Integration of scRNA-seq and Bulk-seq to Analyse the Infiltration of Monocytes in Pancreatic Cancer and Establish a Molecular Risk Model

Wenchao Yao
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Xuxu Liu
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Tianming Liu
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Yi Zheng
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Ziang Meng
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Yifei Hao
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Jinzuo Han
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Qiang Wang
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Zhenyi Lv
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Dongbo Xue
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Zhituo Li
Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Yingmei Zhang
Key Laboratory of Hepatosplenic Surgery, Ministry of Education, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Research Article

Keywords: monocyte, prognosis, pancreatic cancer, single-cell sequencing, immune checkpoint blockade

Posted Date: December 29th, 2022
Abstract

Background

Many researches have confirmed that immunotherapy of tumor immune microenvironment is necessary. In pancreatic cancer, monocytes play an important role in poor prognosis, but the mechanism and prognosis prediction methods are unclear.

Methods

CIBERSORT was used to identify cellular immune score and evaluate the effect of each immune cell on prognosis. The gene modules related to monocytes were obtained by weighted correlation network analysis through WGCNA package. Consensus clustering was used to sort prognostic genes. The regression signature was generated by LASSO Cox analysis and verified by Cox analysis. The ssGSEA and TIDE algorithms were used to predict immune status and sensitivity to ICB. Finally, the expression levels of each gene were verified at tissue level and single-cell level.

Results

High infiltration of monocytes suggests poor prognosis of pancreatic cancer. 262 genes were significantly differentially expressed and prognostic after WGCNA analysis and cluster typing. The related 6 genes prognostic signature established by LASSO Cox analysis was verified to be an independent prognostic factor. The high-risk group had high infiltration of monocytes in the immune microenvironment and was more sensitive to ICBs. At the tissue level, all genes were highly expressed in cancer tissues. At the single-cell level, MET and MYEOV were significantly higher in malignant cells and lower in monocytes.

Conclusions

High infiltration of monocytes affects poor prognosis of pancreatic cancer, suggesting that the immune microenvironment has a certain research prospect for treatment of pancreatic cancer. The monocyte-related genes signature can accurately assess the prognostic risk of pancreatic cancer.

Introduction

Pancreatic adenocarcinoma (PAAD) is a common tumor worldwide and is the fourth and sixth leading cause of cancer-related death in the United States and China, respectively (Lin et al. 2015; Siegel et al. 2022). Due to its high malignancy, PAAD has an extremely poor prognosis, with a 5-year overall survival rate of less than 5% and a 1-year overall survival rate of 24% based on standard treatment (Schizas et al. 2020). Male, advanced age, smoking and family history of PAAD were risk factors associated with
increased incidence and mortality from PAAD (Klein et al. 2004). Due to the lack of early clinical symptoms, most PAAD patients are already too late for tumor resection (Hijioka et al. 2022), and the current chemotherapy regimen is not satisfactory (Oberstein and Olive 2013). Therefore, early diagnosis and timely treatment are essential to reduce the mortality and improve the prognosis of patients with PAAD.

Monocytes are transported through the bloodstream to tissues during homeostasis, and faster during inflammation. Monocytes play a key role in supporting tissue homeostasis, initiating and spreading the host response to pathogens, and resolving immune responses before excessive tissue damage occurs (Mattoscio et al. 2021; Shi and Pamer 2011). More and more reports indicate that monocytes have become important regulatory factors in cancer development, and play an important role in promoting tumor growth and preventing metastasis and spread of cancer cells, because different subpopulations of monocytes may play different roles (Urakawa et al. 2019). The recruitment of monocytes occurs in almost all stages of cancer development and metastasis, which further indicates the indispensable role of monocytes in cancer development (Olingy et al. 2019). Just as they play a double-edged role in cancer, monocytes play a double-edged role in immunotherapy (De Ridder et al. 2022). In addition, monocytes are the main source of tumor-associated macrophages (TAM) and dendritic cells (DC) (Engblom et al. 2016).

The role of monocytes in PAAD is also attracting more and more attention. Previous studies have confirmed that the lymphocyte/monocyte ratio can be used as a marker to predict the prognosis of PAAD (Lin et al. 2020). Single cell analysis showed that PAAD was characterized by high monocytes in blood (Kemp et al. 2021), suggesting the important role of monocytes in PAAD. Recent studies have shown that S100A8 and S100A9 proteins secreted by monocytes can lead to secretion of cytokines such as TNF-α and TGF-β in PAAD cell lines, which in turn lead to secretion of S100A8 and S100A9 by monocytes (Nedjadi et al. 2018). Although this study did not mention the effect of monocytes on the proliferation of PAAD, it suggested that there was crosstalk between monocytes and PAAD cells (Nedjadi et al. 2018). In addition, monocytes are associated with the immunotherapy sensitivity of PAAD. Monocytes may be recruited by CCL2, which leads to the decrease of CD8+ T cells in the tumor microenvironment of PAAD, resulting in increased resistance of PAAD to immune checkpoint inhibitors (Li et al. 2022). In conclusion, these studies suggest the importance of monocytes in PAAD. However, the pathogenesis, diagnosis and therapeutic effects of monocytes in PAAD still need a lot of research.

In this study, we explored PAAD transcriptome data and clinical data from public databases such as TCGA and GEO to develop a monocyte related prognostic signature. This model can effectively predict the prognosis of PAAD patients and its sensitivity to immunotherapy, which provides a basis for personalized management of PAAD.

**Methods**

**Data download**
PAAD scRNA-seq data GSE156405 including 5 PAAD patients, 33694 genes and 17370 cells were downloaded from GEO databases. TCGA-PAAD contains 178 PAAD samples and 4 cases of paracancerous tissue. Bulk RNA-seq data and corresponding clinical information. GSE28735 contains mRNA microarray data and clinical information from 45 PAAD samples and 45 paracancer tissues. GSE62452 contains mRNA microarray data and clinical information from 69 PAAD samples and 61 paracancer tissues. GSE78229 contains mRNA microarray data and clinical information of 50 PAAD samples. 178 PAAD patients from TCGA-PAAD were downloaded from the UCSC Xena link (http://www.genome.ucsc.edu/index.html). The GSE28735, GSE62452, and GSE78229 datasets are all annotated by the GPL6244 platform and downloaded from the NCBI GEO link. The GSE28735, GSE62452 and GSE78229 datasets were combined using the combat algorithm to remove batch effects.

**CIBERSORT estimation and Kaplan-Meier survival analysis**

The CIBERSORT algorithm was utilized to assess the 22 kinds of immune cell types in PAAD. We used Kaplan-Meier survival analysis to appraise diversities in the PAAD (Overall Survival) between the high-abundance and low-abundance of all immune cells. The R packages survMiner and survival were tools to enable this process.

**Weighted gene co-expression network analysis (WGCNA)**

Co-expression networks were created utilizing the "WGCNA" package in R software. The PAAD samples in the TCGA databases were clustered to determine the existence of remarkable outliers. Following this, the co-expression network was developed utilizing the automatic network construction function, and the soft threshold was computed using the pickSoftThreshold function. The co-expression similarity was derived according to the soft threshold, and the adjacency was calculated thereafter. Next, the modules were ascertained using hierarchical clustering as well as dynamic tree-cut functions. Finally, gene significance, as well as module membership, were determined to correlate modules with monocytes cell content.

**Functional enrichment analysis in Metascape**

GO terms and KEGG pathway enrichment analyses play a vital role in identifying characteristic biological attributes for high-throughput transcriptome data. We used Metascape(http://metascape.org/), a gene annotation and analysis resource (Zhou et al. 2019), to perform a functional enrichment analysis, which included cellular component (CC), molecular function (MF), and biological process (BP), and a KEGG pathway analysis of the hub genes.

**Unsupervised subtype discovery based on gene expression analysis**

After the univariate Cox analysis, 296 prognostic mRNAs were left. Consensus Cluster Plus software (1.54.0) was used to subgroup all TCGA-PAAD patients into two clusters based on these prognostic mRNAs’ expression levels. Survival analysis, immune infiltration analysis and KEGG pathway analysis were performed between the two clusters.
Risk model construction

Univariate cox regression analysis and least absolute shrinkage and selection operator (LASSO) Cox regression analysis, were employed to identify the monocyte-related mRNAs prognostic signature in the training dataset. The risk score for each PAAD patient was calculated based on the following formula (1):

\[ \sum_{i=1}^{n} \text{Coef}_i * x_i \]

Where Coefi means the coefficients, xi is the FPKM value of each monocyte-related mRNAs. The receiver operating characteristic (ROC) curve analysis was used to evaluate the accuracy of the signature in the training dataset and testing dataset.

Gene set enrichment analysis

The Gene Set Enrichment Analyses (GSEA) method was used to explore the potential KEGG pathway implicated in the high-risk group and low-risk group. The reference gene set was retrieved from c2.cp.kegg.v7.1.symbols files, and the significant pathways were screened based on the criterion: P < 0.05 and FDR < 0.25.

ssGSEA and tumor immune dysfunction and exclusion (TIDE)

Single sample GSEA (ssGSEA) algorithm was performed to evaluated the immune cell papulations and immune status. We used the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm to assess the sensitivity of PAAD patients to immune checkpoint blockers (ICBs) in TCGA cohort. Generally, patients with high TIDE scores are less sensitive to ICB treatment (Jiang et al. 2018).

scRNA-Seq data processing

The Seurat package SCTransform function was used to preprocess the single-cell transcriptome datasets. The most changed 3000 genes were chosen by Select Integration Features and the FindCluster package used for all cell cluster analysis with the resolution set to 0.15.

Human protein atlas

The Human protein atlas (HPA, https://www.proteinatlas.org/), is a Swedish-based program initiated in 2003 with the aim to map all the human proteins in cells, tissues, and organs using an integration of various omics technologies, including antibody-based imaging, mass spectrometry-based proteomics, transcriptomics, and systems biology (Sjöstedt et al. 2020). In the present study, we evaluated the expression level of 6 genes in PAAD and paracancer tissues with the database.

TISCH analysis
Tumor immune single-cell hub (TISCH, http://tisch.comp-genomics.org/), a database focusing on the tumor microenvironment (TME), provides single-cell level cell-type annotation [20]. In the present study, we evaluated the expression level of 6 genes in each subgroup of cells in the other two PAAD datasets (CRA001160 and GSE111672).

**CBioPortal dataset**

CBioPortal (https://www.cbioportal.org), a comprehensive database, provides analysis and visualization functions to process multi-tumor genomics data (Gao et al. 2013). Based on data in TCGA, genetic alterations and OS analysis of 6 genes were obtained from cBioPortal.

**Statistical analysis**

The statistical analysis was performed in R software (version 4.0.2). The Perl programming language (Version 5.30.2) was used for data processing. Kaplan-Meier survival curve analysis with log-rank test was applied to analyse OS. Univariate, Lasso, and multivariate Cox regression analyses were used to evaluate prognostic significance. ROC curve analysis and its AUC value was used to evaluate the reliability and sensitivity of the prognostic signature. P < 0.05 was regarded as statistically significant.

**Results**

**CIBERSORT analysis**

To evaluate the role of various immune cell infiltrates in PAAD, bulk RNA-seq data were used to analyze the proportion of immune cells and their impact on the prognosis of PAAD. Cibersort is an analytical tool that predicts the ratio of 22 immune cells based on RNA-seq data from the TCGA database (Fig. 2A). In addition, survival analysis results showed that patients with low CD4+ T cell abundance had a poor prognosis (P = 0.016), patients with low MAIT cell abundance had a poor prognosis (P = 0.018), while patients with low Monocyte abundance had a good prognosis (P = 0.0095) (Fig. 2B-D). Depending on the tumor type, monocytes/macrophages may play a dual role as good or bad indicators of cancer recovery (Alwani et al. 2022). We speculate that high infiltration of monocytes may play a bad role in the immune microenvironment of PAAD, which leads to a poor prognosis.

**WGCNA analysis**

To further explore the potential role of monocytes in PAAD, we performed WGCNA analysis based on TCGA data of PAAD and constructed a scale-free co-expression network to identify monocytes’ associated gene characteristics (Fig. 3A). A total of 13 modules were generated, among which the pink module was most significantly positively correlated with the high infiltration of monocytes (Fig. 3B). The module was composed of 296 genes, which were analyzed by functional enrichment. The top ten pathways with significantly enriched KEGG functions include Cell cycle, Cell senescence, P53 signaling Pathway and other signaling pathways (Fig. 3F). Consistent with this, Cell cycle biological process was most significantly enriched in GO-BP (Fig. 3C).
Molecular typing based on genes associated with monocytes

In order to explore the role of genes in this module in PAAD, we further screened out 296 prognostic genes by COX analysis. Consensus clustering was then used to divide the samples into several subgroups according to the expression levels of these 296 mRNAs. K = 2 was taken as the best parameter (Fig. 4A), so the dataset was divided into two subgroups. Then, the consistency matrix (Fig. 4B) was established. The value of the consistency matrix was [0, 1], equal to 1 means multiple clustering and two data points are all in the same class, and 0 represents that multiple clustering is not in the same class. The heatmap showed that these prognostic genes were expressed differently in different subgroups (Fig. 4C), and the prognosis of C2 group was worse than that of C1 group (Fig. 4D). By analysing the degree of immune cell infiltration in different groups, we found that there were significant differences in various types of immune cell infiltration between the two groups. The infiltration degree of monocytes in C2 group was significantly higher than that in C1 group, while MAIT cells were significantly lower than that in C1 group, which was consistent with the previous results (Fig. 4G). We also conducted functional enrichment analysis for the differentially expressed genes between the two groups (Fig. 4E, F). KEGG enrichment showed Pancreatic secretion, Protein Digestion and absorption, BCM-receptor interaction, Cell cycle, P53 signaling pathway and other signaling pathways were significantly enriched. The results of GO-BP showed that there were several biological processes such as digestion, extracellular structure organization and extracellular matrix organization.

Construct a monocyte related prognostic model

In view of the high monocytes infiltration and poor prognosis in C2 subgroup, we performed differential analysis and univariate Cox analysis between C2 and C1 subtypes. 262 prognostic genes were identified that differentially expressed between C2 and C1 subtypes. In the TCGA cohort, LASSO regression analysis was used to construct the model, which was composed of 6 mRNAs (Fig. 5A, B).

RiskScore = 0.266*MET + 0.100*LY6D + 0.053*CEP55 + 0.022*ITGB6 + 0.034*MYEOV + 0.053*NUSAP1.

The risk score of each sample in the TCGA cohort was calculated by the formula and divided into low and high-risk subgroups according to the median risk score. Figure 4B depicts the survival status and duration of patients in the high-risk and low-risk groups, with a scatter plot showing higher mortality in the high-risk group. The expression patterns of the 6 model-related genes were also shown in Fig. 4B. Kaplan Meier survival curve showed that patients in the high-risk group had a worse prognosis than those in the low-risk group (Fig. 5C). ROC curve (Fig. 5D) showed that the monocyte related gene signatures (MGS) had good predictive ability of OS in TCGA cohort (1-year AUC = 0.85 – 0.67, 3-year AUC = 0.92 – 0.72, and 5-year AUC = 0.97 – 0.62). The combined tumor samples of GSE28735, GSE62452, and GSE78229 datasets were used as validation datasets. The risk score of each patient in the validation set was calculated by formula, and the patients were also divided into high and low risk groups according to the median risk scores. Kaplan Meier survival curve showed that patients in the low-risk group had higher OS
than those in the high-risk group (Fig. 5E). Time-dependent ROC curve showed that the AUC values of 1-year, 3-year and 5-year OS were 0.44–0.64, 0.72–0.90 and 0.84–0.97, respectively, indicating that the model was accurate in predicting OS of patients with PAAD (Fig. 5F).

**MGS is an independent prognostic factor for patients with PAAD**

Univariate and multivariate Cox analyses were used to assess whether the MGS was an independent prognostic factor in patients with PAAD. Univariate Cox analysis based on TCGA data of PAAD patients in the training group showed a significant association between MGS and OS [Hazard ratio (HR): 5.511, 95%CI: 3.269–9.292, p < 0.001]. Multivariate Cox analysis further showed that MGS was an independent prognostic factor for OS (HR: 5.387, 95%CI: 3.157–9.192, P < 0.001) (Table 1). In order to evaluate the possible biological functions of MGS, we performed GSEA enrichment analysis. The results indicated that significant enrichment of P53 signaling Pathway, Cell cycle, Pathway in Cancer, ECM receptor interaction, Pancreatic cancer, which are consistent with the previous results (Fig. 5G). Meanwhile, we also analyzed the correlation between the risk score of MGS and monocyte infiltration, and the results showed that the higher the risk score, the higher the level of monocyte infiltration (Fig. 5H, r = 0.480, P < 0.001).
### Table 1
Multivariable Cox analysis of the MRS and routine clinicopathological characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total(N)</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hazard ratio (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>RiskScore</td>
<td>171</td>
<td>5.511 (3.269–9.292)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>age</td>
<td>171</td>
<td>1.027 (1.006–1.048)</td>
<td>0.013</td>
</tr>
<tr>
<td>grade</td>
<td>171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>28</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>G2-4</td>
<td>143</td>
<td>1.922 (1.015–3.640)</td>
<td>0.045</td>
</tr>
<tr>
<td>T</td>
<td>171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>6</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>T2-4</td>
<td>165</td>
<td>5.037 (0.696–36.478)</td>
<td>0.109</td>
</tr>
<tr>
<td>gender</td>
<td>171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>93</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>78</td>
<td>1.120 (0.742–1.692)</td>
<td>0.590</td>
</tr>
<tr>
<td>stage</td>
<td>171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td>20</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Stage -</td>
<td>151</td>
<td>2.735 (1.185–6.316)</td>
<td>0.018</td>
</tr>
</tbody>
</table>

### Characteristic analysis of immunotherapy in MGS

Then we explored differences in immune characteristics between different risk groups. Compared with
the low-risk group, the PAAD immune microenvironment in the high-risk group was infiltrated with more
primitive CD8 naive cells, natural regulatory T cells (nTreg), Th2, Th17, dendritic cells (DC), monocytes,
and neutrophils. The infiltration of central memory T cells (Tc), T Follicular Helper Cells (Tfh), Mucosal
Associated Invariant T (MAIT), NK cells, Tgd cells, CD4 + T cells, and CD8 + T cells was higher in the low-
risk group (Fig. S1A). The ssGSEA algorithm was used to estimate the immune status of patients in the
tumor microenvironment. Compared with the low-risk group, the high-risk group had higher APC
costimulation, parainflammation, type 1 IFN immune response scores, and lower cytolytic toxicity, HLA, T
cell costimulation, type 2 IFN immune response scores (Fig. S1B).
Immunotherapy led by immune checkpoint blockers (ICBs) has played a significant role in the treatment of PAAD in recent years, but a large number of patients are still insensitive to it [23]. Therefore, we predicted the sensitivity of PAAD patients to ICBs treatment based on TIDE algorithm, and compared the differences in the sensitivity to ICBs of patients in different risk groups. As shown in Fig. S1C,D, TIDE score in the low-risk group was significantly lower than that in the high-risk group, and the proportion of patients responding to ICBs treatment in the low-risk group was significantly higher than that in the high-risk group. These results suggest that PAAD patients in different risk groups have different immune cell infiltration and immune status in the immune microenvironment, and different sensitivity to immunotherapy.

Expression of characteristic genes in PAAD tissues

In order to verify the protein level expression levels of the 6 genes in PAAD, we found their immunohistochemical results through the online database of Human Protein Atlas, as shown in Fig. 6. Consistent with the mRNA expression results, these 6 genes were significantly overexpressed in PAAD tissues.

The expression of 6 genes in monocytes was analysed by scRNA-seq

The scRNA-seq dataset GSE156405 was used to characterize the diversity and heterogeneity of various cell subsets in PAAD tissues. The Uniform Manifold Approximation and Projection (UMAP) is used for nonlinear dimensionality reduction, and the cells are clustered by the FindCluster function, resulting in 16 clusters (Fig. 7A). Epithelial cells (Cluster 0, 2, 8, 10, 11; Markers: LCN2, CEACAM5, PLAT) and malignant cells (Cluster 0, 10; Markers: ANKRD36C, NEAT1, CAPN8); Myeloid cells (Cluster 3, 6; Markers: SPP1, S100A8, S100A9) and monocytes (Cluster 3; Markers: S100A8, S100A9, FCN1), obtained by cell marker classification (Fig. 7B).

The expression levels of these 6 genes in each cell are shown in Fig. 7C. MET and MYEOV were significantly lower expressed in monocytes (P < 0.01, logFC<-0.2), but significantly higher expressed in malignant cells (P < 0.01, logFC > 0.5). With GSE111672 and CRA001160 data from TISCH database, we verified again that the expression levels of these 6 genes were low in monocytes but high in malignant cells (Fig. S2).

Meanwhile, functional enrichment analysis was performed for all gene sets that were low expressed in monocytes and high expressed in malignant cells. KEGG enrichment showed pathways associated with cancer such as Tight junction, ECM-receptor interaction, PI3K-Akt signaling pathway, MAPK signaling pathway and other signaling pathways were significantly enriched. The results of GO-BP showed that there were several biological processes such as cell migration, cell adhesion, cell motility was enriched (Fig. 7E). These results reconfirmed that the gene expression level in monocytes is closely related to the function of malignant cells.

Mutation analysis of 6 genes in PAAD model
To better understand the genomic characterization of the MGS in PAAD, we performed mutation analysis using the cBioPortal database. Fig. S3A shows a summary of the genetic mutation in the six genes, which have a high mutation frequency in the PAAD sample. According to TCGA database, mutations occurred in 45% of 178 PAAD samples, most of which belonged to amplified mutation, among which mutation frequency of LY6D was the highest (14%).

Grouping according to whether all six genes mutated or not, we found the mutated group had a worse overall survival prognosis (Fig. S3B, p < 0.001). Grouping according to whether each gene mutated or not, we found the same results (Fig. S3C, p < 0.01). These results suggest that there is a significant mutation trend in 6 genes in PAAD, which may lead to a worse prognosis of patients through such adverse mutations.

**Discussion**

The poor prognosis of PAAD is mainly due to lack of early symptoms, rapid tumor progression, and limited effective drugs (Schizas et al. 2020). Therefore, there is an urgent need for more effective treatment strategies such as immunotherapy. Immunotherapy of tumor immune microenvironment has attracted more and more attention now. Although PAAD is a non-immunogenic cancer, a large number of infiltrating immunogenic cells, such as tumor-associated macrophages, myeloid derived suppressor cells (MDSCs) and neutrophils, have been identified (Schizas et al. 2020).

MDSCs are a relatively stable form of pathologic activation of monocytes and do not represent an actual subpopulation of myeloid cells, but rather an activated state (Grover et al. 2021). The presence of these cells has been associated with adverse clinical outcomes in a variety of solid tumors, including PAAD (Grover et al. 2021). We also found that high infiltration of monocytes was associated with poor prognosis in CIBERSORT analysis of immune cell infiltration.

In this study, we firstly analyzed the most significant positive correlation gene modules in the high infiltration group of monocytes by WGCNA, and screened out 296 genes related to prognosis by COX analysis. Consensus clustering was then used to divide the samples into two subgroups based on the expression levels of these genes. The degree of monocyte infiltration in C2 group was significantly higher than that in C1 group, and the prognosis was worse than that in C1 group. We continued to screen out 262 genes that were significantly differentially expressed and prognostic after clustering and typing. LASSO Cox analysis was used to generate regression models and coefficients, which was composed of 6 mRNAs: MET, ITGB6, CEP55, NUSAP1, MYEOV, LY6D.

Univariate and multivariate Cox analyses were performed to verify that the signature could be used as an independent prognostic factor for patients with PAAD. More importantly, we also analyzed the characteristics of immunotherapy in the signature, reconfirming the high infiltration of monocytes in the immune microenvironment of PAAD in the high-risk group. We also found that the high-risk group was more sensitive to ICBs treatment. On the other hand, this signature may provide guidance for immunotherapy of PAAD through monocytes (Ugel et al. 2021).
In the process of validation of signature genes, we not only verified the expression of 6 genes at the tissue level, but also verified the expression levels in monocytes and malignant cells at the single cell level. Interestingly, at the tissue level, all 6 genes were highly expressed in PAAD tissues; at the single-cell level, MET and MYEOV were significantly overexpressed in malignant cells, but significantly underexpressed in monocytes. Mutation analysis of 6 genes confirmed that the overall survival prognosis of each gene mutation group was worse, which also confirmed the reliability of the MGS from the perspective of gene mutation.

Recently, several articles have reported the involvement of these six genes in PAAD. MET was identified as PAAD-specific receptor tyrosine kinases (RTKs), which are significantly associated with the prognosis of immune “hot” and “cold” PAADs. It's found in the mechanism, MET inhibition combined with PD-1/PD-L1 blockade can effectively improve the efficacy of immunotherapy for PAAD(Li et al. 2021). The coding protein c-MET, as the receptor of HGF, is expressed in PAAD cells and endothelial cells, and plays an important role in the progression and spread of PAAD(Pothula et al. 2020; Xu et al. 2020). Through prospective and retrospective serum and pancreatic tissue collections, Daniela et al. found that ITGB6 can be used as a novel tumor marker for the refined diagnosis and prognosis of PAAD. Meanwhile, ITGB6 expression in tissues is associated with that in serum levels, and CA19-9 and prognosis, which is consistent with our findings(Lenggenhager et al. 2021). In vitro and in vivo, CEP55 overexpression accelerated the proliferation, migration and invasion of PAAD cells, and is a valuable prognostic factor and potential therapeutic target(Peng et al. 2017). Meanwhile, CEP55 is also one of the downstream genes affected by RACGAP1, which is involved in enhancing the migration and invasion ability of PAAD cells(Khalid et al. 2019). Guo et al. demonstrated in vitro and in vivo that microRNA-569 inhibits PAAD metastasis by directly targeting NUSAP1(Guo et al. 2022). Experiments showed that MYEOV interacts with oncogenic transcription factor SOX9 to promote the expression of its target gene HES1, thereby enhancing the ability of PAAD cells to transfer and proliferate in vivo and in vitro(Liang et al. 2020). Shen et al. found that MYEOV promotes oncogenic miR-17/93-5p expression through its association with MYC, thereby promoting the progression of PAAD(Shen et al. 2021). Steve E Kalloger collected epithelial samples from 96 cases of PAAD through virtual fiber anatomy and found that LY6D gene could be used as one of the genes to predict prognosis, which was consistent with our research results(Kalloger et al. 2021). In addition, some research on bioinformatics has also found that these six genes MET(Wu et al. 2019), ITGB6(Zhuang et al. 2020), CEP55(Tang et al. 2021), NUSAP1(Deng et al. 2020), MYEOV(Wu et al. 2022), LY6D(Wang et al. 2021; Xu et al. 2021) are one of the genes for prognostic prediction. These studies fully reflect the importance of these 6 genes in PAAD, and also confirm the rationality and importance of the prognostic signature in this study.

In conclusion, based on the effect of immune cell infiltration on prognosis in PAAD, we constructed MGS through multiple screening of monocyte related prognostic genes. We not only verified the power and reliability of this signature, but also combined with multiple single-cell sequencing datasets such as GSE156405, GSE111672 and CRA001160 to confirm the important prognostic and immunotherapeutic significance of these six genes in PAAD at tissue and single-cell level.
Abbreviations

ICBs
immune checkpoint blockade
PAAD
Pancreatic adenocarcinoma
TAM
tumor-associated macrophages
TCGA
The Cancer Genome Atlas
GEO
Gene Expression Omnibus
WGCNA
Weighted Gene Co-expression Network Analysis
GO
Gene ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
CC
cellular component
MF
molecular function
BP
biological process
LASSO
least absolute shrinkage and selection operator
ROC
receiver operating characteristic
GSEA
Gene Set Enrichment Analyses
TIDE
Tumor Immune Dysfunction and Exclusion
HPA
Human Protein Atlas
TME
tumor microenvironment
MAIT
Mucosal Associated Invariant T
MGS
monocyte related gene signatures
OS
overall survival
nTreg
natural regulatory T cells
Tfh
T Follicular Helper Cells
MDSCs
myeloid derived suppressor cells
RTKs
receptor tyrosine kinases.

Declarations

Author Contributions

WY, XL, WQ, YZ and ZL (Zhenyi Lv) contributed to the conception. WY, XL, TL, ZM and ZL (Zhituo Li) designed the study. YH and JH participated in data collection. WY and XL performed data analysis, prepared the figures and tables and wrote the initial manuscript. TL, ZM and DX performed proofreading and deep editing of the manuscript. ZL (Zhituo Li) and YZ supervised the project and approved the final manuscript. ZL (Zhituo Li) contributed to funding acquisition. All authors have read and approved the manuscript for publication.

Funding

This research was supported by National Natural Science Foundation of China (No. 82170654) and Scientific Research and Innovation Fund of the First Affiliated Hospital of Harbin Medical University, No. 2021M10.

Data Availability

This work benefited from previous cohort studies. The data used to analyse the findings of this study and related materials are available from the corresponding author (Zhituo Li) upon request.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval

The patients involved in the database have obtained ethical approval.

Consent to Participate
Not applicable.

Consent to Publish

Not applicable.

References


Figures
Figure 1

Flow chart showing the analysis process in this study.
Figure 2

The proportion of immune cells. A The proportion of 22 immune cells built on RNA-seq data. The Kaplan-Meier curves (overall survival) of patients with different proportions of CD4+ T cells B MAIT cells C and Monocyte D.
Figure 3

Screening of monocyte-related genes. A Genes are clustered into discrete modules. B The correlation between different modules and the proportion of Monocyte-high and low infiltration. C-F Functional enrichment analysis of genes in pink module C GO-BP D GO-CC E GO-MF F KEGG
Figure 4

Monocyte-related Molecular typing. **A** Consensus Map of NMF Clustering. **B** Sample cluster of TCGA-PAAD. **C** Expression of relating genes of Monocyte. **D** Kaplan-Meier curves of two PAAD molecular subtypes. **E, F** GO-BP and KEGG functional enrichment analysis of differentially expressed genes in two molecular subtypes. **G** Immune cell infiltration analysis in two different groups.
Construct a risk model: **A** The trajectory of each independent variable (horizontal axis represents the log value of the independent variable lambda) and the confidence interval under each lambda. **B** Distributions of risk scores and survival status of PAAD patients in the train dataset. **C, E** Survival curve of high and low risk group in train and test datasets. **D, F** ROC curves of signature for predicting 1/3/5-year survival in
train and test datasets. **G** GSEA functional enrichment analysis of differentially expressed genes in high and low risk groups. **H** Correlation of risk score with monocyte infiltration in this signature.

**Figure 6**

Protein expression lever of 6 genes in the human protein atlas (HPA) database.
Figure 7

Expression of 6 genes in PAAD scRNA-seq. **A**, BUMAP plot of 17370 cells from 5 PAAD patients. (Label colors according to separate clusters and cells.) **C** Expression of 6 genes in each cell. **D, E** KEGG and GO-BP functional enrichment analysis of genes which with low expression in monocytes and high expression in malignant cells.
Supplementary Files

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

- Supplementfigure1.tif
- Supplementfigure2.tif
- Supplementfigure3.tif