

Gene Co-Expression Network Characterizing Microenvironmental Heterogeneity and Intercellular Communication in Pancreatic Ductal Adenocarcinoma: Implications of Prognostic Significance and Therapeutic Target

Chengsi Wu

state key laboratory of oncogenes and related genes, shanghai cancer institute, renji hospital, shanghai jiao tong university school of medicine, shanghai 200032, china

Yizhen Liu

Fudan University Shanghai Cancer Center Department of Medical Oncology

Kun Cai

state key laboratory of oncogenes and related genes, shanghai cancer center, renji hospital, shanghai jiao tong university school of medicine

Li Tao

305 hospital of people's liberation army no.13 department of emergency

Dianhui Wei

state key laboratory of oncogenes and related genes, shanghai cancer center, renji hospital, shanghai jiao tong university school of medicine

Tiantian Jing

state key laboratory of oncogenes and related genes, shanghai cancer center, renji hospital, shanghai jiao tong university school of medicine

Boshi Wang (✉ wbs137@shsci.org)

Shanghai Cancer Institute, Renji Hospital <https://orcid.org/0000-0001-6923-2588>

Research

Keywords: PDAC, Tumor Micro-environment, Cell-cell communication, Integrin

Posted Date: November 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1054804/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Pancreatic ductal adenocarcinoma (PDAC) is characterized by intensive stroma involvement and heterogeneity. Pancreatic cancer cells interplay with surrounding tumor micro-environment (TME), leading to exacerbated tumorigenesis, dismal prognosis and tenacious therapy resistance. Herein, we aim to ascertain a gene-network indicative of vicious features of TME, then find a vulnerability for pancreatic cancer.

Methods: Single cell RNA sequencing data was processed by Seurat package, retrieving the cell component marker genes (CCMGs). Correlation networks/modules of CCMGs were determined by WGCNA algorithm in a combined PDAC mRNA expression dataset. The gene modules that statistically associate with prognosis were chosen for classifying TME subgroups, constructing neural network and designing the risk score system. Cell-cell communication analysis was achieved by NATMI software. The tumor suppressive effect of ITGA2 inhibitor was evaluated *in vivo* by using a *Kras*^{G12D}-driven murine pancreatic cancer model.

Results: WGCNA analysis categorized cell component marker genes into eight co-expression networks. From gene modules with the maximum and minimum hazard ratio, we stratify PDAC samples based on TME gene patterns, resulting in two main TME subclasses with contrasting survival periods. Furthermore, we generated a neural network model and a risk score model which robustly predict prognosis and therapeutic outcomes. The hub genes in both gene modules were also gathered for functional enrichment analysis, elucidating a crucial role of cell communication-mediating integrins in TME associated PDAC malignancy. To perform a confirmatory experiment underpinning the significance of hub gene targeting, the mice with spontaneously developed pancreatic cancer were orally treated with an integrin inhibitor. The *in vivo* assays unraveled that pharmacologically inhibiting ITGA2 counteracts cancer-promoting micro-environment, and ameliorates pancreatic lesions.

Conclusions: By recapitulating gene-network across various cell types, we exploited novel PDAC prognosis-predicting strategies. Medically interfering ITGA2, a key factor guiding cellular reciprocal interaction, attenuated tumor development. These findings may open new avenue about PDAC targeting therapy.

Background

Pancreatic ductal adenocarcinoma (PDAC) is one of the most aggressive and fatal disease, which consists of almost 95% pancreatic malignant cancers. Over the last decade, intense efforts to improve the survival rates have so far failed. Effective treatment for PDAC patients is largely limited by low early-diagnosis rate, high relapse probability, and the therapy-refractory nature of PDAC, making it featured by the lowest 5-year survival rate among cancer types [1, 2]. Statistical projection result predicts that PDAC will reach the second leading cause of cancer-related death at 2030[3]. Facilitated by the advances in

high-throughput technologies, the understanding of molecular landscape may be beneficial for solving PDAC.

Molecular classification of PDAC based on the genome and transcriptome data help to identify clinically relevant gene signatures, actionable genetic variations and/or prognostic biomarkers[4–8]. However, the conventional molecular analyses may be inefficient in fully dissect microenvironment dynamics. One hallmark of PDAC is extensive inclusion of stroma. The high heterogeneity of cell components within PDAC stroma make it difficult to map distinguishable changes into specific microenvironmental components. One way to resolve this challenge is the application of deconvolution algorithms, which let to evaluate the abundance or relative proportion of certain pre-defined cell type[9, 10]. But the information about intracellular diversity and intercellular interactions still lacks.

On the other hand, the emergence of single-cell RNA sequencing technology enables assessing expression profiles at single-cell level. In addition to the cell type-specific gene expression, single-cell transcriptome data also expose how different cell types cooperate and work together. Through collected literature-supported ligand-receptor pairs, computational methods can estimate the degree of inter- or intra- cell type linkage, constructing a complex cellular interaction network [11, 12]. Despite the advantage of high resolution, several single-cell expression profiling assays in PDAC merely enrolled limited samples[13, 14]. The utilization of this method may be restricted by its cost-inefficient property. To compromise this issue, extracting cell-type specific information from bulk RNA-seq data by incorporating single-cell data is an attractive and more practical manner to mine a large body of sequencing data.

To propose a scalable approach finding clinical relevance of cell-type-specific information, we firstly obtained cell component marker genes from single-cell sequencing data. Then, the gene-network modules reflecting correlations of micro-environmental factors in the public available transcriptome datasets were identified. The defined co-expression signatures were eventually used to classify PDAC samples, followed by generating neural network-based or risk score-based outcome predictors. Especially, by using specific cell-cell communication algorithm, an intercellular connection network was determined in single-cell transcriptome, which could also be extended to the bulk gene expression data for clinical analyses. In the end, central molecule mediating prognosis-related cell-cell communication network was tested for the *in vivo* druggable value, emphasizing the tumor-promoting role of a key integrin in PDAC.

Methods

Data acquisition and processing

A single-cell RNA sequencing (scRNA-seq) dataset of pancreatic adenocarcinoma samples was obtained from the Genome Sequence Archive (GSA) database under the accession code CRA001160 [13]. The count matrix was directly downloaded from the website. Low-quality cells were removed according to the results of 'calculateQCMetrics' function in 'Scater' package. The gene expression profiles of PDAC transcriptome assays, GSE28735[15], GSE62452[16] and GSE71729[17] were obtained from the Gene Expression Omnibus (GEO) database using the R package 'GEOquery'. The count matrix of PDAC RNA-

seq dataset GSE79668[18] was downloaded from supplementary file on GEO website, then the counts were transformed to transcripts per kilobase million (TPM) values for further usage. All clinical information from GEO database were acquired by 'GEOquery' package. In Cancer Genome Atlas (TCGA) database, the transcriptome data, genetic copy number variation data, simple nucleotide variation data and the clinical features of pancreatic adenocarcinoma samples were downloaded and integrated through 'TCGAbiolinks' package[19]. From the International Cancer Genome Consortium (ICGC) database[20], we downloaded the RNA-seq data of PACA-AU and PACA-CA, and array-based gene expression profiling data (exp_array) together with clinical information. To construct a training set, we integrated three RNA-seq datasets PACA-AU, PACA-CA and GSE79668 into a combined PDAC dataset using 'combat' R package. The samples lacking prognostic information were excluded from the combined PDAC dataset.

Re-analysis of Single-cell RNA-seq data

The single cell RNA-seq data in CRA001160 dataset was consecutively analyzed by 'Seurat' package[21]. Firstly, the count matrix was converted to $\log_2(\text{TPM}+1)$ values. Then, the top 2000 variable features were selected to perform PCA dimension reduction, followed by dimension reduction through Uniform Manifold Approximation and Projection (UMAP). Finally, the seurat clusters were determined by 'FindNeighbors' and 'FindClusters' functions in the Seurat package. Cellular identity of each cluster was identified by the expression of cell type specific genes: Epithelial cells (EPCAM, KRT19), Pancreatic islet (INS), Pancreatic acinar cells (CPA1), Immune cells (PTPRC/CD45), B cells (MS4A1/CD20, CD79A), T cells (CD3E), Myeloid cells (ITGAX/CD11C), Endothelial cells (CDH5) and Fibroblasts (COL1A2).

The highly expressed genes in each cell type were calculated by the 'FindConservedMarkers' function. The Cell Component Marker Genes were defined by those genes Fold Change > 2 and p value < 0.05, the overlapped genes within at least two cell types were excluded. The Tumor Micro-Environment Marker genes (MEMGs) were defined by the cell component marker genes representing B cell, T cell, myeloid cell, endothelial cell and fibroblast.

Weighted correlation network analysis (WGCNA)

Weighted correlation network analysis (WGCNA) was performed through 'WGCNA' R package. The cell component marker genes were used as input genes for WGCNA. Input genes and samples were filtered by good genes samples test via the 'goodSamplesGenes' function. The soft thresholding power β was chosen as the lowest power when scale free fit R^2 nears 0.85. In this study, $\beta = 5$ was selected to construct the scale-free network, generating eight non-grey gene modules. Eigengene values of the gene modules were calculated by 'moduleEigengenes' function.

The prognostic significance of each module was determined by univariate cox regression analyses and Kaplan–Meier analyses using 'survival' and 'survminer' package. The optimal cutoff values were estimated by R package 'maxstat'. Hub genes in each module were determined using both Intramodular connectivity (kWithin) and module membership (kME) scores. The functional enrichment of hub genes

was performed by 'enricher' function within 'ClusterProfiler' package, using genesets in Reactome database (msigdb, version 7.1.1).

Unsupervised transcriptome clustering

Consensus clustering of PDAC transcriptome data was performed via R package 'CancerSubtypes'[22], based on the expression of the MEMGs in blue and green modules, under the parameters: clusterAlg="km", distance="euclidean". The samples in each TME cluster were further filtered by silhouette score calculated by 'silhouette_SimilarityMatrix' function. Gene expression in TME clusters were visualized by the 'ComplexHeatmap' package. In the TCGA-PAAD dataset, genetic variations in TME clusters were summarized and visualized by 'oncoplot' function in 'maftools' package.

Estimation of tumor-microenvironmental infiltrating cells

To quantify the abundance of immune cells and other tumor-microenvironmental cells, we used the R package 'quanTIseq' to deconvolute RNA-seq data of PDAC samples[10]. Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was used to calculate tumor sample-infiltrating myeloid-derived suppressor cells (MDSC), and predicting immunotherapy responsiveness in PDAC patients[23]. The python script tidepy-1.3.7 was used to perform TIDE program.

Neural network model and risk score

To construct a prognosis-predicting model, we employed PyTorch to build a five-layer deep neural network (DNN) model[24] based on the expression of MEMGs. To train the DNN model, a randomly selected 2/3 subset of combined PDAC samples was used as training set. The optimizer learning rate was set to 0.05. The batch normalization was conduct in each layer. The Relu function was used as the activation function and the sigmoid function was applied in the output layer. The trained model was applied in other 1/3 subset of combined PDAC samples for internal testing and also subjected to external testing in other PDAC datasets. The probability value generated by DNN program was also used in prognostic analyses. Alternatively, another MEMGs based risk model was defined as weighted average expression of MEMGs. The Cox coefficient was used as the weight for each gene. The risk score was established in combined PDAC dataset and tested in other datasets. Meta-analysis was performed in R using the 'metafor' package [25] with random-effects models.

Cell-cell communication analysis

The intercellular cell-cell communication network in CRA001160 single cell RNA-seq data was constructed by the Network Analysis Toolkit for the Multicellular Interactions (NATMI) [11]. The ligand-receptor pairs were restricted to the cell junction molecules within WGCNA hub genes and extracted from a published ligand-receptor interaction list connectomeDB2020 [26]. The cell-cell communication score in bulk RNA-seq data was defined as the geometric mean of $(TPM_{Ligand}/TPM_{Ligand_reference})$ and $(TPM_{Receptor}/TPM_{Receptor_reference})$. The gene used as reference for each cell type was as follows: Tumor

cells (EPCAM), Endothelial cells (CDH5), Fibroblasts (COL1A2), Myeloid cells (ITGAM), Pancreatic acinar cells (CPA1), Pancreatic islet (NEUROD1), B cells (MS4A1) and Cytotoxic cells (CD3E).

Tissue samples and immunohistochemistry

A set of tissue microarrays (TMA) containing 66 PDAC samples purchased from Shanghai OUTDO Biotech CO.,LTD were used for immunohistochemistry (IHC) staining. This study has been approved by the Ethics Committee of Renji Hospital, Shanghai Jiao Tong University School of Medicine. For IHC analysis, the slides were rehydrated and then immersed in 3% hydrogen peroxide solution for 15 min. Slides were pretreated by microwave for 25 min in 0.01 mol/L citrate buffer, pH 6.0, at 95 °C; and cooled naturally to room temperature. Between each incubation step, the slides were washed with PBS, pH 7.4. Then, the tissues were incubated overnight at 4 °C with diluted anti-ITGA2 antibody (abcam, ab133557). After washing with PBS, the sections were visualized using VECTASTAIN® Elite ABC-HRP Kit, Peroxidase (Vectorlabs, PK-6104) as the manufacturer's instructions.

Mice and treatment

The *Pdx1-Cre* mice were crossed with *Kras*^(LSL-G12D) mice (Shanghai Model Organisms Center, Inc.) to generate mice with genotype *Pdx1-Cre*⁺, *Kras*^{G12D} (KC). The 12-16-week-old mice were orally treated with E7820 (100mg/kg bodyweight) once a day, for 15 consecutive days. The mice were sacrificed after management, the pancreas was fixed and subjected to hematoxylin and eosin (H&E)–staining. The tumoral lesions within pancreas were diagnosed and statistically analyzed. The fixed mice pancreas specimens were also subjected to immunohistochemical staining with primary antibodies, including anti-CK19 (Servicebio, GB12197), anti-Ki67 (Servicebio, GB111141), anti- α SMA (Servicebio, GB13044), anti-CD31 (Servicebio, GB113151) and anti-Gr1 (Servicebio, GB11229). The alcian blue staining was performed using the alcian blue staining kit (Servicebio, GP1040). All animal experiments were approved by the Institutional Animal Care and Use Committee at the Renji Hospital, Shanghai Jiao Tong University School of Medicine.

Cell culture and viability assay

The PDAC cell line SW1990 and PANC1 were acquired from the American Type Culture Collection (ATCC, Manassas VA, USA), and were maintained at 37°C in 5% CO₂ in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum. Cells were seeded at 1000 cells in 200 μ L DMEM per well in 96-well plates. At the indicated time points, 20 μ l Cell Counting Kit-8 reagent (Beyotime, C0039) was added to each well and incubated at 37°C for 3 h. The absorbance was measured by spectrophotometer at 450 nm with a reference wavelength of 600 nm.

Western blot

Cells were lysed by RIPA buffer (Thermo Fisher Scientific, 89901) with protease inhibitors cocktail (Roche Diagnostics, 05892970001) and phosphatase inhibitor cocktail (Roche Diagnostics, 04906845001). The

lysates were clarified by centrifugation at 12 000 g for 20 min at 4°C. Protein concentrations were measured by BCA protein assay kit (Thermo Fisher Scientific, 23225) and the samples were boiled with loading buffer. Protein samples (50–150 µg) were separated through SDS-PAGE, then transferred to nitrocellulose filter membrane (Pall Corporation) blocked and incubated with the primary antibodies. After washing with TBST three times, the blots were incubated with IRDye 800CW Secondary Antibody (licor, 926-32211) and visualized by Odyssey Sa Infrared Imaging System (LI-COR).

Real-time PCR

Total RNA from cells were extracted through RNAiso Plus kit (Takara Bio Inc.). The cDNA preparation was finished through primeScript RT Master kit (Takara Bio Inc.). Real-time PCR was performed by SYBR gGreen quantitative PCR kit (Life Technology) using the 7500 Real Time PCR System or ViiA7 System (AB Applied Biosystems). The primers include human ITGA2-F: GGCTGGCCCAGAGTTTACAT, human ITGA2-R: ATCGCCCCCTCTCCTAACTT. human GAPDH-F: CATGAGAAGTATGACAACAGCCT, human GAPDH-R: AGTCCTTCCACGATACCAAAGT.

Results

Weighted gene co-expression network analysis (WGCNA) identifies eight cell-marker-gene modules

Pancreatic ductal adenocarcinoma (PDAC) is preponderated by intricate tumor micro-environmental (TME) cells. To evaluate the prognostic importance of TME variations and key molecule events underlying tumor cell-TME interactions, a set of cell-cluster-marker genes was defined based on PDAC single-cell RNA-sequencing data. Then, the cell-cluster-marker genes were used in gauging TME status in bulk sequencing data to mine prognostic prediction systems in PDAC patients. (**Figure 1A**). A single-cell RNA-seq dataset CRA001160 was consecutively analyzed by Seurat software and dimension reduced by UMAP methods to obtain consensus clusters (**Figure 1B**). According to the expression of a handful of well-documented cell type specific genes (**Supplementary figure 1A**), we distinguished cellular identity of each cluster and integrated them into eight major components including B cells, cytotoxic cells, endothelial cells, fibroblasts, myeloid cells, acinar cells, islet cells and tumor cells (**Figure 1C**, **Supplementary figure 1B**). To acquire the Cell Component Marker Genes (CCMGs), the conserved markers were determined by the non-overlapped genes with Fold Change>2 and P value < 0.05, in each cluster (**Supplementary table 1**, **Supplementary figure 1C**). The genes representing B cell, T cell, myeloid cell, endothelial cell and fibroblast were then picked up and defined as Tumor Micro-Environment Marker genes (MEMGs).

To generate a training set for deciphering how tumor-microenvironment status links prognosis, we merged three cohorts of PDAC bulk RNA-sequencing datasets (PACA-CA, PACA-AU and GSE79668) into one combined PDAC RNA-seq dataset, using 'Combat' function to eliminate batch effect (**Supplementary figure 1D**). The single-cell data derived CCMGs were subjected as input genes to WGCNA analysis in the combined PDAC dataset under soft threshold 0.85, resulting in eight non-grey gene modules (**Figure 1D**, **Supplementary figure 2A-B**). When each gene in every module was traced back to the represented cell

type, the results showed that gene module-distribution across cell types (**Figure 1E**), and the cell type constitutes among gene modules (**Figure 1F**) are both highly heterogeneous. Tumor cell-derived marker genes were enriched and dominant in blue and green models, while the turquoise model was mainly contributed by endothelial cells and fibroblast-derived genes. Each CCMGs were also tested for their prognostic potential respectively, by Cox regression analyses. According to Cox regression results, The CCMGs were classified into good ($HR < 1$, $p < 0.05$) and poor ($HR > 1$, $p < 0.05$) outcome associated genes (**Supplementary table 2**). Considering the attribution of prognosis-significant genes to respective gene modules (**Figure 1G**) and cell types (**Figure 1H**), we found that the majority of genes within most of the gene modules and cell types indicated good survival. alternatively, the green gene module was preoccupied by the genes indicating deteriorated prognosis.

Censuses clustering PDAC samples by MEMGs into prognostic-related subtypes

Superior to calculating the prognostic value of a single gene, it is meaningful to evaluate the overall effects of one gene module to obtain the impact of gene co-expression networks. To this end, the Module Eigengene (ME) values were calculated and used for survival analysis for each module (**Supplementary table 3**). The Cox regression analyses and Kaplan-Meier analyses under optimal cutoff values revealed that high MEgreen values and low MEblue values inferred unfavorable outcomes (**Figure 2A-B**). To unwind the gene co-expression networks in green and blue modules, we considered the genes with both high Module membership (MM) value and intramodular connectivity value ($MM > 0.5$ & $Connectivity > 0.5$) as the hub genes (**Figure 2C, Supplementary table 4**). The results indicated that all 63 hub genes, excluding 2 endothelial genes, in both two modules derived from tumor cells, in agreement of the notion that tumor cells provide the the driving force in shaping tumor microenvironment.

To categorize PDAC patients in respect to tumor microenvironmental heterogeneity, we picked up MEMGs in green and blue modules for consensus clustering of PDAC patients. After filtering samples using silhouette score ($silhouette\ width > 0$), the remaining samples were assigned to three TME classes (**Figure 2D**). The vast majority of patients fell into C1 and C2 classes. Most poor-prognostic MEMGs and green module derived MEMGs were highly expressed in TME 2 classes. When analyzing the expression of hub genes in the TME classes, the expression of hub genes in green and blue modules were remarkably segregated along with C1 and C 2 classes (**Figure 2E**). The green module was correlated with worse survival, however, by contrast, the blue module was associated with benefited survival. Unexpectedly, patients in C1 and C2 were effectively separated by Kaplan-Meier curves (**Figure 2F**). To extend this finding, we referred the same consensus clustering method in TCGA-PAAD cohort, resulting in a similar result (**Supplementary figure 3**). These results demonstrated that distinct tumor-microenvironmental status, reflecting by a MEMG subset, was related to patients' outcomes. To compare tumor microenvironmental components between main TME classes, we used the 'quantiseq' and 'TIDE' deconvolution methods to estimate the abundance of some crucial tumor-infiltrating cells. The results showed that the C1 class were infiltrated by more cytotoxic cells, such as NK cells and $CD8^+$ T cells. In contrary, the C2 class were enriched of cancer-associated fibroblasts (CAF) and immunosuppressive cells MDSC (**Figure 2G**). To assess the relationship of genetic variations with TME classes, we mapped the

gene variation landscape in different TME classes. The results showed that the occurrence of concurrent mutations, *KRAS* and *TP53*, were more frequently in C2 class than C1 class, which may lead to tumor microenvironment remodeling.

Neural network model and risk score in predicting overall survival and chemo-responsiveness

To utilize the prognosis-significant gene modules in predicting overall survival of PDAC patients, we used PyTorch platform to construct a deep neural network (DNN) model based on MEMGs in blue and green modules [24]. The DNN model contains five layers as shown in **Figure 3A**. Firstly, the DNN model was trained using randomly selected 2/3 samples in the combined PDAC dataset. Then, the remaining 1/3 samples were used as internal testing set (AUC = 0.90). Since the median survival time is closed to one year, we applied this model to predict one-year survival. Consequently, we got an AUC = 0.88 in the training set, and AUC = 0.9 in the testing set (**Figure 3B**). When using the DNN predictor-derived probability score in Kaplan-Meier survival analysis, the patients with higher score indeed had shorter survival time (**Figure 3C**). In another external testing set, the TCGA-PAAD dataset, the DNN model was also successful in predicting patients' outcomes (AUC = 0.81; Kaplan-Meier survival analysis, $p < 0.0001$) (**Figure 3D-E**). Furthermore, the meta-analysis was performed to review the prognostic efficiency of DNN probability score in independent datasets. The results showed that DNN score associated with poor prognosis in all tested PDAC datasets. Meta-analysis through DL (DerSimonian and Laird) model resulted in a positive hazard ratio (HR=3.076, $p < 0.00001$), corroborating the general prognostic effect (**Figure 3F**). We also utilize the DNN model to predict chemo-sensitivity of PDAC. From the ICGC datasets, we selected the PDAC patients receiving chemotherapy for training DNN model. Then, the DNN model accurately predicted chemo-responsiveness (**Figure 3G-H**). Concurrently, the DNN probability score level stratified therapeutic success in PDAC patients (**Figure 3I**).

In addition to DNN model, we also developed a MEMG-based risk score model in prognosis prediction. The risk score was defined as weighted average expression of MEMGs. The Cox coefficient was used as the weight. In multiple independent PDAC datasets (**Supplementary table 5**), high risk score positively correlated with unfavorable prognosis (**Figure 4A**). Meta-analysis showed that the risk score had a positive hazard ratio in most individual datasets and in overall testing by DL model (**Figure 4B**). Moreover, the risk score related to the outcomes after chemotherapy (**Figure 4C**). When the risk scores were dimidiated by maxstat-derived cutoff value, the risk score level was significantly correlated with actual chemo-responses (Fisher's exact test, $P=0.017$) (**Figure 4D**). To exam the relationship between risk score and immunotherapeutic response, we showed that MEMG-based risk score positively correlated with TIDE-estimated MDSC abundance, but negatively correlated with Dysfunction score (**Figure 4E**). Importantly, the risk score level was correlated with TIDE-estimated immune checkpoint blockage response (Fisher's exact test, $P=0.007$) (**Figure 4F**).

Cell junction molecule-mediating cell-cell communications govern the prognosis-related gene-network

To understand the molecular basis of prognosis-related gene networks, we performed functional enrichment analysis of hub genes in REACTOME database. The results revealed that "cell junction" and

"cell communication" were on the top of enriched molecular function terms (**Figure 5A, Supplementary table 6**). We selected thirteen cell junction hub genes (in green and blue modules), which serve as cell communication ligand/receptor, as central cell-cell communication mediators (**Figure 5B**). As shown in **Figure 5C**, the selected central cell communication mediators weaved a co-expression correlation network with convoluted tumor cell-tumor cell or tumor cell-TME connections. The ligand-receptor pairs involving the above cell-cell communication mediators were extracted from the connectomeDB2020 database (**Figure 5D**) and subjected for following assays. The cell-cell communication analysis was performed in the single-cell RNA-seq data (CRA001160) through NATMI algorithm. Because certain ligand/receptor molecule may be expressed in more than one cell type. One ligand-receptor pair may modulate communications between several cell pairs. on the other hand, one cell pairs could also be bridged by multiple ligand-receptor pairs (**Figure 5E**). The cell pair with the largest average expression value was considered as the top cell-cell connection for each ligand-receptor pair (**Supplementary table 7**). By assigning the ligands/receptors to sending cell types and target cell types according to the top connections, we got an cells-ligand-receptor-targets alluvial plot, in which molecules were weighted by average expression and cell types were weighted by sum average expression of contributing molecules. This network figure out that tumor cells, fibroblasts and endothelial cells were the most vital cell types engaged in cell-cell communication in pancreatic cancers (**Figure 5F**).

Integrins are key mediators critical for tumor cell to micro-environment communications

To interrogate the prognostic effect of cell-cell communication linked by particular ligand-receptor pair, we calculated the cell-cell communication score in bulk RNA-seq data for Cox regression analyses in the combined PDAC and TCGA-PAAD cohorts (**Figure 6A**). The strength of most ligand-receptor-induced cell-cell connection remarkably correlated with poor prognosis (HR>1, P<0.05). Integrin-mediated tumor cell-fibroblast communication and tumor cell-endothelial cell communication were prior in the ordered hazard ratio ranks. Besides, the majority of cell-cell communication scores were correlated with DNN model-generated probability score (correlation coefficient > 0, P<0.05), capable of indicting patient's outcome (**Figure 6B**). By integrating the cell-cell communication networks in both combined PDAC dataset and TCGA-PAAD dataset, we found that tumor cell, fibroblast and endothelial cells were the most significant cellular component, meanwhile, integrins such as ITGA2, ITGA6 were main contributing molecules (**Figure 6C**). Correlation analyses showed that the mean expression of the integrins in hub genes was largely in parallel with cell-cell communication score, as well as DNN probability score and risk score (**Figure 6D, Supplementary figure 4**). More than that, the PDAC samples highly expressing the hub integrins were accumulated in TME C1 class, and possessed more driver genetic variations (**Figure 6E**). These results revealed that integrins may key mediators participated in the dialogue between tumor cell and micro-environmental cells, potentiating the aggressive phenotype of PDAC.

Pharmacological blocking of ITGA2 orchestrates micro-environmental changes and limits PDAC initiation

Intrigued by the pivotal roles of integrins in intercellular network, we seek out to choose the most significant integrin, ITGA2, testing for the protein expression and druggable potential. In the tissue

microarray containing 66 PDAC samples, we performed immunohistochemical analysis against ITGA2. ITGA2 was shown to be clearly expressed on the membrane of tumor cells (**Figure 7A**). In accordance with the bioinformatic analyses, the PDAC patients highly expressing ITGA2 implied inferior prognosis (**Figure 7B**). To explore whether and to what extent targeting ITGA2 influence PDAC development, we employed *Pdx1-Cre⁺, Kras^{G12D}* (KC) mice for *in vivo* inhibitor management. The mice were orally treated with ITGA2 inhibitor E7820[27], and then, the effected pancreas area was quantified. It appeared that E7820 essentially diminished pancreatic lesions without profoundly influence body weights (**Figure 7C**, **Supplementary Figure 5A**). The efficiency of E7820 treatment was also revealed by the detection of ductal biomarker CK19, mucin content (Alcian blue staining) and proliferation marker Ki67 (**Figure 7D**). In cultured PDAC cells, E7820 indeed suppressed the mRNA and protein expression of ITGA2 (**Supplementary Figure 5B-C**), which in accordance with the molecular mechanism of this inhibitor[28], but it does not significantly alter the proliferation rate of PDAC cells *in vitro* (**Supplementary Figure 5D**). Furthermore, in mice model, the lesioned area within pancreas from mice receiving E7820 contained less α SMA-positive fibroblasts, CD31-positive micro-vessels and Gr1-positive MDSCs (**Figure 7E**). Collectively, our *in vivo* pharmacology experiment claimed that targeting the key molecule in the cell communication network reshaped the tumor-benefiting micro-environment to decelerate PDAC growth.

Discussion

In this work, we set out to utilize cell type-specific information for uncovering clinically valuable gene co-expression networks in a large amount of bulk transcriptome data. By performing the WGCNA analysis through cell component marker genes, we gained eight non-grey gene modules, representing eight gene co-expression networks. Since every single module was constituted by the genes derived from more than one cell type, each network actually incorporates the co-expression relationship inside certain cell type and across separated cell types. Here, we mainly focus on the two prognosis-related gene networks, the green and the blue ones (Figure 2A-B). According to the concepts of WGCNA, the hub genes within the co-expression network may play a central role in the functional group[15]. So, it is reasonable for us to assume that hub genes of these two networks are responsible for the clinical behaviors of PDAC. Because the majority of hub genes shall trace back to tumor cells (Figure 2C). Our findings reinforce the notion that tumor intrinsic genes shape the micro-environment, regulating tumor development and therapeutic responses [29–31].

In the context of pancreatic cancer, there are also increasing evidences propose that tumor intrinsic factors may contribute to stroma remodeling[32, 33]. One excellent work demonstrated that distinct clones of cancer cells give rise to heterogenous tumor micro-environments[34]. To screen the key cancer cell-intrinsic molecules, we performed functional enrichment analysis. The result highlights an accumulation of cellular communication and/or cell junction molecules in the hub genes (Figure 5A-B). Intriguingly, these molecules wire a ligand-receptor web, guiding the cross-talk between cancer cells and non-tumoral cells (Figure 5D-E). In both single-cell RNA-seq data and bulk expression data, the main participators are fibroblasts and endothelial cells, which frequently connected with tumor cells.

Extraordinarily, a group of integrins govern the fibroblast-tumor cell or endothelial cell-tumor cell communications that are prioritized in the significant risk factors (Figure 6A).

Integrins belongs to a family of transmembrane receptors, mediating cell-cell adhesion and cell to extracellular matrix (ECM) interaction. On the cell surface, integrins form heterodimeric complex, composed of one α subunit and one β subunit, that recognizes ECM on one side, while links cell skeleton and/or intercellular signaling pathways on the other side [35, 36]. By attaching to ECM, integrins response to these micro-environmental components, transmitting outer signals to inner compartment. Comparable to the cytokine- cytokine receptor system, certain integrin binds to distinct ECM molecule, exhibits exquisite specificity. Based on this property, the integrins are also capable of configuring intracellular dialoguing ligand-receptor pairs (Figure 5D).

In this work, through data mining in single-cell RNA-seq data, the cell type displaying each integrin was identified. In the top interlinks, ITGA2, ITGA3, ITGB4 and ITGB6 are dominantly expressed in tumor cells. ITGA6 is mainly expressed in endothelial cells, which is one of the two non-tumoral cell derived hub genes (Figure 5C). ITGB1 is overwhelmed localized in fibroblasts. Some of these integrins have been implicated in tumorigenesis of PDAC [37–39]. The main contributing cell types of some integrins are also endorsed by literatures[40, 41]. As the ligands for specific integrin, the ECM components may also be secreted by many cell types. In our hub gene-associated ligand-receptor pairs, ITGA2 is partnered with the maximum number of ligands (Figure 5D), such as fibroblast-secreting collagens, fibronectins, or laminins (COL1A2, COL1A1, COL8A1, FN1, LAMA1)[42] and endothelial cell-producing HSPG2[43]. More importantly, ITGA2 is shown to be the most remarkable hub in the cell-cell communication network, weighted by integrated hazard ratio (Figure 6C). Therefore, it is much more likely that ITGA2 is an Achilles' heel in cell-cell communication network. Some comprehensive works have shown that targeting cell-cell communication may be an attractive way in moderating PDAC [40, 44]. It is instructive to test the impact of blocking ITGA2.

To do with a clinical compatible way for targeting ITGA2, we selected a ITGA2 inhibitor E7820 which has been undertaken clinical trials [27, 45–47]. In PDAC, the effect of E7820 is unclear. To pre-clinically mimic the *in vivo* performance of this reagent, we used an oncogene-driven spontaneous pancreatic tumor model, in which oral delivery of E7820 substantially alleviates pancreatic lesions. This *in vivo* pharmacological experiment highlights the medical prospect of this ITGA2 inhibitor. ITGA2 was initially used as an angiogenesis antagonist, since E7820 apparently slows down proliferation and attenuates tube formation ability of endothelial cells[28]. Here, we clearly show that ITGA2 is basically expressed in the tumor cells, demonstrating by both single-cell transcriptome data and immunohistochemistry detection (Figure 7A). The high protein expression of ITGA2 in tumor cells was associated with terrible outcomes (Figure 7B). Hence, the primary target of E7820 is tumor cell itself, rather than endothelial cells. *In vitro* assays show that, similar to cultured endothelial cells, E7820 also reduces mRNA expression of ITGA2 in PDAC cells, whereas the suppressive effects on cell growth are merely slight, disproportionate with the inhibitory effect to the tumors *in situ*. Moreover, E7820 considerably reduces the fibroblasts and micro-vessels around tumor foci. These observations suggest that E7820 shrinks pancreatic tumors in a

micro-environment dependent way. Targeting ITGA2 centered cell-cell communication network may be a perspective strategy to cure PDAC.

Conclusions

Taken together, our work reveals that a group of micro-environmental genes classifies PDAC samples into prognosis-associated subtypes. This gene set are capable of predicting patient's outcome or therapeutic sensitivity based on neural network and risk score models. Underneath extracellular heterogeneity, a set of tumor cell-intrinsic hub factors, exemplified by integrins, govern a clinical-relevant cell-cell communication network. Pharmacologically inhibiting ITGA2 dramatically alleviates pancreatic tumor development *in vivo*, providing an appealing drug target for pancreatic cancer.

Abbreviations

PDAC: Pancreatic ductal adenocarcinoma

TME: Tumor micro-environment

CCMG: Cell component marker genes

NATMI: Network Analysis Toolkit for the Multicellular Interactions

TPM: Transcripts per kilobase million

UMAP: Uniform Manifold Approximation and Projection

MEMGs: Tumor Micro-Environment Marker genes

WGCNA: Weighted correlation network analysis

kWithin: Intramodular connectivity

kME: Module membership

TIDE: Tumor Immune Dysfunction and Exclusion

MDSC: Myeloid-derived suppressor cells

DNN: Deep neural network

ME: Module Eigengene

MM: Module membership

CAF: Cancer-associated fibroblasts

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

A single-cell RNA sequencing (scRNA-seq) dataset of pancreatic adenocarcinoma samples was obtained from the Genome Sequence Archive (GSA) database under the accession code CRA001160 (<https://ngdc.cncb.ac.cn/gsa/browse/CRA001160>). The gene expression profiles of PDAC transcriptome assays, GSE28735, GSE62452 and GSE71729 were obtained from the Gene Expression Omnibus (GEO) database (GEO; <https://www.ncbi.nlm.nih.gov/geo/>). RNA-seq data of PACA-AU and PACA-CA, and array-based gene expression profiling data (exp_array) together with clinical information were from International Cancer Genome Consortium (ICGC) database (<https://dcc.icgc.org/>). The TCGA dataset was available in the GDC portal (<https://portal.gdc.cancer.gov/>).

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China (82172905, 81972209, 82060041) and Shanghai Natural Science Fund (21ZR1461500).

Authors' contributions

WBS, WCS, JTT and LYZ designed the research, analyzed data, performed experiments and wrote the manuscript. CK, LT and WDH performed some experiments and provided intellectual inputs.

Acknowledgements

Not applicable

References

1. Mizrahi JD, Surana R, Valle JW, Shroff RT: **Pancreatic cancer**. *The Lancet* 2020, **395**:2008-2020.
2. Grossberg AJ, Chu LC, Deig CR, Fishman EK, Hwang WL, Maitra A, Marks DL, Mehta A, Nabavizadeh N, Simeone DM, et al: **Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma**. *CA Cancer J Clin* 2020, **70**:375-403.

3. 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.
4. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, Miller DK, Christ AN, Bruxner TJ, Quinn MC, et al: **Genomic analyses identify molecular subtypes of pancreatic cancer.** *Nature* 2016, **531**:47-52.
5. Dreyer SB, Upstill-Goddard R, Legrini A, Biankin AV, Jamieson NB, Chang DK: **Genomic and molecular analyses identify molecular subtypes of pancreatic cancer recurrence.** *Gastroenterology* 2021.
6. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, et al: **Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy.** *Nat Med* 2011, **17**:500-503.
7. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, Mollaei M, Wagner KU, Koduru P, Yopp A, et al: **Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets.** *Nat Commun* 2015, **6**:6744.
8. Connor AA, Denroche RE, Jang GH, Lemire M, Zhang A, Chan-Seng-Yue M, Wilson G, Grant RC, Merico D, Lungu I, et al: **Integration of Genomic and Transcriptional Features in Pancreatic Cancer Reveals Increased Cell Cycle Progression in Metastases.** *Cancer Cell* 2019, **35**:267-282 e267.
9. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, List M, Aneichyk T: **Comprehensive evaluation of transcriptome-based cell-type quantification methods for immunology.** *Bioinformatics* 2019, **35**:i436-i445.
10. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, Krogsdam A, Loncova Z, Posch W, Wilflingseder D, et al: **Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data.** *Genome Med* 2019, **11**:34.
11. Hou R, Denisenko E, Ong HT, Ramilowski JA, Forrest ARR: **Predicting cell-to-cell communication networks using NATMI.** *Nat Commun* 2020, **11**:5011.
12. Kumar MP, Du J, Lagoudas G, Jiao Y, Sawyer A, Drummond DC, Lauffenburger DA, Raue A: **Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics.** *Cell Rep* 2018, **25**:1458-1468 e1454.
13. Peng J, Sun BF, Chen CY, Zhou JY, Chen YS, Chen H, Liu L, Huang D, Jiang J, Cui GS, et al: **Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma.** *Cell Res* 2019, **29**:725-738.
14. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, Teinor JA, Belleau P, Biffi G, Lucito MS, et al: **Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts.** *Cancer Discov* 2019, **9**:1102-1123.
15. Zhang G, He P, Tan H, Budhu A, Gaedcke J, Ghadimi BM, Ried T, Yfantis HG, Lee DH, Maitra A, et al: **Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer.** *Clin Cancer Res* 2013, **19**:4983-4993.

16. Yang S, He P, Wang J, Schetter A, Tang W, Funamizu N, Yanaga K, Uwagawa T, Satoskar AR, Gaedcke J, et al: **A Novel MIF Signaling Pathway Drives the Malignant Character of Pancreatic Cancer by Targeting NR3C2.** *Cancer Res* 2016, **76**:3838-3850.
17. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, et al: **Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma.** *Nat Genet* 2015, **47**:1168-1178.
18. Kirby MK, Ramaker RC, Gertz J, Davis NS, Johnston BE, Oliver PG, Sexton KC, Greeno EW, Christein JD, Heslin MJ, et al: **RNA sequencing of pancreatic adenocarcinoma tumors yields novel expression patterns associated with long-term survival and reveals a role for ANGPTL4.** *Mol Oncol* 2016, **10**:1169-1182.
19. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, et al: **TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data.** *Nucleic Acids Res* 2016, **44**:e71.
20. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, Stein LD, Ferretti V: **The International Cancer Genome Consortium Data Portal.** *Nat Biotechnol* 2019, **37**:367-369.
21. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: **Integrating single-cell transcriptomic data across different conditions, technologies, and species.** *Nat Biotechnol* 2018, **36**:411-420.
22. Xu T, Le TD, Liu L, Su N, Wang R, Sun B, Colaprico A, Bontempi G, Li J: **CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization.** *Bioinformatics* 2017, **33**:3131-3133.
23. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, et al: **Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response.** *Nat Med* 2018, **24**:1550-1558.
24. Bao X, Shi R, Zhao T, Wang Y: **Mast cell-based molecular subtypes and signature associated with clinical outcome in early-stage lung adenocarcinoma.** *Mol Oncol* 2020, **14**:917-932.
25. Viechtbauer W: **Conducting Meta-Analyses in R with the metafor Package.** *Journal of Statistical Software* 2010, **36**:1-48.
26. Ramilowski JA, Goldberg T, Harshbarger J, Kloppmann E, Lizio M, Satagopam VP, Itoh M, Kawaji H, Carninci P, Rost B, Forrest AR: **A draft network of ligand-receptor-mediated multicellular signalling in human.** *Nat Commun* 2015, **6**:7866.
27. Hulskamp MD, Kronenberg D, Stange R: **The small-molecule protein ligand interface stabiliser E7820 induces differential cell line specific responses of integrin alpha2 expression.** *BMC Cancer* 2021, **21**:571.
28. Funahashi Y, Sugi NH, Semba T, Yamamoto Y, Hamaoka S, Tsukahara-Tamai N, Ozawa Y, Tsuruoka A, Nara K, Takahashi K, et al: **Sulfonamide derivative, E7820, is a unique angiogenesis inhibitor suppressing an expression of integrin alpha2 subunit on endothelium.** *Cancer Res* 2002, **62**:6116-6123.
29. Yang L, Li A, Lei Q, Zhang Y: **Tumor-intrinsic signaling pathways: key roles in the regulation of the immunosuppressive tumor microenvironment.** *J Hematol Oncol* 2019, **12**:125.

30. Li L, Yang L, Cheng S, Fan Z, Shen Z, Xue W, Zheng Y, Li F, Wang D, Zhang K, et al: **Lung adenocarcinoma-intrinsic GBE1 signaling inhibits anti-tumor immunity.** *Mol Cancer* 2019, **18**:108.
31. Yang Z, Xu G, Wang B, Liu Y, Zhang L, Jing T, Tang M, Xu X, Jiao K, Xiang L, et al: **USP12 downregulation orchestrates a protumorigenic microenvironment and enhances lung tumour resistance to PD-1 blockade.** *Nat Commun* 2021, **12**:4852.
32. Thomas D, Radhakrishnan P: **Tumor-stromal crosstalk in pancreatic cancer and tissue fibrosis.** *Mol Cancer* 2019, **18**:14.
33. Neesse A, Bauer CA, Ohlund D, Lauth M, Buchholz M, Michl P, Tuveson DA, Gress TM: **Stromal biology and therapy in pancreatic cancer: ready for clinical translation?** *Gut* 2019, **68**:159-171.
34. Li J, Byrne KT, Yan F, Yamazoe T, Chen Z, Baslan T, Richman LP, Lin JH, Sun YH, Rech AJ, et al: **Tumor Cell-Intrinsic Factors Underlie Heterogeneity of Immune Cell Infiltration and Response to Immunotherapy.** *Immunity* 2018, **49**:178-193 e177.
35. Desgrosellier JS, Cheresch DA: **Integrins in cancer: biological implications and therapeutic opportunities.** *Nat Rev Cancer* 2010, **10**:9-22.
36. Kechagia JZ, Ivaska J, Roca-Cusachs P: **Integrins as biomechanical sensors of the microenvironment.** *Nat Rev Mol Cell Biol* 2019, **20**:457-473.
37. Shimomura H, Okada R, Tanaka T, Hozaka Y, Wada M, Moriya S, Idichi T, Kita Y, Kurahara H, Ohtsuka T, Seki N: **Role of miR-30a-3p Regulation of Oncogenic Targets in Pancreatic Ductal Adenocarcinoma Pathogenesis.** *Int J Mol Sci* 2020, **21**.
38. Dey S, Liu S, Factora TD, Taleb S, Riverahernandez P, Udari L, Zhong X, Wan J, Kota J: **Global targetome analysis reveals critical role of miR-29a in pancreatic stellate cell mediated regulation of PDAC tumor microenvironment.** *BMC Cancer* 2020, **20**:651.
39. Ren D, Zhao J, Sun Y, Li D, Meng Z, Wang B, Fan P, Liu Z, Jin X, Wu H: **Overexpressed ITGA2 promotes malignant tumor aggression by up-regulating PD-L1 expression through the activation of the STAT3 signaling pathway.** *J Exp Clin Cancer Res* 2019, **38**:485.
40. Zhao W, Ajani JA, Sushovan G, Ochi N, Hwang R, Hafley M, Johnson RL, Bresalier RS, Logsdon CD, Zhang Z, Song S: **Galectin-3 Mediates Tumor Cell-Stroma Interactions by Activating Pancreatic Stellate Cells to Produce Cytokines via Integrin Signaling.** *Gastroenterology* 2018, **154**:1524-1537 e1526.
41. Xu H, Pumiglia K, LaFlamme SE: **Laminin-511 and alpha6 integrins regulate the expression of CXCR4 to promote endothelial morphogenesis.** *J Cell Sci* 2020, **133**.
42. Kalluri R: **The biology and function of fibroblasts in cancer.** *Nat Rev Cancer* 2016, **16**:582-598.
43. Apostolidis SA, Stifano G, Tabib T, Rice LM, Morse CM, Kahaleh B, Lafyatis R: **Single Cell RNA Sequencing Identifies HSPG2 and APLNR as Markers of Endothelial Cell Injury in Systemic Sclerosis Skin.** *Front Immunol* 2018, **9**:2191.
44. Shi Y, Gao W, Lytle NK, Huang P, Yuan X, Dann AM, Ridinger-Saison M, DelGiorno KE, Antal CE, Liang G, et al: **Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring.** *Nature* 2019, **569**:131-135.

45. Semba T, Funahashi Y, Ono N, Yamamoto Y, Sugi NH, Asada M, Yoshimatsu K, Wakabayashi T: **An angiogenesis inhibitor E7820 shows broad-spectrum tumor growth inhibition in a xenograft model: possible value of integrin alpha2 on platelets as a biological marker.** *Clin Cancer Res* 2004, **10**:1430-1438.
46. Mita M, Kelly KR, Mita A, Ricart AD, Romero O, Tolcher A, Hook L, Okereke C, Krivelevich I, Rossignol DP, et al: **Phase I study of E7820, an oral inhibitor of integrin alpha-2 expression with antiangiogenic properties, in patients with advanced malignancies.** *Clin Cancer Res* 2011, **17**:193-200.
47. Milojkovic Kerklaan B, Slater S, Flynn M, Greystoke A, Witteveen PO, Megui-Roelvink M, de Vos F, Dean E, Reyderman L, Ottesen L, et al: **A phase I, dose escalation, pharmacodynamic, pharmacokinetic, and food-effect study of alpha2 integrin inhibitor E7820 in patients with advanced solid tumors.** *Invest New Drugs* 2016, **34**:329-337.

Figures

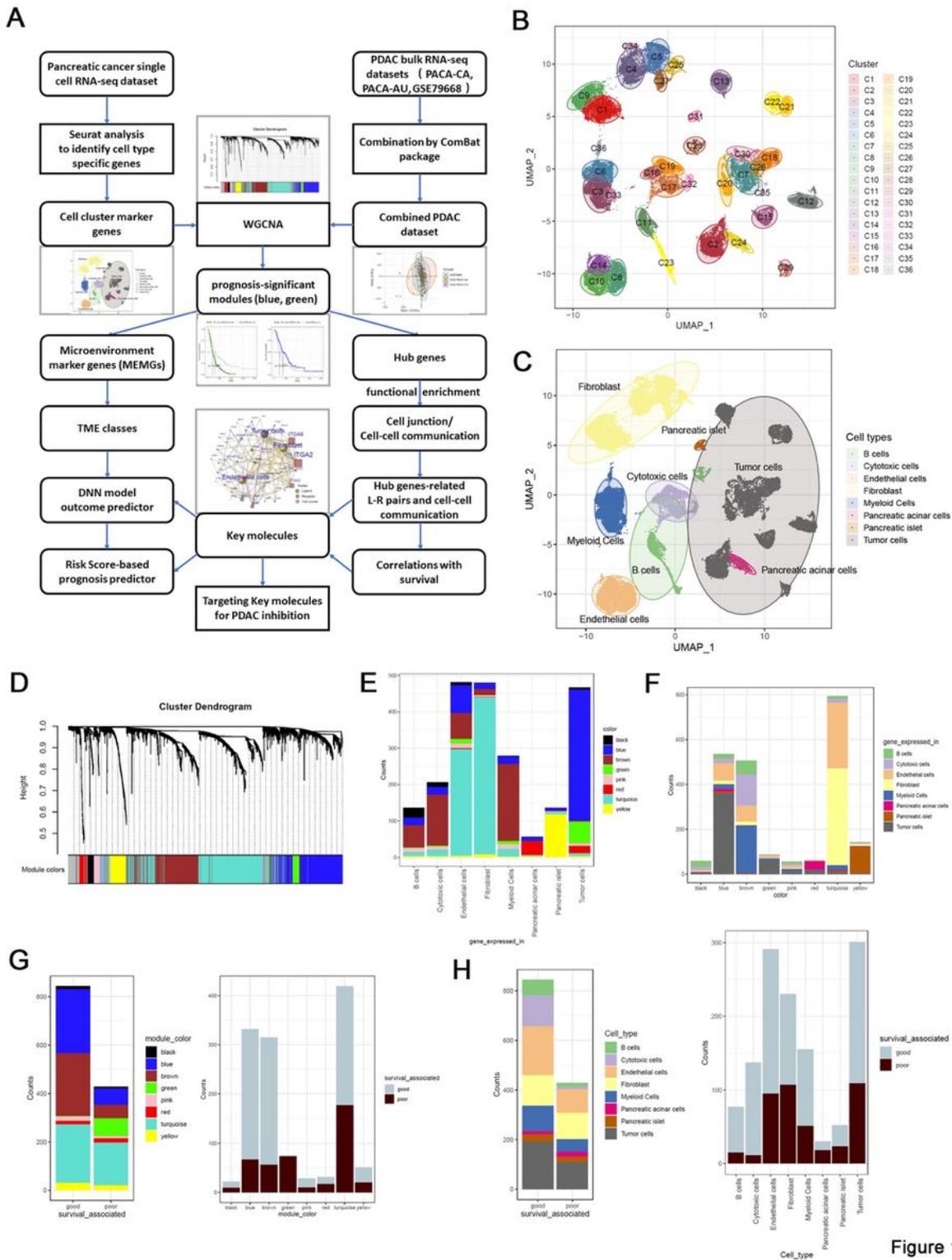


Figure 1

Figure 1

Weighted gene co-expression network analysis (WGCNA) classifies cell component marker genes into eight gene modules in pancreatic cancer (A) The workflow diagram summarizes the study design in this work. (B-C) Uniform Manifold Approximation and Projection (UMAP) Visualization of pancreatic ductal adenocarcinoma (PDAC) single-cell Sequencing data exhibits cell clusters discovered by Seurat analysis (B), and cellular identities of every cluster (C). Cell type of each cluster was distinguished through the

expression of well-documented markers (shown in Supplementary figure 1A) and grouped into eight major components. (D) Cell component marker genes found in PDAC single-cell dataset (Supplementary figure 1C and Supplementary table 1A) were subjected to Weighted gene co-expression network analysis (WGCNA) in a combined PDAC bulk RNA-sequencing dataset. In this process, eight gene-network modules were detected. (E) Stacked barplots show the cell origins of the genes constituting each module. (F) Distribution of different cell component marker genes in the gene modules. (G-H) Barplots show the prognostic significant genes in every gene module (G) and each cell type (H). The prognosis-significance was determined by cox regression analysis.

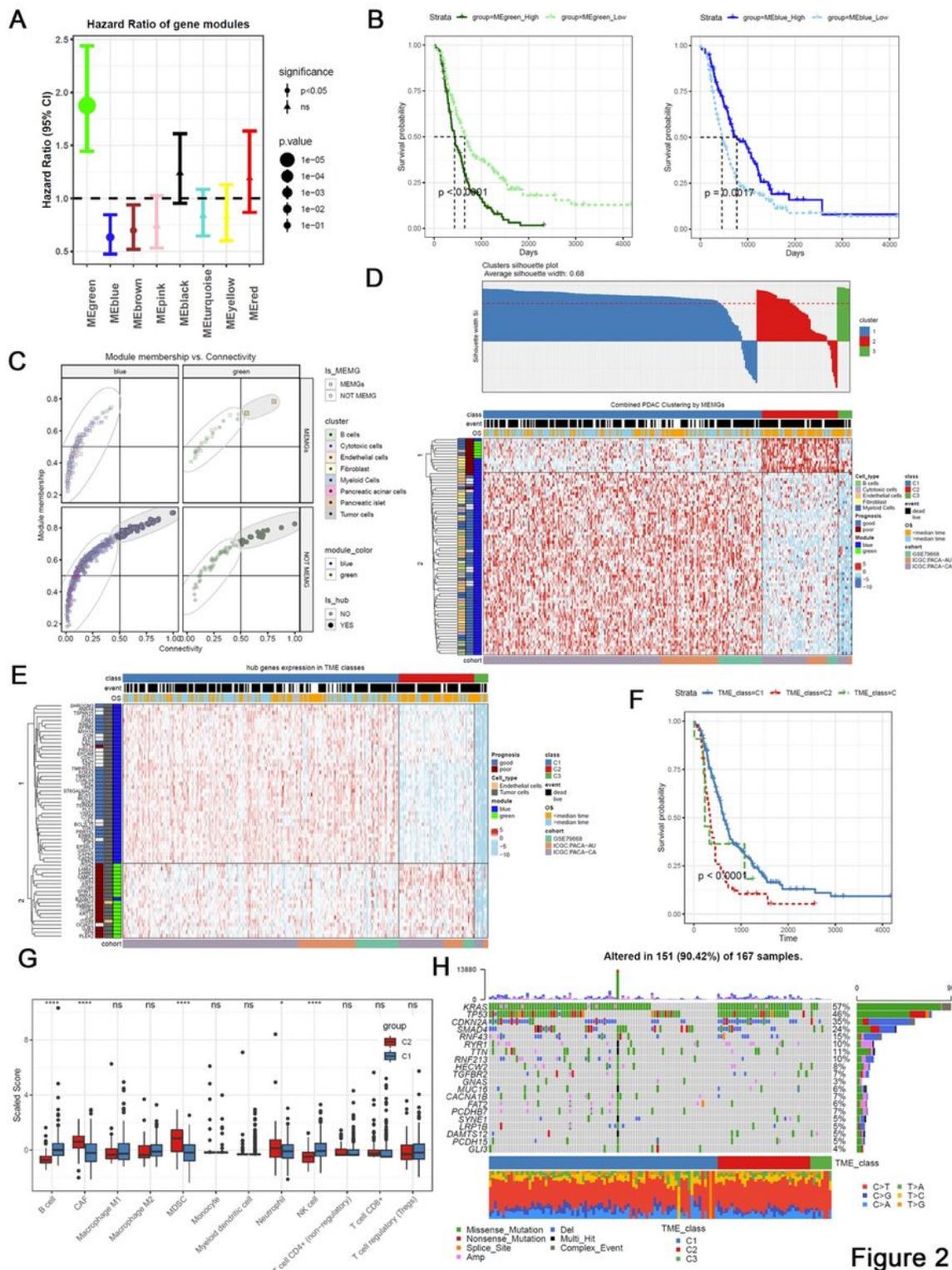


Figure 2

Figure 2

Tumor micro-environment marker genes (MEMGs) stratify PDAC patients into subtypes with distinct outcomes (A) The eigengene (ME) values for the gene modules were calculated in PDAC samples using the transcriptome data, then the cox regression analyses were performed to estimate the hazard ratio (HR) for every module. Forest plot shows the HR and and 95% confidence interval (95% CI). The optimal cutoff value in one module was estimated by maxstat function. (B) Kaplan–Meier curves indicate the

eigengene values of green and blue modules correlate with poor and favorite prognosis respectively, P values were evaluated by log-rank test. (C) Micro-environment marker genes (MEMGs) and hub genes in the prognosis-related modules (blue and green modules) were defined by the expression in cell components, module membership and intramodular connectivity. (D) MEMGs stratifies PDAC patients into three TME classes by consensus cluster algorithm. Different expressed MEMGs among subclasses were shown in heatmap. (E) Heatmap shows the expression pattern of hub genes in different TME classes. (F) Kaplan–Meier curves show the variations of survival rates among TME classes, statistical analysis was taken by log-rank test. (G) Relative abundance and differentiations of tumor environmental cells in two major TME classes. The abundance of tumor infiltrating immune cells, fibroblasts and endothelial cells were estimated by Microenvironment Cell Populations-counter (MCP-counter) algorithm or Tumor Immune Dysfunction and Exclusion (TIDE) methods. Differences were tested by student's t test, ***P<0.001, *P<0.05. (H) The waterflow plot displays recurrent genetic variations in each TME class.

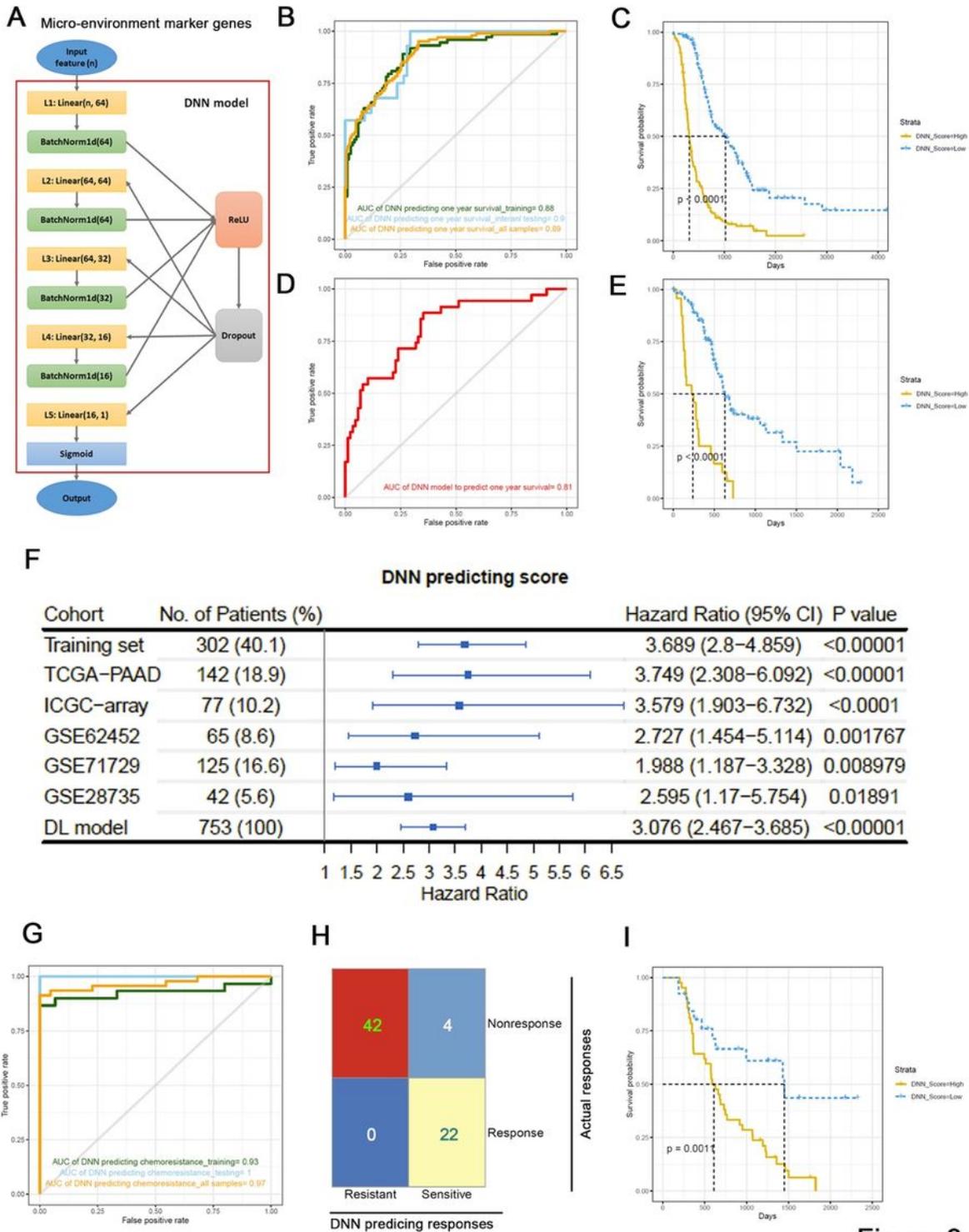


Figure 3

Figure 3

A deep neural network (DNN) model predicts outcomes of PDAC patients based on the expression of MEMGs. (A) The framework of the deep neural network (DNN) containing five layers. (B) The samples in combined PDAC dataset were randomly divided into training groups (2/3) and internal testing groups (1/3). The DNN model was trained in training samples by 400 iterations, and tested by internal testing samples and all samples. ROC curves show the one-year survival predicting accuracies in each group.

The Area Under Curve (AUC) values were shown. (C) The DNN probability score generated by DNN predictor in the all-sample-testing process was correlated with overall survival of patients. the Kaplan–Meier plot shows the survival rates, and statistical significance was tested by log-rank test. (D-E) The prognostic capability of DNN model was examined in the external testing set. The ROC curve (D) and Kaplan–Meier plot stratified by DNN scores (E) were shown. (F) Systematical cox analyses estimate the Hazard Ratio of DNN scores in multiple PDAC datasets from TCGA, ICGC or GEO databases. The meta-analysis was performed using DL (DerSimonian and Laird) model to estimate the general prognostic effect of DNN score in pancreatic cancer patients. (G) The DNN framework was also training to be a chemotherapy response predictor using a cohort of PDAC patients undergone chemotherapy. ROC curves show the AUC values using a DNN model to predict chemotherapeutic responsiveness in training, testing or all samples. (H) A confusion matrix demonstrates the accuracy of DNN model in predicting chemotherapy response. (I) DNN score correlates with patient's survival after chemotherapy. The survival rates were shown in Kaplan-Meier plot, the differentiation was tested by log-rank test.

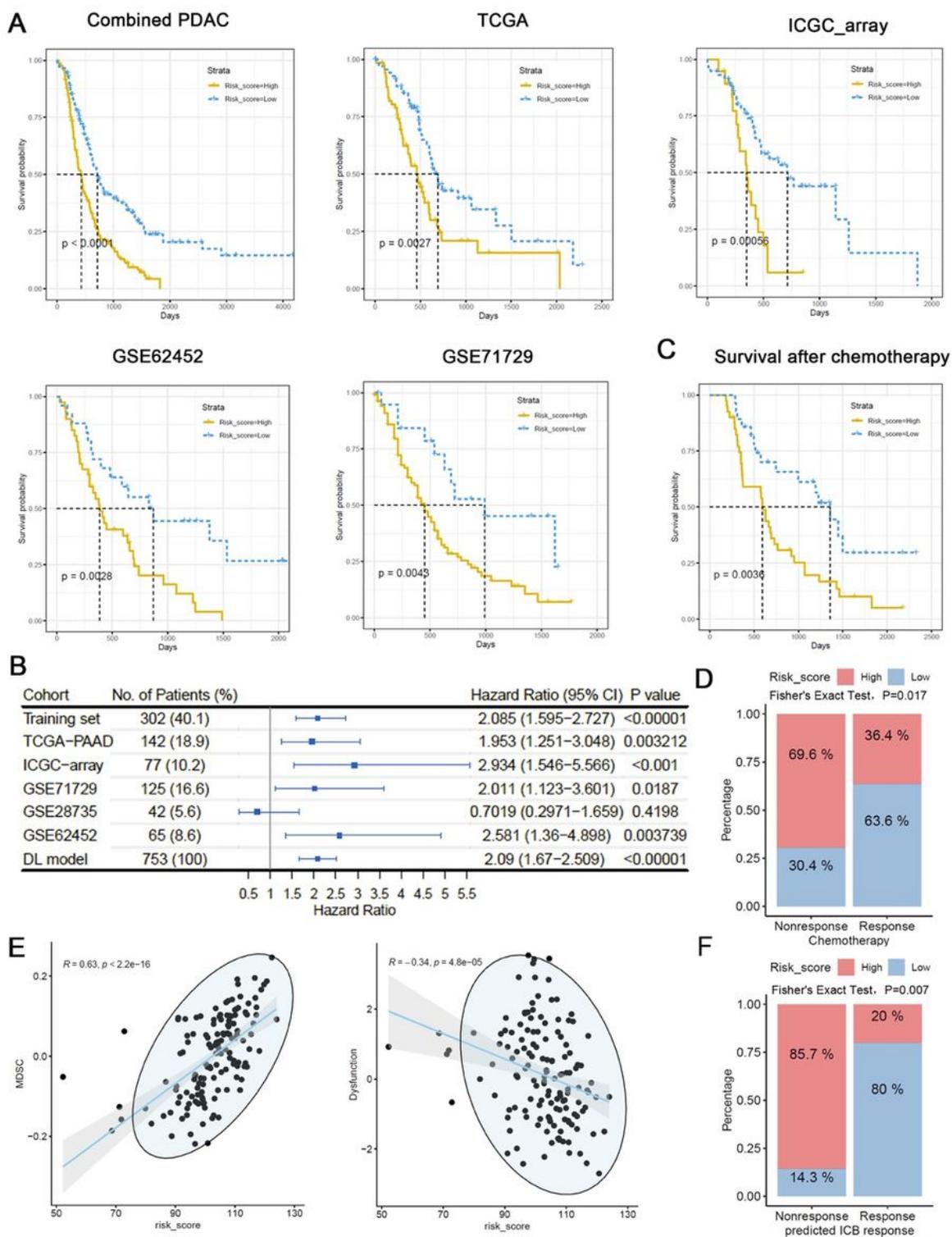


Figure 4

Figure 4

MEMGs-based risk score infers therapy efficiencies in pancreatic cancer patients. (A) A risk score algorithm based on the expression pattern of MEMGs was established in combined PDAC (training) dataset, and tested for its prognostic relationship in the multiple testing sets. (B) The hazard ratios of risk scores in different PDAC datasets were shown and a meta-analysis with DL model was performed to estimate the overall effect. (C) A Kaplan–Meier plot shows the correlation with risk score with PDAC

Cell junction molecule-mediated cell-cell communications between tumor cell and micro-environmental cells dominate the prognosis-related gene-network. (A) Functional enrichment analysis was performed in REACTOME database to unveil the dominant molecular function of hub genes constituting prognosis-related networks. (B) Hub genes functioned as cell junction molecule to modulate external cellular communication (ligands/receptors) were selected for the following study. (C) A gene correlation network shows the links between gene pairs initiated from central cell-cell communication mediators (named in the graph) to their closely connected ($R > 0.5$) genes within modules. colored edges represent the links involving MEMGs. (D) The hub genes related ligand-receptor pairs were extracted from connectomeDB2020 database for cell-cell communication prediction in PDAC single cell dataset. (E) The cell-cell communications mediated by each ligand-receptor pair were determined using Network Analysis Toolkit for the Multicellular Interactions (NATMI) software. The top cell-cell communication pattern bridged by each ligand-receptor pair was noted. (F) Alluvial diagram represents the hub gene-associated ligand-receptor pairs engaging tumor cell to TME communications.

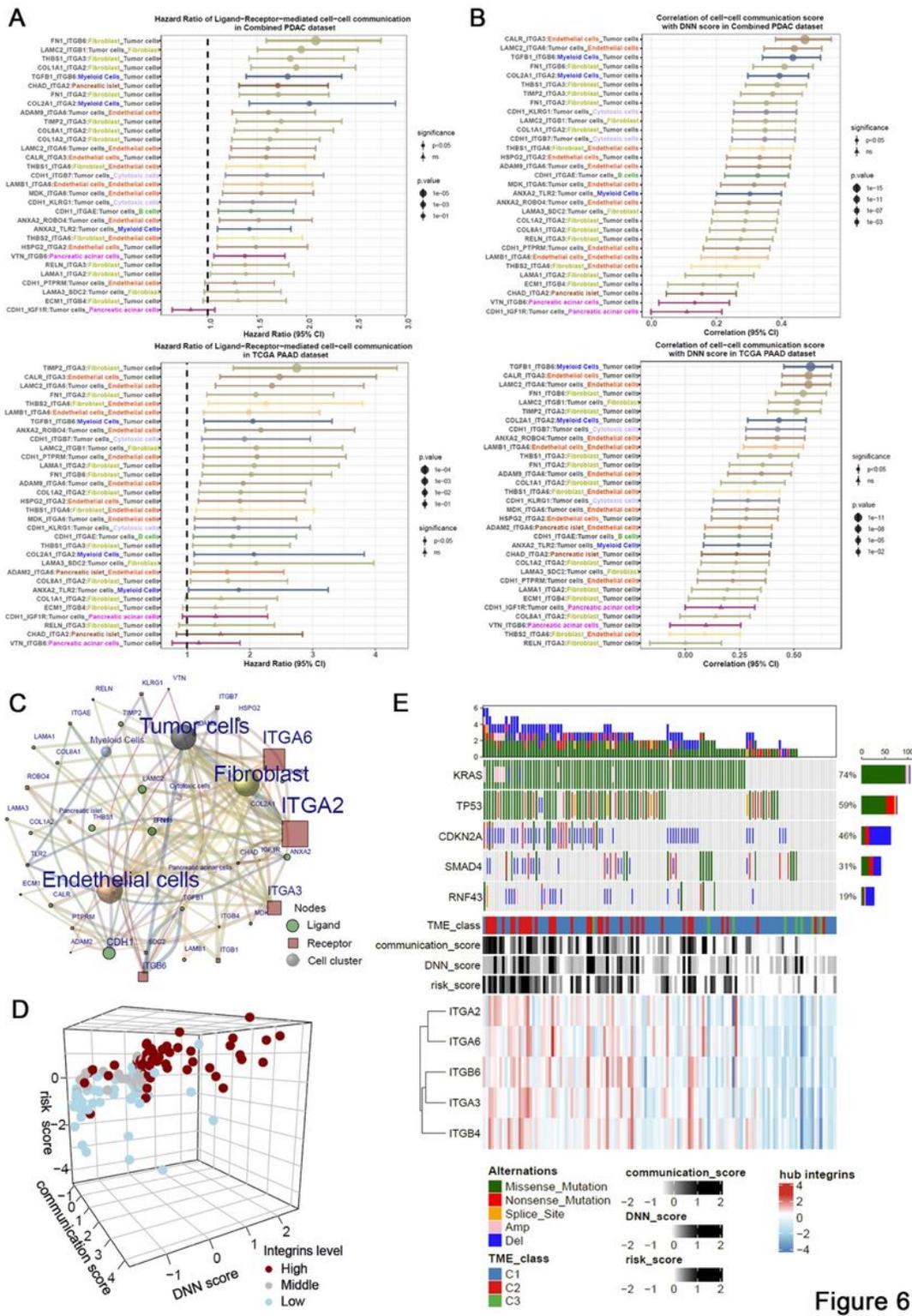


Figure 6

Integrins are key mediators critical for tumor cell-fibroblast and tumor cell-endothelial cell communications (A) The hazard ratios of cell-cell communication scores of PDAC patients in both combined PDAC dataset and TCGA dataset, analyzed by cox regression model. (B) Correlation of cell-cell communication score with DNN scores in both combined PDAC and TCGA datasets. Forest plots show the correlations together with the 95% confidence intervals. (C) A sending cell-ligand-receptor-target cell

network shows the key contributors within it. The nodes and edges are weighted by the related total hazard ratio in combined PDAC and TCGA datasets. The result indicates that integrins-mediated tumor cell-fibroblast and tumor cell-endothelial cell communications are dominant in all cell-ligand-receptor-cell linkages. (D) 3D plot shows the relationships with cell-cell communication score, DNN score, risk score and expression levels of integrins. (E) Waterflow plot plus a heatmap indicates the enrichment of genetic variations in pancreatic cancers with high levels of cell-cell communications.

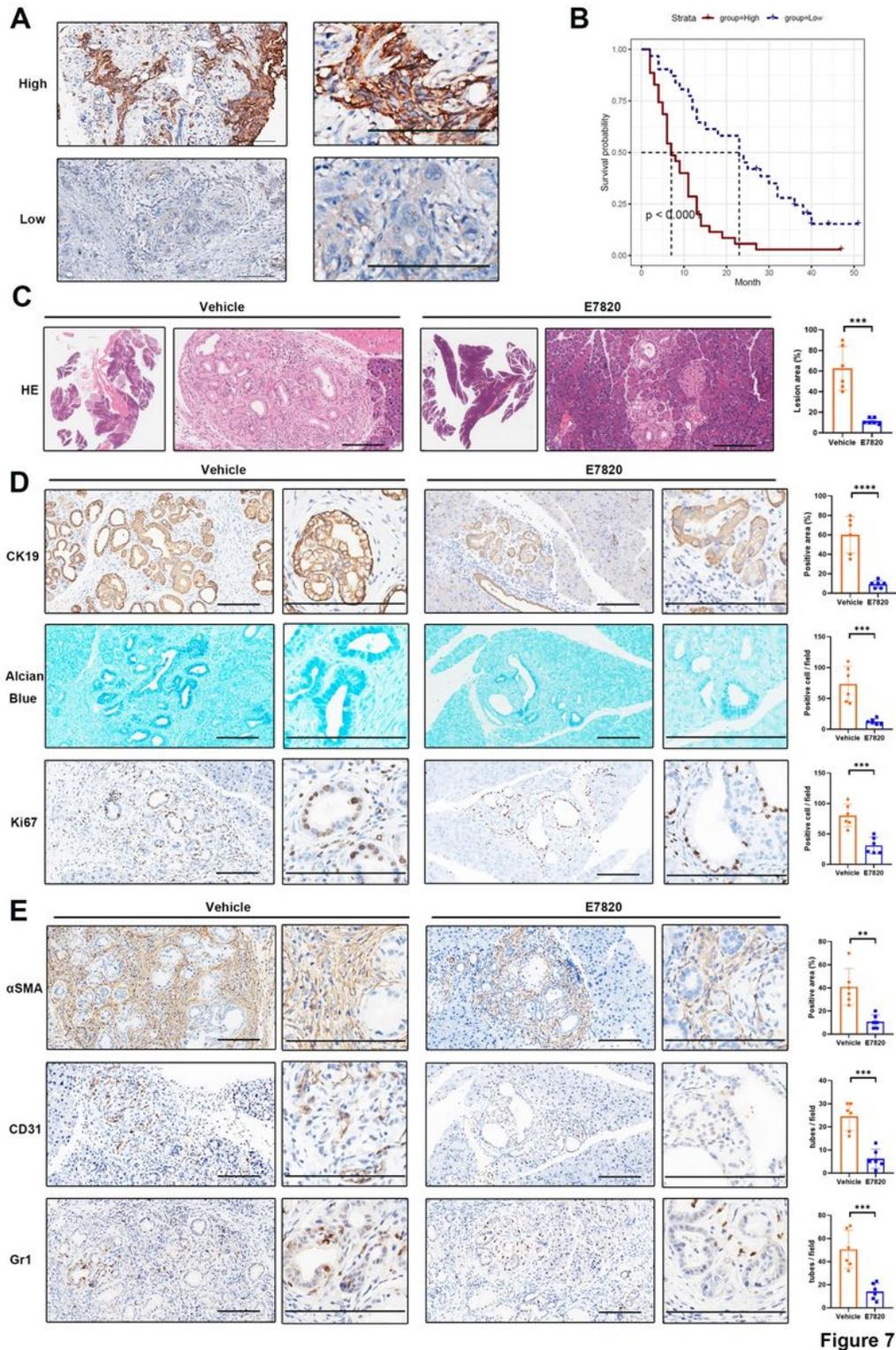


Figure 7

Inhibition of ITGA2 prevents PDAC growth (A) Immunohistochemical analyses of ITGA2 expression in PDAC samples. The representative images of high and low ITGA2 staining in PDAC specimens. (B) Kaplan–Meier plots show the overall survival of PDAC patients expressing high and low levels of ITGA2 protein. (C-E) Pdx1-Cre+, KrasG12D (KC) mice were orally treated with E7820 (100mg/kg bodyweight) once a day, for 14 consecutive days. (C) The H&E staining of mice pancreatic lesions. The percentage of lesion area was statistically compared between vehicle and E7820 treatment groups, t test, ***P<0.001, n=6 fields. (D) Alcian blue staining of pancreas tissues. And the representative images and statistically analysis results of CK19 and Ki67 immunohistochemical staining slides, t test, ***P<0.001, ****P<0.0001, n=6 fields. (E) Immunohistochemistry analyses of pancreas tissues from vehicle and E7820 treated KC mice by anti- α SMA, anti-CD31 and anti-Gr1 antibodies, t test, **P<0.01, ***P<0.001, n=6 fields. Bar=100 μ m.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigures.docx](#)
- [Supplementarytable1.Cellcomponentmarkergenes.xls](#)
- [Supplementarytable2.CellcomponentmarkergenesOS.xls](#)
- [Supplementarytable3.SurvivalanalysesofmoduleEigengene.xls](#)
- [Supplementarytable4.MEMGsandhubgenes.xls](#)
- [Supplementarytable5.MEMGgenesandweights.xls](#)
- [Supplementarytable6.Hubgenesenrichementanalysis.xls](#)
- [Supplementarytable7.Keycellcellcommunications.xls](#)