Circulating EVs long RNA-based subtyping and deconvolution enable prediction of immunogenic signatures and clinical outcome for PDAC

Yuchen Li,[1,2,3] Ye Li,[1,2,3] Shulin Yu,[1,2,3] Ling Qian,[1,2] Kun Chen,[1,2] Hongyan Lai,[1,2] Hena Zhang,[1,2] Yan Li,[1,2] Yalei Zhang,[1,2] Sijia Gu,[1,2] Zhiqiang Meng,[1,2] Shenglin Huang,[1,2] and Peng Wang[1,2]

[1]Department of Integrative Oncology, Fudan University Shanghai Cancer Center, and the Shanghai Key Laboratory of Medical Epigenetics, Institutes of Biomedical Sciences, Fudan University, 270 Dong An Road, Shanghai 200032, China; [2]Department of Oncology, Shanghai Medical College, Fudan University, 130 Dong An Road, Shanghai 200032, China

Identification of clinically applicable molecular subtypes of pancreatic ductal adenocarcinoma (PDAC) is crucial to improving patient outcomes. However, the traditional tissue-dependent transcriptional subtyping strategies are invasive and not amenable to routine clinical evaluation. In this study, we developed a circulating extracellular vesicle (cEV) long RNA (exLR)-based PDAC subtyping method and provided exLR-derived signatures for predicting immunogenic features and clinical outcomes in PDAC. We enrolled 426 individuals, among which 227 PDACs served as an internal cohort, 118 PDACs from two other medical centers served as an independent validation cohort, and 81 healthy individuals served as the control. ExLR sequencing was performed on all plasma samples. We found that PDAC could be categorized into three subtypes based on plasma exLR profiles. Each subpopulation showed its own molecular features and was associated with patient clinical prognosis. The immunocyte-derived cEV fractions were altered among PDAC subtypes and interconnected with tumor-infiltrating lymphocytes in cancerous tissue. Additionally, we found a significant concordance of immunoregulators between tissue and blood EVs, and we harvested potential PDAC therapeutic targets. Most importantly, we constructed a nine exLR-derived, tissue-applicable signature for prognostic assessment of PDAC. The circulating exLR-based features may offer an attractive platform for personalized treatment and predicting patient outcomes in multiple types of cancer.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is currently the fourth leading cause of cancer-related death worldwide.[1] The treatment efficiency of PDAC is restricted by the aggressive tumor-intrinsic characteristics and suppressive tumor microenvironment (TME).[2] Previous studies showed that personalized treatment approaches based on PDAC genotypic and phenotypic characteristics may potentially improve patient outcome.[3,4] However, most of these reports focused on gene subtyping and exploring biomarkers in resectable tumors. Resection is possible only in a minority of patients, as most pa-

tients present at an advanced stage at the time of diagnosis,[5] and safely obtaining sufficient quantities of pancreatic tumor tissue for molecular analysis is difficult at unresectable stages. Therefore, the development of specific, robust, and noninvasive molecular subtyping strategies for the stratification and prognostic evaluation of PDAC are required.

Extracellular vesicles (EVs) have become a hotspot in liquid biopsy for cancer precision medicine. EVs are nanometer-sized membrane-bound particles that are shed from both tumor and nonneoplastic cells into the peripheral circulation. EVs deliver a milieu of cargo material, such as proteins, genomic DNA, and lipids, as well EV long RNAs (exLRs), from parent cells to receptive cells.[6] Circulating EVs (cEVs) contain approximately 10,000 exLRs, mainly mRNAs, long non-coding RNAs, and circular RNAs, which are stably expressed and have shown clinical significance.[7] Tissue- and tumor-originated exLRs are specifically enriched in cEVs and reflect the biological activities, metabolic conditions, and immunogenic features of their host cells.[8] Previous studies showed that PDAC tumor-derived EVs regulate the immunosuppressive TME and deliver anti-tumoral transcriptomic signals, and that plasma-originated exLRs are critical indicators of PDAC antitumor immune activities.[9,10] Additionally, exLRs are involved in critical processes of PDAC carcinogenesis, including tumor growth, metastasis, drug resistance, and immune modulation.[11,12] Therefore, the mixture of non-enriched cEVs containing both tumor-derived and non-tumor-derived exLRs[13] may

reflect tumor-intrinsic and non-tumor signatures, and evaluation of exLRs may represent a new strategy for noninvasive molecular stratification, prognostic evaluation, and treatment monitoring in PDAC.

We previously determined the abundance of exLRs in the plasma of cancer patients[14] and developed a strategy for exLR sequencing (exLR-seq) in human plasma.[15] Our study indicated that cEVs contain approximately 10,000 exLRs, mainly mRNAs, which are stably expressed, harbor tissue-cellular specificity, and could serve as biomarkers for PDAC early screening.[13]

In this study, we evaluated the potential usefulness of the circulating exLR profile for PDAC subtyping. We traced the tissue-cellular source contribution of cEVs and explored the correlation of the immune landscape between tumors and cEVs. Our study established a prognostic signature enabling noninvasive prediction of immunogenic features and clinical outcome for PDAC.

RESULTS

Circulating exLR-based subtyping of PDAC

The exLR-seq approach was implemented on plasma samples from an internal cohort of 227 PDAC patients (Table S1). Approximately 14,000 exLRs were reliably detected, including protein-coding genes (10,253, 74.91%), circRNAs (1,322, 9.66%), pseudogenes (1,218, 8.9%), lncRNA genes (764, 5.58%), and other small non-coding RNAs (130, 0.95%) (Figure S1A). We included the top 60% aberrantly expressed exLRs as candidate targets to explore underlying exLR-based PDAC subtypes. We assessed clustering stability using the consensus-clustering algorithm, which indicated the existence of three major clusters (Figures S1B–S1F; see Materials and methods); 60, 97, and 70 cases were stratified to clusters C-I, C-II, and C-III, respectively (Figure 1A). Principal-component analysis demonstrated a clear distinction between each subgroup (Figure 1B). The number of detected exLR species shows a significant difference between three clusters. In particular, the amount of enriched protein-coding genes in C-III was greater than in other group sets (C-I and C-II), which indicates that plasma EVs of the third cases were additionally enriched with exLRs, which could conduct exclusive biological functions (Figure 1C).

Evaluation of the clinical characteristics of the three clusters revealed a significantly increased number of patients with advanced tumor stage and shorter overall survival (OS) time in C-III; the other characteristics (age, sex, and CA19-9) were not associated with exLR-based stratification (Figure 1D). Survival analysis showed that C-III correlated with the shortest OS, while C-I correlated with the longest OS (log-rank test, p < 0.001; Figure 1E). Therefore, our results suggested that circulating exLR-based subtyping enables noninvasive prediction of clinical outcome for pancreatic cancer.

Molecular features and pathway annotation of exLRs in exLR-based clusters

To assemble these dysregulated circulating exLRs and explore their underlying biological functions in each exLR-based cluster, the weighted correlation network analysis (WGCNA) approach was used to generate co-expressing exLR modules based on circulating exLR-seq expression profiles, and it correlated each module with phenotypes and clinical parameters (see Materials and methods and Figure S2A). The variably expressed exLRs were distributed in eight independent modules according to their degree of connectivity. Six gene modules (M2–M7) were significantly associated with C-III and one module (M1) was particularly referred to C-I (Spearman's correlation analysis, p < 0.001, Rho > 0.45; Figures 2A and S2B). The exLRs that could not be contained in any modules were placed into M8. All C-III-related modules were positively correlated with tumor aggressiveness, whereas the inverse was observed with the C-I-specific module; none of the modules showed any demographic bias (Figure 2A).

Gene Ontology (GO) enrichment analysis was conducted to annotate exLR genes in each module (Figure S2B). The exLRs of M1 (corresponding to C-I) were involved in the adhesion and migration of hemopoietic cells and represented favorable indicators for PDAC prognosis (Figures 2B and S2C). Conversely, the exLRs of M2–M7 (C-III) were involved in the activation of tumoral and immunogenic features, including RNA anabolic and catabolic processes, energy metabolism, cellular and immune responses, and DNA replication (Figures 2B and S2C; Tables S2–S8). The pre-ranked gene set enrichment analysis (GSEA) strategy identified 191 pathways that were significantly different in each cluster (C-I, n = 57; C-II, n = 42; C-III, n = 92; false discovery rate [FDR] < 0.1, normalized enrichment score > 1.5; Table S9). Unsupervised hierarchical clustering of the absolute enrichment score determined two distinct pathway signatures enriched in C-I and C-III samples (Figure 2C). C-I was exclusively enriched in vesicle-mediated molecule transport, GPR signal transduction, and cell motility-related pathways, while C-III was enriched in cell proliferation and metastasis, metabolism, and immune-related pathways (Table S9). The activation of chemokines, T cell receptor (TCR)/B cell receptor (BCR) and antigen processing and presentation signaling, as well as complement component genes overexpressed in cEVs suggests that patients in the C-III group showed primary and adaptive immune reactions in response to PDAC development.

Correlation analysis revealed 77 signaling pathways that showed similar patterns in cEVs and tissue profiles, including PDAC- and immune-related hallmark signatures, such as JAK-STAT, transforming growth factor β (TGF-β), and WNT/β-catenin pathways and TLR-9/10 cascades (Figure 2D; Table S10). These results suggest that circulating exLRs serve as noninvasive indicators to jointly reflect tumor-intrinsic and antitumor immune signatures.

Comprehensive landscape of tissue-cellular origins for cEVs

We previously developed a deconvoluting approach, termed EV-origin, as a computational framework to infer the tissue-cellular origins of cEVs and cell type-specific genes from transcriptomic profiles of cEVs.[16] In this study, we expanded the number of traceable hemopoietic cell fractions. A total of 39 types of tissue-cellular fractions were relatively or absolutely quantified, including 23 hematopoietic
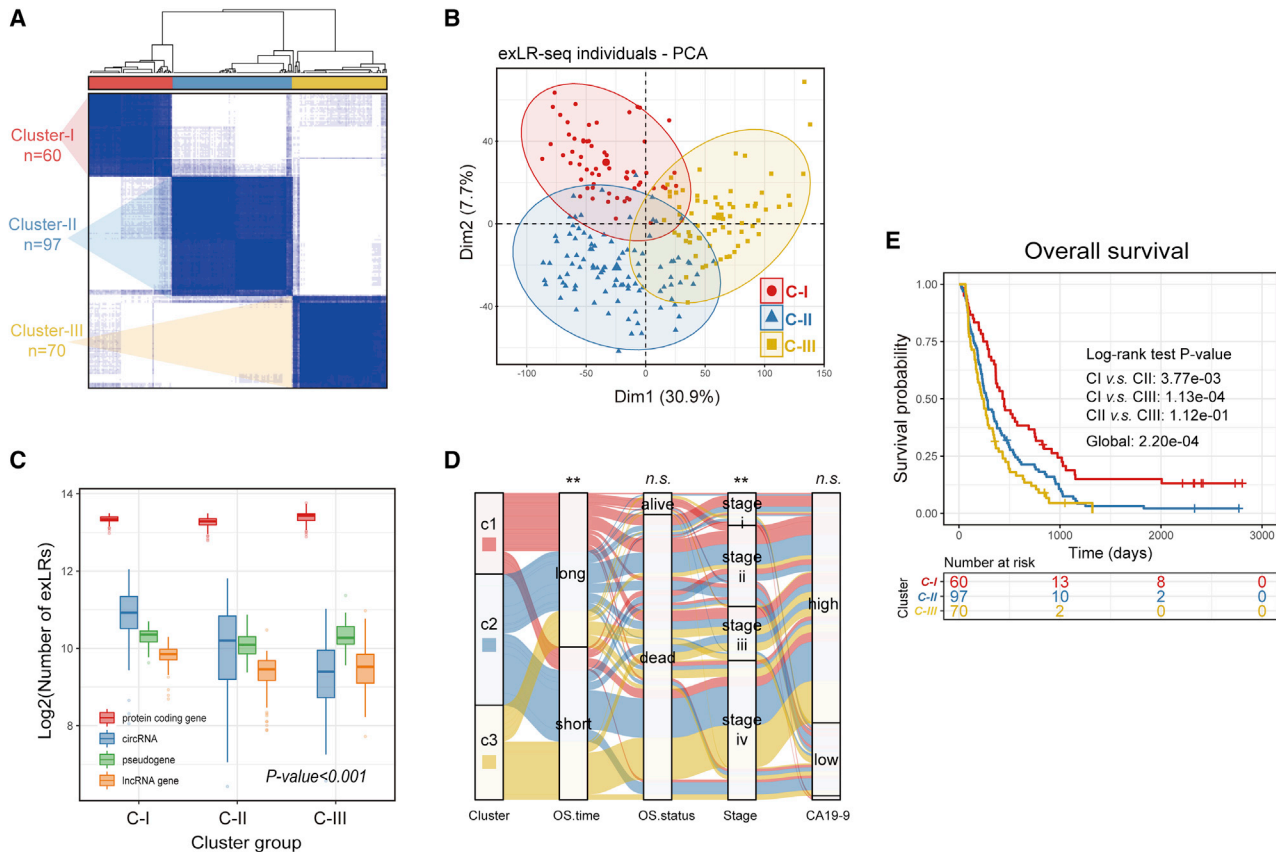
Figure 1. PDAC prognostic subtyping based on circulating exLR-seq profile

(A) Consensus clustering matrix of exLR-seq data from the internal cohort for aggregating number k (k = 3), representing the clustering stability using 500 iterations of the hierarchical clustering approach. The heatmap depicts the relative abundance of signature exLRs. (B) Two-dimensional scatterplot of principal-component analysis (PCA); first two principal components derived from expression intensities of variable exLRs, with samples connected by centroids according to exLR-based PDAC subsets (red, C-I; blue, C-II; yellow, C-III), are shown. The ellipse indicates the 90% confidence interval for each clustered group. (C) The relative distribution of exLR species per sample among PDAC clusters. The diagram shows exLRs reliably enriched in plasma vesicles, with a frequency > 0.1 and mean expression value > 1. The center of the boxplots indicates median values; the boundaries of the boxes show 25% and 75% quantiles. (D) Alluvial diagram shows differences among PDAC clusters according to overall survival (OS), patient status, tumor stage, and carbohydrate antigen 19-9 (CA19-9). CA19-9 level greater than 100 U/mL was defined as high. **p < 0.01; n.s., not significant. (E) Kaplan-Meier curves for OS in the 227 PDAC patients in the internal exLR-seq cohort indicate that exLR-based phenotypes strongly correlated with PDAC prognosis (log-rank test, p < 0.001).

components and 16 types of tissue constituents. Detailed information of the candidate tissue/blood components with their hierarchical relationships are shown in Table S11. Using the updated EV-origin strategy, we portrayed an integrated atlas of a traceability system of cEVs in exLR-seq samples from the internal cohort. In evaluation of hemopoietic components, cEVs predominantly originated from the hematopoietic stem cell fraction (platelets, 57%; red blood cells, 13%), followed by immunocyte components (30%) (Figure 3A, left). Concerning the proportion of immune cell sources, lymphocytes accounted for 81%, and the remaining myeloid component was 19% (Figure 3A, right). These immune cell constituents were categorized into adaptive (71%) and innate (29%) immune-related components (Figure 3A, right; Table S11). CD8/CD4+ T cells accounted for 66% of all immunocyte-derived constituents, and there was also a considerable abundance of immunosuppressive cell components (regulatory

T cells, 5%; monocytes, 16%). We also evaluated the abundance of tissue/solid organ-originated cEVs in PDAC plasma. Adipose tissue was the most prominent tissue contributor to plasma EVs (72%), followed by muscle (10%), lung (4%), and liver (4%); the pancreas accounted for only 0.24% (Figure 3B; Table S8).

We additionally evaluated the cEV fractions from 81 healthy individuals. In the comparison of absolute fractions of components, there was no statistically significant difference in blood cell/tissue-derived fractions between healthy and PDAC individuals (Mann-Whitney U test, p > 0.05; Figure 3C). Source contribution heterogeneity analysis revealed that 80% (12/15) of the tissue-derived components were increased in the PDAC group, while 65% (15/23) of hemopoietic-sourced fractions were highly enriched in the healthy cohort (absolute score, mean fold change > 1; Figure 3C). The enriched level of the
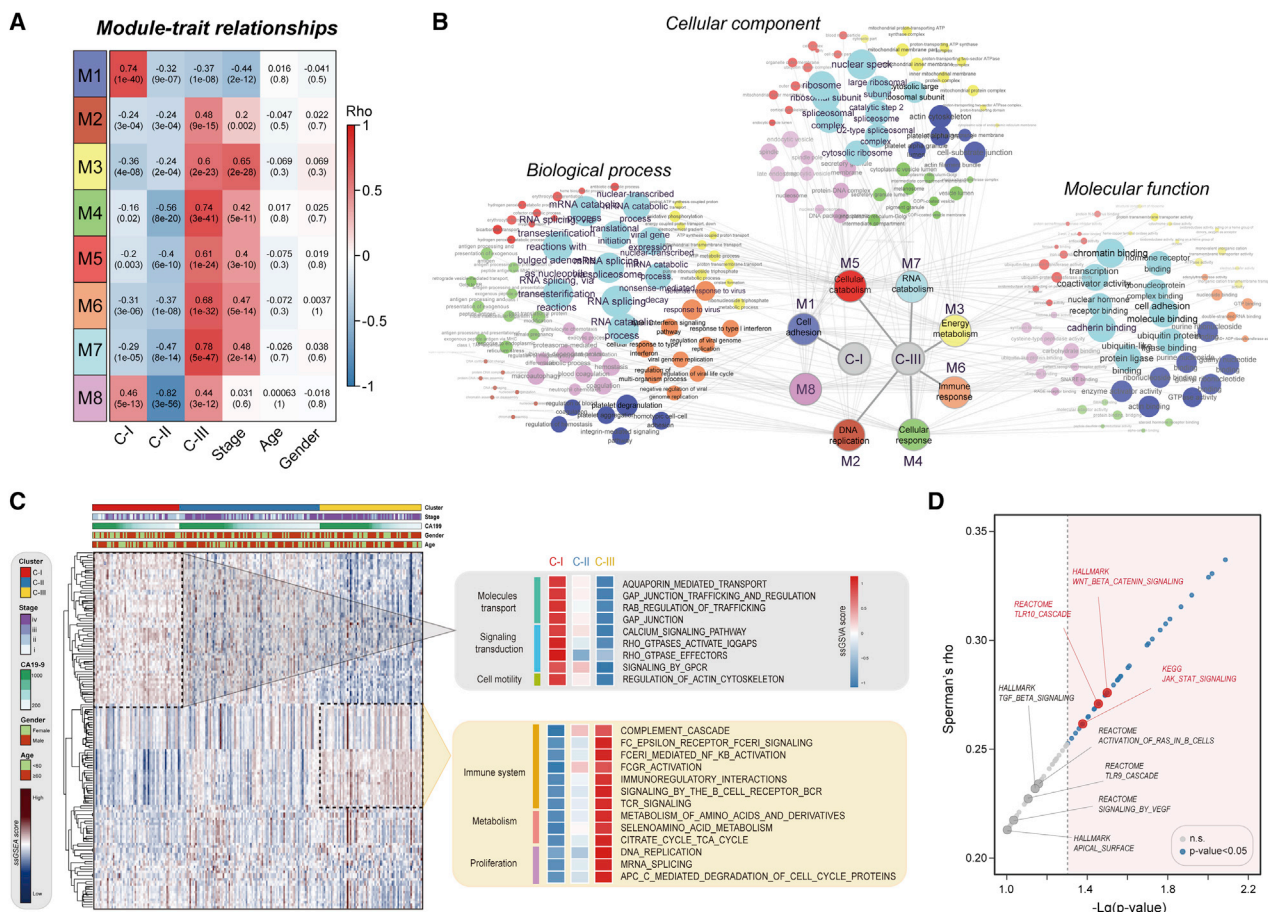
Figure 2. Systematic annotation of molecular features and pathway activation of exLR-based clusters

(A) The module-trait correlating matrix displays the correlation coefficients and significance levels of PDAC clusters and clinical and demographical factors for each exLR module. (B) Systematic Gene Ontology (GO) annotation network for genes in the eight modules in biological process, cellular component, and molecular function aspects. We filtered the top 10 significant annotating terms of each module and displayed these terms on the network diagram. The annotated GO terms are shown on each ellipse; the size of the ellipse reflects the degree of enrichment (−log10[p value]), and the transparency of the text indicates the number of exLR genes from each enriched GO term (log₂ transferred). The gray connection communicates each gene module with its corresponding annotation information. (C) Unsupervised clustering heatmap of ssGSEA enrichment score of 191 cluster-specific signaling pathway signatures in the exLR-seq data from the internal cohort. Hierarchical clustering was implemented with Euclidean distance and Ward linkage. Rows represent the signaling pathways derived from the current MSigDB 7.2 database that includes three major gene set collections (Hallmark, Reactome, and KEGG sets), and columns represent plasma samples. The right panels display the average normalized enrichment score (NES) of represented signaling pathways in PDAC clusters I and III. (D) The scatterplot demonstrates the tissue-cEV correlated signaling pathways from 61 PDAC patients with cEV and matched tissue transcriptomic expression profiles.

pancreatic tissue-sourced cEV fraction demonstrated a modest upregulation in the PDAC cohort (Figure 3D).

Immunocyte-derived EVs perturbed in PDAC subtypes and interconnected with tumor-infiltrating lymphocytes (TILs) in tissue

We next investigated the underlying heterogeneity of cEVs and their immunogenic source contribution among the PDAC phenotypes. The immunocyte-derived EV fractions can be categorized into innate and adaptive immunity categories to decode complex TME components. The three clusters showed significant differences in four types of innate immune constituents (i.e., monocytes, natural killer [NK]

cells, dendritic cells [DCs], and γδ T cells). C-III exhibited the highest proportion of all four innate immune cell types while C-I showed the lowest proportions (Kruskal-Wallis rank-sum test, p < 0.05; Figure 4A, top). Among the adaptive immunity constituents, T helper (Th)2 and Th17 cell fractions were upregulated in C-I, while high proportions of T regulatory (Treg) cells, naive Th cells, central memory CD8 T cells, memory B cells, and plasmablasts were observed in C-III (Figure 4A, bottom). We further examined the absolute numbers to determine the true abundance of cEVs derived from each immune cell population. The immunogenic composition of cEVs gradually increased from C-I to C-III (mean absolute score: C-I, 0.38; C-II, 0.40; C-III, 0.51; Figure 4B). The levels of innate and adaptive
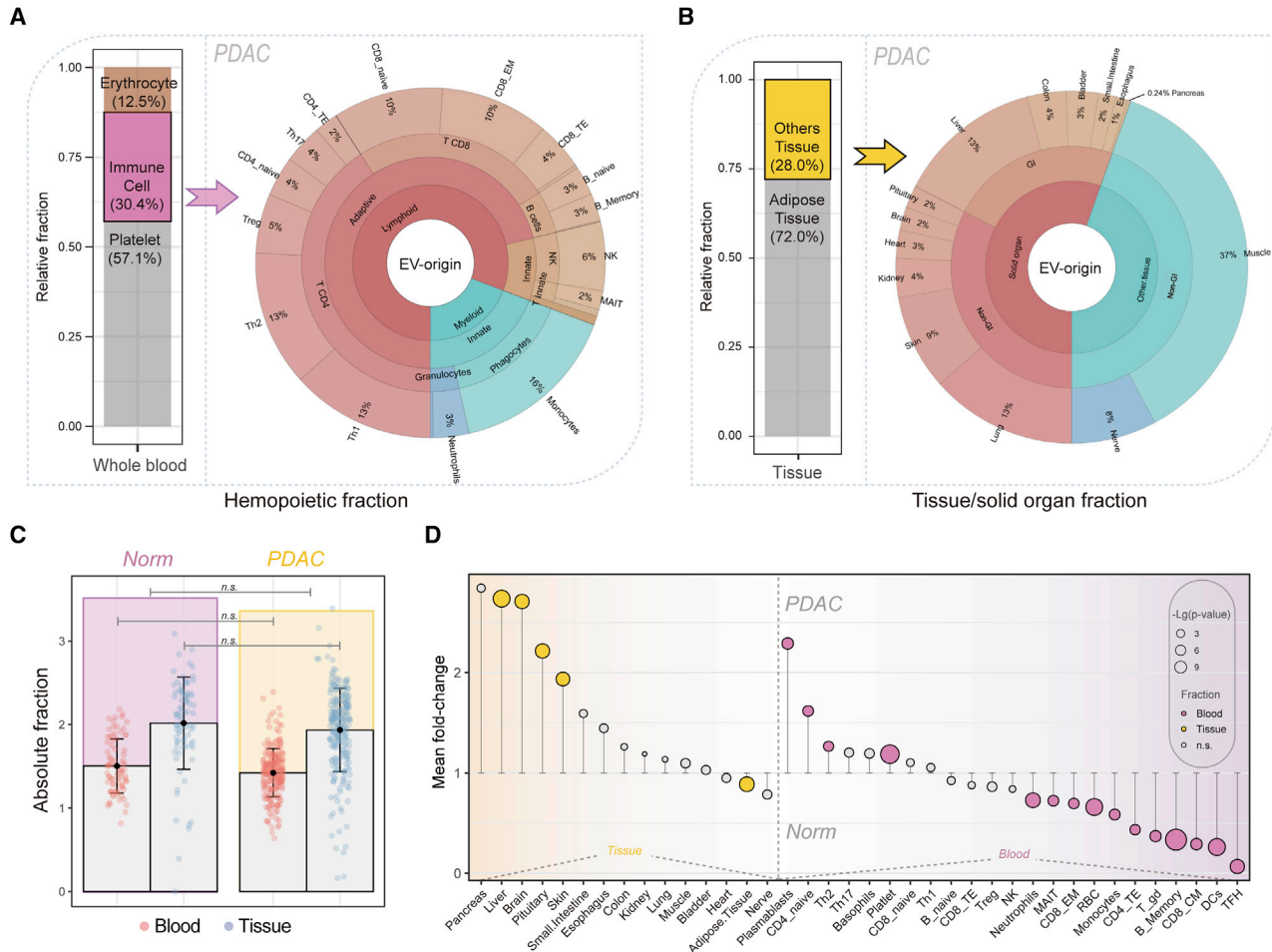
Figure 3. Comprehensive landscape of tissue-cellular origins for cEVs

(A) Relative distribution of all blood cell components estimated using the EV-origin approach from exLR-seq data from the internal PDAC cohort. The bar chart on the left displays the relative abundance of cEVs from erythrocytes, immune cells, and platelets. The multilayer pie chart on the right shows the relative proportions of immunocyte-originated cEV fractions. The diagram represents the categorization of immune cells belonging to each hemopoietic constituent and indicates the relative proportion of each subset in the total immune fraction. (B) Relative distribution of tissue/solid organ-derived EV fractions from the internal PDAC exLR-seq cohort. n.s., not significant. (C) Integral comparison of hemopoietic- and tissue-derived cEV absolute fractions from healthy control and PDAC plasma exLR-seq samples. The mean and upper and lower standard deviation values are indicated by bars in each panel; the pink and yellow boxes show the absolute scores of hemopoietic- and tissue-derived EV fractions for healthy control and PDAC exLR-seq samples, respectively. (D) Overview of the results of tissue-cellular originated cEVs in the PDAC and healthy control exLR-seq samples. The y axis shows the mean fold change of all absolute fractions for PDAC versus healthy individuals.

immunogenic cell fractions were both significantly upregulated in C-III compared with C-I and C-II (Mann-Whitney U test, p < 0.05; Figure 4B).

In addition, we dichotomized the absolute enrichment score of each fraction and identified six innate immune subpopulations, including five risk components (monocytes, basophils, DCs, NK cells, and neutrophils; hazard ratio [HR] > 1) and one potential protective factor (γδ T cells; HR < 1), that were significantly related to OS in PDAC (p < 0.05; Figure 4C, left). In addition, 8 of 12 adaptive immune populations were significantly correlated with OS in PDAC, including five unfavorable (naive Th cells, Treg cells, plasmablasts, effector memory

CD8 T cells, and terminal effector CD8 T cells) and three protective cell subsets (memory B, naive CD8 T, and Th2 cells) (Figure 4C, right). Univariate regression analysis indicated that 4 of 7 innate cell fractions and 9 of 12 adaptive-lymphocyte-derived components were likely to be associated with exLR-based PDAC phenotypes (p < 0.05; Figure 4D). We then used a stepwise variable selection procedure from Cox's model to identify the most prognostic immunogenic fractions as variables, including Th2 cells, monocytes, Treg cells, neutrophils, and effector memory CD8 T cells. By taking the intersection of the variables and the significant cell sets with regard to OS (Figure 4C) and subtypes (Figure 4D), we identified three hub immunocyte-originated cEV fractions, including two adaptive immune
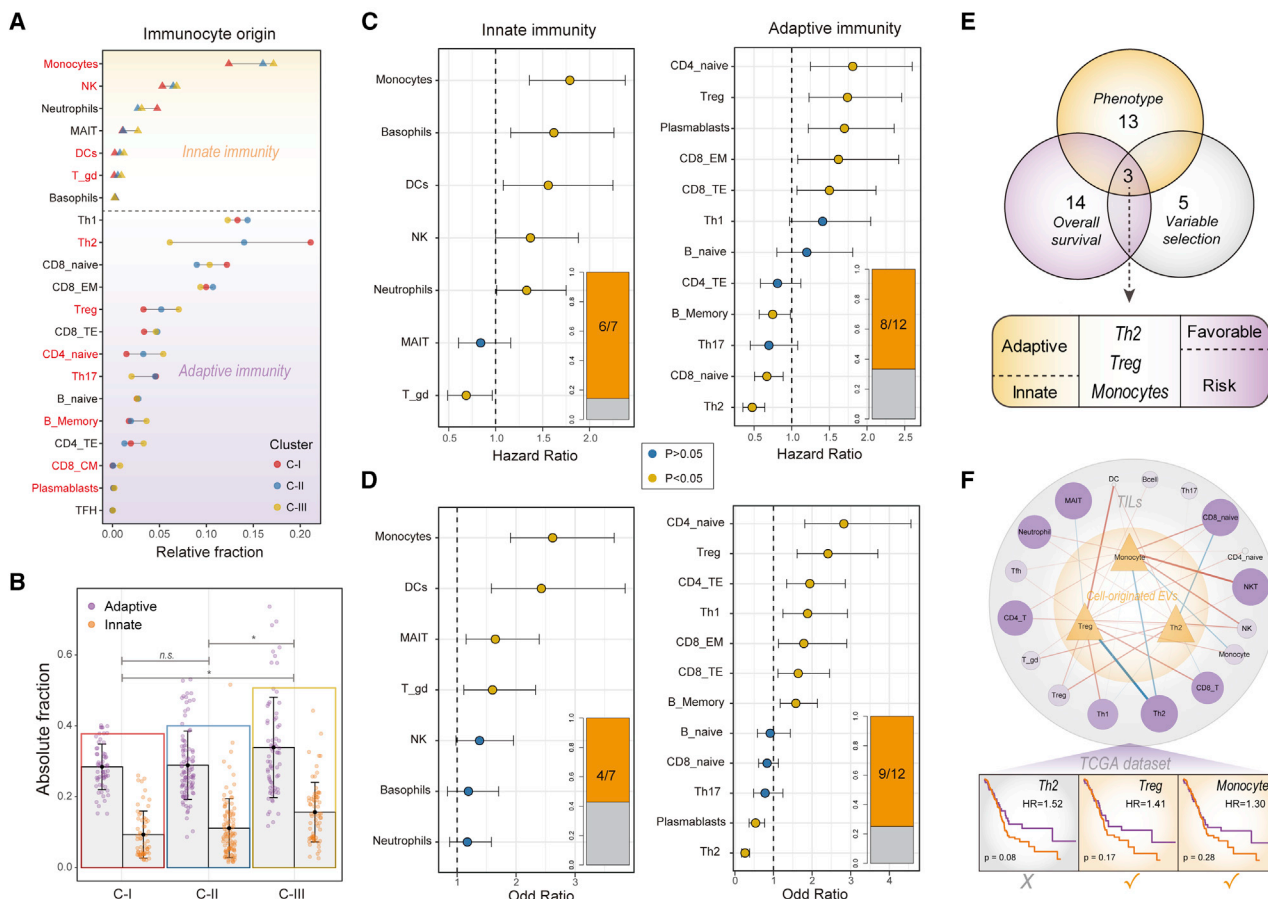
Figure 4. Immunocyte-derived EVs perturbed in PDAC subtypes and interconnect with TILs in tissue

(A) The graph indicates the relative enrichment results of EV fractions from 21 types of immunocytes in the three PDAC subtypes from the discovery exLR-seq dataset. The graph shows 7 types of innate immunity components on the top and 14 adaptive immune subsets on the bottom. The significantly different immune constituents are highlighted in red color. (B) Comparison of the absolute fractions of cEVs from innate and adaptive immunocytes in the three exLR-based subtypes. The points and error bars indicate the mean value and standard deviation of the absolute score, respectively. *p < 0.05; n.s., not significant. (C) OS in high- and low-enriched groups for absolute immunocyte fractions by calculating the hazard ratio (HR) through Cox regression analysis. High and low expression was defined by the optimal cutoff of the quantitative score. The dots represent the HR of univariate Cox test and the error bars show the 95% confidence intervals. (D) The panel displays the correlation between exLR-based PDAC classifications and immune cell-originated EV fractions. The dots show the odds ratios of univariate logistic regression for immune cells in innate and adaptive categories. The error bars show 95% confidence levels of odds ratios. (D) Venn diagram illustrating the number of immunocyte-originated EV constituents that are significantly perturbed in PDAC subtypes and may function as prognostic indicators for PDAC. The bottom panel lists the three hub immune cell components along with their cell origin and prognostic impact. (F) Co-enrichment network of tissue/cEV immune-relevant signatures. The three candidate immunocyte-originated EV constituents are shown in the inside track (yellow triangles). The outside panel refers to the TILs in paired PDAC tissue. Each line represents the correlation coefficient of the immune cell signature between tissue and circulating vesicles among the same PDAC cases. The bottom panel shows the prognostic value (high level, orange line; low level, purple line) of the three hub TIL signatures in TCGA cohort.

components (Treg and Th2 cells) and one innate immune fraction (monocytes) that displayed critical roles for PDAC clinical outcome, and these components showed significant differences among PDAC subsets (Figure 4E).

The enrichment results of immunogenic cEVs raises questions regarding whether EVs from blood could represent the abundance of TILs in PDAC tissue. By deciphering the detailed portraits of immunocyte disaggregation from transcriptomic profiles of cEVs and paired tissues, we generated a full-scale correlation network

between TILs and hub cEV fractions (Figure 4F). The Treg cell-originated EV component positively correlated with DC and CD8[+] T cell infiltration, whereas it was inversely associated with Th2 cell abundance in paired tissue. There was a parallel enrichment pattern between the congenital immunity cell-derived vesicle components (monocytes) and the infiltration of NKT cells in tissues.

We externally validated the prognostic significance of the three hub constituents on PDAC patient survival using The Cancer Genome
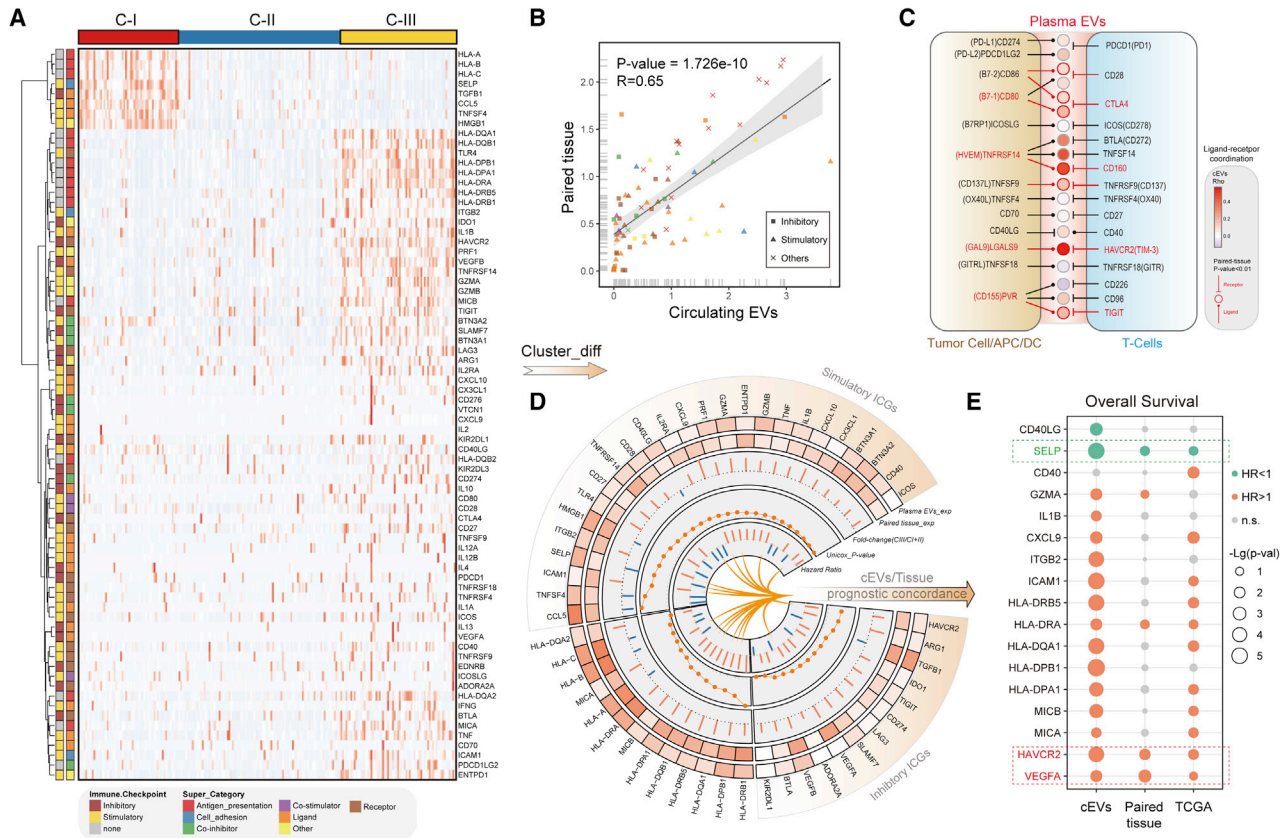
Figure 5. Concordance of gene expression of immunoregulatory targets in cEVs and neoplastic tissues

(A) The heatmap of the expressional heterogeneity of immunomodulator (IM) genes in exLR-seq data from the internal PDAC cohort, scaled by row. Each row refers to an immune-related gene and columns represent exLR-seq samples. Detailed classification and underlying functions of these immunoregulatory targets are listed in Table S12. (B) Scatterplot displaying the relationship of IM gene expression values between cEVs and paired tissue in 61 PDAC patients. (C) Correlation of ligand-receptor immune checkpoint genes in cEVs and corresponding paired tissues. (D) Circular chart represents the differently expressed IMs among the three PDAC subtypes. From the outside to the inside, the diagram displays the average gene expression level of IMs in cEVs and matched tumor tissue, the mean fold-change enrichment of IMs among PDAC subtypes, the p value (−log transformed), and the hazard ratio from univariate Cox analysis. The innermost arc-shaped curve highlights the prognostically consistent targets for tissue and plasma vesicles. (E) Bubble plot indicates the prognostic concordance of 17 IM genes in cEVs, paired tumor tissue, and TCGA expression profiles. Three of the IMs (dotted box marked) predicted PDAC prognosis (p < 0.05) in both blood and tissue datasets.

Atlas (TCGA) dataset. The enrichment of immunogenic signatures from Treg and monocyte cells exhibited equivalent prognostic value in tissue and cEVs, which is consistent with their immunosuppressive properties, whereas the Th2 cell-derived fraction showed opposite results (Figure 4F, bottom). The two candidate immune cell (Treg cells and monocytes)-originated cEV components could jointly serve as unprotective factors to forecast the clinical outcome for PDAC. Taken together, these data suggested perturbed immunogenic cEVs that were associated with immunosuppressive TILs in the TME.

Concordance of immunoregulatory targets in cEVs and cancer tissues

Immunomodulators (IMs) are critical regulators for cancer chemoradiotherapy and immunotherapy, and numerous IM agonists and antagonists have been explored and applied in clinical oncology.[17,18] We collected a list of 76 IMs from the literature that exhibit multiple

effects on the tumor immunoreaction, with functions in antigen presentation, cell adhesion, and immune checkpoint responses[19] (Table S12). The exLR level of IMs perturbed and unsupervised clustering results demonstrated that the IM profiles in cEVs could distinctly segregate patients into three groups, which precisely matched the exLR subtypes aforementioned in this study (Figure 5A). Eight immunogenic exLRs that are upregulated in C-I show antitumor effects such as major histocompatibility complex (MHC) class I antigen presentation and stimulatory immune checkpoint responses. C-III was over-enriched with both antitumor and immunosuppressive IMs, with an expression pattern that was completely different from that of C-I. In addition, we compared candidate immune-related targets in paired cEV and tumor transcriptomes from 61 patients and observed close correlations of expression between the two IM profiles (Rho = 0.65; Figures 5B and 5D, top two tracks).

Considering that the ligand-receptor pairs of immune checkpoint genes (ICGs) are key factors that reflect the magnitude and duration of the immune checkpoint response,[20] we evaluated the expression correlation of 19 ICG pairs in cEVs and matched cancer tissue. All 19 pairs were detected in tumors and cEVs (Table S13), and 7 of 19 ICG pairs showed coincident expression patterns (p < 0.01; Figure 5C). The expression levels of nine ligands displayed a significantly positive correlation with their receptors in cEV profiles (p < 0.05; Figure 5C). Overall, four ligand-receptor ICG pairs were activated both in cEVs and tissue, including three inhibitory pairs (LGALS9-HAVCR2, CD80-CTLA4, and PVR-TIGIT) and one stimulatory pair (TNFRSF14-CD160) (Table S14). No significant expression correlation was observed between CD274 (which encodes PD-L1) and PDCD1LG2 (PD-L2) and their co-receptor PDCD1 (PD1) in cEVs (Figure 5C). The inhibitory immune checkpoint receptor HAVCR2 (TIM-3) and its ligand LGALS9 (GAL9) represented the most significant co-expressed targets activated both in cancer tissue and cEVs (Figures 5C, S3A, and S3B).

We next investigated whether these immunogenic targets can be noninvasively used as biomarkers to track patient clinical outcomes. We identified 52 IM targets that were differentially expressed among the three PDAC clusters (Kruskal-Wallis rank-sum test, p < 0.05; Figure 4A) and we correlated their expression level with OS in PDAC using HRs from univariate Cox regression model. Higher levels of 34 (65%) IMs indicated an increased risk of death in PDAC (HR > 1), whereas 18 (35%) IMs showed the opposite prognostic trend (HR < 1; Figure 5D, last track). Notably, HAVCR2 (TIM-3) was significantly upregulated in C-III and represented the best survival predictor (p < 0.05; Figures 5D and S3C). Figure 5E summarizes the risk assessment power of all IMs in PDCA from both internal and independent cohorts. Seventeen IMs showed similar prognostic patterns between cEVs and cancer tissue. Among the 17 IMs, two immunosuppressive targets, TIM-3 and VEGFA, were significantly correlated with poor prognosis, while one stimulatory target, SELP, was correlated with good prognosis (Figures 5D, S3D, and S3E).

Construction of a tissue/cEV prognostic signature for tracking PDAC clinical outcome

We next tried to simplify the above stratification strategy. The sample sizes of our included internal and external cohorts enable adequate power (1-β, more than 0.9) to predict patient survival and prognosis (see Materials and methods). The workflow used to identify a tissue/cEV prognostic signature for PDAC is shown in Figure 6A. We first eliminated exLRs with an expression frequency less than 25% and evaluated the differentially expressed genes among the three PDAC subtypes from exLR-seq profiles in the internal cohort (Kruskal-Wallis rank-sum test, FDR < 0.05; Table S15). We further harvested markers with prognostic concordance in blood and lesions; the univariate Cox approach was used to investigate the blood-tissue prognostic valuable markers in patients with cEV and paired-tissue transcriptomic profiles (p < 0.05, see Materials and methods). We obtained 986 exLRs, most of which (n = 802) were associated with poor

prognosis of PDAC (HR > 1); a small number of candidate exLRs (n = 184) were indicated as protective prognostic factors (HR < 1) (Table S16). GO enrichment analysis revealed that the exLRs were involved in regulation of lymphocyte activation and differentiation, innate/adaptive immune responses, RNA transport, and activation of cancer-related cell signaling (Table S17).

We further conducted a variable constriction in the internal exLR-seq dataset; the random-forest algorithm and the LASSO-Cox method were used to reduce the dimensionality. Nine exLR targets (HAVCR2, ADK, TP53I11, TBL3, ACD, LGALS9, CANT1, LYL1, and PKIG) were used to construct a model for predicting PDAC prognosis. The immunoregulatory factors HAVCR2 and LGALS9, which were identified as receptor-ligand ICGs, were also selected as candidate prognostic factors (Figure 6B). Using the multivariate Cox algorithm, we established a prognostic model and generated an aggregate circulating exLR-based prognostic score for PDAC (cp-score; Figure 6B, left). C-III exhibited a high median cp-score compared with the other exLR-based clusters (C-I, −0.39; C-II, −0.05; CIII, 0.40; Kruskal-Wallis test, p < 0.001). Kaplan-Meier curves were constructed for PDAC patients separated into high- and low-risk subgroups relative to the optimal cutoff for cp-score (see Materials and methods). The median OS in the low-risk group was significantly higher than that of the high-risk group in the internal and external independent exLR-seq cohorts (p < 0.001, internal cohort, n = 227; p = 0.013, external independent cohort, n = 118; log-rank test; Figures 6D and 6E). Multivariate Cox regression analysis showed that the cp-score was highly associated with the risk of death and was an independent prognostic factor of survival in all exLR-seq datasets (p < 0.001; Figure 6F).

We next explored whether the cp-score can be used for evaluation of prognosis using resected tissue. Using the same markers, we reconstructed a prognostic model and generated a tissue-specific p-score (tp-score) from internal paired tissue profiles. The nine candidates showed equivalent power for risk estimation in tissue consistent with that in cEVs (Figure 6B, right). Patients with a high tp-score showed a shortened survival time (p < 0.001, HR = 2.9; 95% confidence interval [CI], 1.66–5.09; Figure 6G). The predictive ability of the tp-score was externally validated in tissue transcriptomic profiles from three independent datasets (TCGA-PDAC, p < 0.01, HR = 2.08; ICGC-PDAC-AU, p < 0.05, HR = 2.99; ICGC-PDAC-CA, p < 0.05, HR = 1.64; Figures 6H, S4A, and S4B).

In addition, we evaluated whether the p-score could predict PDAC therapeutic effect by assigning the patients into high or low p-score groups both in cEVs and tissue datasets. The cp-score predicted tumor recurrence for resectable PDAC (stage I/II) cases after surgery (TTR, HR, 1.71; 95% CI, 1.02–2.85; p < 0.05; Figure S4C). Consistent with these findings, a low tp-score in patients showed significant association with prolonged TTR both in the internal (HR, 2.69; 95% CI, 1.49–4.83; p < 0.001; Figure S4D) and independent tissue cohort (HR, 1.94; 95% CI, 1.17–3.22; p < 0.01; TCGA dataset; Figure 6I). Collectively, our results suggested that the cEV-derived prognostic signature
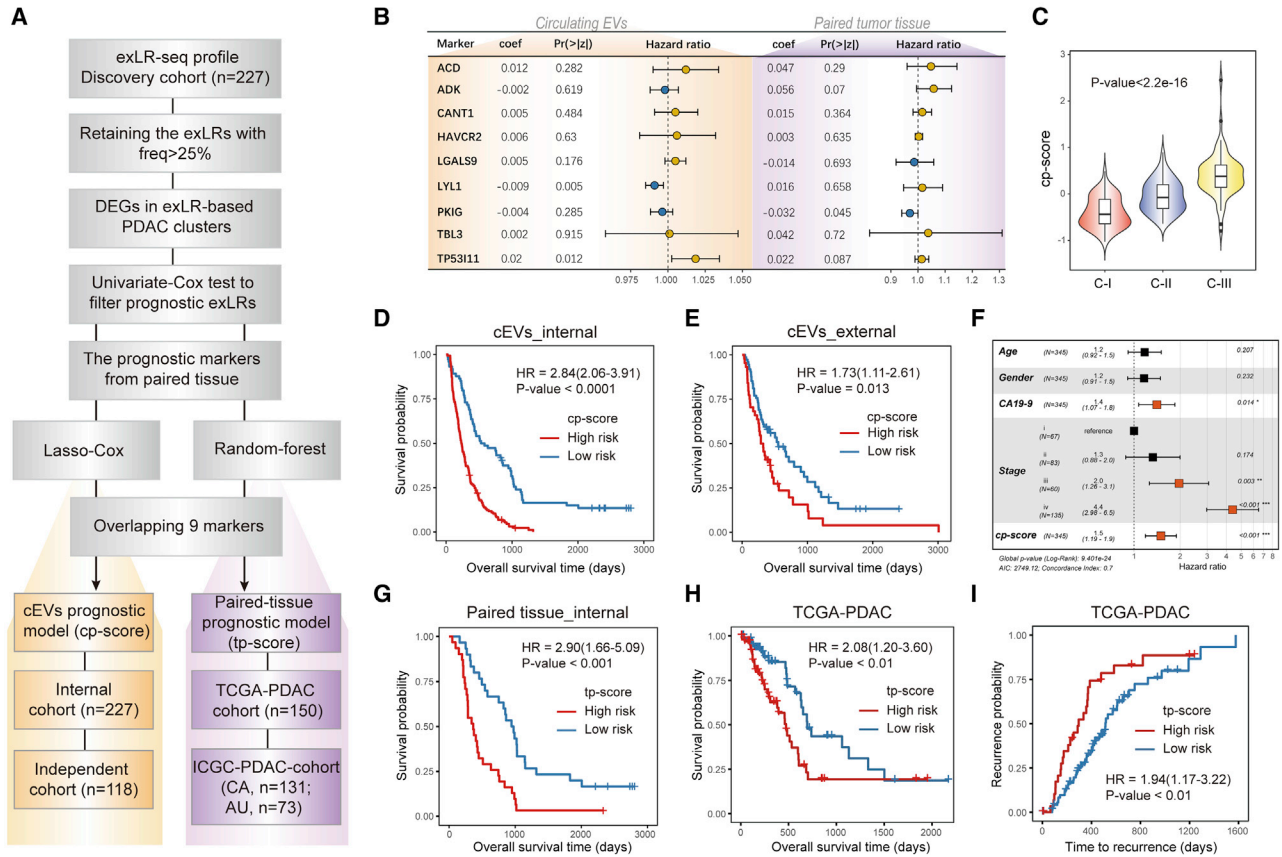
Figure 6. Construction of a tissue/cEV prognostic signature for tracking PDAC clinical outcome

(A) Workflow of prognostic marker selection and prognostic model construction. DEG, differentially expressed gene. (B) Characteristics of the nine markers with their coefficient and hazard ratios of a multivariate Cox model for PDAC prognosis prediction. (C) Comparison of cp-scores among the exLR-based PDAC subtypes. (D and E) OS curves of PDAC patients were stratified according to high- and low-cp-score in the discovery (D) and independent datasets (E). (F) Multivariate Cox analysis of clinico-pathologic factors and cp-score with OS in the entire cohort. Variables with independently prognostic significance are highlighted in red. (G) Analysis of the tissue prognostic score (tp-score) using the paired tissue of the internal cohort; the number of patients in the high-tp-score and low-tp-score groups are n = 31 and n = 30, respectively. (H and I) Kaplan-Meier curve shows the OS (H) and time to recurrence (I) in patients according to tp-score in the TCGA-PDAC cohort. All of the high and low classifications were determined for each patient based on the dichotomized p-score.

harbor equivalent as that of tumor tissue derived for predicting PDAC clinical outcome, including for predicting risk of recurrence after surgery.

DISCUSSION

PDAC is characterized by a high degree of intra-tumoral heterogeneity and an immunosuppressive TME, which represent the main obstacles to accurate stratification and treatment efficiency.[21,22] The difficulty in obtaining tumor specimens from most PDAC patients as well as the molecular heterogeneity of PDAC tumors make liquid biopsy-based biomarker assays crucially important to individualize management.[5,13,23] In this study, we deciphered the cEV transcriptomic information using exLR-seq profiles from 345 PDAC plasma samples, representing the first study of noninvasive prognostic subtyping of PDAC based on large-scale plasma EV transcriptome analysis. We comprehensively characterized the heterogeneity of tissue-cellular origins and the immunomodulatory landscape from cEV

profiles, identified novel therapeutic targets, and constructed a nine exLR-based prognostic signature for PDAC.

We identified three PDAC subtypes based on plasma exLR profiles. Each subtype has its own molecular features and is related to patient prognosis. C-III patients had the worst prognosis and exhibited high expression of multiple cancer/immune-related exLRs in plasma. In addition, many signaling pathways shared similar concordance both in cEVs and cancer tissue, which includes PDAC- and immune-related hallmark signatures. Wnt/β-catenin signaling plays a critical role in pancreatic carcinogenesis and therapeutic resistance and regulates cell cycle progression, apoptosis, epithelial-mesenchymal transition, angiogenesis, stemness, and the tumor immune microenvironment.[24] TGF-β signaling pathways play a central role in the immunosuppressive response to cancer.[25] PDAC-derived EVs could transform TGF-β signaling relative signatures into Kupffer cells and are involved in liver pre-metastatic niche formation and

TME re-construction to increase PDAC metastatic burden.[9] The crosstalk between TGF-β and WNT/β-catenin signaling in EVs promotes PDAC carcinogenesis. Thus, targeting these signaling pathways is an actively pursued strategy in PDAC treatment.[26–28]

We previously demonstrated that plasma exLRs derived from pancreatic cancer were significantly over-enriched in PDAC cases compared with benign samples and healthy controls.[13] Based on the enumerating results of our EV-origin approach, the quantification of cEVs from specific tissue sources may reflect organ status during disease progression. In this study, cEVs in PDAC patients had significantly dysregulated tissue/organ sources, such as pancreas and liver, and were more likely severely affected by neoplastic invasion from primary and metastatic lesions, suggesting that the tissue/solid organ tracing strategy of our approach may noninvasively monitor the occurrence and development of PDAC in real time. We also observed that the cytotoxic immune cell components with a tumor-killing effect in PDAC were significantly reduced while immunosuppressive-related components were increased, reflecting neoplastic education to the immune system in response to carcinogenesis.[29–31]

PDAC comprises tumor cells within a microenvironment of extracellular matrix, stroma cells, endothelial cells, and immune cells.[32] The therapeutic limitations of chemotherapy and immunotherapy have been attributed to the PDAC-specific TME.[21] We hypothesize that analysis of the immune cell source contribution of cEVs could partially trace the body's anti-tumor immune response and the dynamic alternation of the TME. A previous study reported that the PDAC-specific TME was predominantly affected by immunosuppressive cells, which may lead to the dysfunction of infiltrated CD8$^+$ T cells.[33] These findings are consistent with our results and show that although cEV components from anti-tumor immune cells (such as terminal effector CD8/CD4$^+$ T cells) were enriched in C-III, the patient prognosis and clinical outcome were still dismal. At the same time, monocyte-derived suppressor components, as well as Treg cells, are partnering factors that jointly contribute to local immunosuppression and tumor immune escape.[34,35] In agreement with this, we found that monocyte- and Treg cell-originated EV fractions were significantly associated with aggressive phenotype and indicated poor patient outcome. In addition, pancreatic neoplasia cells may produce high levels of cytokines to recruit Treg cells, which in turn inhibit the anti-tumor effects of cytotoxic T lymphocytes.[36] Treg cells engage in extended interactions with tumor-infiltrated DCs and restrain their immunogenic function by suppressing the expression of costimulatory ligands necessary for CD8$^+$ T cell activation.[37] Consistent with the results observed in the TME, we found that the abundance of EVs derived from Treg cells was positively correlated with DCs and CD8 T cells infiltrating in tissues, suggesting that these interactions might be relevant to the immunosuppressive function of Treg cells. The immunosuppressive cEV components partially explain the poor prognosis in C-III. Therefore, exploration of these immune cell-derived components can provide a noninvasive and joint approach to dissect TME alterations and provide evidence for PDAC treatment.

Advances in immune checkpoint blockade have brought a renewed appreciation of the value for reversing the immunosuppressive TME,[38] and recent studies have focused on targeting PD-1/PDL1 and CTLA-4 in cancer treatment.[39,40] However, it appears that other key regulators also suppress the host's immune system and function in immunotherapy resistance in cancers, including PDAC. By determining the heterogeneous expression pattern of ICGs within PDAC subtypes, we identified inhibitory immune checkpoint receptors, including TIM-3 and TIGIT, which were upregulated in cEVs and cancer tissue. TIM-3 and TIGIT expressed on the surface of T lymphocytes are bound by their ligands, Galectin-9 (encoded by LGALS9) and CD155 (PVR), to induce T cell exhaustion and immune escape.[41,42] Several anti-TIM-3 agents have been evaluated in clinical trials for solid and hemopoietic malignances.[43] Targeting the TIM-3/LGALS9 and PVR/TIGIT axis may represent a promising approach for improvement of current immune checkpoint therapies.[44,45] We confirmed that this pattern of co-expression activation of ligand-receptor pairs can be detected in cEVs and harbored prognostic significance as in tissue. In addition, these suppressive ICGs were strictly enriched in the aggressive subsets, which may also explain why most patients with advanced PDAC are resistant to immunotherapies.[46]

Our study had several strengths. First, this is the first PDAC subtyping study using cEV long RNA profiling. Second, the EV-origin deconvolution study has provided the first comprehensive landscape of tissue-cellular origins for cEVs in PDAC and suggested that immune cell-derived cEV components could reflect the TME heterogeneity, thus allowing dynamic monitoring of TME remodeling in a real-time, noninvasive fashion. However, this study has some limitations. The subtyping study was based on retrospective cohorts, and therefore most patients did not have enough gene mutation and cancer tissue-based molecular information to annotate exLR-based molecular subtypes. Moreover, the tumor samples matched with plasma EVs included in our study were derived from early resectable PDAC tumor tissue, which may insufficiently explain the characteristic correlation in advanced PDAC.

Conclusions

Our findings demonstrated the potential of circulating exLR for PDAC subtyping. Based on a deconvolution study and paired cancer tissue profiling, we identified a profound correlation of the immune landscape between tumors and cEVs and identified novel PDAC therapeutic targets. Importantly, we constructed a robust nine exLR-based tissue-applicable signature for prognostication assessment of PDAC. Therefore, cEV long RNA-based subtyping and deconvolution enable noninvasive prediction of immunogenic signatures and clinical outcomes for PDAC.

MATERIALS AND METHODS

Study design

This study employed a two-phase design. First, the subtyping, deconvoluting, and biomarker discovery phase was conducted, where we comprehensively analyzed the circulating exLR-seq transcriptomic

data from patients with PDAC in the internal cohort for the identification of exLR-based PDAC phenotypes and clinical significance. We deconvoluted the tissue-cellular origin of cEVs between healthy controls and PDAC samples and investigated the potential clinical efficacy of immunogenic components at tissue and circulatory levels. In the second phase, we identified a clinically translatable signature from cEVs that predicts patient survival and tumor relapse. We performed validation of the discovered biomarkers both in cEVs and tissue profiles from multiple clinical cohorts of patients with PDAC.

Patient cohorts, sample characteristics, and data generation

A total of 426 individuals were enrolled in this study, including 345 patients with pathologically confirmed PDAC and 81 healthy controls receiving routine healthcare. Among the 345 PDAC cases, 227 plasma specimens were collected from Fudan University Shanghai Cancer Center (Shanghai, China) (as an internal cohort) between January 31, 2012 and June 6, 2017, and 118 PDAC cases (as an external independent cohort) were recruited from Shanghai Hospital of the Second Military Medical University (Shanghai, China; n = 100) and Xi'an Jiaotong University Affiliated Medical Center (Xi'an, China; n = 18) between September 11, 2018 and April 2, 2019. Plasma was collected from PDAC patients before surgery for patients with resectable tumors and before chemotherapy for patients with inoperable tumors. The available plasma EV transcriptomic profiles originated from our previously published study[13] and can be downloaded from the Gene Expression Omnibus (GEO) dataset with the access no. GEO: GSE133684. For tissue specimen collection, we obtained 61 formalin-fixed, paraffin-embedded exLR-seq-matched PDAC tumor tissues from the internal cohort. The transcriptomic expression profiles of plasma EVs and matched tissue specimens from 61 PDAC patients from the internal cohort with OS and TTR information were determined. We downloaded tissue transcriptomic profiles from three external PDAC datasets (TCGA-PDAC, n = 150; ICGC-PDAC-CA, n = 131; ICGC-PDAC-AU, n = 73) to further validate the candidate exLR signature at the tissue level. The demographics and clinicopathologic characteristics for all patient cohorts are shown in Table S1.

We isolated plasma EVs by affinity-based binding to a spin column using the exoRNeasy serum/plasma kit (QIAGEN, Hilden, Germany). A SMARTer stranded total RNA-seq kit (Clontech, Palo Alto, CA, USA) was used to prepare the strand-specific RNA-seq libraries. The library quality of the 10 plasma exLR-seq samples per lane was sequenced with adding 20% PhiX to control base balance. The previously published Assembling Splice Junctions Analysis (ASJA) approach was used to conduct quality control, reads mapping, and transcripts quantification.[47] The raw count matrixes were transformed into transcripts per kilobase million (TPM) values with gencodeV29 annotation to make gene expression profiles more comparable between cEVs and paired tissues. The expression of a circRNA gene for each exLR-seq sample was calculated as spliced reads per billion mapping (SRPBM = number of circular reads/total mapped reads [units in billion]). For the external TCGA-PDAC cohort, the RNA sequencing data and updated clinicopathological information were downloaded from UCSCXenaTools package (https://github.com/cran/UCSCXenaTools).

Digital quantification of tissue-cellular-originated cEVs and TILs

We used the EV-origin approach to quantify the abundance of EV fractions from blood and tissue origins from plasma exLR expression data.[16] The EV-origin method was constructed from two representative signature matrices and a core SVM regression algorithm to deconvolute 7 hemopoietic cells and 16 tissue components from exLR-seq transcriptomic profiles. In this study, we expanded the number of traced hemopoietic fractions to 23, including 21 types of immunocytes and 2 hematopoietic stem cell constituents (see Table S11). The new hemopoietic signature matrix was constructed by the normalized expression profiles of isolated blood cells from two datasets.[16,48] We selected cell-specific genes and re-built the comparable blood signature matrix by the same method as EV-origin. The relative proportions and absolute fractions of 23 hemopoietic cells and 16 types of tissue/solid organ-derived cEV constituents were detected in each exLR-seq sample. To determine the abundance of infiltrating immune cells in PDAC tissue, we used a single-sample GSEA (ssGSEA) integrated with TIL gene signatures,[49,50] which allows for reliably discriminating the enrichment of 17 human immune cell phenotypes from tissue expression profiles.

exLR-based PDAC subtyping

We used the ConsensusClusterPlus package to select the optimal cluster number and determined the clustering stability for the internal cohort.[51] To capture stabilized exLR-based subtypes of PDAC, we first removed the exLRs with expression frequency no more than 50% to reduce the influence of low-frequency noises. Then we used a median absolute deviation (MAD) algorithm (a robust measure of sample bias for univariate numerical data) to fully characterized the exLR expressional viability in the internal exLR-seq cohort. We finally harvested 8,922 exLR targets (top 60% targets of MAD quantitative ranking) as input markers for consensus clustering. These exLR-seq samples were grouped using hierarchical agglomerative clustering based on Manhattan distance and the strategy of partition around medoids (PAM). We selected the iteration number as 500 and the subsampling ratio as 0.8 to identify optimal clusters from a consensus similarity matrix generated by consensus clustering. Three completely segregated subtypes from the exLR-seq data in the internal cohort were obtained. These discovered clusters were used as representative subpopulations for further analysis.

Building a predictive model for PDAC clinical outcome

On the basis of the exLR-seq analysis, we developed a combined circulating prognostic score (cp-score) for noninvasively estimating PDAC prognosis. We used the 227 cases in the internal cohort for discovery and 118 cases in the external independent cohort for validation. In the exLR-seq data from the internal cohort, we removed exLRs with a frequency less than 25% to obtain reliably expressed variables. A Kruskal-Wallis test was used to assess the differential expression of exLRs among the three clusters, and the p value of each marker was adjusted by the Benjamini-Hochberg method to control the

FDR.[52] We retained exLRs with an FDR less than 0.05 and carried out a consecutive model-based variable selection approach to identify prognostic valuable markers for predicting survival outcome.[53] We first performed a univariate pre-screening procedure to remove excessive noise and obtain cancer-related targets to facilitate the subsequent computational analysis. We used whole-transcriptome markers as covariates, and the univariate Cox proportional hazards model was fitted to explore prognostic targets both in cEVs and paired tumor tissue profiles. Only markers with a p value <0.05 from the Wald statistic were retained. Overall, we obtained 986 exLRs that concurrently showed prognostic concordance in blood and tissue, in which the expression level indicates the same prognostic risk both in patient cEVs and tissues. We then used LASSO-Cox and the random forest algorithm to further shrink the marker number to a reasonable range.[54,55] For the LASSO-Cox constraint, 80% of exLR-seq samples from the training cohort were randomly subsampled to conduct LASSO regularization in 1,000 repetitions. We used 10-fold cross-validation and the Akaike information criterion to select an optimal value of "1-se" lambda parameter to construct an adaptive prognostic model for marker selection. The Boruta algorithm (a wrapper built around the random forest classification algorithm) was used to concurrently conduct dimension reduction and screen prognostically significant exLRs. This process generated nine final markers for a prognostic signature in the internal and validation cohorts. By fitting a multivariable Cox proportional hazards model on these nine targets from the internal EVs and tissue dataset, we determined the coefficients of each marker and generated two combined prognostic scores (p-score, designated as the cp-score for blood EVs and the tp-score for PDAC tumor tissue) for each individual. We then validated the significance of the prognostic scores in the independent datasets for the assessment of clinical outcome of PDAC.

Functional and pathway enrichment analysis

Weighted gene co-expression network analysis (WGCNA) was performed to construct co-expressed networks and identify co-expression of exLR modules in the exLR-seq data from the internal cohort.[56] We used median absolute deviation to estimate the degree of variation for exLR expression and applied WGCNA based on the exLR-seq expression data constructed by the top 5,000 aberrantly expressed exLRs. We first created a matrix of adjacencies by conducting the Pearson correlation analysis between gene pairs to determine the concordances of gene expression; this matrix was then transformed into a topological overlap matrix. The resulting topological overlap is a biologically meaningful measure of gene similarity based on co-expression relationships for the comparison of each gene pair. To identify the relatively balanced scale independence and mean connectivity of the WGCNA, we carefully selected the scale-free topology fit index to find an optimal soft-thresholding power. As shown in Figure S2A, the power value 16 was used to produce a hierarchical clustering dendrogram of 5,000 variable exLR genes. The module eigengenes were determined as the primary key components computed by principal-component analysis that represents the manifestation of exLRs of a functional module into a unitary characteristic quantitative value. The manifestation configurations of exLR modules related to the

traits of internal individuals were quantified by gene significance (GS) and module significance (MS). The GS measure was defined as the results of the Spearman correlation analysis among the ith gene profile xi and the sample trait T, and MS was defined as the average GS for all genes in the module:

$$GSi \ = \ |cor(xi, \ T)|.$$

The exLR genes were separated into eight co-expression modules according to their degree of connectivity. Each subunit was autonomously authenticated with another, which revealed a high level of individuality among the modules and the comparative independence of gene manifestation in each module. The exLRs that could not be contained in any modules were placed into module 8, which was removed in the subsequent GO analysis. All modules revealed a high correlation with PDAC phenotypes (C-I and C-III) and stage (Figure 2A). We represented seven modules (M1–M7) as the most significant modules to reflect cEV inner transcriptomic heterogeneity among PDAC clusters. Gene annotation enrichment results were estimated from the clusterProfiler package[57] for each module-specific exLR gene. We used a strict cutoff of FDR <0.05 to identify GO terms that were significantly enriched in each module (Tables S2–S8). We identified the aberrantly enriched signaling pathways among PDAC subtypes by running a pre-ranked GSEA of the expression data for exLR genes. Gene sets were downloaded from the MSigDB platform of the Broad Institute that includes current annotations of Hallmark, Reactome, and KEGG datasets (Molecular Signatures Database v7.2).[58,59] The p value calculations were based on 5,000 permutations of GSEA, and the multiple testing was subsequently adjusted by Benjamini-Hochberg method to regulate the FDR.[52]

Statistical analysis

We used the Shapiro-Wilk normality test to examine the normality of quantitative variables. For the comparison of variables in two groups, the statistical significance for normally distributed variates was evaluated by a two-side Student's t tests, and the group with non-normal distributions was tested by a Mann-Whitney U test (Wilcoxon rank-sum test). For the comparison of unreplicated data in more than two groups, the Kruskal-Wallis test and Friedman rank-sum test were implemented as non-parametric statistical approaches. We applied the Benjamini-Hochberg strategy to adjust the p value of multiple DEG analysis to control the FDR.[52] The logistic regression was used to determine the statistical significance of the correlation between qualitative variables and PDAC subtypes. The optimal cutoff values of each dataset were calculated based on the correlation between patient survival data and continuous variables in each separate dataset using the survminer package. The Kaplan-Meier approach was used to construct cumulative survival curves for subgroups of each cohort, and the log-rank test was implemented to determine the statistical difference between groups. We used the univariate Cox proportional hazards regression model to analyze HRs for univariate prognostic analysis. We hypothesized that the power of prognostic evaluation could be more than 0.9 with the introduction of exLRs, EV-derived components, and p-signature. With a two-side alpha (significance level) at 0.05, an overall event rate set at 0.93 for the internal cohort

and 0.72 for the independent cohort, and an HR at 1.5, the included sample size was sufficient in the internal and validation cohorts to ensure a power of more than 95% for predicting PDAC prognosis (according to the PASS 11.0 software of power $[1-\beta]$ calculation for the Cox proportional hazards regression analysis with covariates).[60] The prognostic independence of the p-score was determined by the multivariate Cox regression model. All statistical analyses were conducted by R program (v3.6.1, https://www.r-project.org/). The p values were two-sided, and a p value less than 0.05 was considered statistically significant.

Ethical approval and consent to participate

This study was approved by the Ethics Committee of the Fudan University Shanghai Cancer Center, Shanghai, China (approval no. 050432-4-1212B), and written informed consent was obtained from each participant in accordance with institutional guidelines.

Consent for publication

All of the authors have read and approved the final manuscript for publication.

Availability of supporting data

The transcriptomic data have been deposited at the GEO. The raw sequencing reads and integrated count matrix can be fully accessed under GEO: GSE133684 and GSE181273.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2021.08.017.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Conception and design, P.W. and S.H.; provision of study materials or patients, Yuchen Li, Ye Li, S.Y., L.Q., K.C., Y.Z., S.G., and Z.M.; collection and assembly of data, Yuchen Li, Ye Li, S.Y., L.Q., K.C., H.L., Yan Li, Y.Z., S.G., Z.M., and H.Z.; data analysis and interpretation, P.W., S.H., Yuchen Li, S.Y., and Ye Li; manuscript writing, Yuchen Li, P.W., and S.H.; final approval of manuscript, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Siegel, R.L., Miller, K.D., and Jemal, A. (2018). Cancer statistics, 2018. CA Cancer J. Clin. 68, 7–30.

2. Mahajan, U.M., Langhoff, E., Goni, E., Costello, E., Greenhalf, W., Halloran, C., Ormanns, S., Kruger, S., Boeck, S., Ribback, S., et al. (2018). Immune cell and stromal signature associated with progression-free survival of patients with resected pancreatic ductal adenocarcinoma. Gastroenterology 155, 1625–1639.e2.

3. Aguirre, A.J., Nowak, J.A., Camarda, N.D., Moffitt, R.A., Ghazani, A.A., Hazar-Rethinam, M., Raghavan, S., Kim, J., Brais, L.K., Ragon, D., et al. (2018). Real-time genomic characterization of advanced pancreatic cancer to enable precision medicine. Cancer Discov. 8, 1096–1111.

4. Puleo, F., Nicolle, R., Blum, Y., Cros, J., Marisa, L., Demetter, P., Quertinmont, E., Svrcek, M., Elarouci, N., Iovanna, J., et al. (2018). Stratification of pancreatic ductal adenocarcinomas based on tumor and microenvironment features. Gastroenterology 155, 1999–2013, e3.

5. Mellby, L.D., Nyberg, A.P., Johansen, J.S., Wingren, C., Nordestgaard, B.G., Bojesen, S.E., Mitchell, B.L., Sheppard, B.C., Sears, R.C., and Borrebaeck, C.A.K. (2018). Serum biomarker signature-based liquid biopsy for diagnosis of early-stage pancreatic cancer. J. Clin. Oncol. 36, 2887–2894.

6. Whitham, M., Parker, B.L., Friedrichsen, M., Hingst, J.R., Hjorth, M., Hughes, W.E., Egan, C.L., Cron, L., Watt, K.I., Kuchel, R.P., et al. (2018). Extracellular vesicles provide a means for tissue crosstalk during exercise. Cell Metab. 27, 237–251.e4.

7. Möller, A., and Lobb, R.J. (2020). The evolving translational potential of small extracellular vesicles in cancer. Nat. Rev. Cancer 20, 697–709.

8. Murillo, O.D., Thistlethwaite, W., Rozowsky, J., Subramanian, S.L., Lucero, R., Shah, N., Jackson, A.R., Srinivasan, S., Chung, A., Laurent, C.D., et al. (2019). exRNA atlas analysis reveals distinct extracellular RNA cargo types and their carriers present across human biofluids. Cell 177, 463–477.e15.

9. Costa-Silva, B., Aiello, N.M., Ocean, A.J., Singh, S., Zhang, H., Thakur, B.K., Becker, A., Hoshino, A., Mark, M.T., Molina, H., et al. (2015). Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat. Cell Biol. 17, 816–826.

10. Kimura, H., Yamamoto, H., Harada, T., Fumoto, K., Osugi, Y., Sada, R., Maehara, N., Hikita, H., Mori, S., Eguchi, H., et al. (2019). CKAP4, a DKK1 receptor, is a biomarker in exosomes derived from pancreatic cancer and a molecular target for therapy. Clin. Cancer Res 25, 1936–1947.

11. Capello, M., Vykoukal, J.V., Katayama, H., Bantis, L.E., Wang, H., Kundnani, D.L., Aguilar-Bonavides, C., Aguilar, M., Tripathi, S.C., Dhillon, D.S., et al. (2019). Exosomes harbor B cell targets in pancreatic adenocarcinoma and exert decoy function against complement-mediated cytotoxicity. Nat. Commun. 10, 254.

12. Li, Z., Yanfang, W., Li, J., Jiang, P., Peng, T., Chen, K., Zhao, X., Zhang, Y., Zhen, P., Zhu, J., and Li, X. (2018). Tumor-released exosomal circular RNA PDE8A promotes invasive growth via the miR-338/MACC1/MET pathway in pancreatic cancer. Cancer Lett. 432, 237–250.

13. Yu, S., Li, Y., Liao, Z., Wang, Z., Wang, Z., Li, Y., Qian, L., Zhao, J., Zong, H., Kang, B., et al. (2020). Plasma extracellular vesicle long RNA profiling identifies a diagnostic signature for the detection of pancreatic ductal adenocarcinoma. Gut 69, 540–550.

14. Li, S., Li, Y., Chen, B., Zhao, J., Yu, S., Tang, Y., Zheng, Q., Li, Y., Wang, P., He, X., and Huang, S. (2018). exoRBase: A database of circRNA, lncRNA and mRNA in human blood exosomes. Nucleic Acids Res. 46 (D1), D106–D112.

15. Li, Y., Zhao, J., Yu, S., Wang, Z., He, X., Su, Y., Guo, T., Sheng, H., Chen, J., Zheng, Q., et al. (2019). Extracellular vesicles long RNA sequencing reveals abundant mRNA, circRNA, and lncRNA in human blood as potential biomarkers for cancer diagnosis. Clin. Chem. 65, 798–808.

16. Li, Y., He, X., Li, Q., Lai, H., Zhang, H., Hu, Z., Li, Y., and Huang, S. (2020). EV-origin: Enumerating the tissue-cellular origin of circulating extracellular vesicles using exLR profile. Comput. Struct. Biotechnol. J. 18, 2851–2859.

17. Hengge, U.R., Benninghoff, B., Ruzicka, T., and Goos, M. (2001). Topical immunomodulators—Progress towards treating inflammation, infection, and cancer. Lancet Infect. Dis. 1, 189–198.

18. Thotathil, Z., and Jameson, M.B. (2007). Early experience with novel immunomodulators for cancer treatment. Expert Opin. Investig. Drugs 16, 1391–1403.

19. Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. Immunity 48, 812–830, e14.

20. Hu, F.F., Liu, C.J., Liu, L.L., Zhang, Q., and Guo, A.Y. (2021). Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. Brief. Bioinform. 22, bbaa176.

21. Ho, W.J., Jaffee, E.M., and Zheng, L. (2020). The tumour microenvironment in pancreatic cancer—Clinical challenges and opportunities. Nat. Rev. Clin. Oncol. *17*, 527–540.

22. Peng, J., Sun, B.F., Chen, C.Y., Zhou, J.Y., Chen, Y.S., Chen, H., Liu, L., Huang, D., Jiang, J., Cui, G.S., et al. (2019). Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res. *29*, 725–738.

23. Bernard, V., Kim, D.U., San Lucas, F.A., Castillo, J., Allenson, K., Mulu, F.C., Stephens, B.M., Huang, J., Semaan, A., Guerrero, P.A., et al. (2019). Circulating nucleic acids are associated with outcomes of patients with pancreatic cancer. Gastroenterology *156*, 108–118.e4.

24. Zhan, T., Rindtorff, N., and Boutros, M. (2017). Wnt signaling in cancer. Oncogene *36*, 1461–1473.

25. David, C.J., Huang, Y.H., Chen, M., Su, J., Zou, Y., Bardeesy, N., Iacobuzio-Donahue, C.A., and Massagué, J. (2016). TGF-β tumor suppression through a lethal EMT. Cell *164*, 1015–1030.

26. Alvarez, M.A., Freitas, J.P., Mazher Hussain, S., and Glazer, E.S. (2019). TGF-β inhibitors in metastatic pancreatic ductal adenocarcinoma. J. Gastrointest. Cancer *50*, 207–213.

27. Krishnamurthy, N., and Kurzrock, R. (2018). Targeting the Wnt/beta-catenin pathway in cancer: Update on effectors and inhibitors. Cancer Treat. Rev. *62*, 50–60.

28. Ram Makena, M., Gatla, H., Verlekar, D., Sukhavasi, S., Pandey, M.K., and Pramanik, K.C. (2019). Wnt/-catenin signaling: the culprit in pancreatic carcinogenesis and therapeutic resistance. Int. J. Mol. Sci *20*, 4242.

29. Best, M.G., Sol, N., In 't Veld, S.G.J.G., Vancura, A., Muller, M., Niemeijer, A.N., Fejes, A.V., Tjon Kon Fat, L.A., Huis In 't Veld, A.E., Leurs, C., et al. (2017). Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. Cancer Cell *32*, 238–252.e9.

30. Nolte-'t Hoen, E.N., Buermans, H.P., Waasdorp, M., Stoorvogel, W., Wauben, M.H., and 't Hoen, P.A. (2012). Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. Nucleic Acids Res. *40*, 9272–9285.

31. Campanella, R., Guarnaccia, L., Cordiglieri, C., Trombetta, E., Caroli, M., Carrabba, G., La Verde, N., Rampini, P., Gaudino, C., Costa, A., et al. (2020). Tumor-educated platelets and angiogenesis in glioblastoma: Another brick in the wall for novel prognostic and targetable biomarkers, changing the vision from a localized tumor to a systemic pathology. Cells *9*, 294.

32. Hessmann, E., Buchholz, S.M., Demir, I.E., Singh, S.K., Gress, T.M., Ellenrieder, V., and Neesse, A. (2020). Microenvironmental determinants of pancreatic cancer. Physiol. Rev. *100*, 1707–1751.

33. Siret, C., Collignon, A., Silvy, F., Robert, S., Cheyrol, T., André, P., Rigot, V., Iovanna, J., van de Pavert, S., Lombardo, D., et al. (2020). Deciphering the crosstalk between myeloid-derived suppressor cells and regulatory T cells in pancreatic ductal adenocarcinoma. Front. Immunol. *10*, 3070.

34. Nielsen, S.R., Quaranta, V., Linford, A., Emeagi, P., Rainer, C., Santos, A., Ireland, L., Sakai, T., Sakai, K., Kim, Y.S., et al. (2016). Macrophage-secreted granulin supports pancreatic cancer metastasis by inducing liver fibrosis. Nat. Cell Biol. *18*, 549–560.

35. Wang, X., Lang, M., Zhao, T., Feng, X., Zheng, C., Huang, C., Hao, J., Dong, J., Luo, L., Li, X., et al. (2017). Cancer-FOXP3 directly activated CCL5 to recruit FOXP3⁺Treg cells in pancreatic ductal adenocarcinoma. Oncogene *36*, 3048–3058.

36. Tung, S.L., Boardman, D.A., Sen, M., Letizia, M., Peng, Q., Cianci, N., Dioni, L., Carlin, L.M., Lechler, R., Bollati, V., et al. (2018). Regulatory T cell-derived extracellular vesicles modify dendritic cell function. Sci. Rep. *8*, 6065.

37. Jang, J.E., Hajdu, C.H., Liot, C., Miller, G., Dustin, M.L., and Bar-Sagi, D. (2017). Crosstalk between regulatory T cells and tumor-associated dendritic cells negates anti-tumor immunity in pancreatic cancer. Cell Rep. *20*, 558–571.

38. Kim, H.J., Cantor, H., and Cosmopoulos, K. (2020). Overcoming immune checkpoint blockade resistance via EZH2 inhibition. Trends Immunol. *41*, 948–963.

39. Burrack, A.L., Spartz, E.J., Raynor, J.F., Wang, I., Olson, M., and Stromnes, I.M. (2019). Combination PD-1 and PD-L1 blockade promotes durable neoantigen-specific T cell-mediated immunity in pancreatic ductal adenocarcinoma. Cell Rep. *28*, 2140–2155.e6.

40. Royal, R.E., Levy, C., Turner, K., Mathur, A., Hughes, M., Kammula, U.S., Sherry, R.M., Topalian, S.L., Yang, J.C., Lowy, I., and Rosenberg, S.A. (2010). Phase 2 trial of single agent ipilimumab (anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. J. Immunother. *33*, 828–833.

41. Blake, S.J., Dougall, W.C., Miles, J.J., Teng, M.W., and Smyth, M.J. (2016). Molecular pathways: Targeting CD96 and TIGIT for cancer immunotherapy. Clin. Cancer Res *22*, 5183–5188.

42. Yasinska, I.M., Sakhnevych, S.S., Pavlova, L., Teo Hansen Selnø, A., Teuscher Abeleira, A.M., Benlaouer, O., Gonçalves Silva, I., Mosimann, M., Varani, L., Bardelli, M., et al. (2019). The Tim-3-Galectin-9 pathway and its regulatory mechanisms in human breast cancer. Front. Immunol. *10*, 1594.

43. Zeidan, A.M., Komrokji, R.S., and Brunner, A.M. (2021). TIM-3 pathway dysregulation and targeting in cancer. Expert Rev. Anticancer Ther. *21*, 523–534.

44. Chauvin, J.M., and Zarour, H.M. (2020). TIGIT in cancer immunotherapy. J. Immunother. Cancer *8*, e000957.

45. Du, W., Yang, M., Turner, A., Xu, C., Ferris, R.L., Huang, J., Kane, L.P., and Lu, B. (2017). TIM-3 as a target for cancer immunotherapy and mechanisms of action. Int. J. Mol. Sci. *18*, 645.

46. Fan, J.Q., Wang, M.F., Chen, H.L., Shang, D., Das, J.K., and Song, J. (2020). Current advances and outlooks in immunotherapy for pancreatic ductal adenocarcinoma. Mol. Cancer *19*, 32.

47. Zhao, J., Li, Q., Li, Y., He, X., Zheng, Q., and Huang, S. (2019). ASJA: A program for assembling splice junctions analysis. Comput. Struct. Biotechnol. J. *17*, 1143–1150.

48. Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y.Y., Carré, C., Burdin, N., Visan, L., Ceccarelli, M., Poidinger, M., et al. (2019). RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. Cell Rep *26*, 1627–1640.e7.

49. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: Digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

50. Miao, Y.R., Zhang, Q., Lei, Q., Luo, M., Xie, G.Y., Wang, H., and Guo, A.Y. (2020). ImmuCellAI: A unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. Adv. Sci. (Weinh.) *7*, 1902880.

51. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

52. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B *57*, 289–300.

53. Xu, R.H., Wei, W., Krawczyk, M., Wang, W., Luo, H., Flagg, K., Yi, S., Shi, W., Quan, Q., Li, K., et al. (2017). Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat. Mater. *16*, 1155–1161.

54. Kursa, M.B., and Rudnicki, W.R. (2010). Feature selection with the Boruta package. J. Stat. Softw. *36*, 1–13.

55. Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. Stat. Med. *16*, 385–395.

56. Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

57. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

58. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

59. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

60. Hsieh, F.Y., and Lavori, P.W. (2000). Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. Control. Clin. Trials *21*, 552–560.