Reveal the Heterogeneity in the Tumor Microenvironment of Pancreatic Cancer and Analyze the Differences in Prognosis and Immunotherapy Responses of Distinct Immune Subtypes

Xiaoqin Wang[1*†], Lifang Li[2†], Yang Yang[3†], Linlin Fan[1], Ying Ma[1] and Feifei Mao[4*]

[1] Department of Clinical Laboratory, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2] Emergency Department, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [3] Department of Hepatobiliary and Pancreatic Surgery, The First People's Hospital of Changzhou, Changzhou, China, [4] Tongji University Cancer Center, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China

Purpose: The current clinical classification of pancreatic ductal adenocarcinoma (PDAC) cannot well predict the patient's possible response to the treatment plan, nor can it predict the patient's prognosis. We use the gene expression patterns of PDAC patients to reveal the heterogeneity of the tumor microenvironment of pancreatic cancer and analyze the differences in the prognosis and immunotherapy response of different immune subtypes.

Methods: Firstly, use ICGC's PACA-AU PDAC expression profile data, combined with the ssGSEA algorithm, to analyze the immune enrichment of the patient's tumor microenvironment. Subsequently, the spectral clustering algorithm was used to extract different classifications, the PDAC cohort was divided into four subtypes, and the correlation between immune subtypes and clinical characteristics and survival prognosis was established. The patient's risk index is obtained through the prognostic prediction model, and the correlation between the risk index and immune cells is prompted.

Results: We can divide the PDAC cohort into four subtypes: immune cell and stromal cell enrichment (Immune-enrich-Stroma), non-immune enrichment but stromal cell enrichment (Non-immune-Stroma), immune-enriched Collective but non-matrix enrichment (Immune-enrich-non-Stroma) and non-immune enrichment and non-stromal cell enrichment (Non-immune-non-Stroma). The five-year survival rate of immune-enrich-Stroma and non-immune-Stroma of PACA-CA is quite different. TCGA-PAAD's immune-enrich-Stroma and immune-enrich-non-Stroma groups have a large difference in productivity in one year. The results of the correlation analysis between the risk index and immune cells show that the patient's disease risk is significantly related to epithelial cells, megakaryocyte-erythroid progenitor (MEP), and Th2 cells.

Conclusion: The tumor gene expression characteristics of pancreatic cancer patients are related to immune response, leading to morphologically recognizable PDAC subtypes with prognostic/predictive significance.

Keywords: pancreatic cancer, immune subtypes, heterogeneity, prognosis, microenvironment

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the lethal malignant neoplasms around the world (1–4), and its genetic and phenotypic heterogeneity makes generally effective therapies ineffective (5–9). The salient feature of pancreatic cancer is that it has an immunosuppressive microenvironment, the prognosis of patients is poor, and most of the patients' tumors will metastasize (10, 11). Research on the immune microenvironment of pancreatic cancer may help improve the therapeutic effect (12, 13). By detecting the expression of anti-tumor immune genes, markers that can predict patient response to treatment have been screened (14). In addition, mutations in genes such as PIK3CA, FGFR3, and TP53 have been shown to be related to tumor immune infiltration (15–18). Although we have a better understanding of the molecular mechanism and genetic background of pancreatic cancer, the 5-year survival rate for this disease is approximately 10% in the USA (19). Several phase III clinical trials that are effective for other cancers have not worked well in pancreatic cancer patients (7). Tumor heterogeneity and host differences will affect the characteristics of its tumor microenvironment. It is necessary to identify new biomarkers and explore new treatment approaches to provide more and more effective references for overcoming the immunosuppressive mechanism in the pancreatic cancer microenvironment.

The immune microenvironment plays an important role in tumor cell invasion and pancreatic cancer progression (20), and immune expression characteristics may affect the degree of inhibition of cancer cells. Invasive PDAC has epithelial-to-mesenchymal transition (EMT)-like characteristics and has been shown to be a poor prognostic factor for pancreatic cancer (21). The immune microenvironment with EMT-like tumors is conducive to tumor growth. Researchers reported on three subtypes of pancreatic cancer: classic, quasi-mesenchymal, and exocrine, and clarified the genetic markers of different subtypes, which may help to carry out more targeted treatments for patients (22). Other researchers have identified two tumor-specific subtypes based on gene expression: basal-like subtype and classical subtype (23). The classic subtype is consistent with the subtype described by Collisson et al. Tumor subtypes defined by exocrine-like genes have not been validated in its data set, and may be related to tissue contamination. Recently, researchers classified pancreatic cancer into four subtypes based on genomic studies— squamous cells, pancreatic progenitor cells, immunogenicity and abnormally differentiated endocrine and exocrine-identified the differences between pancreatic cancer subtypes and provided Different subtypes of treatment options (22, 24). Among them, squamous cells, pancreatic progenitor cells, and

abnormally differentiated endocrine and exocrine (ADEX) subtypes correspond to the quasi-mesenchymal, classical, and exocrine-like subtypes reported by Collisson et al. (22). Recently, studies have shown that ADEX and immunogenic subtypes are related to the lower purity of the sample (24, 25). Although researchers have basically determined the characteristics of some pancreatic cancer subtypes, research conclusions about exocrine differentiation or immunogenic subtypes are still inconsistent.

Therefore, we aim to redefine the subtypes of PDAC and clarify its immune expression patterns, provide useful clues for exploring the different immunosuppressive mechanisms of PDAC, and use it in the stratification of patient clinical trials, so as to provide patients with PDAC more precise treatment.

RESULTS

Classification of Distinct Tumor Microenvironment Subtypes

Single sample gene set enrichment analysis (ssGSEA) defines an enrichment score to indicate the absolute enrichment degree of the gene set in each sample in a given data set. The enrichment score of each immune category can be found in the R package GSVA In the realization (26). Firstly, ssGSEA algorithm (27) was utilized to analyze the expression profiling database of the PACA-AU pancreatic cancer in the International Cancer Genome Consortium (ICGC). We obtained the immune enrichment of the tumor microenvironment of each patient's tumor tissue. And the tumor microenvironment-related genes come from the following references (**Table 1**).

Subsequently, we apply the spectral clustering algorithm to extract different categories based on the ssGSEA scores (**Figure 1A**). Meanwhile, we used t-distributed stochastic neighbor embedding (tSNE) to show the groups (**Figure 1B**), and revealed an immune-enriched subtype (Immune-enrich) exists in the cohort, and the rest are of the Non-immune type, that is, less immune infiltration (**Figure 1C**). In addition, even in the presence of a large population of immune cells, stromal cells also play vital roles in tumor immunity evasion. Therefore, we further dissected the enrichment of stromal cells in the patient's gene expression profile. Also using ssGSEA analysis, we found that the cohort had characteristics of activated stromal response (**Figure 1C**). Based on the above classification, we can divide the pancreatic cancer cohort into four subtypes: immune cell and stromal cell enrichment (Immune-enrich-Stroma), non-immune enrichment but stromal cell enrichment (Non-immune-Stroma), Immune enrichment but non-matrix enrichment (Immune-enrich-non-Stroma) and non-immune enrichment and non-stromal cell enrichment (Non-immune-non-Stroma). Immune-

TABLE 1 | Immune-related gene signatures and their references.

Signature name	Reference
Immune enrichment score	Yoshihara et al. Nat Commun. 2013 (28)
6-gene IFN-γsignature	Chow et al. J Clin Oncol. 2016 (suppl) (29)
Activated stroma	Moffitt et al. Nat Genet. 2015 (30)
Immune cell subsets	Cancer Genome Atlas Network. Cell. 2015 (31)
T cells	Bindea et al. Immunity. 2013 (32)
CD8 Tcells	Bindea et al. Immunity. 2013 (32)
T. NK. metagene	Alistar et al. Genome Med. 2014 (33)
B-cell cluster	Iglesia et al. Clin Cancer Res. 2014 (34)
Macrophages	Bindea et al. Immunity. 2013 (32)
Cytotoxic cells	Bindea et al. Immunity. 2013 (32)
Immunophenoscore	Charoentong et al. Cell Rep. 2017 (35)
T cell-inflamed GEP	Cristescu et al. Science. 2018 (36)
Expanded immune signature	Ayers et al. J Clin Invest. 2017 (37)
TGF-β-associated ECM	Chakravarthy et al. Nat Commun. 2018 (38)
MDSC	Yaddanapudi et al. Cancer Immunol Res. 2016 (39)
CAF	Calon et al. Cancer Cell. 2012 (40)
TAM M2/M1	Beyer et al. PLoS One. 2012 (41)
CD8 T cell exhaustion	Giordano et al. EMBO J. 2015 (42)
T cell exhaustion early/late stage	Philip et al. Nature. 2017 (43)
Nivolumab responsive	Riaz et al. Cell. 2017 (44)

enrich-Stroma subtypes mainly enrich tumor immune-related molecular signatures, including T cell-inflamed GEP, Expanded immune signature, Immunophenoscore, Immune enrichment score, CD8 T cell exhaustion, myeloid-derived suppressor cells (MDSC), cytotoxic cells, Immune cell subset, etc. At the same time, it also enriches PD1 and stroma related signatures, including anti-PD-1 resistant, nivolumab responsive and normal stroma. The signatures of Non-immune-Stroma subtypes mainly include anti-PD-1 resistant, activated stroma, CAF-stimulated, and normal stroma, while its immune-related family features are very low. Immune-enrich-non-Stroma subtypes mainly enrich tumor immune-related signatures, including T cell-inflamed GEP, Expanded immune signature and cytotoxic cells, etc., while its stromal signatures expression is very low. Non-immune-non-Stroma subtypes, as the name suggests, are rarely enriched in tumor immunity and stromal signatures.

## Comparison of the Striking Differences in the Immune Microenvironment of the Four Subtypes

As follow, the four subtypes have the following immune differences (**Figure 2A**). Patients with immuno-enriched subtypes (**Figure 2A** red and light blue boxes) showed significant enrichment in the characteristics of recognizing immune cells or immune responses (all P <0.05). We further compared the difference in gene expression between immune-enriched and non-immune-enriched patients, mainly using the limma algorithm, and P<0.05 as the standard of significant difference (**Table S1**). At the same time, the significantly different genes of stromal cell enrichment and non-stromal enrichment was compared (**Table S2**).In order to verify the accuracy and consistency of the analysis method, we use the same strategy to predict the enrichment of other data. The first 50 genes that are differentially up-regulated are selected to construct a gene set, and the ssGSEA algorithm is used to predict the

enrichment of other data. In addition, select significantly different immune activity or immune cell-related genes to verify their enrichment. The analysis results show that the GSE124231 data set (**Figure 2B**, n=48), the GSE131050 data set (**Figure 2C**, n=66), the PACA-CA data set (**Figure 2D**, n=234) and the TCGA-PAAD database (**Figure 2E**, n = 177) can be divided into immune enrichment and stromal cell enrichment groups. According to the constructed gene set, samples of different data sets can be divided into immune-enrich-Stroma, immune-enrich-non-stroma, non-immune-stroma and non-immune-non-stroma types. And immune enrichment type samples are mainly enriched for immune-related signatures, such as immune enrichment score, immunophenoscore, Immune cell subsets, etc. Stromal cell enrichment types mainly enrich stroma-related signatures, such as normal stroma, activated stromanivolumab responsive, etc. The above results show that the accuracy and consistency of our classification and research methods are trustworthy.

Four Immune Subtypes Are Related to Clinical Characteristics and Survival Prognosis

Based on the previous results, we have divided patients into 4 different subtypes of immune enrichment and stromal cell enrichment. Therefore, we need to further compare the clinical characteristics of different types and try to explore the relationship between each type and patient survival prognosis. Firstly, we sequentially compared the clinical information between different subtypes in the PACA-AU, PACA-CA and TCGA-PAAD cohorts. Statistics showed that there were significant differences among subtypes in the PACA-AU cohort, which included donor_sex, donor_vital_status, donor_relapse_type, donor_age_at_diagnosis and enrollment, donor_survival_time, donor_interval_up, donor_interval_up, donor_interval_up (**Table S3**). In the PACA-CAcohort, clinical markers such as donor_age_at_diagnosis and enrollment, donor_age_at_last_followup, donor_survival_time, donor_interval_of_last_followup are significantly different among subgroups (**Table S4**). Age_at_initial_pathologic_diagnosis, family_history_of_cancer (%), history_of_chronic_pancreatitis (%), history_of_diabetes (%) and other clinical indicators were significantly different among 4 subsets in the TCGA-PAAD cohort (**Table S5**).

Then, we successively explored the relationship between different subgroups in the cohort and the survival prognosis of patients. In the PACA-AU cohort, the 1-year (**Figure 3A**) and 5-year (**Figure 3C**) survival rates between different subtypes are significantly different (p.value <0.05), and the survival rate of the Immune_enrich_Stroma subgroup is higher than that of the other three groups. However, the difference in 3-year survival rates between patient groups was not significant (**Figure 3B**). Finally, we compared the survival rates of all PACA-AU patients (8 years) and found that the survival rates of different subgroups are still significantly different (p.value <0.05) (**Figure 3D**). Similarly, we compared the survival rates of patients in the PACA-CA cohort for 1 year (**Figure 3E**), 3 years (**Figure 3F**), 5 years (**Figure 3G**) and all patients (12 years) (**Figure 3H**) in detail, and found There are no significant differences between
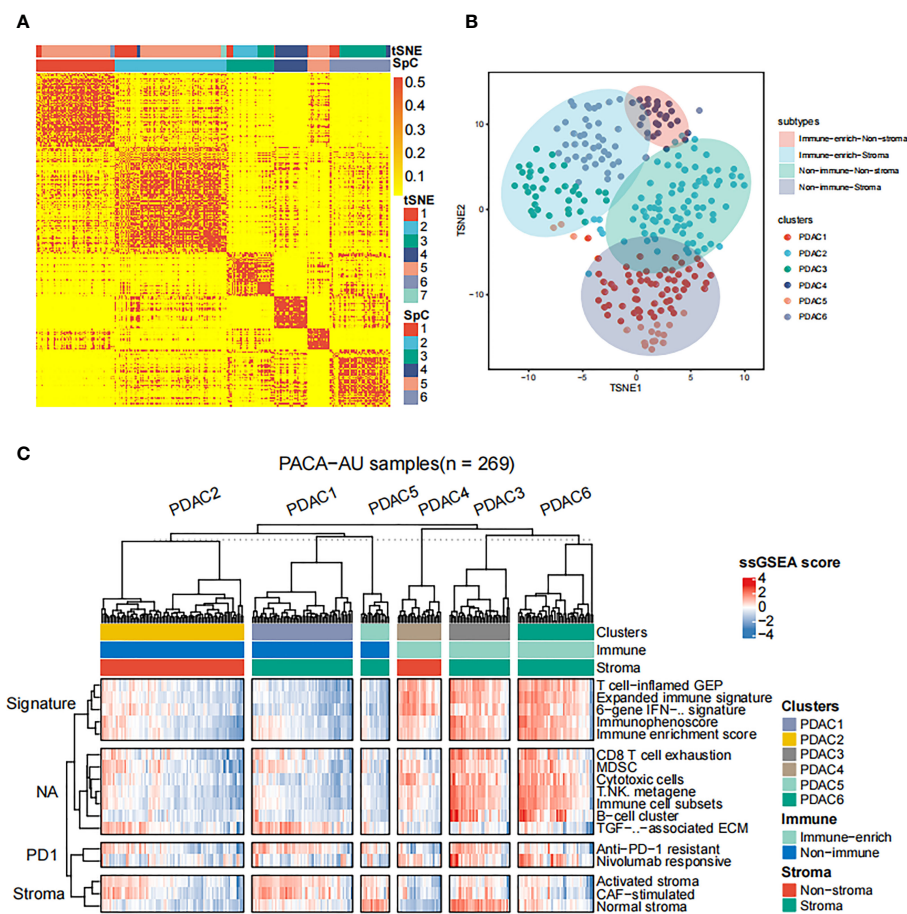
FIGURE 1 | Classification of distinct tumor microenvironment subtypes **(A)** Spectral classification of tumor microenvironment in PACA-AU alignment. This plot shows a heat map of the ssGSEA score, estimated using the gene set from the ICGC database. Based on tSNE cluster analysis, 7 subgroups were obtained, namely PDAC1, PDAC2, PDAC3, PDAC4, PDAC5, PDAC6, PDAC7. Based on Spectral classification, 6 subgroups were obtained, namely PDAC1, PDAC2, PDAC3, PDAC4, PDAC5, PDAC6. **(B)** tSNE classification of tumor microenvironment in PACA-AU cohort. **(C)** This figure shows the 4 immune subtypes of the PACA-AU cohort based on ssGSEA analysis and the main signatures of each subtype.

different subgroups. It is worth mentioning that there is a relatively large difference in the five-year survival rate between the immune-enrich-stroma and non-immune-stroma groups of PACA-CA (**Figure 3I**). The analysis results of the TCGA-PAAD cohort showed that the survival rates of patients in different subgroups were 1 year (**Figure 3J**), 3 years (**Figure 3K**), 5 years (**Figure 3L**) and all patients (8 years) (**Figure 3M**). It was found that there were no significant differences between the different subgroups. However, the one-year survival rate difference between immune-enrich-stroma and immune-enrich-non-stroma groups is relatively large (**Figure 3N**). In general, the classification of the PACA-AU cohort can provide an important reference for their clinical survival prognosis.

Prognostic Prediction Model Based on Signatures of Tumor Microenvironment

Since the subtype classification in the PACA-AU cohort has a strong correlation with survival prognosis, we use PACA-AU data as training data, and PACA-CA and TCGA-PAAD as test data to construct a prognostic prediction model. Firstly, PACA-AU data is treated as training data for parameter training of prediction models and selection of related gene sets. PACA-CA and TCGA-PAAD are regarded as testing data to test the parameters given by the training set and the predictive ability of the gene set. Then, use the cox regression algorithm to initially screen the genes that are significantly related to the patient's overall survival ($P<0.05$), and use the LASSO algorithm to further screen these genes. In the end, the best gene panel is obtained, and the forest diagram of the multivariate COX regression model is drawn (**Figure 4A**). In detail, those genes are KRT6C, PRR11, LTC4S, FGG, SERPINB3, CACNA2D3, FLT3LG, FDCSP, C5ORF46, FAM107A, CCL19, BLK, SLAMF1 and their multiple regression coefficients are 0.58, 0.89, -0.68, 0.69, 0.27, -0.56, -0.83, -0.54, 0.73, 0.97, -0.42, 0.62, 0.78. Subsequently, based on the expression level and multiple regression coefficients of gene panel obtained above, calculate their risk score. We further divided patients into high-risk groups and low-risk groups based on the risk index of the sample. Kaplan-Meier survival analysis was performed and showed in survival curve. There is a
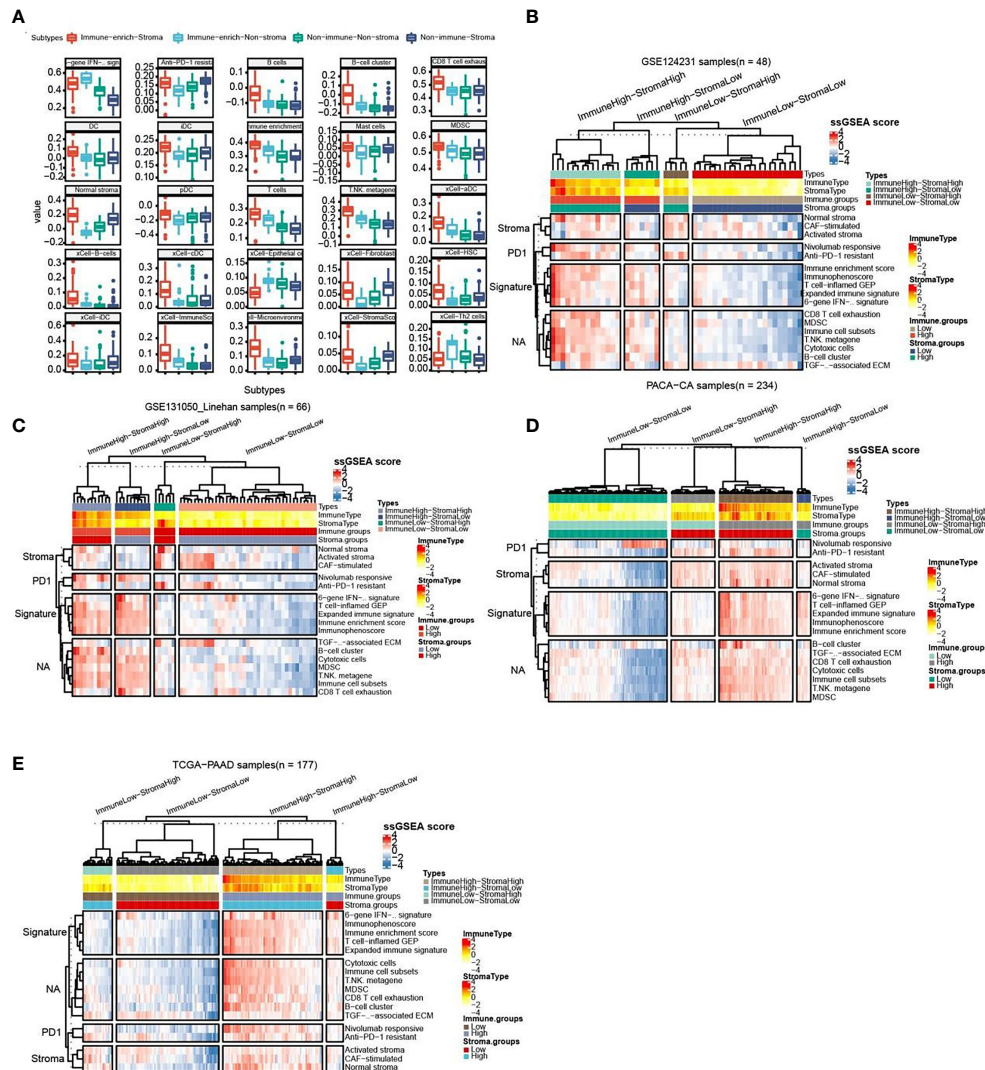
FIGURE 2 | Comparison of the striking differences in the immune microenvironment of the four subtypes. **(A)** Comparison of the striking differences in the immune microenvironment of the four subtypes. Red represents immune-enrich-stroma subtype, Light_blue represents immune-enrich-non-stroma subtype, Green represents non-immune-non-stroma subtype, and Navy blue represents non-immune-stroma subtype. **(B)** Immune-enrich-Stroma, Immune-enrich-non-Stroma, Non-immune-Stroma and Non-immune-non-Stroma types in the GSE124231 data set (n=48). **(C)** Four types in the GSE131050 data set (n=66). **(D)** Four types in the PACA-CA data set (n=234). **(E)** Four types in the TCGA-PAAD database (n = 177).

significant difference in survival probability between the high-risk group and the low-risk group in PACA-AU cohort (p <0.05) (**Figure 4B**). At the same time, we drew the ROC curve of the one-year, three-year, and five-year survival period of the patients in the training set based on the risk index (**Figure 4C**). However, there was no significant difference in the survival probability between the high-risk group and the low-risk group in the TCGA-PAAD testing set.

At the same time, we drew the ROC curve of patient survival in PACA-CA (**Figure 4E**) cohorts based on the risk index. The ROC curve of the prediction model of the PACA-CA training set shows that the prediction model is relatively ideal, and the prediction of the 1-year survival period is slightly better than the 3-year and 5-year

survival periods. In addition, the prediction effect of the PACA-CA testing set (**Figure 4D**) is slightly inferior to that of the PACA-AU training set, except for the 5-year survival period of the TCGA-PAAD testing set. Overall, the prognosis prediction model can better predict the grouping of patients based on the risk index, which provides guidance for the prognosis prediction of patients.

Immune Cells Related to Risk Index
Based on the previous results, we want to know which immune cells are specifically related to the risk index of the PACA-AU cohort. Therefore, we used the sample risk index to make further correlation analysis with the expression of various immune cells and immune molecules. The results showed that the patient's risk
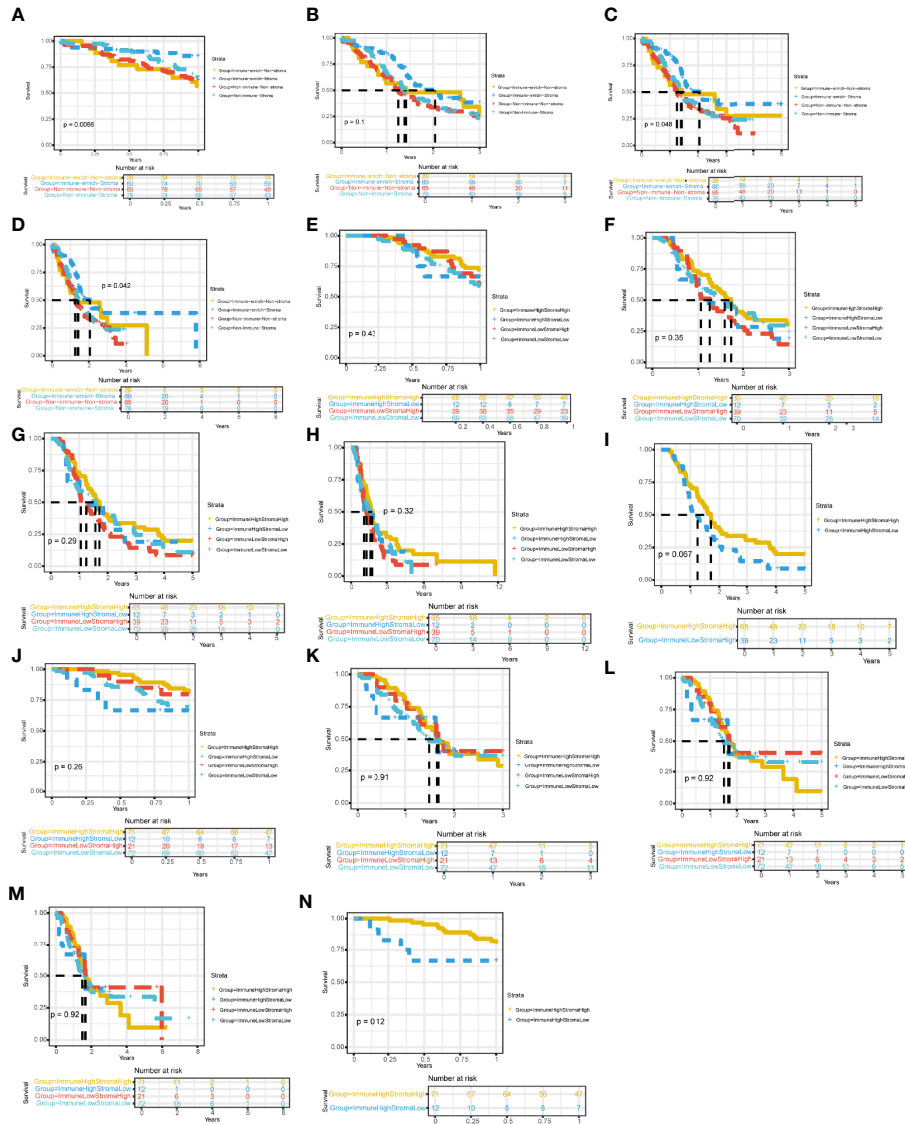
FIGURE 3 | Four immune subtypes are related to clinical characteristics and survival prognosis Comparison of survival rates between subgroups in different cohorts **(A)** Comparison of 1-year survival rate of PACA-AU cohort. **(B)** Comparison of 3-year survival rate of PACA-AU cohort. **(C)** Comparison of 5-year survival rate of PACA-AU cohort. **(D)** Comparison of survival rates of all PACA-AU cohort. **(E)** Comparison of 1-year survival rate of PACA-CA cohort. **(F)** Comparison of 3-year survival rate of PACA-CA cohort. **(G)** Comparison of 5-year survival rate of PACA-CA cohort. **(H)** Comparison of survival rates of all PACA-CA cohort. **(I)** Comparison of survival rates of the Immune-enrich-Stroma and Non-immune-Stromasubtypes. **(J)** Comparison of 1-year survival rate of TCGA-PAAD cohort. **(K)** Comparison of 3-year survival rate of TCGA-PAAD cohort. **(L)** Comparison of 5-year survival rate of TCGA-PAAD cohort. **(M)** Comparison of survival rates of all TCGA-PAAD cohort. **(N)** Comparison of survival rates of the Immune-enrich-Stroma and Immune-enrich-non-Stroma subtypes.

index and epithelial cells, megakaryocyte-erythroid progenitor (MEP), and Th2 cells showed a positive correlation with p<0.01. In addition, T cells, NK cells, memory B-cells, mast cells and other immune cells have a negative correlation with p <0.01 (**Figure 5**).

DISCUSSION

In this study, we used the ssGSEA algorithm to calculate the ssGSEA scores of PACA-AU pancreatic cancer patients, and

then combined the Spectral clustering algorithm to extract the 4 subtypes in the cohort. We further compared the differences in the immune microenvironment of the four subtypes, and screened the immune enrichment and stromal enrichment molecular markers. Genes with significant differences are mostly related to immunity in (**Table S1**). For example, changes in the expression of PTPRCAP affect the survival rate of cancer patients (45), and the single nucleotide polymorphism (SNP) of PTPRCAP is associated with the susceptibility of gastric cancer (46). Natural killer cell granule protein 7 (NKG7) is
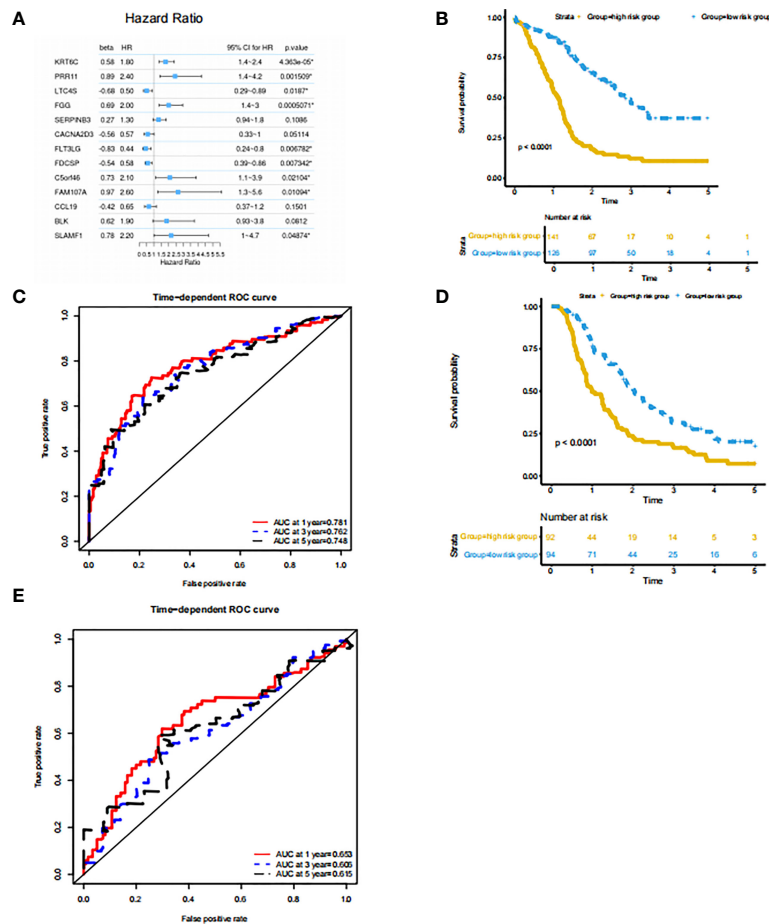
FIGURE 4 | Prognostic prediction model based on signatures of tumor microenvironment **(A)** features: significant factor name; multi_beta: Cox multiple regression coefficient; multi_HR: Cox multiple regression risk ratio; multi 95% CI for HR: Cox multiple regression risk ratio 95% confidence interval; Forest diagram: horizontal line shows the confidence interval interval, and the dot represents the hazard ratio; multi_p.value: Cox multiple regression proportional hazard hypothesis test P value. **(B)** Survival curve of the high and low risk groups in the training set. The horizontal axis represents time (unit: day), the vertical axis represents survival rate. A flat curve represents a high survival rate or a longer survival period, and a steep curve represents a low survival rate or a shorter survival period. **(C)** ROC curve of the training set prediction model. The horizontal axis is the false positive rate FP, and the vertical axis is the true positive rate TP. The legend in the upper left corner corresponds to the AUC value of the ROC curve for different survival periods. **(D)** Survival curves of the high- and low-risk groups in the PACA-CA testing set. **(E)** ROC curve of PACA-CA test set prediction model. The horizontal axis is the false positive rate FP, and the vertical axis is the true positive rate TP. The legend in the upper left corner corresponds to the AUC value of the ROC curve for different survival periods. The horizontal axis is the false positive rate (FP), and the vertical axis is the true positive rate (TP). The legend in the upper left corner corresponds to the AUC value of the ROC curve for different survival periods.

related to inflammatory diseases (47), and its lack will result in a significant reduction in IFN-γ produced by T cells and NK cells. In addition, NKG7 is related to the cytotoxic degranulation of CD8+ T cells (48). Researchers have discovered that CD96 can serve as a new immune checkpoint receptor target for T cells and natural killer cells (49). Similarly, we observed the top genes and their stromal functions in (**Table S2**). For example, Slits3 is expressed in primary bone marrow stromal and bone marrow-derived endothelial cells and stromal cell lines, and plays a role in *in vitro* migration and *in vivo* homing of hematopoietic stem and progenitor cells (50). SPARC is a stromal cell protein, which can be produced by cells associated with tumor stromal cells and has high expression levels in many cancers. It plays an important role in the fibroproliferative reaction of tumors (51).

Using the same research method, it was verified in the GSE124231 (n=48), GSE131050_Linahan (n=66), PACA-CA (n=234), TCGA-PAAD (n=177) cohorts, and the typing was accurate in different cohorts. Further compare the clinical information of patients in the cohort, and in-depth exploration of the difference in survival of patients with different subgroups. We found that in the PACA-AU cohort, the 1-year, 5-year, and 8-year survival times of different subsets patients were significantly correlated. Next, cox regression combined with Lasso algorithm was performed to construct a multivariate COX model. Calculate the patient's risk index based on gene expression level and multiple regression coefficients, and divide the patients into high-risk groups and low-risk groups based on the risk index. Interestingly, the PACA_AU and PACA-CA risk
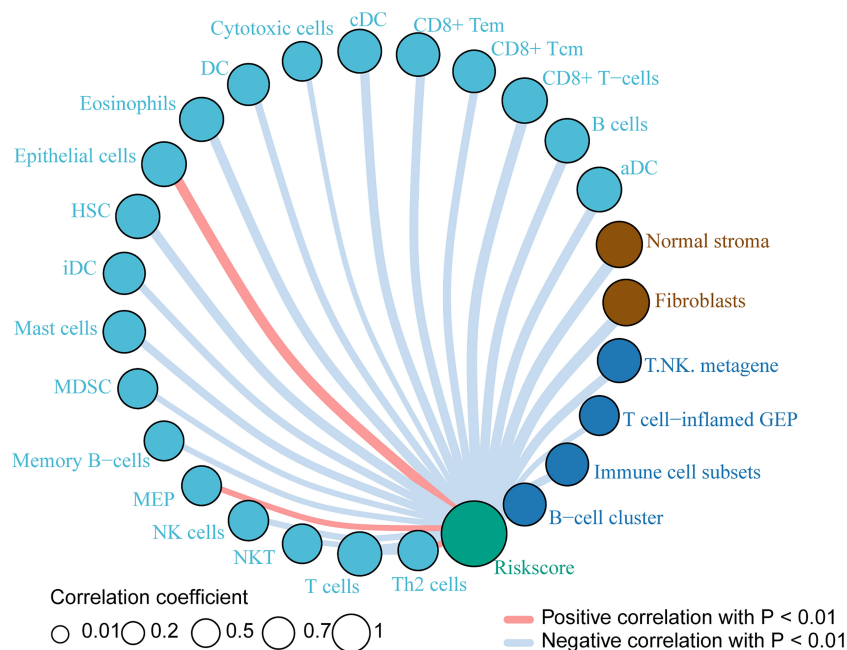
FIGURE 5 | Immune cells related to risk index Immune cells associated with the risk index of PACA-AU patients. The red line indicates a positive correlation between the risk index and immune cells, and the gray line indicates a negative correlation between the risk index and immune cells. The size of the circle indicates different correlation coefficients, and the larger the area of the circle, the larger the correlation coefficient.

indexes are significantly correlated with the survival level of patients.

In PADA-AC and TCGA-PAAD, the survival time difference between different immune subgroups is not significant. Only the five-year survival of immune-enrich-stroma and non-immune-stroma group in PACA-CA cohort and the one-year survival of immune-enrich-Stroma and immune-enrich-non-Stroma group in TCGA-PAAD cohort are relatively large. On the one hand, the cohort clustering algorithm may not cover all patients in the cohort, on the other hand, it may also be because the cohort samples are not large enough, and the representativeness of the statistical results needs to be further improved.

We initially explored the types of immune cells related to the risk index, and we identified immune cells that are positively and negatively related to the risk index. This research lays the foundation for the subsequent in-depth exploration of the correlation mechanism between immune cells and patient disease risk. However, only analyzing the types of immune cells is insufficient for the study of the mechanism. In the later stage, we will conduct more in-depth analysis and verification of important immune cells and their molecular signatures.

METHODS

Project and Sample

Dataset of 461 PACA-AU donors were downloaded from ICGC database (https://dcc.icgc.org/projects/PACA-AU) with detailed clinical information. The independent datasets used for

verification come from GSE124231, GSE131050_Linehan, PACA-CA and TCGA-PAAD projects, including 48, 66, 234 and 177 donors respectively. Moreover, patients in the PACA-CA and TCGA-PAAD cohorts had detailed clinical information.

Bioinformatics Analysis

1) ssGSEA algorithm: Use the R package "GSVA" and use ssGSEA to explore the PACA-AU pancreatic cancer expression profile data of the ICGC database, and analyze the immune enrichment of each patient's tumor microenvironment. Additionally, the gene expression of all samples were took as the input and ssGSEA algorithms were occupied to determine the proportion of the various immune cells of all PDAC samples. The immune gene signatures were listed in the **Table 1**. According to the immune enrichment status of PACA-AU samples, they are divided into immune cells and stromal cell enriched (immune-enrich-stroma), non-immune enrichment but stromal cell enrichment (non-immune-stroma), and immune-enriched but Non-matrix enrichment (immune-enrich-non-stroma) and non-immune enrichment and non-stromal cell enrichment (non-immune-non-stroma). According to the ssGSEA score obtained by each sample, the Spectral clustering algorithm is used to extract different classifications. In addition, the R package "limma" was used to analyze immuno-enriched and non-immune-enriched patients, as well as the significantly different genes of stromal cell enrichment and non-matrix enrichment, and P<0.05 was taken as the significant difference.

2) The unsupervised clustering of the data set was performed mainly based on tSNE which embedded in t-distributed random

neighborhoods (45). In this study, we use tSNE to show the different subgroups of the PACA-AU cohort.

3) We performed Kaplan-Meier survival analysis on the samples and plotted survival curves. Survival analysis divided the samples into high-index groups and low-index groups based on the median. Data visualization is mainly done in the R environment (version 4.1.0). Kaplan-Meier survival analysis relies on the use of the "survival" package. The ROC curve is drawn based on the 'survivalROC' package.

4) Prognosis prediction model establishment process: a). Use the training set to perform unit cox regression on each gene to initially screen disease-related genes; b). After obtaining all cox significant genes in all units, perform 1000X LASSO regression to calculate the frequency of each gene and rank it; c). According to the sorting result of the previous step, build the gene set incrementally. Use each gene set to perform multiple cox regression to get the contribution of each gene; d). Obtain the optimal gene set according to the gene contribution degree, and perform multiple cox regression analysis on these genes. Finally, we determined the regression coefficient of each gene; e). Calculate the death risk score of each patient through regression coefficients; f). The death risk score model is tested in the training set (comparing the predicted situation with the actual situation); g). The same model is tested in the independent testing set at the beginning (comparison of the predicted situation with the actual situation).

5) Construct the optimal multivariate COX model based on the Lasso algorithm. This analysis uses the LASSO algorithm for gene screening: In the field of statistics and machine learning, Lasso algorithm (least absolute shrinkage and selection operator, also translated as minimum absolute shrinkage and selection operator, lasso algorithm) is a regression analysis method that simultaneously performs feature selection and regularization (mathematics).It aims to enhance the predictive accuracy and interpretability of statistical models. Lasso adopts the linear regression method of L1-regularization, so that the weight of some learned features is 0, so as to achieve the purpose of sparseness, selection of variables, and construction of the best model. The characteristic of LASSO regression is to perform variable selection and regularization while fitting a generalized linear model. Therefore, regardless of whether the target dependent variable (dependent/response variable) is continuous, binary or discrete, it can be modeled by LASSO regression and then predicted.

6) We use the Lasso algorithm (glmnet package) to select the best gene model based on the COX multiple regression model, and finally draw the unit cox regression model forest diagram based on the gene Panel as follows: We calculate the risk score (Risk Score) of each patient based on the expression of the gene Panel and the multiple regression coefficient. The formula is as follows:

$$Riskscore = \sum_{i=1}^{n} \beta i * xi$$

xi represents the expression level of each gene in the Panel, βi is the multivariate COX regression beta value (multi_beta) corresponding to each gene.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

FM and XW conceived this project. XW, LL, LF, and YM collected the data. YY and FM analyzed and interpreted the data. XW and FM performed the statistical analyses and wrote the manuscript. All authors have reviewed the manuscript and approved the final version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.832715/full#supplementary-material

REFERENCES

1. Ansari D, Tingstedt B, Andersson B, Holmquist F, Sturesson C, Williamsson C, et al. Pancreatic Cancer: Yesterday, Today and Tomorrow. *Future Oncol* (2016) 12(16):1929–46. doi: 10.2217/fon-2016-0010

2. Goral V. Pancreatic Cancer: Pathogenesis and Diagnosis. *Asian Pac J Cancer Prev* (2015) 16(14):5619–24. doi: 10.7314/APJCP.2015.16.14.5619

3. Gupta R, Amanam I, Chung V. Current and Future Therapies for Advanced Pancreatic Cancer. *J Surg Oncol* (2017) 116(1):25–34. doi: 10.1002/jso.24623

4. Hidalgo M, Cascinu S, Kleeff J, Labianca R, Lohr JM, Neoptolemos J, et al. Addressing the Challenges of Pancreatic Cancer: Future Directions for Improving Outcomes. *Pancreatology* (2015) 15(1):8–18. doi: 10.1016/j.pan.2014.10.001

5. Schlesinger Y, Yosefov-Levi O, Kolodkin-Gal D, Granit RZ, Peters L, Kalifa R, et al. Single-Cell Transcriptomes of Pancreatic Preinvasive Lesions and Cancer Reveal Acinar Metaplastic Cells' Heterogeneity. *Nat Commun* (2020) 11(1):4516. doi: 10.1038/s41467-020-18207-z

6. Dąbkowski K, Bogacka B, Tarnowski M, Starzyńska T. Pancreatic Cancer Microenvironment. *Pol Merkur Lekarski* (2016) 41((246):296–302.

7. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch Repair Deficiency Predicts Response of Solid Tumors to PD-1 Blockade. *Science* (2017) 357(6349):409–13. doi: 10.1126/science.aan6733

8. Knudsen ES, Vail P, Balaji U, Ngo H, Botros IW, Makarov V, et al. Stratification of Pancreatic Ductal Adenocarcinoma: Combinatorial Genetic, Stromal, and Immunologic Markers. *Clin Cancer Res* (2017) 23(15):4429–40. doi: 10.1158/1078-0432.CCR-17-0162

9. Dreyer SB, Chang DK, Bailey P, Biankin AV. Pancreatic Cancer Genomes: Implications for Clinical Management and Therapeutic Development. *Clin Cancer Res* (2017) 23(7):1638–46. doi: 10.1158/1078-0432.CCR-16-2411

10. Ren B, Cui M, Yang G, Wang H, Feng M, You L, et al. Tumor Microenvironment Participates in Metastasis of Pancreatic Cancer. *Mol Cancer* (2018) 17(1):108. doi: 10.1186/s12943-018-0858-1

11. Lin QJ, Yang F, Jin C, Fu DL. Current Status and Progress of Pancreatic Cancer in China. *World J Gastroenterol* (2015) 21(26):7988–8003. doi: 10.3748/wjg.v21.i26.7988

12. Chronopoulos A, Robinson B, Sarper M, Cortes E, Auernheimer V, Lachowski D, et al. ATRA Mechanically Reprograms Pancreatic Stellate Cells to Suppress Matrix Remodelling and Inhibit Cancer Cell Invasion. *Nat Commun* (2016) 7:12630. doi: 10.1038/ncomms12630

13. Sunami Y, Kleeff J. Immunotherapy of Pancreatic Cancer. *Prog Mol Biol Transl Sci* (2019) 164:189–216. doi: 10.1016/bs.pmbts.2019.03.006

14. Yamazaki K, Masugi Y, Effendi K, Tsujikawa H, Hiraoka N, Kitago M, et al. Upregulated SMAD3 Promotes Epithelial-Mesenchymal Transition and Predicts Poor Prognosis in Pancreatic Ductal Adenocarcinoma. *Lab Invest* (2014) 94(6):683–91. doi: 10.1038/labinvest.2014.53

15. Sivaram N, McLaughlin PA, Han HV, Petrenko O, Jiang YP, Ballou LM, et al. Tumor-Intrinsic PIK3CA Represses Tumor Immunogenecity in a Model of Pancreatic Cancer. *J Clin Invest* (2019) 129(8):3264–76. doi: 10.1172/JCI123540

16. Siemers NO, Holloway JL, Chang H, Chasalow SD, Ross-MacDonald PB, Voliva CF, et al. Genome-Wide Association Analysis Identifies Genetic Correlates of Immune Infiltrates in Solid Tumors. *PloS One* (2017) 12(7): e0179726. doi: 10.1371/journal.pone.0179726

17. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic Analyses Identify Molecular Subtypes of Pancreatic Cancer. *Nature* (2016) 531(7592):47–52. doi: 10.1038/nature16965

18. Hashimoto S, Furukawa S, Hashimoto A, Tsutaho A, Fukao A, Sakamura Y, et al. *ARF6 and AMAP1 are Major Targets of KRAS and TP53 Mutations to Promote Invasion, PD-L1 Dynamics, and Immune Evasion of Pancreatic Cancer. Proc Natl Acad Sci USA* (2019) 116(35):17450–9. doi: 10.1073/pnas.1901765116

19. Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic Cancer. *Lancet* (2020) 395((10242):2008–20. doi: 10.1016/S0140-6736(20)30974-0

20. Fabris L, Perugorria MJ, Mertens J, Bjorkstrom NK, Cramer T, Lleo A, et al. The Tumour Microenvironment and Immune Milieu of Cholangiocarcinoma. *Liver Int* (2019) 39(Suppl 1):63–78. doi: 10.1111/liv.14098

21. Wartenberg M, Zlobec I, Perren A, Koelzer VH, Gloor B, Lugli A, et al. Accumulation of FOXP3+T-Cells in the Tumor Microenvironment Is Associated With an Epithelial-Mesenchymal-Transition-Type Tumor Budding Phenotype and Is an Independent Prognostic Factor in Surgically Resected Pancreatic Ductal Adenocarcinoma. *Oncotarget* (2015) 6(6):4190–201. doi: 10.18632/oncotarget.2775

22. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of Pancreatic Ductal Adenocarcinoma and Their Differing Responses to Therapy. *Nat Med* (2011) 17(4):500–3. doi: 10.1038/nm.2344

23. O'Kane GM, Grunwald BT, Jang GH, Masoomian M, Picardo S, Grant RC, et al. GATA6 Expression Distinguishes Classical and Basal-Like Subtypes in Advanced Pancreatic Cancer. *Clin Cancer Res* (2020) 26(18):4901–10. doi: 10.1158/1078-0432.CCR-19-3724

24. Puleo F, Nicolle R, Blum Y, Cros J, Marisa L, Demetter P, et al. Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features. *Gastroenterology* (2018) 155(6):1999–2013 e3. doi: 10.1053/j.gastro.2018.08.033

25. Cancer Genome Atlas Research Network and Electronic address: andrew_aguirre@dfci.harvard.edu; Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* (2017) 32(2):185–203.e13. doi: 10.1016/j.ccell.2017.07.007

26. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA Interference Reveals That Oncogenic KRAS-Driven Cancers Require TBK1. *Nature* (2009) 462((7269):108–12. doi: 10.1038/nature08460

27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. *Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. Proc Natl Acad Sci USA* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

28. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. *Inferring Tumour Purity and Stromal and Immune Cell Admixture From Expression Data*. 2013: Nat Commun. doi: 10.1038/ncomms3612

29. Chow LQM, Haddad R, Gupta S, Mahipal A, Mehra R, Tahara M, et al. Antitumor Activity of Pembrolizumab in Biomarker-Unselected Patients With Recurrent and/or Metastatic Head and Neck Squamous Cell Carcinoma: Results From the Phase Ib KEYNOTE-012 Expansion Cohort. *J Clin Oncol* (2016) 34(32):3838–45. doi: 10.1200/JCO.2016.68.1478

30. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual Microdissection Identifies Distinct Tumor- and Stroma-Specific Subtypes of Pancreatic Ductal Adenocarcinoma. *Nat Genet* (2015) 47 (10):1168–78. doi: 10.1038/ng.3398

31. Cancer Genome Atlas Research Network. *The Molecular Taxonomy of Primary Prostate Cancer. Cell* (2015) 163(4):1011–25. doi: 10.1016/j.cell.2015.10.025

32. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* (2013) 39(4):782–95. doi: 10.1016/j.immuni.2013.10.003

33. Alistar A, Chou JW, Nagalla S, Black MA, D'Agostino RJr., Miller LD. Dual Roles for Immune Metagenes in Breast Cancer Prognosis and Therapy Prediction. *Genome Med* (2014) 6(10):80. doi: 10.1186/s13073-014-0080-8

34. Iglesia MD, Vincent BG, Parker JS, Hoadley KA, Carey LA, Perou CM, et al. Prognostic B-Cell Signatures Using Mrna-Seq in Patients With Subtype-Specific Breast and Ovarian Cancer. *Clin Cancer Res* (2014) 20(14):3818–29. doi: 10.1158/1078-0432.CCR-13-3368

35. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-Cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019

36. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-Tumor Genomic Biomarkers for PD-1 Checkpoint Blockade-Based Immunotherapy. *Science* (2018) 362(6411):eaar3593. doi: 10.1126/science.aar3593

37. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-Gamma-Related Mrna Profile Predicts Clinical Response to PD-1 Blockade. *J Clin Invest* (2017) 127(8):2930–40. doi: 10.1172/JCI91190

38. Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD. TGF-Beta-Associated Extracellular Matrix Genes Link Cancer-Associated Fibroblasts to Immune Evasion and Immunotherapy Failure. *Nat Commun* (2018) 9 (1):4692. doi: 10.1038/s41467-018-06654-8

39. Yaddanapudi K, Rendon BE, Lamont G, Kim EJ, Al Rayyan N, Richie J, et al. MIF is Necessary for Late-Stage Melanoma Patient MDSC Immune Suppression and Differentiation. *Cancer Immunol Res* (2016) 4(2):101–12. doi: 10.1158/2326-6066.CIR-15-0070-T

40. Calon A, Espinet E, Palomo-Ponce S, Tauriello DV, Iglesias M, Cespedes MV, et al. Dependency of Colorectal Cancer on a TGF-Beta-Driven Program in Stromal Cells for Metastasis Initiation. *Cancer Cell* (2012) 22(5):571–84. doi: 10.1016/j.ccr.2012.08.013

41. Beyer M, Mallmann MR, Xue J, Staratschek-Jox A, Vorholt D, Krebs W, et al. High-Resolution Transcriptome of Human Macrophages. *PloS One* (2012) 7 (9):e45466. doi: 10.1371/journal.pone.0045466

42. Giordano M, Henin C, Maurizio J, Imbratta C, Bourdely P, Buferne M, et al. Molecular Profiling of CD8 T Cells in Autochthonous Melanoma Identifies Maf as Driver of Exhaustion. *EMBO J* (2015) 34(15):2042–58. doi: 10.15252/embj.201490786

43. Philip M, Fairchild L, Sun L, Horste EL, Camara S, Shakiba M, et al. Chromatin States Define Tumour-Specific T Cell Dysfunction and Reprogramming. *Nature* (2017) 545(7655):452–6. doi: 10.1038/nature22367

44. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and Microenvironment Evolution During Immunotherapy With Nivolumab. *Cell* (2017) 171(4):934–49. doi: 10.1016/j.cell.2017.09.028

45. van der Maaten L. *Visualizing Data Using T-SNE. J Mach Learn Res* (2008) 1:1–48.

46. Ju H, Lim B Fau - Kim M, Kim M Fau - Kim YS, Kim Ys Fau - Kim WH, Kim Wh Fau - Ihm C, Ihm C Fau - Noh S-M, et al. A Regulatory Polymorphism at

Position -309 in PTPRCAP is Associated With Susceptibility to Diffuse-Type Gastric Cancer and Gene Expression. *Neoplasia* (2009) 11(12):1340–7. doi: 10.1593/neo.91132

47. van Es LA, de Heer E, Vleming LJ, van der Wal A, Mallat M, Bajema I, et al. GMP-17-Positive T-Lymphocytes in Renal Tubules Predict Progression in Early Stages of Iga Nephropathy. *Kidney Int* (2008) 73(12):1426–33. doi: 10.1038/ki.2008.66

48. Ng SS, De Labastida Rivera F, Yan J, Corvino D, Das I, Zhang P, et al. The NK Cell Granule Protein NKG7 Regulates Cytotoxic Granule Exocytosis and Inflammation. *Nat Immunol* (2020) 21(10):1205–18. doi: 10.1038/s41590-020-0758-6

49. Dougall WC, Kurtulus S, Smyth MJ, Anderson AC. TIGIT and CD96: New Checkpoint Receptor Targets for Cancer Immunotherapy. *Immunol Rev* (2017) 276(1):112–20. doi: 10.1111/imr.12518

50. Geutskens SB, Andrews WD, van Stalborch AM, Brussen K, Holtrop-de Haan SE, Parnavelas JG, et al. Control of Human Hematopoietic Stem/Progenitor Cell Migration by the Extracellular Matrix Protein Slit3. *Lab Invest* (2012) 92 (8):1129–39. doi: 10.1038/labinvest.2012.81

51. Framson PE, Sage EH. SPARC and Tumor Growth: Where the Seed Meets the Soil? *J Cell Biochem* (2004) 92(4):679–90. doi: 10.1002/jcb.20091