and healthy controls, respectively, $P = .023$); chronic pancreatitis patients were more likely to be current smokers ($P = .009$) and have diabetes ($P < .001$). No statistically significant difference was found in sex.

3.2 | Protein distributions

The 93 proteins were clustered using hierarchical clustering analysis in both study populations. Two clusters were observed in both
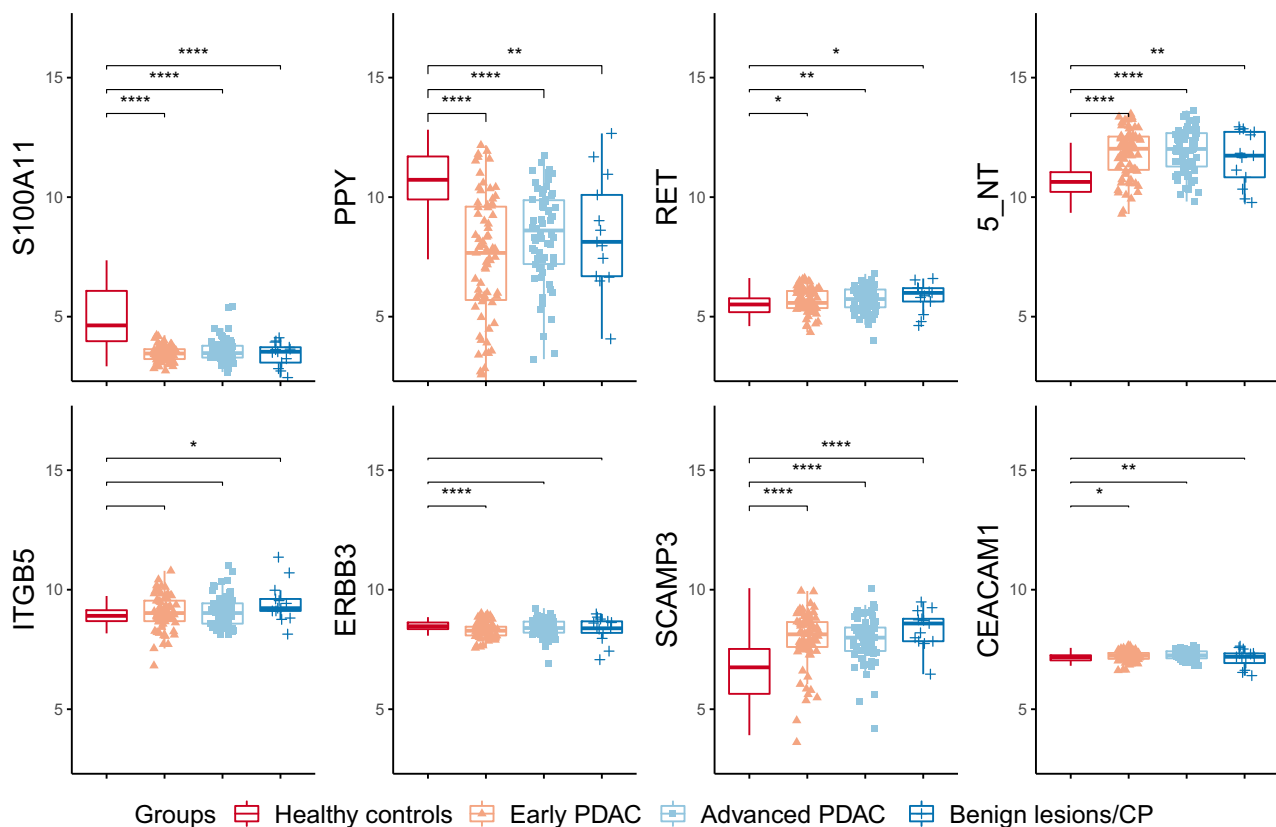

FIGURE 3 Eight selected plasma protein levels for discriminating healthy controls from early PDAC in the Swedish participants, across four groups. Levels are shown as jittered-level plots by groups, overlaid with boxplots. For each protein, the global null hypothesis of no difference across all four groups was tested via a Kruskal-Wallis test; group-wise comparisons for all pairs of groups were conducted via Wilcoxon rank-sum tests, with statistically significantly different pairs indicated by stared horizontal lines. *, Bonferroni-adjusted $P < .05$; **, Bonferroni-adjusted $P < .01$; ***, Bonferroni-adjusted $P < .001$; ****, Bonferroni-adjusted $P < .0001$

populations: in the Swedish samples, one cluster consisted predominantly of healthy controls, and the other mainly of PDAC patients (Supplementary Figure 3a); in the Spanish samples, the separation was less clear, with one cluster composed predominantly of PDAC patients, and the other a mixture of healthy controls and PDAC patients (Supplementary Figure 3b).

For early PDAC vs healthy controls, we found that among Swedish samples 27 of 93 and among Spanish samples 6 of 93 protein levels were different (adjusted P value <.001 and .05, in Figure 2A,B, respectively) with statistical significance, of which three proteins overlapped: stem cell factor (SCF), 5′-nucleotidase ecto (5′-NT) and CA19-9. For advanced PDAC, 25 of
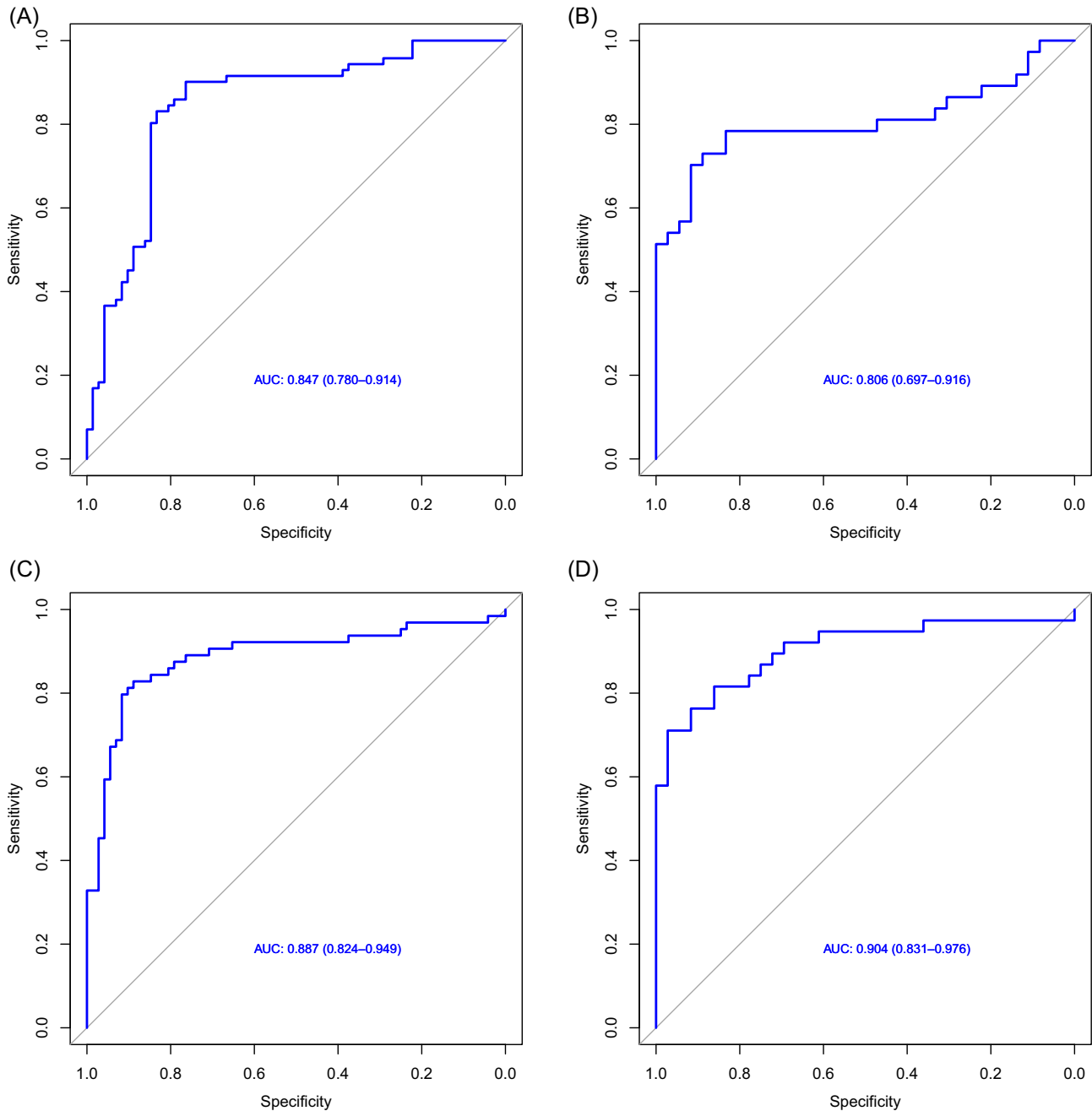

FIGURE 4 Predictive performance of selected plasma protein biomarkers. The results are presented with receiver operating characteristic (ROC) curves and corresponding cross-validated area under the curve (AUC) values with 95% confidence intervals (CIs) for early pancreatic ductal adenocarcinoma (PDAC) vs healthy control in test set of the Swedish participants (A) and external validation set of the Spanish participants (B), and for advanced PDAC vs healthy control in test set of the Swedish participants (C) and external validation set of the Spanish participants (D). Feature selection was validated in the Swedish participants, using support vector machine (SVM) with 5-fold cross-validation. The identified proteins for discriminating healthy controls from PDAC in the Swedish participants were externally validated in the Spanish participants, using SVM with leave-one-out cross-validation [Color figure can be viewed at wileyonlinelibrary.com]

93 and 8 of 93 protein levels were significantly different from healthy controls in the two study populations (adjusted *P* value <.001 and .05, respectively, in Figure 2C,D), and six proteins overlapped: SCF, 5′-NT, mucin 16 (MUC-16), CA19-9, pancreatic prohormone (PPY) and carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5).

3.3 | Feature selection and performance assessment

We used a backward elimination algorithm to select features from 93 proteins, age, sex, and smoking status in the Swedish participants, which were able to discriminate early PDAC from healthy controls. Features that did not improve the classification of outcomes were dropped, resulting in a panel of eight proteins, consisting of S100 calcium-binding protein A11 (S100A11), PPY, proto-oncogene tyrosine-protein kinase receptor Ret (RET), 5′-NT, integrin subunit beta 5 (ITGB5), receptor tyrosine-protein kinase erbB-3 (ERBB3), secretory carrier-associated membrane protein 3 (SCAMP3) and CEACAM1. For comparing distributions of these eight protein levels between four groups, boxplots are displayed in Figure 3. The cross-validated AUCs in test set of the Swedish participants and external validation set of the Spanish participants were 0.85 (95% CI 0.78-0.91) and 0.81 (95% CI 0.70-0.92), respectively (Figure 4A,B, sensitivities at given specificities are shown in Supplementary Table 3). We also used the same panel to discriminate early PDAC from chronic pancreatitis in the Spanish participants, with a cross-validated AUC of 0.76 (95% CI 0.62-0.91, Supplementary Figure 4), since statistical power was insufficient to identify a relevant protein panel for discriminating specifically benign lesions/chronic pancreatitis (only 13 cases) from early PDAC in the Swedish participants.

Moreover, we identified a panel of features from 96 variables in the Swedish participants to discriminate advanced PDAC from healthy controls, including tyrosine-protein kinase Lyn (LYN), ITGB5, CEACAM1, secreted protein acidic and cysteine rich (SPARC), alpha-taxilin (TXLNA), cyclin-dependent kinase inhibitor 1 (CDKN1A), annexin (ANXA) 1 and CA19-9, with two biomarkers appearing in both panels (ITGB5 and CEACAM1). The eight-protein panel achieved a cross-validated AUC of 0.89 (95% CI 0.83-0.95, Figure 4C) in Swedish participants, with the corresponding external AUC of 0.90 (95% CI 0.83-0.98, Figure 4D) among Spanish participants.

In comparison with the established protein marker CA19-9, the predictive performance of our eight-protein panel for discriminating early PDAC from healthy controls tended to be higher in both populations, but not the panel for discriminating late PDAC from healthy controls (Supplementary Figure 5).

4 | DISCUSSION

This study was designed to identify plasma protein biomarkers to discriminate PDAC patients in early stage from healthy controls. A panel of eight protein biomarkers predicting early PDAC status was selected in a Swedish case-control study, then validated in an independent Spanish case-control study. In the same manner, another eight-protein panel for discriminating advanced PDAC patients from healthy controls was selected in Swedish participants and then validated in Spanish participants.

The 93 tumor-related candidate proteins in this study are involved in multiple biological processes: angiogenesis, apoptotic process, cell adhesion, differentiation, proliferation, immune response and so on. Early PDAC patients expressed less PPY and S100A11 than healthy controls using log2-fold change values of normalized values (Figure 2A) and normalized values (Figure 3). PPY acts as a function regulator of the digestive system, and its concentration in plasma has been found to be lower in pancreatic cancer patients compared to noncancer participants, irrespective of the degree of glucose intolerance,[23] which is consistent with our finding. However, no different expression of this protein among PDAC, chronic pancreatitis and healthy controls regardless of glycemic status was reported in another study.[24] Ohuchida et al[25] reported that expression of S100A11, involved in regulation of cellular processes, was increased in the early stage of pancreatic carcinogenesis and decreased during subsequent progression to cancer; while in our study, PDAC patients had lower level of S100A11 than controls. On the other hand, 5′-NT and SCAMP3 were overexpressed in early PDAC patients compared to healthy controls in our results: 5′-NT is a membrane protein, and its overexpression in tumor tissues or blood is associated with several types of cancer and tumor progression[26,27]; SCAMP3 is found to be overexpressed in hepatocellular carcinoma patients with poor prognosis.[28] Of note, CA19-9 was not included in this panel, since it was eliminated in the feature selection process due to insufficient contribution of discriminatory power compared to the other candidate features. This is however in line with the study from Mellby et al,[20] which did not include CA19-9 in the identified protein panel either. Although the blood-based protein biomarker CA19-9 is widely used clinically, it lacks accuracy for early detection of PDAC; however, some studies reported improved performance for a combination of CA19-9 with other biomarkers, such as thrombospondin-2 (THBS2),[29] plasma tissue factor pathway inhibitor (TFPI), tenascin C (TNC),[30] trefoil factors (TFFs),[31] or circulating tumor DNA (ctDNA),[32] over CA19-9 alone. Our results also revealed that the eight-protein panel improved the diagnostic performance of discriminating early PDAC, a stage which has the possibility for receiving surgical resection, from healthy controls as compared to CA19-9 alone. Not surprisingly, the eight-protein panel for advanced PDAC vs healthy controls included CA19-9, since advanced PDAC patients had much higher level of CA19-9 than healthy controls.

Recently, Mellby and colleagues[20] selected 29 protein biomarkers in serum by using an antibody microarray technique for discriminating early PDAC patients (n = 148) and advanced PDAC patients (n = 295) from controls (n = 888) in a Scandinavian cohort, with validated performance in an independent US cohort. In our study, we had the same geographic population for feature selection (Scandinavian country), but had different biological nature of samples (serum vs plasma),

protein detection methods (microarray technology vs PEA), candidate protein pools (156 vs 93), the final selected protein biomarkers and sample sizes. AUC value was 0.96 for the classification of early PDAC from healthy controls and 0.98 for discriminating advanced PDAC from healthy controls in the reference study;[20] while in our findings, the cross-validated AUC values were 0.85 (95% CI, 0.78-0.91) and 0.89 (95% CI, 0.83-0.95), respectively, lower than both comparisons mentioned earlier, which is partly attributable to less proteins included in the identified panel in our study.

When validating the identified protein panels in the Spanish participants, predictive performances were similar to those in the Swedish participants, with the absolute difference of AUCs small compared to the overall statistical uncertainty, despite differences in source population and pre-analytical handling of samples. However, the actual shapes of ROC curves, and the implied tradeoff between sensitivity and specificity, were different in the two populations for early PDAC vs healthy controls. For example, at a fixed sensitivity of 80%, the eight-protein panel can distinguish early PDAC patients from healthy individuals with a specificity of 83% in the Swedish participants, but only 50% in the Spanish participants, with a wide CI. This suggests that protein expressions between Swedish and Spanish participants were different from each other; however, to some degree, this variation somewhat increases our confidence in the generalizability of our results to other populations.

Generally, predictive performance of the protein panel for early PDAC vs healthy controls tended to be lower than that for advanced PDAC vs healthy controls, which demonstrates that it is harder to distinguish early PDAC patients from healthy individuals compared to advanced PDAC patients. Cancer progression is dynamic and somatic mutation events happen in different stages of carcinogenesis with different protein expressions,[5] so the molecular characteristics may be less sensitive for discriminating early PDAC patients from healthy individuals when there are no specific clinical symptoms. Finding protein biomarkers for advanced PDAC is clinically and etiologically less interesting, but we chose to include these results to demonstrate how biomarker profiles can change during disease progression, possibly also reflecting changes in the underlying biology of the disease. Additionally, we found that the early PDAC protein panel could also be used to discriminate early PDAC from chronic pancreatitis with an AUC value of 0.76 in the Spanish participants; this suggests that differential diagnosis between early PDAC and patients with chronic pancreatitis or benign lesions is sometimes difficult. Therefore, targeted screening and regular follow-up in high-risk populations are important.

The results of our protein panels should be interpreted with caution. First, about one-third of PDAC and chronic pancreatitis patients in the Swedish data had missing diabetes status. This variable was consequently not informative, and therefore not considered for feature selection. Studies have shown that new-onset diabetes may indicate a substantial burden of malignancy,[33] which could lead to early detection of asymptomatic PDAC. However, there is no pronounced difference in distributions of diabetes status between PDAC patients and healthy individuals in the Spanish participants, which should limit the effect on predictive performance of the proposed protein panels. Second, we found an eight-protein panel to discriminate early PDAC among participants with known disease status, and it may not perform as well in samples before any diagnosis. However, it will most likely be useful in high-risk individuals (with chronic pancreatitis, older age, etc.) rather than in the general population, since general population-based screening would result in a large number of false positives given the low prevalence of pancreatic cancer.[34] At last, further work is needed to define the decision points of the identified panel for early detection of PDAC, and to determine how these decision points change with tumor progression, as well as to determine the specificity for distinguishing PDAC from other cancers, since the selected protein biomarkers are not organ-specific.

Identification of biomarkers in liquid (blood, juice, urine, etc.) for early detection of cancers has been widely studied recently, because it can not only provide evidence for therapeutic intervention of cancers but can also provide deeper insight into carcinogenesis and progression. Besides protein biomarkers, ctDNA, microRNAs, circulating tumor cells and exosomes have been evaluated for early detection of cancers; however, several challenges still exist before clinical applications, such as technique standardization, cost controlling, and stability and accuracy of the detection approach, while protein biomarkers still have advantages. Our eight-protein panel is promising for early detection of pancreatic cancer, although the comparatively low diagnostic accuracy is a point of concern. Multicenter prospective studies and validations are needed before establishing a diagnosis panel with high clinical value.

In summary, we identified and validated an eight-protein panel of biomarkers for discriminating early PDAC patients from healthy individuals; protein biomarkers require further study and if possible, validation in multicenter prospective studies.

CONFLICT OF INTEREST

Co-author Maximilian Kordes played a role of consulting in Alligator Bioscience and Roche, and received funds for traveling and accommodation from Molecular Health. All other authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

ORCID

Jingru Yu https://orcid.org/0000-0002-4157-4647
Alfredo Carrato https://orcid.org/0000-0001-7749-8140
Francisco X Real https://orcid.org/0000-0001-9501-498X
Núria Malats https://orcid.org/0000-0003-2538-3784
Weimin Ye https://orcid.org/0000-0002-6859-4648

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin*. 2017;67:7-30.
2. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res*. 2014;74:2913-2921.
3. Gillen S, Schuster T, Meyer Zum Buschenfelde C, et al. Preoperative/neoadjuvant therapy in pancreatic cancer: a systematic review and meta-analysis of response and resection percentages. *PLoS Med*. 2010;7:e1000267.
4. Yachida S, Jones S, Bozic I, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*. 2010;467:1114-1117.
5. Yachida S, Iacobuzio-Donahue CA. Evolution and dynamics of pancreatic cancer progression. *Oncogene*. 2013;32:5253-5260.
6. Berger AC, Garcia M Jr, Hoffman JP, et al. Postresection CA 19-9 predicts overall survival in patients with pancreatic cancer treated with adjuvant chemoradiation: a prospective validation by RTOG 9704. *J Clin Oncol*. 2008;26:5918-5922.
7. Passerini R, Cassatella MC, Boveri S, et al. The pitfalls of CA19-9: routine testing and comparison of two automated immunoassays in a reference oncology center. *Am J Clin Pathol*. 2012;138:281-287.
8. Capello M, Bantis LE, Scelo G, et al. Sequential validation of blood-based protein biomarker candidates for early-stage pancreatic cancer. *J Natl Cancer Inst*. 2017;109(4):djw266.
9. Resovi A, Bani MR, Porcu L, et al. Soluble stroma-related biomarkers of pancreatic cancer. *EMBO Mol Med*. 2018;10:e8741.
10. Radon TP, Massat NJ, Jones R, et al. Identification of a three-biomarker panel in urine for early detection of pancreatic adenocarcinoma. *Clin Cancer Res*. 2015;21:3512-3521.
11. Chari ST, Kelly K, Hollingsworth MA, et al. Early detection of sporadic pancreatic cancer: summative review. *Pancreas*. 2015;44:693-712.
12. Birgisson H, Tsimogiannis K, Freyhult E, Kamali-Moghaddam M. Plasma protein profiling reveal Osteoprotegerin as a marker of prognostic impact for colorectal cancer. *Transl Oncol*. 2018;11:1034-1043.
13. Abrahamsson A, Rzepecka A, Dabrosin C. Equal pro-inflammatory profiles of CCLs, CXCLs, and matrix Metalloproteinases in the extracellular microenvironment in vivo in human dense breast tissue and breast cancer. *Front Immunol*. 2017;8:1994.
14. Al-Talabany S, Mordi I, Graeme Houston J, et al. Epicardial adipose tissue is related to arterial stiffness and inflammation in patients with cardiovascular disease and type 2 diabetes. *BMC Cardiovasc Disord*. 2018;18:31.
15. Huang J, Lohr JM, Nilsson M, et al. Variant profiling of candidate genes in pancreatic ductal adenocarcinoma. *Clin Chem*. 2015;61:1408-1416.
16. Edge SB, Byrd DR, Compton CC, et al., eds. *AJCC Cancer Staging Manual*. 7th ed. New York, NY: Springer; 2010.
17. Molina-Montes E, Coscia C, Gómez-Rubio P, et al. Deciphering the complex interplay between pancreatic cancer, diabetes mellitus subtypes and obesity/BMI through causal inference and mediation analyses. *Gut*. 2021;70:319-329.
18. Olink. https://www.olink.com/question/data-normalization. Accessed June 24, 2020.
19. Ploner A. Heatplus: Heatmaps with row and/or column covariates and colored clusters. R package version 2.32.1. https://github.com/alexploner/Heatplus. Accessed June 15, 2020.
20. Mellby LD, Nyberg AP, Johansen JS, et al. Serum biomarker signature-based liquid biopsy for diagnosis of early-stage pancreatic cancer. *J Clin Oncol*. 2018;36:2887-2894.
21. Thomas Lumley based on Fortran code by Alan Miller. leaps: Regression Subset Selection. R package version 3.0. https://CRAN.R-project.org/package=leaps. Accessed June 15, 2020.
22. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(5):1-26.
23. Škrha J, Bušek P, Uhrová J, et al. Lower plasma levels of glucose-dependent insulinotropic peptide (GIP) and pancreatic polypeptide (PP) in patients with ductal adenocarcinoma of the pancreas and their relation to the presence of impaired glucoregulation and weight loss. *Pancreatology*. 2017;17:89-94.
24. Nagpal SJS, Bamlet WR, Kudva YC, Chari ST. Comparison of fasting human pancreatic polypeptide levels among patients with pancreatic ductal adenocarcinoma, chronic pancreatitis, and type 2 diabetes mellitus. *Pancreas*. 2018;47:738-741.
25. Ohuchida K, Mizumoto K, Ohhashi S, et al. S100A11, a putative tumor suppressor gene, is overexpressed in pancreatic carcinogenesis. *Clin Cancer Res*. 2006;12:5417-5422.
26. Wang G, Fu S, Li D, Chen Y. Expression and clinical significance of serum NT5E protein in patients with colorectal cancer. *Cancer Biomark*. 2019;24:461-468.
27. Bertoni APS, Bracco PA, de Campos RP, et al. Activity of ecto-5′-nucleotidase (NT5E/CD73) is increased in papillary thyroid carcinoma and its expression is associated with metastatic lymph nodes. *Mol Cell Endocrinol*. 2019;479:54-60.
28. Naboulsi W, Bracht T, Megger DA, et al. Quantitative proteome analysis reveals the correlation between endocytosis-associated proteins and hepatocellular carcinoma dedifferentiation. *Biochim Biophys Acta*. 2016;1864:1579-1585.
29. Kim J, Bamlet WR, Oberg AL, et al. Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers. *Sci Transl Med*. 2017;9(398):eaah5583.
30. Balasenthil S, Huang Y, Liu S, et al. A plasma biomarker panel to identify surgically Resectable early-stage pancreatic cancer. *J Natl Cancer Inst*. 2017;109(8):djw341.
31. Jahan R, Ganguly K, Smith LM, et al. Trefoil factor(s) and CA19.9: a promising panel for early detection of pancreatic cancer. *EBioMedicine*. 2019;42:375-385.
32. Cohen JD, Javed AA, Thoburn C, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc Natl Acad Sci U S A*. 2017;114:10202-10207.
33. Chari ST, Klee GG, Miller LJ, Raimondo M, DiMagno EP. Islet amyloid polypeptide is not a satisfactory marker for detecting pancreatic cancer. *Gastroenterology*. 2001;121:640-645.

34. Lindquist CM, Miller FH, Hammond NA, Nikolaidis P. Pancreatic cancer screening. *Abdom Radiol (NY)*. 2018;43:264-272.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.