scientific reports

Check for updates

OPEN

Characterization of pancreatic cancer stem cells and construction of a stem cell-based prognostic signature

Junwei Fang[1], Huitao Ji[2], Meiping Wang[1], Baohua Zheng[2], Junwei Zhou[2] & Chunhong Xiao[1]✉

Pancreatic cancer is a gastrointestinal malignancy with a dismal prognosis. Cancer stem cells (CSCs) are considered key drivers of its aggressiveness, metastatic capacity, and resistance to treatment. However, the intratumoral heterogeneity of CSCs and their roles within the tumor microenvironment remain poorly characterized. We integrated single-cell transcriptomic data (GSE214295) with large-scale bulk RNA-seq datasets (TCGA-PAAD, GEO and CPTAC) for comprehensive analysis. Malignant epithelial subpopulations were identified and their stemness evaluated using Seurat, CopyKAT, and CytoTRACE. Monocle2, scMetabolism, decoupleR, and CellChat were applied to investigate differentiation trajectories, metabolic characteristics, transcription factor activities, and intercellular communication. Finally, CSC-like prognostic index (CSCLPI) was constructed and validated with multiple machine-learning approaches. The robustness and generalizability of the model were validated using the GEO and CPTAC dataset. Seven malignant epithelial subpopulations were identified. Among them, the C2 cluster displayed the highest stemness score, lowest metabolic activity, and prominent CSC-marker expression. Pseudotime analysis placed C2 at the origin of the differentiation trajectory. C2 also exhibited activation of key stemness-related transcription factors (e.g., SOX9, MYC) and robust cell–cell communication via WNT and TGF-β signaling. The CSCLPI model, derived from C2-specific genes, stratified patients into distinct risk groups. The CSCLPI score correlated significantly with tumor mutation burden, stemness indices, immune escape, and chemoresistance, and showed consistent prognostic accuracy across multiple cohorts. A nomogram incorporating CSCLPI and clinical parameters demonstrated strong prognostic value in pancreatic cancer. This study systematically characterized the heterogeneity and functional landscape of CSC-related subpopulations in pancreatic cancer. The stemness-based CSCLPI model offers a robust prognostic tool and potential therapeutic targets for personalized treatment and CSC-directed strategies.

Keywords Cancer stem cells, Immune cells, Machine learning, Nomogram, Pancreatic cancer

Pancreatic adenocarcinoma (PAAD) ranks among the most aggressive and lethal gastrointestinal malignancies[1]. Worldwide, an estimated 460,000 new cases and 430,000 deaths occur annually; the 5-year survival rate is < 10%, and incidence is slightly higher in men than in women[2]. Histologically, cancer is heterogeneous: pancreatic ductal adenocarcinoma (PDAC) constitutes ~ 90% of cases, while neuroendocrine tumors, acinar cell carcinoma, and other rare subtypes make up the rest[3]. Clinically, pancreatic tumors are classified as resectable, borderline-resectable, or unresectable, a scheme that directly dictates therapy and prognosis[4]. Radical surgery offers the only chance of cure, yet most patients present with locally advanced or metastatic disease. Consequently, multidisciplinary therapy—including chemotherapy, radiotherapy, targeted therapy, and immunotherapy—is routinely adopted[5]. Neoadjuvant chemotherapy and precision-medicine strategies have recently offered new hope, but pancreatic cancer's distinctive genetic and epigenetic landscape still impedes early diagnosis and effective treatment.

CSCs constitute a malignant subpopulation endowed with quiescence, self-renewal, and multipotency[6]. Growing evidence implicates CSCs in pancreatic cancer initiation, progression, chemoresistance, and

[1]Department of General Surgery, 900th Hospital of Joint Logistics Support Force of People's Liberation Army, Fuzhou, Fujian, China. [2]Department of General Surgery, Fuzhou General Teaching Hospital, Fujian University of Traditional Chinese Medicine (900th Hospital of Joint Logistics Support Force), Fuzhou, China. ✉email: xiao84chun@163.com

recurrence[7]. In 2007, Li et al. first isolated pancreatic CSCs, identifying a small number of cells co-expressing CD24, CD44 and ESA with pronounced tumorigenicity and self-renewal[8]。 Ercan et al. later showed that co-expressing CD133 and CXCR4 CSCs exhibit enhanced invasiveness and metastasis, which can be curtailed by CXCR4 inhibition[9]. These findings offer fresh insight into the pathogenesis and therapeutic hurdles of pancreatic cancer. Nevertheless, the pancreatic tumor microenvironment's complexity continues to hinder full characterization of CSC traits. Single-cell RNA-seq has proved invaluable for dissecting this heterogeneity; Domenichini et al. showed that under stress, pancreatic CSCs sustain low oxygen consumption, whereas non-CSCs markedly increase it[10].

Here, we integrate single-cell and bulk transcriptomes to map the developmental trajectory of pancreatic CSCs, build a machine-learning stemness prognostic index, and explore its molecular basis and translational relevance.

Methods and materials

Data acquisition

The PAAD scRNA-seq dataset (GSE214295) was obtained from the GEO database. This dataset contains sequencing data from multiple primary tumors and paired organoids of pancreaticobiliary carcinoma. In this study, only 2 of the PAAD primary tumor samples were extracted for subsequent analysis[11]. Bulk RNA-seq data for PAAD were obtained from the TCGA and GEO databases under the accession numbers TCGA-PAAD, GSE28735, and GSE68465. The TCGA-PAAD dataset comprises 178 tumor samples and 4 normal samples, which were used as the training cohort. Corresponding clinical information included overall survival time, survival status, age, sex, pathological stage, and TNM stage. After batch effect correction, the two GEO datasets were integrated to yield 114 pancreatic cancer samples, which were used as the external validation cohort. In addition, we incorporated the CPTAC-PAAD dataset as an additional external validation cohort.

Analyze the PAAD scRNA-seq data

Cells were filtered using the Seurat package, with criteria set to retain cells expressing between 200 and 7500 genes and a mitochondrial gene expression percentage below 10%. The top 3000 highly variable genes were identified using the FindVariableFeatures function in Seurat. After normalization using the ScaleData function, principal component analysis (PCA) was performed on the highly variable genes. Clustering was performed using the FindClusters function at a resolution of 0.6[12]. Cell clusters were annotated based on known markers reported in the literature: epithelial cells (EPCAM, KRT18, KRT19), immune cells (CD3D, CD3E, CD163), and stromal cells (VWF, COL1A1, MYL9). Copy number variation (CNV) analysis of epithelial cell clusters was performed using the CopyKAT R package to distinguish malignant epithelial cells (ECs) from non-malignant ECs[13]. To further characterize malignant epithelial cells, functional enrichment analysis was conducted on differentially expressed genes between malignant ECs and other ECs.

Pseudotime trajectory analysis for malignant ECs

Pseudotime trajectories were constructed based on the gene expression profiles of malignant epithelial cells using the Monocle2 R package[14]. Malignant ECs were projected onto trajectories with multiple branches, where cells within the same branch shared similar differentiation states. Branch-dependent or state-specific genes were identified by clustering genes at branch points, enabling investigation of the underlying mechanisms of cellular differentiation.

CytoTRACE

CytoTRACE is a novel computational method designed to measure cellular differentiation potential, representing a major advance in predicting differentiation at the single-cell level[15]. It has been shown to outperform existing computational approaches in assessing stemness. After processing the gene expression matrix, CytoTRACE was used to assign a stemness score to each malignant epithelial cell (EC). A higher CytoTRACE score indicates greater stemness and a lower level of differentiation.

Analysis of malignant ECs regulatory network and metabolism

Transcription factor (TF) activity responsible for gene regulation in each malignant epithelial cell (EC) phenotype was inferred using the decoupleR package[16]. Metabolic activity in malignant ECs was quantified using the scMetabolism package. Metabolic scores for each cell were calculated using the sc.metabolism.Seurat function, with the pathway database set to "KEGG"[17].

Cell–cell communication analysis

CellChat is a novel computational tool designed to quantify intercellular communication networks using single-cell RNA sequencing (scRNA-seq) data[18]. Intercellular communication among different cell subtypes in PAAD was predicted by analyzing ligand–receptor interactions. The netVisual_circle function was used to visualize the strength of cell–cell communication networks. The netVisual_bubble function was applied to generate bubble plots highlighting key ligand–receptor interactions.

Machine learning-based screening of CSCs-related prognostic genes

Univariate Cox regression analysis was first performed to identify potential prognostic genes within the C2 cluster gene set. Subsequently, 117 combinations of 10 machine learning algorithms were applied to select CSCs-related genes. The algorithms included Random Survival Forest (RSF), Elastic Net (Enet), Stepwise Cox (StepCox), CoxBoost, Partial Least Squares Regression for Cox (plsRcox), Gradient Boosting Machine (GBM), survival support vector machine (survival-SVM), Ridge, Lasso, and Supervised Principal Components (SuperPC). The

TCGA cohort was set as the training cohort, and the GEO and CPTAC cohorts were set as the testing cohort. The C-index of all cohorts was calculated, and the model with the highest average c-index was the optimal model, and the 5 genes that contributed the most to the model were finally filtered out based on the gene importance assessment of the optimal model.

Construction and validation of a CSC-like prognostic index (CSCLPI)

These five key genes were included in a multivariate Cox regression model, and the Cancer Stem Cell-Like Prognostic Index (CSCLPI) was constructed based on the gene expression values and their correlation coefficients.Patients were stratified into high- and low-risk groups based on the median CSCLPI score. Kaplan–Meier survival analysis was performed to evaluate the prognostic differences between the two risk groups. Time-dependent receiver operating characteristic (ROC) curve analysis was used to assess the predictive performance of the CSCLPI for 2-, 3-, and 4-year overall survival in PAAD patients.

Analysis of CSCLPI subgroups

The mutation landscape of the TCGA cohort was profiled by counting gene-level mutations, and the 20 genes with the highest mutation counts were selected as significantly mutated genes. Mutation frequencies of these genes across all samples were assessed with the maftools package in R 4.4.0. Tumor mutational burden (TMB) was calculated for each tumor sample, and differences between CSCLPI subgroups were compared. The mRNA stemness index (mRNAsi), which reflects stem-cell–like features, was retrieved for PAAD using the TCGAbiolinks package[19]. The prognostic value of mRNAsi and its correlation with CSCLPI were evaluated. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to investigate the biological functions and potential regulatory pathways associated with different CSCLPI subgroups. The analysis was conducted using the R package clusterProfiler (v4.14.0)[20], with annotations obtained from the org.Hs.eg.db and KEGG.db databases. GO terms and KEGG pathways were referenced from the Gene Ontology Consortium[21] and KEGG database[22], respectively.

Evaluation of tumor microenvironment based on CSCLPI

The ESTIMATE algorithm was used to assess immune and stromal scores in high- and low-CSCLPI groups[23]. Single-sample gene set enrichment analysis (ssGSEA) was applied to evaluate immune cell infiltration and immune-related functional activity scores[24]. The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm is an effective tool for evaluating immune evasion and has shown potential in predicting patients' responses to immunotherapy[25]. Drug sensitivity to conventional chemotherapies was assessed in high- and low-CSCLPI groups using the oncoPredict R package by comparing the half-maximal inhibitory concentration (IC50) values[26].

Development and validation a nomogram

Univariate and multivariate Cox regression analyses were conducted to evaluate the impact of clinical features—including age, sex, and pathological stage—and CSCLPI on the overall survival of PAAD patients. A nomogram was constructed based on these variables to predict 2-, 3-, and 4-year overall survival. The predictive performance of the nomogram model was further evaluated using ROC curves, survival analysis, and calibration plots. A heatmap was used to visualize the relationships between selected CSC-related genes, clinical features, and CSCLPI scores.

Statistical analysis

All statistical analyses and visualizations were performed using R software. Continuous variables were compared using either the independent t-test or the Wilcoxon rank-sum test. Univariate and multivariate prognostic analyses were conducted using the Cox proportional hazards model. Kaplan–Meier survival analysis and the log-rank test were used to compare survival differences. Time-dependent ROC curves were generated using the "timeROC" R package. A p-value of less than 0.05 was considered statistically significant.

Results

Cell subtypes

The analytical workflow of this study is illustrated in Fig. 1. After rigorous quality control, a total of 8,037 high-quality single cells were obtained from two PAAD samples. At a resolution of 0.6, 20 distinct clusters were identified (Fig. 2A). Clusters 11 and 15 were identified as immune cells, while clusters 4, 7, 10, 14, 16, and 18 were classified as stromal cells. The remaining clusters were identified as epithelial cells (Fig. 2B and C). Based on scRNA-seq data from PAAD samples and without using normal cells as a reference, CopyKAT identified 4,106 epithelial cells as malignant epithelial cells (ECs) (Fig. 2D). The malignant epithelial cells (ECs) were re-normalized, dimensionally reduced, and clustered. At a resolution of 0.15, seven subclusters were identified (Fig. 2E).

Identification of stem cells cluster for PAAD

The major differentially expressed genes (DEGs) of each malignant epithelial cell (EC) subcluster are shown in Fig. 3A. GO and KEGG enrichment analyses of the seven clusters revealed distinct biological characteristics. Genes in Cluster 1 (C1) were enriched in heterochromatin formation, negative regulation of gene expression, and epigenetic-related processes. C2 genes were enriched in the regulation of mitotic nuclear division, chromosome segregation, and nuclear division. C3 genes were associated with the regulation of myoblast proliferation, endothelial cell regulation, and sprouting angiogenesis. C4 genes were enriched in apoptotic processes and the establishment or maintenance of transmembrane electrochemical gradients. C5 genes were involved in digestion

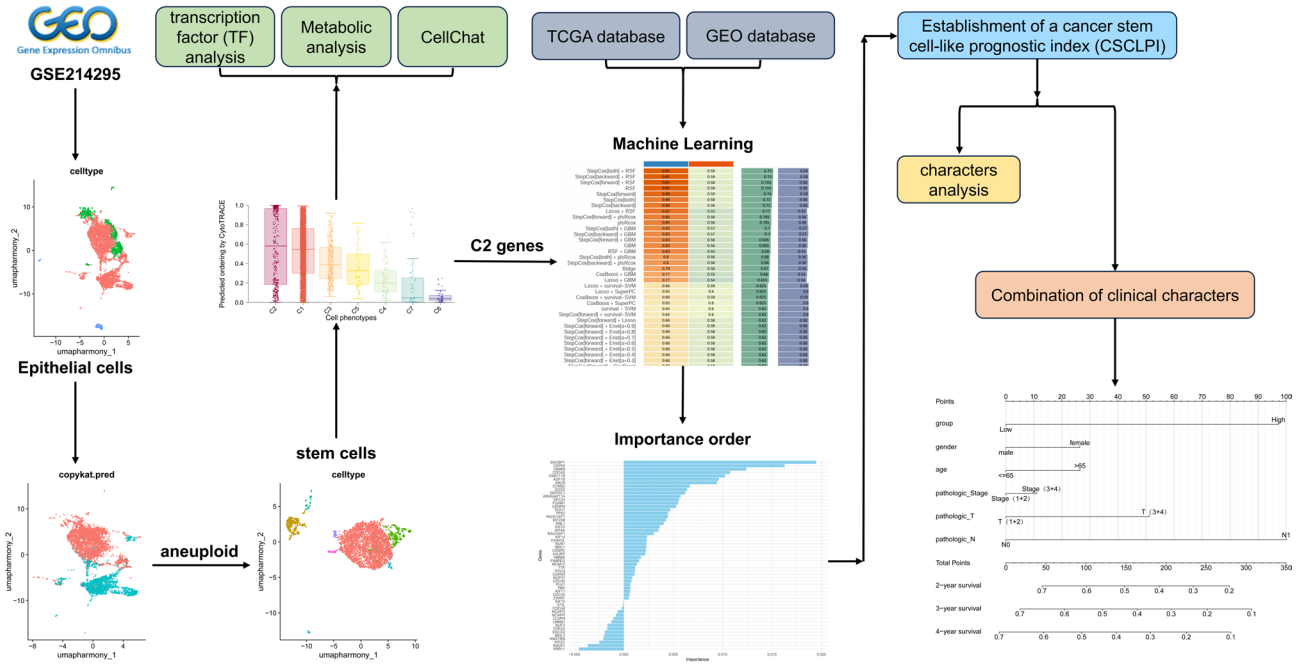


Fig. 1. The flowchart of analysis.

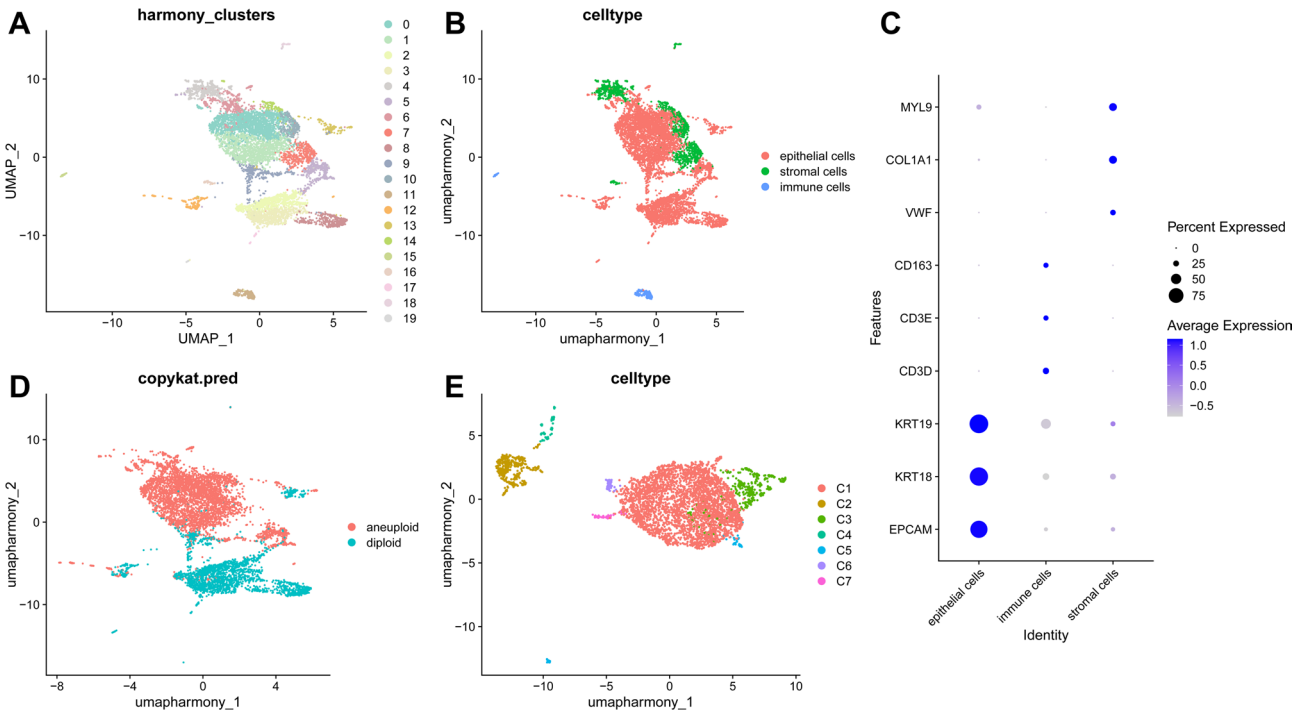


Fig. 2. Identification of cell subtypes. **A** 20 clusters are identified at a resolution of 0.6. **B** The UAMP plot displays the distribution of epithelial cells, stromal cells, and immune cells. **C** Markers that identify different cells are expressed. We identify three major cell clusters: epithelial cells, stromal cells, and immune cells. **D** The results of CopyKAT for predicting malignant ECs. **E** 6 clusters of malignant ECs are identified with a resolution of 0.15.

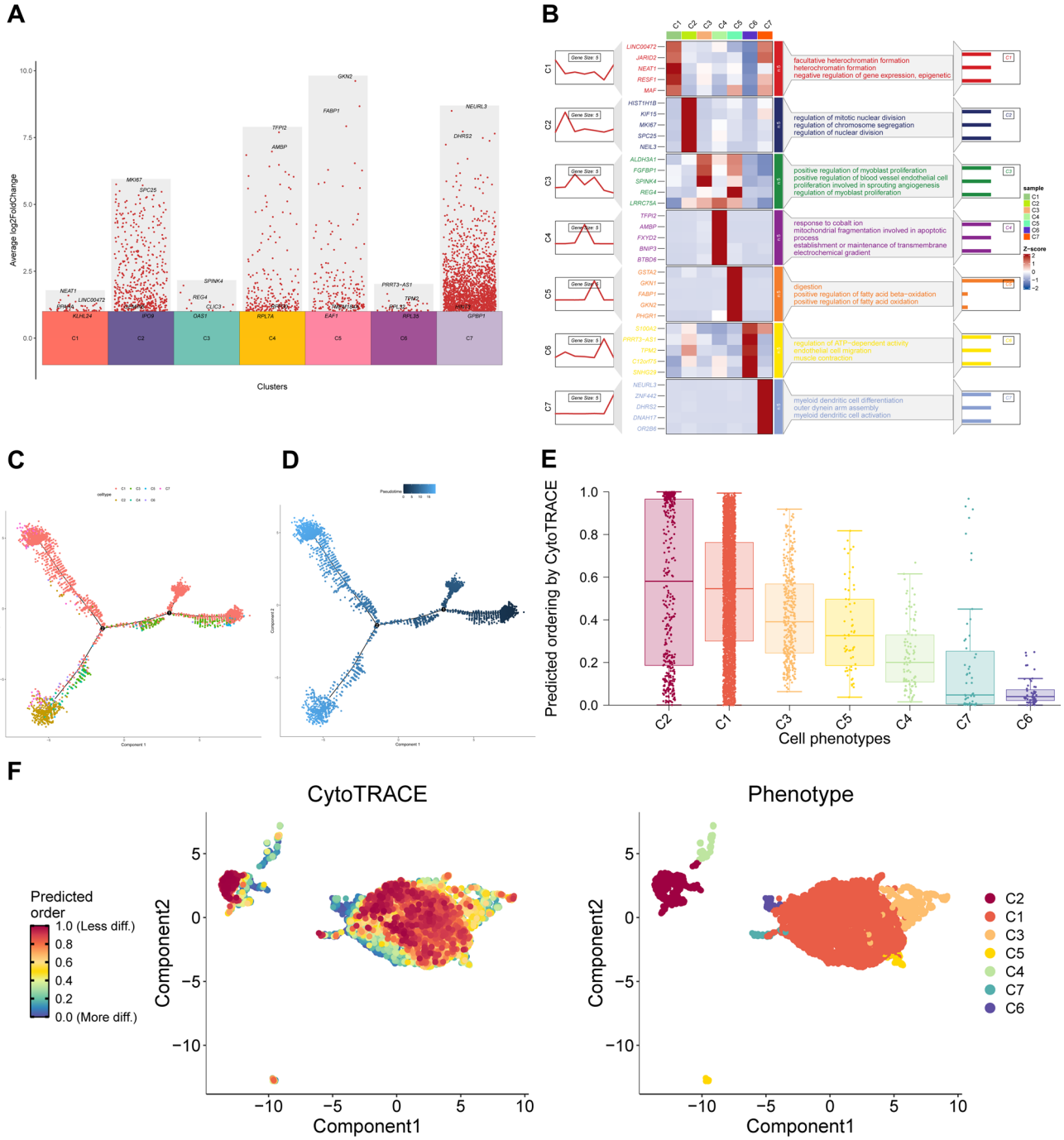


Fig. 3. Cell differentiation trajectory. **A** Some of positively differential genes in each cluster. **B** Display the differential genes and KEGG pathway enrichment results of each cluster simultaneously. **C** Differentiation trajectory of malignant ECs in PAAD for sub-cell types. **D** A color-coded pseudotime plot depicts the differentiation trajectory of malignant ECs in PAAD. **E**, **F** CytoTRACE predicts the cell differentiation potential of malignant ECs in PAAD, and identified C2 as the cancer stem cell cluster.

and the regulation of fatty acid oxidation. C6 genes were enriched in the regulation of ATP-dependent activity, endothelial cell migration, and muscle contraction. Lastly, C7 genes were mainly enriched in myeloid dendritic cell differentiation and activation (Fig. 3B).

Pseudotime trajectory analysis of malignant epithelial cells (ECs) was performed using Monocle2 to explore their developmental trajectory (Fig. 3C), revealing that distinct subclusters resided at different stages of differentiation (Fig. 3D). Based on the pseudotime trajectory, Cluster 2 (C2) and Cluster 1 (C1) appeared to represent the early stages of EC differentiation. Furthermore, CytoTRACE analysis indicated that C2 was located at the beginning of the differentiation path and exhibited the highest stemness potential, whereas C6 was positioned at the terminal end of the trajectory, representing a more differentiated state (Fig. 3E–F). Taken

together, these results suggest that C2 harbors cancer stem cell-like properties and was therefore selected for further analysis in the context of PAAD.

In addition, we investigated the expression of classical pancreatic CSC markers (CD44, CD24, ESA/EPCAM, SOX9) in seven subpopulations (Supplementary Fig. 1). The results showed that the C2 subpopulation exhibited the highest expression levels of the above CSC markers, further strengthening our rationale for considering C2 as a stem cell subpopulation.

Characteristics of PAAD stem cells

First, we investigated transcription factor (TF) activity and major pathway activation in Cluster 2 (C2). The results revealed significant activation of E2F1, AR, and EGR1 (Fig. 4A). Additionally, Estrogen and MAPK signaling pathways were found to be involved in regulating the biological functions of C2 (Fig. 4B).

Second, metabolic analysis identified key metabolic pathways enriched in C2, including ether lipid metabolism, glyoxylate and dicarboxylate metabolism, and glycolysis/gluconeogenesis (Fig. 4C), indicating notable metabolic reprogramming in CSCs.

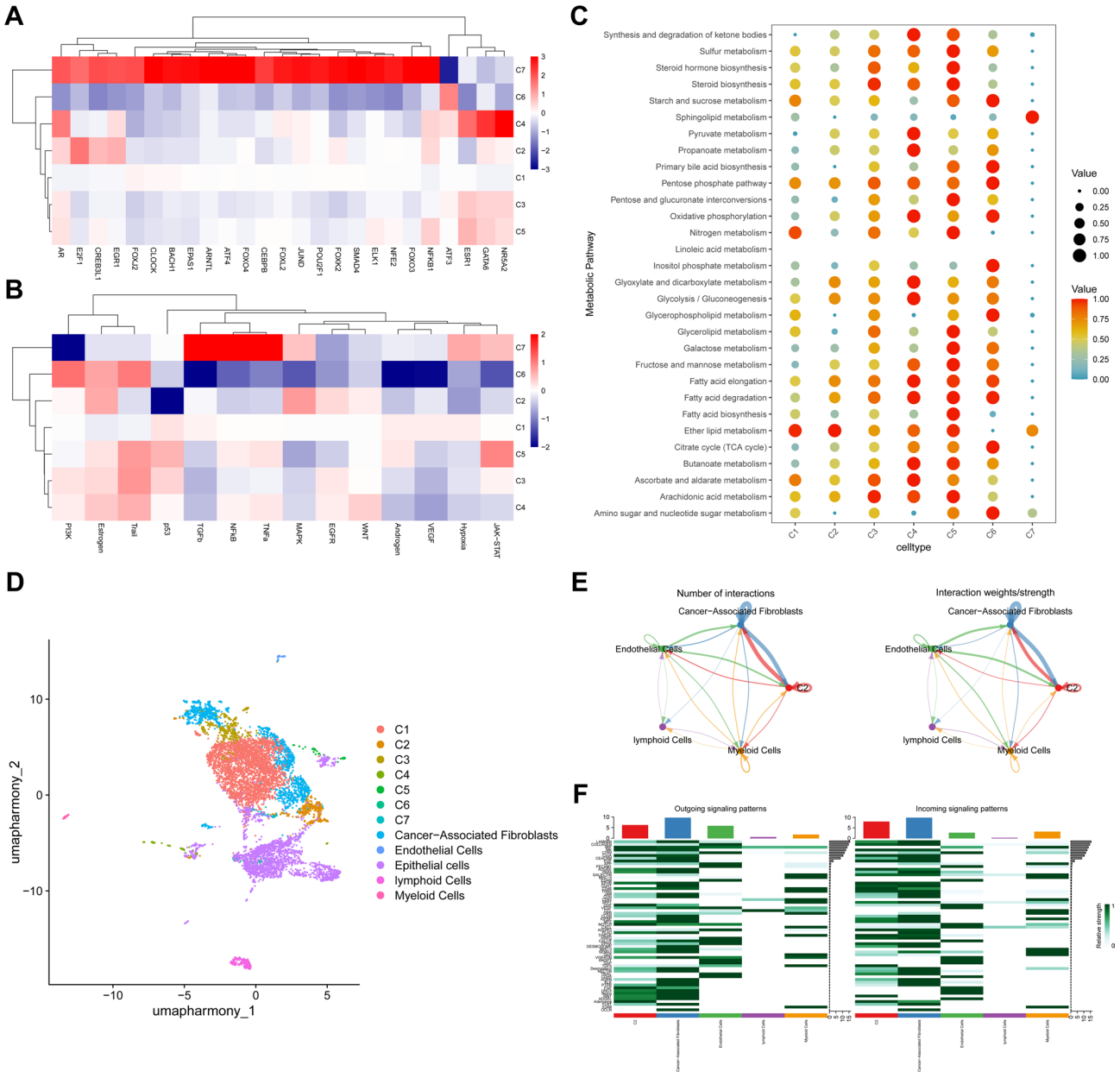


Fig. 4. Examining the stem cell landscape. **A** The activity of TFs in C2 and other clusters. **B** The activation of main pathways in C2 and other clusters. **C** The changes in metabolism of C2 and other clusters. **D** The final clusters and annotations for PAAD. **E** CellChat analysis displays communications among these cell types. The impact is greater with a thicker line. **F** The heatmap exhibits the incoming and outgoing signaling patterns between cells. The deeper color indicates more weight and strength of the regulation of the signaling pathway.

Third, based on the expression of canonical marker genes, Cluster 11 was classified as myeloid cells (CD14, CD163), and Cluster 15 as lymphoid cells (CD3D, CD3E). Among the stromal populations, Clusters 4, 7, 10, 14, and 16 were identified as tumor-associated fibroblasts (CXCL1, MYL9, COL1A1), while Cluster 18 was classified as vascular endothelial cells (VWF, KDR). The final clustering and annotation results are shown in Fig. 4D.

CellChat analysis revealed a complex network of intercellular communications among the various cell types (Fig. 4E). Distinct ligand–receptor interaction patterns were observed between Cluster 2 (C2) and cancer-associated fibroblasts (CAFs), as well as among immune cell populations (Supplementary Fig. 2). In CAFs, secreted factors such as PPIA, MDK, LAMB3, and CEACAM6 were predominantly targeted to receptors on C2 cells, including BSG, NCL, and CD44. The ligand SPP1, primarily produced by myeloid cells, interacted with the integrin complex (ITGAV + ITGB6) on the surface of C2 cells. Further signaling pathway analysis by CellChat revealed that multiple microenvironment-derived pathways—including MK, CypA, EGF, ADGRE, EPHB, CD46, TWEAK, TGF-β, WNT, and adenosine—were involved in modulating the cellular characteristics and behavior of C2 (Fig. 4F).

Building CSCLPI based on machine learning

Differentially expressed genes (DEGs) in Cluster 2 (C2) were extracted and used to generate a gene expression matrix for the TCGA-PAAD cohort. Univariate Cox regression analysis identified 97 candidate CSC-related genes significantly associated with prognosis in PAAD. These genes were subsequently used to construct prognostic models via 117 algorithmic combinations of 10 machine learning methods. The concordance index (C-index) of each model across different cohorts is presented in Fig. 5A. Among all combinations, the model integrating Stepwise Cox regression (both directions) and Random Survival Forest (RSF) achieved the highest average C-index (0.75), indicating the best overall performance (Fig. 5A).

To identify the most informative features, variable importance scores were assessed (Fig. 5B). The top contributors included SHCBP1, CEP55, GMNN, CDCA5, and FAM111B. These five genes were subsequently included in a multivariate Cox regression analysis, which revealed that CEP55 was significantly associated with poor prognosis in PAAD patients (HR = 1.482, 95% CI: 1.056–2.081, $P = 0.023$; Fig. 5C).

The CSCLPI was calculated as follows: CSCLPI = 0.02709 × SHCBP1 + 0.39334 × CEP55 + 0.17426 × GMNN + 0.05385 × CDCA5 + 0.14058 × FAM111B. TCGA-PAAD cases were dichotomized into high- and low-CSCLPI groups using the median score as the cutoff. Kaplan–Meier survival analysis demonstrated a significant survival difference between the two CSCLPI subgroups (Fig. 5D). In the TCGA cohort, time-dependent receiver operating characteristic (ROC) analysis yielded area under the curve (AUC) values of 0.69, 0.75, and 0.78 for predicting 2-, 3-, and 4-year overall survival, respectively (Fig. 5E). The same five-gene CSCLPI formula was applied to the external validation cohort. Kaplan-Meier survival analysis demonstrated a significant difference in survival prediction by our model in both the GEO and CPTAC cohorts. In the GEO cohort, the AUCs for predicting 2-, 3-, and 4-year overall survival were 0.74, 0.85, and 0.80(Fig. 5F). In the CPTAC cohort, the AUCs for predicting 1-, 2-, and 3-year overall survival were 0.70, 0.74, and 0.78 (Supplementary Fig. 3).

Differences in gene mutations, mRNAsi, TMB and functional enrichment among CSCLPI subpopulations

Genomic mutation analysis revealed a significantly higher mutation frequency in the high-CSCLPI group compared to the low-CSCLPI group (95.18% vs. 70.89%). Most genes exhibited higher mutation rates in the high-CSCLPI group, such as KRAS and TP53, whereas mutations in SMAD4 and TTN were more prevalent in the low-CSCLPI group (Fig. 6A and B).The mRNA expression-based stemness index (mRNAsi) is a computational score derived from tumor transcriptomes that reflects the stemness characteristics of cancer cells. Tumors with high stemness typically possess enhanced self-renewal capacity, pluripotency, and therapeutic resistance, and are often associated with aggressive behavior, metastasis, and treatment failure[27]. By analyzing transcriptomic profiles, we calculated mRNAsi scores for each PAAD patient. The low-CSCLPI group showed significantly lower mRNAsi levels ($P < 0.001$) (Fig. 6C), and mRNAsi scores were positively correlated with CSCLPI scores (Spearman correlation = 0.328, $P < 0.001$) (Fig. 6D). Furthermore, we calculated the tumor mutational burden (TMB) for all tumor samples. The high-CSCLPI group exhibited significantly elevated TMB compared to the low-CSCLPI group ($P < 0.001$) (Fig. 6E), and a strong positive correlation was observed between TMB levels and CSCLPI scores (Spearman correlation = 0.391, $P < 0.001$) (Fig. 6F). To explore the molecular features underlying the CSCLPI stratification, GO and KEGG enrichment analyses were performed on differentially expressed genes between the two groups. The results indicated significant enrichment in biological processes related to protein synthesis, digestion, antigen processing, and immune response, suggesting that CSCLPI subgroups differ substantially in multiple biological dimensions.

To further verify the expression of CSCLPI gene at the protein level, we queried the immunohistochemical data of CSCLPI gene in Human Protein Atlas (HPA) database. The results showed that CEP55, GMNN and CDCA5 were expressed at very low or undetectable levels in normal pancreatic tissues, whereas they all showed moderate to strong positive staining in pancreatic cancer tissues. In contrast, SHCBP1 and FAM111B showed weak staining in both normal and tumor tissues, and no significant differences were observed (Supplementary Fig. 4).

Effect of tumor microenvironment on the biological characteristics of CSCLPI subpopulations

To assess differences in tumor-infiltrating components between CSCLPI subgroups, we first applied the ESTIMATE algorithm. The results showed that tumor purity was significantly higher in the high-CSCLPI group, whereas the stromal and immune scores were lower compared to the low-CSCLPI group (Fig. 7A). Using ssGSEA, we further evaluated immune cell infiltration and found that, with the exception of regulatory T cells (Tregs), all other immune cells had significantly higher enrichment scores in the low-CSCLPI group

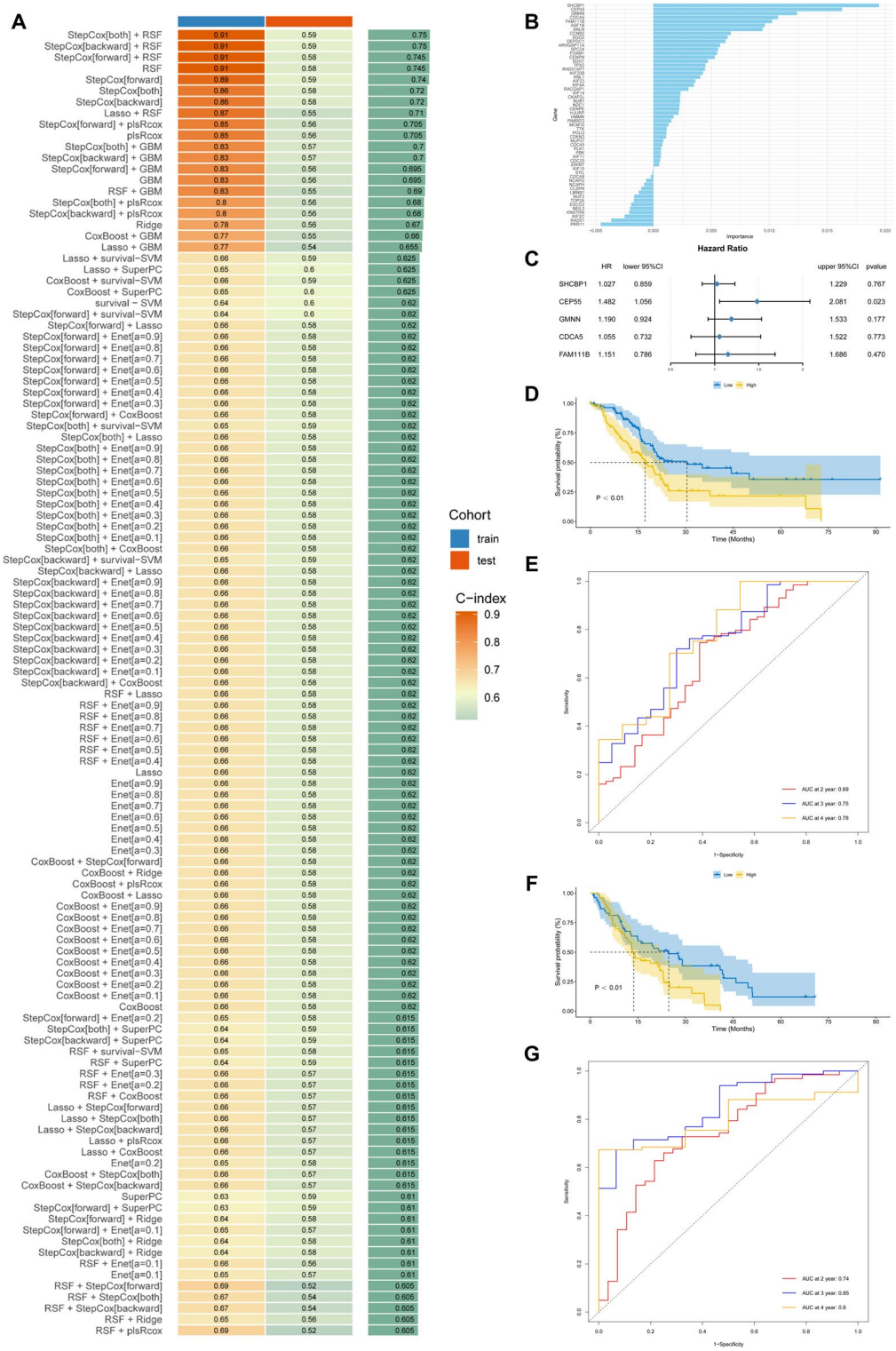


Fig. 5. Construction and validation of CSCLPI. **A** The C-index of a total of 117 prediction models. The StepCox[both] + RSF method was utilized to construct the model, which had the highest C-index of 0.75. **B** Ranking of variables' importance. **C** The results of multivariate Cox regression analysis. **D** The result of the KM survival analysis of CSCLPI in the TCGA cohort. **E** The results of the ROC curve for CSCLPI in predicting 2-year, 3-year, and 4-year survival in the TCGA cohort. **F** The result of the KM survival analysis of CSCLPI in the GEO cohort. **G** The results of the ROC curve for CSCLPI in predicting 2-year, 3-year, and 4-year survival in the GEO cohort.

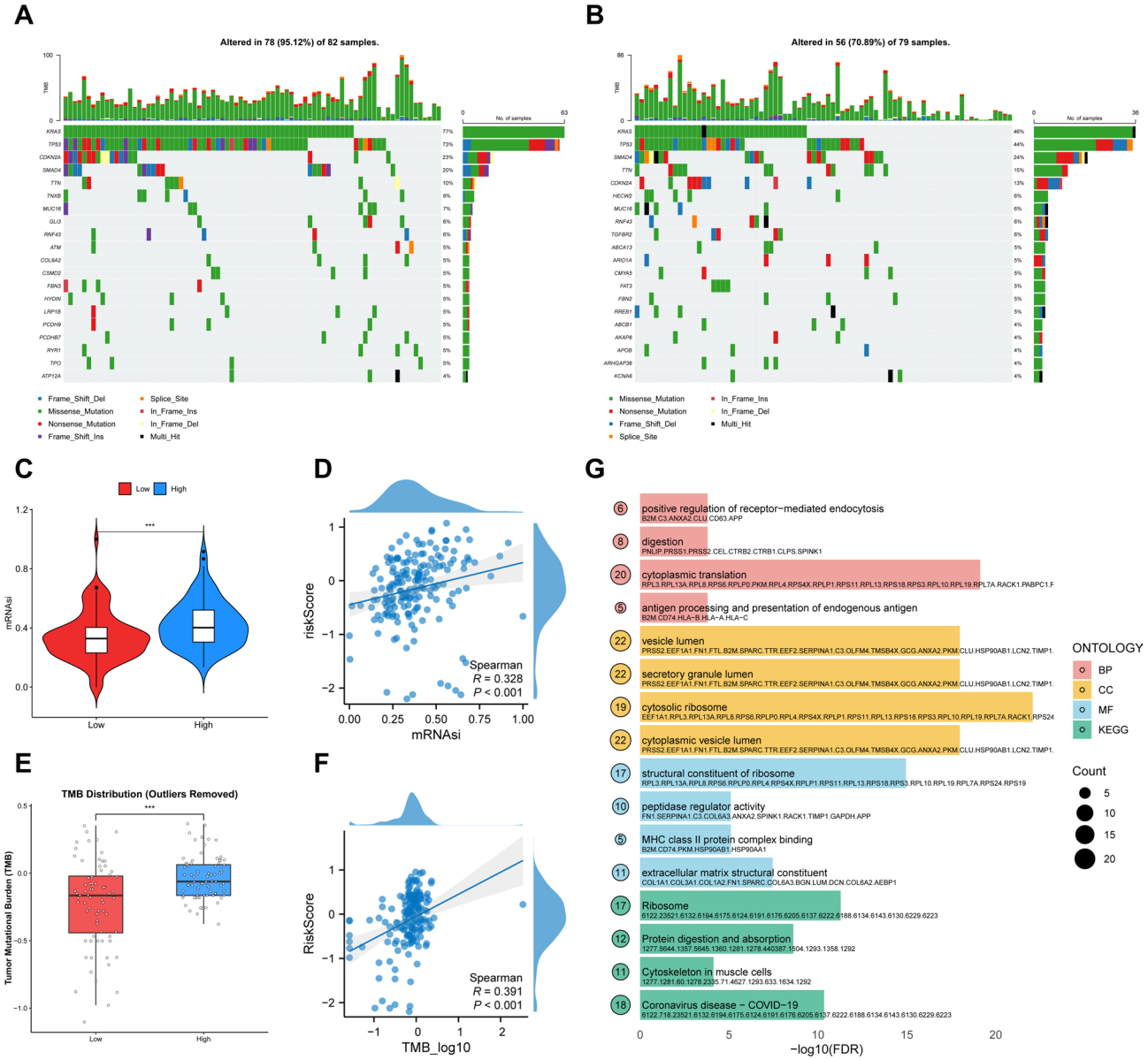


Fig. 6. Characteristics analysis for CSCLPI subgroups. **A**, **B** The waterfall plot demonstrated mutational genes in the high CSCLPI group (**A**) and in the low CSCLPI group (**B**). **C** The comparison of the mRNAsi between CSCLPI subgroups. **D** The Spearman correlation of the CSCLPI score with mRNAsi. **E** Comparison of TMB between CSCLPI subgroups. **F** The Spearman correlation of the CSCLPI score with TMB levels. **G** GO and KEGG analysis of differential genes between CSCLPI subgroups.

(Fig. 7B). Similarly, immune pathway analysis revealed that, aside from MHC class I, parainflammation, and the type I interferon response, most immune-related pathways were more active in the low-CSCLPI group, with statistically significant differences (Fig. 7C). The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm, which predicts responses to immune checkpoint blockade, indicated that both TIDE and CD274 scores were significantly elevated in the high-CSCLPI group compared to the low-CSCLPI group (Fig. 7D, E). Drug sensitivity analysis using oncoPredict revealed that low-CSCLPI patients were more sensitive to commonly used chemotherapeutic agents such as gemcitabine, navitoclax-gemcitabine, fluorouracil, and oxaliplatin (Fig. 7E).

Nomogram structure and inspection

Clinical characteristics are key determinants of patient prognosis. A heatmap was generated to visualize the association between the five key genes and various clinical features (Fig. 8A). Both univariate and multivariate Cox regression analyses were performed to evaluate the prognostic significance of CSCLPI, age, gender, tumor stage, T stage, and N stage, identifying independent prognostic factors (Fig. 8B, C). Multivariate Cox analysis revealed that CSCLPI (HR: 0.534, 95% CI: 0.350–0.817, $P = 0.004$) and N stage (HR: 1.905, 95% CI: 1.101–3.295, $P = 0.021$) were independent predictors of overall survival in PAAD patients. These variables, including

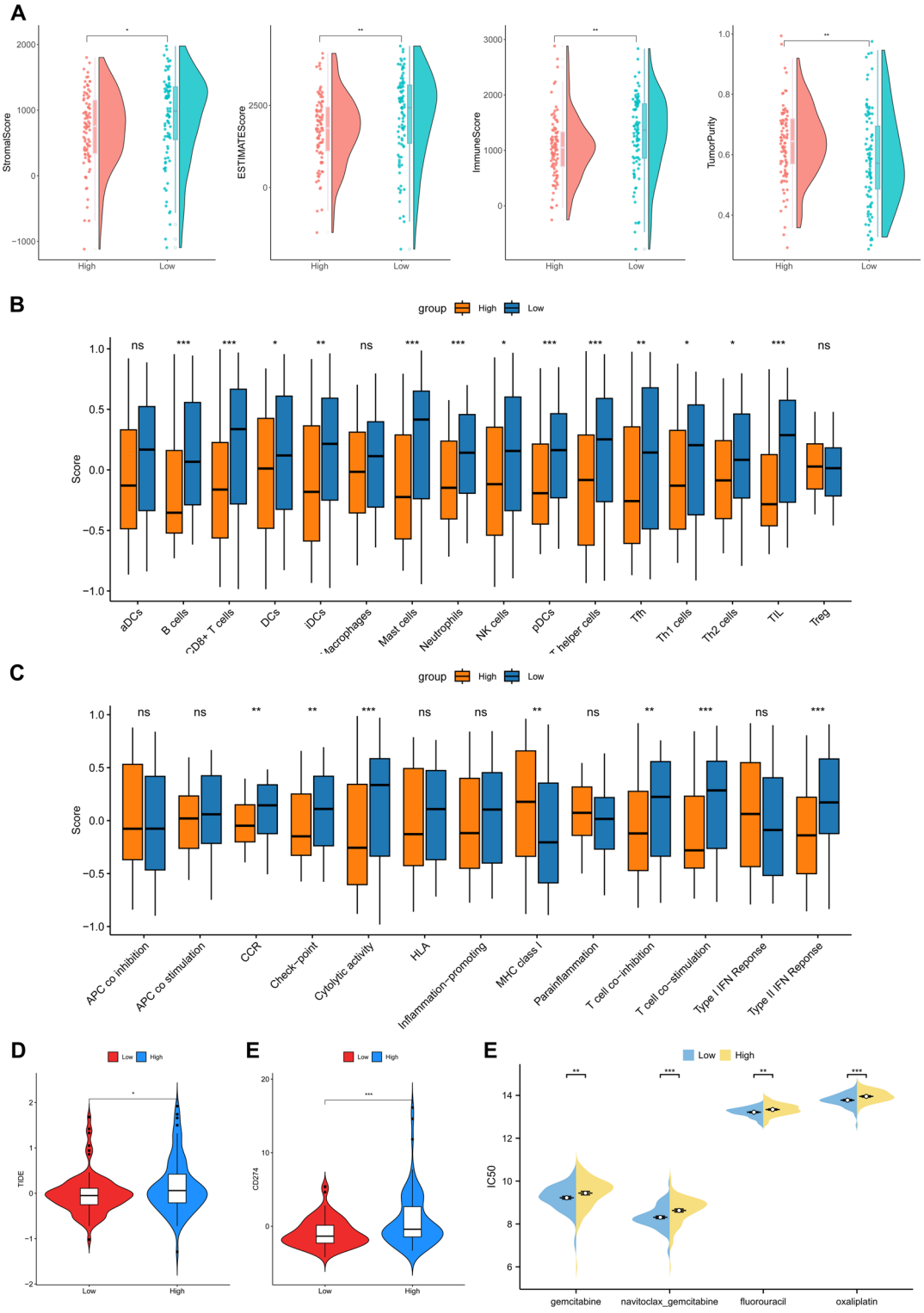


Fig. 7. Tumor microenvironment analysis. **A** ESTIMATE analysis between CSCLPI subgroups. **B**, **C** The immune cell infiltration and immune-related pathways made distinctions between CSCLPI subgroups. **D**, **E** TIDE analysis between CSCLPI subgroups. The differences in the TIDE (**D**) and CD274 (**F**). A Drug sensitivity analysis between CSCLPI subgroups.

CSCLPI, age, gender, tumor stage, T stage, and N stage, were subsequently integrated to construct a nomogram for predicting 2-, 3-, and 4-year overall survival in PAAD (Fig. 8D). Calibration plots for 2-, 3-, and 4-year survival showed good agreement between predicted and observed outcomes (Fig. 8E). Kaplan–Meier survival analysis demonstrated significant differences in survival between high- and low-risk groups based on the

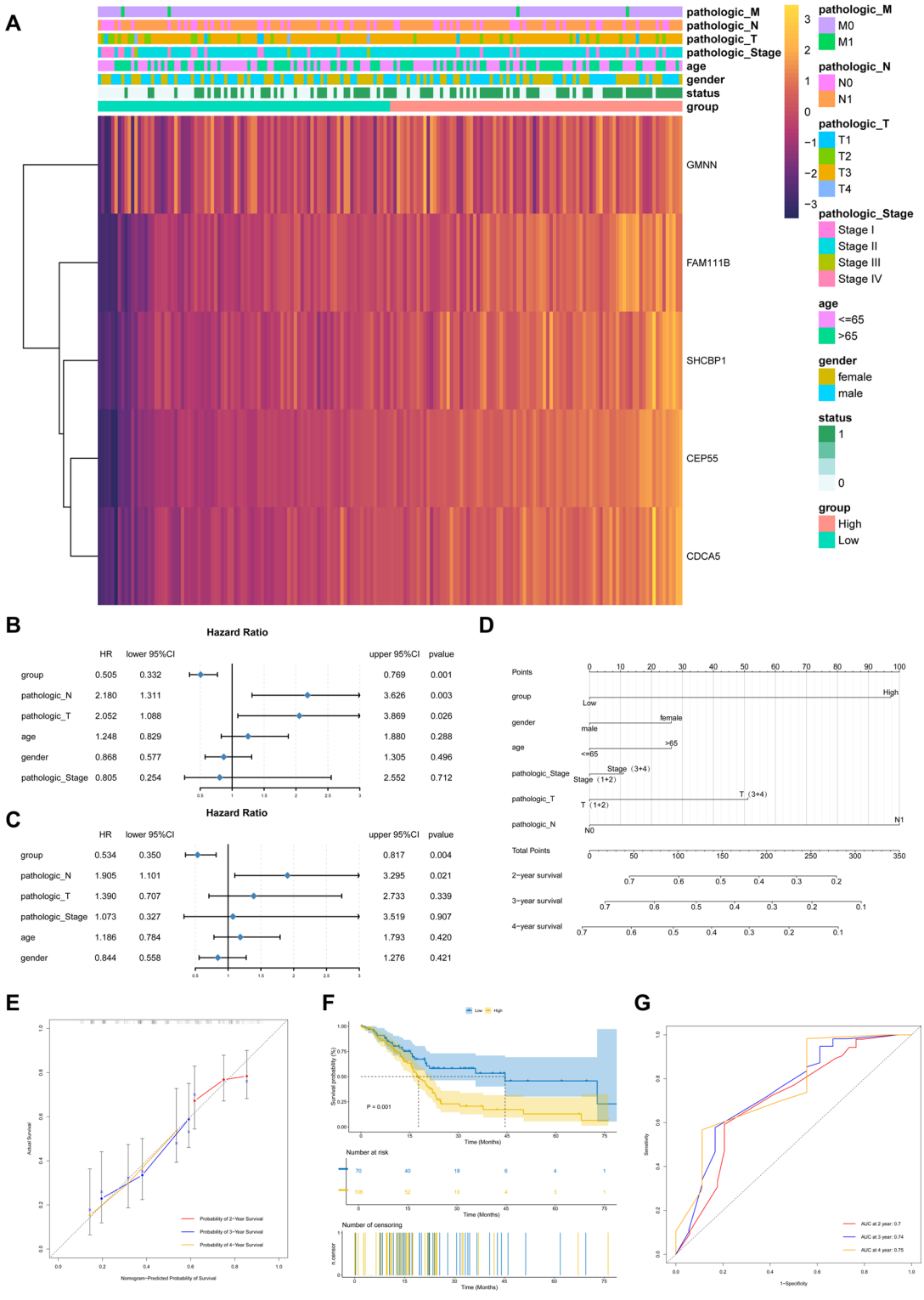


Fig. 8. Construction of a predictive nomogram. **A** The heatmap showed the relationship between five key genes and clinical characteristics. **B**, **C** Univariate and multivariate Cox regression analyses results. **D** A nomogram for predicting the prognosis of 2, 3, and 4 years for patients with PAAD. **E** Calibration curve. **F** Survival analysis. **G** ROC curve of nomogram.

total nomogram score (Fig. 8F). The AUCs for 2-, 3-, and 4-year survival prediction were 0.70, 0.74, and 0.75, respectively (Fig. 8G), indicating that the nomogram had good predictive accuracy.

Discussion

CSCs have been identified in a variety of solid tumors, including breast cancer, glioma, hepatocellular carcinoma, head and neck cancer, colorectal cancer, and melanoma. These cells possess the ability to self-renew, differentiate, and initiate tumorigenesis, and they play a crucial role in tumor development, heterogeneity, and therapeutic resistance[28]. However, the biological characteristics of CSCs in PAAD and their dynamic interactions with other cells in the tumor microenvironment (TME) remain largely uncharacterized.

In this study, we identified CSCs from malignant epithelial cells in PAAD using single-cell transcriptomic data and the CytoTRACE algorithm. We systematically investigated their activated transcription factors, signaling pathways, and metabolic features, and elucidated their interactions with other cell types in the TME. Furthermore, we applied multiple machine learning approaches to screen CSC-related genes and constructed a prognostic model, the CSCLPI, which demonstrated robust predictive performance across cohorts. Our findings suggest that CSCLPI may serve as a reliable biomarker for assessing disease progression, therapeutic response, and prognosis in PAAD.

Aberrant activation of transcription factors (TFs) is a critical step in tumor initiation and progression. In this study, we found that E2F1, AR, and EGR1 were significantly activated in CSCs of PAAD. Previous studies have confirmed that TFs regulate CSCs and thereby influence tumorigenesis and cancer progression.

In breast cancer, miR-20b-5p enhances CSC malignancy by bidirectionally regulating CCND1 and E2F1[29]. In prostate cancer, MUC1-C integrates E2F1 to induce NOTCH1 and NANOG, promoting CSC self-renewal[30]. C.-H. Chang et al. demonstrated that E2F1 and E2F4 suppress WNT ligand expression (e.g., WNT7A, WNT7B, WNT10A, WNT4) by downregulating pRb/RBL2, thereby modulating CSC self-renewal, chemoresistance, and invasiveness in pancreatic and breast cancers[31]. Jiang et al. reported that E2F1 enhances the tumorigenic capacity of glioblastoma (GBM) CSCs by regulating USP5 and stabilizing OCT4, and targeting the E2F1–USP5–OCT4 axis may offer a novel therapeutic strategy for GBM[32].

Androgen deprivation is a cornerstone of prostate cancer therapy, yet CSCs can resist this treatment by upregulating AR expression[33]. In endometrial cancer, the WTAP–EGR1–PTEN axis is essential for CSC self-renewal, chemoresistance, and tumorigenesis. WTAP knockdown reduces EGR1 expression, subsequently downregulating PTEN and disrupting regulatory homeostasis, thereby enhancing CSC aggressiveness and stemness[34]. In hepatocellular carcinoma, EGR1 downregulation suppresses miR-101 transcription, leading to ANXA2 release and ERK pathway activation, which enhances CSC properties[35]. Conversely, in breast cancer CSCs, upregulation of EGR1 promotes invasion, migration, and stemness[36] Collectively, these findings underscore the pivotal roles of TFs in CSC maintenance and cancer-associated phenotypes, highlighting the importance of elucidating their regulatory functions in PAAD.

In addition, numerous signaling pathways have been shown to regulate CSC stemness, survival, proliferation, and differentiation, thereby influencing tumor progression and therapeutic resistance. In our study, the Estrogen and MAPK signaling pathways were found to be abnormally activated in CSCs of PAAD.

SIX1 is significantly upregulated in breast cancer and promotes CSC expansion, epithelial–mesenchymal transition (EMT), and tumor progression via activation of the Estrogen and MAPK pathways[37]. MAPK is a downstream effector of various CSC-specific signaling pathways and plays a vital role in CSC development and maintenance. It is also involved in tumor metastasis and chemoresistance; therefore, targeting MAPK signaling may sensitize tumors to chemotherapy and prevent metastatic progression[38]. One study demonstrated that tipranavir targets PRSS23 and induces apoptosis in gastric CSCs through the MKK3/p38 MAPK–IL24 pathway, thereby inhibiting tumor growth[39]. However, the precise regulatory mechanisms of the Estrogen and MAPK pathways in PAAD CSCs remain to be fully elucidated.

Notably, although the C2 subgroup showed strong activation of stemness-associated transcription factors such as E2F1, AR, and EGR1, as well as aberrant activation of Estrogen and MAPK signaling, we did not observe clear upregulation of canonical embryonic developmental pathways such as WNT or NOTCH in this cluster[40,41]. This phenomenon suggests substantial heterogeneity in the stemness-regulatory mechanisms of PAAD CSCs: not all CSC subpopulations rely on the traditional WNT/NOTCH axis to sustain self-renewal. Instead, C2 may acquire its stemness and proliferative advantages through "non-canonical stemness pathways," including MAPK and Estrogen signaling[42]. This feature aligns with recent concepts of CSC pathway plasticity and compensatory signaling mechanisms, further supporting the notion that pancreatic cancer CSCs may adopt multiple parallel or alternative routes to maintain stemness[43].

Our study also identified surface molecular interactions between CSCs, cancer-associated fibroblasts (CAFs), and immune cells. Several signaling pathways originating from the tumor microenvironment—including MK, CypA, EGF, ADGRE, EPHB, CD46, TWEAK, TGF-β, WNT, and adenosine—were found to be involved in the regulation of C2 characteristics. These findings provide deeper insights into how microenvironmental cues influence CSC behavior.

Collectively, our results suggest that the regulation of CSC phenotypes is a complex process involving multiple molecular interactions and signaling cascades. The roles of CAFs and immune cells in this regulation warrant further attention and investigation.

Using the StepCox[both] + RSF method, we identified five key genes (SHCBP1, CEP55, GMNN, CDCA5, and FAM111B). Among them, CEP55 has been previously implicated in poor prognosis and/or chemoresistance across various cancer types[44]. In glioma, CEP55 significantly enhances tumor cell invasiveness and migratory capacity, while also maintaining the self-renewal potential and stemness phenotype of glioma stem-like cells, thereby contributing to tumorigenicity and treatment resistance[45].

Colorectal CSCs are highly sensitive to both cytotoxic T lymphocytes (CTLs) and CEP55-targeted immunotherapy. CEP55 may facilitate the adaptive immune response, enabling the precise elimination of colorectal CSCs[46]. Tumor vaccine strategies targeting CEP55 have been proposed as a promising therapeutic approach against chemoresistant colorectal CSCs[47]. In the context of pancreatic cancer, CEP55 is closely associated with CSC regulation and patient prognosis, and may serve as a potential therapeutic target.

Notably, we further validated the protein expression of the genes constituting the CSCLPI. Querying the HPA database revealed that CEP55, GMNN, and CDCA5 exhibited markedly higher protein levels in pancreatic cancer tissues than in normal tissues, consistent with our mRNA-level findings. This provides more direct evidence supporting their pro-tumorigenic or stemness-related roles in pancreatic cancer and suggests their potential utility as IHC-based diagnostic or prognostic biomarkers. In contrast, SHCBP1 and FAM111B did not show obvious differences in protein abundance, which may be attributed to insufficient antibody sensitivity, low protein abundance, or post-transcriptional/post-translational regulatory mechanisms that destabilize the proteins. Overall, the concordant protein-level expression patterns of several CSCLPI genes strengthen the biological credibility of our model and provide preliminary support for translating this gene-expression-based scoring system into a clinically feasible, IHC-based testing strategy.

Based on machine learning, we constructed a prognostic model named CSCLPI, which demonstrated strong performance in survival prediction and revealed its potential clinical value in molecular characterization and treatment response.

Further mutational landscape analysis showed that the high CSCLPI group exhibited a significantly higher mutation frequency compared to the low CSCLPI group, with several key driver genes displaying notably higher mutation rates in the high-score subgroup. Additionally, CSCLPI scores were positively correlated with tumor mutational burden (TMB), a well-established biomarker for predicting immune checkpoint blockade (ICB) efficacy across various cancer types[48]. In terms of the immune microenvironment, tumors in the high CSCLPI group showed elevated expression of CD274 (encoding PD-L1) and significantly higher TIDE scores. Upregulation of PD-L1 typically reflects a tumor's ability to suppress T cell activity via immune checkpoint pathways[49], while an increased TIDE score suggests potential T cell dysfunction or exclusion, both of which contribute to weakened antitumor immune responses[50]. These findings imply that the high CSCLPI subtype, despite its higher immunogenic potential, may possess more intricate immune evasion mechanisms. For such tumors, combined treatment strategies may be required to enhance antitumor immunity and overcome immune escape.

Notably, drug sensitivity prediction indicated that the low CSCLPI group was more responsive to several commonly used chemotherapeutic agents, suggesting that patients in this subgroup may derive greater benefit from standard chemotherapy.

In conclusion, the CSCLPI score serves not only as a prognostic indicator but also reflects the immune status and drug response characteristics of tumors. Patients in the high CSCLPI group may have limited responses to immunotherapy and should be considered for combination treatment strategies to optimize therapeutic outcomes. Conversely, those in the low CSCLPI group are more likely to benefit from standard chemotherapy regimens. These findings provide a theoretical foundation and potential clinical guidance for the development of individualized precision treatment strategies.

To facilitate clinical application, we developed a nomogram integrating the CSCLPI score with clinicopathological characteristics. This nomogram demonstrated favorable predictive accuracy for prognosis, providing robust evidence for the survival prognostic value of the CSCLPI score and its potential utility in guiding individualized treatment strategies.This study has several limitations. First, as the data were obtained from publicly available databases rather than institutional datasets, external validation of the CSCLPI using our center's clinical cohort is recommended. Second, the biological functions of CEP55 related to CSCs properties in PAAD require further exploration through in vivo and in vitro models.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 4 July 2025; Accepted: 15 December 2025
Published online: 23 December 2025

References

1. Li, X. et al. Integrated transcriptomics, proteomics and metabolomics-based analysis uncover TAM2-associated Glycolysis and pyruvate metabolic remodeling in pancreatic cancer. *Front. Immunol.* **14**, 1170223. https://doi.org/10.3389/fimmu.2023.1170223 (2023).
2. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249. https://doi.org/10.3322/caac.21660 (2021).
3. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *J. Clin.* **73**, 17–48. https://doi.org/10.3322/caac.21763(2023) (2023).
4. Labori, K. J. et al. Neoadjuvant FOLFIRINOX versus upfront surgery for resectable pancreatic head cancer (NORPACT-1): a multicentre, randomised, phase 2 trial. *Lancet Gastroenterol. Hepatol.* **9**, 205–217. https://doi.org/10.1016/S2468-1253(23)00405-3 (2024).
5. Mizrahi, J. D., Surana, R., Valle, J. W. & Shroff, R. T. Pancreatic cancer. *Lancet* **395**, 2008–2020. https://doi.org/10.1016/S0140-6736(20)30974-0 (2020).
6. Biserova, K., Jakovlevs, A., Uljanovs, R. & Strumfa, I. Cancer stem cells: significance in Origin, pathogenesis and treatment of glioblastoma. *Cells* **10**, 621. https://doi.org/10.3390/cells10030621 (2021).
7. Bubin, R., Uljanovs, R. & Strumfa, I. Cancer stem cells in pancreatic ductal adenocarcinoma. *Int. J. Mol. Sci.* **24**, 7030. https://doi.org/10.3390/ijms24087030 (2023).

8. Li, C. et al. Identification of pancreatic cancer stem cells. *Cancer Res.* **67**, 1030–1037. https://doi.org/10.1158/0008-5472.CAN-06-2030 (2007).
9. Ercan, G., Karlitepe, A. & Ozpolat, B. Pancreatic cancer stem cells and therapeutic approaches. *Anticancer Res.* **37**, 2761–2775. https://doi.org/10.21873/anticanres.11628 (2017).
10. Domenichini, A. et al. Pancreatic cancer tumorspheres are cancer stem-like cells with increased chemoresistance and reduced metabolic potential. *Adv. Biol. Regul.* **72**, 63–77. https://doi.org/10.1016/j.jbior.2019.02.001 (2019).
11. Chen, K. et al. Single-cell transcriptome profiling of primary tumors and paired organoids of pancreatobiliary cancer. *Cancer Lett.* **582**, 216586. https://doi.org/10.1016/j.canlet.2023.216586 (2024).
12. Stuart, T. et al. Comprehensive integration of Single-Cell data. *Cell* **177**, 1888–1902e21. https://doi.org/10.1016/j.cell.2019.05.031 (2019).
13. Gao, R. et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **39**, 599–608. https://doi.org/10.1038/s41587-020-00795-2 (2021).
14. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods.* **14**, 979–982. https://doi.org/10.1038/nmeth.4402 (2017).
15. Kang, M. et al. Mapping single-cell developmental potential in health and disease with interpretable deep learning. *BioRxiv* https://doi.org/10.1101/2024.03.19.585637 (2024). 2024.03.19.585637.
16. Badia-I-Mompel, P. et al. DecoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv.* **2**, vbac016. https://doi.org/10.1093/bioadv/vbac016 (2022).
17. Wu, Y. et al. Spatiotemporal immune landscape of colorectal cancer liver metastasis at Single-Cell level. *Cancer Discov.* **12**, 134–153. https://doi.org/10.1158/2159-8290.CD-21-0316 (2022).
18. Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat. Protoc.* **20**, 180–219. https://doi.org/10.1038/s41596-024-01045-4 (2025).
19. Mounir, M. et al. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Comput. Biol.* **15**, e1006701. https://doi.org/10.1371/journal.pcbi.1006701 (2019).
20. Wu, T. et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. (Camb).* **2**, 100141. https://doi.org/10.1016/j.xinn.2021.100141 (2021).
21. Gene Ontology Consortium. Gene ontology consortium: going forward. *Nucleic Acids Res.* **43**, D1049–1056. https://doi.org/10.1093/nar/gku1179 (2015).
22. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551. https://doi.org/10.1093/nar/gkaa970 (2021).
23. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612. https://doi.org/10.1038/ncomms3612 (2013).
24. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A.* **102**, 15545–15550. https://doi.org/10.1073/pnas.0506580102 (2005).
25. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558. https://doi.org/10.1038/s41591-018-0136-1 (2018).
26. Maeser, D., Gruener, R. F. & Huang, R. S. OncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* **22**, bbab260. https://doi.org/10.1093/bib/bbab260 (2021).
27. Weng, M. et al. mRNAsi-related metabolic risk score model identifies poor prognosis, immunoevasive contexture, and low chemotherapy response in colorectal cancer patients through machine learning. *Front. Immunol.* **13**, 950782. https://doi.org/10.3389/fimmu.2022.950782 (2022).
28. Liu, Y. & Wang, H. Biomarkers and targeted therapy for cancer stem cells. *Trends Pharmacol. Sci.* **45**, 56–66. https://doi.org/10.1016/j.tips.2023.11.006 (2024).
29. Xia, L. et al. Oncogenic miR-20b-5p contributes to malignant behaviors of breast cancer stem cells by bidirectionally regulating CCND1 and E2F1. *BMC Cancer.* **20**, 949. https://doi.org/10.1186/s12885-020-07395-y (2020).
30. Hagiwara, M. et al. MUC1-C activates the BAF (mSWI/SNF) complex in prostate cancer stem cells. *Cancer Res.* **81**, 1111–1122. https://doi.org/10.1158/0008-5472.CAN-20-2588 (2021).
31. Chang, C-H. et al. The pRb/RBL2-E2F1/4-GCN5 axis regulates cancer stem cell formation and G0 phase entry/exit by paracrine mechanisms. *Nat. Commun.* **15**, 3580. https://doi.org/10.1038/s41467-024-47680-z (2024).
32. Jiang, X. et al. E2F1-regulated USP5 contributes to the tumorigenic capacity of glioma stem cells through the maintenance of OCT4 stability. *Cancer Lett.* **593**, 216875. https://doi.org/10.1016/j.canlet.2024.216875 (2024).
33. Mu, J. et al. Human intermediate prostate cancer stem cells contribute to the initiation and development of prostate adenocarcinoma. *Stem Cell. Res. Ther.* **15**, 296. https://doi.org/10.1186/s13287-024-03917-8 (2024).
34. Wang, B. et al. WTAP/IGF2BP3 mediated m6A modification of the EGR1/PTEN axis regulates the malignant phenotypes of endometrial cancer stem cells. *J. Exp. Clin. Cancer Res.* **43**, 204. https://doi.org/10.1186/s13046-024-03120-w (2024).
35. Ma, S. et al. A novel regulatory loop miR-101/ANXA2/EGR1 mediates malignant characteristics of liver cancer stem cells. *Carcinogenesis* https://doi.org/10.1093/carcin/bgaa055 (2021).
36. Tang, T. et al. Protease nexin I is a feedback regulator of EGF/PKC/MAPK/EGR1 signaling in breast cancer cells metastasis and stemness. *Cell. Death Dis.* **10**, 649. https://doi.org/10.1038/s41419-019-1882-9 (2019).
37. Guo, L. et al. SIX1 amplification modulates stemness and tumorigenesis in breast cancer. *J. Transl Med.* **21**, 866. https://doi.org/10.1186/s12967-023-04679-2 (2023).
38. Kudaravalli, S., den Hollander, P. & Mani, S. A. Role of p38 MAP kinase in cancer stem cells and metastasis. *Oncogene* **41**, 3177–3185. https://doi.org/10.1038/s41388-022-02329-3 (2022).
39. Xiong, J-X. et al. Targeting PRSS23 with Tipranavir induces gastric cancer stem cell apoptosis and inhibits growth of gastric cancer via the MKK3/p38 MAPK-IL24 pathway. *Acta Pharmacol. Sin.* **45**, 405–421. https://doi.org/10.1038/s41401-023-01165-9 (2024).
40. Zhao, H. et al. Wnt signaling in colorectal cancer: pathogenic role and therapeutic target. *Mol. Cancer.* **21**, 144. https://doi.org/10.1186/s12943-022-01616-7 (2022).
41. Shi, Q. et al. Notch signaling pathway in cancer: from mechanistic insights to targeted therapies. *Signal. Transduct. Target. Ther.* **9**, 128. https://doi.org/10.1038/s41392-024-01828-x (2024).
42. Ma, H. et al. HMGB3 promotes the malignant phenotypes and stemness of epithelial ovarian cancer through the MAPK/ERK signaling pathway. *Cell. Commun. Signal.* **21**, 144. https://doi.org/10.1186/s12964-023-01172-7 (2023).
43. Bhat, G. R. et al. Cancer cell plasticity: from cellular, molecular, and genetic mechanisms to tumor heterogeneity and drug resistance. *Cancer Metastasis Rev.* **43**, 197–228. https://doi.org/10.1007/s10555-024-10172-z (2024).
44. Jeffery, J., Sinha, D., Srihari, S., Kalimutho, M. & Khanna, K. K. Beyond cytokinesis: the emerging roles of CEP55 in tumorigenesis. *Oncogene* **35**, 683–690. https://doi.org/10.1038/onc.2015.128 (2016).
45. Li, F. et al. CEP55 promoted the migration, invasion and neuroshpere formation of the glioma cell line U251. *Neurosci. Lett.* **705**, 80–86. https://doi.org/10.1016/j.neulet.2019.04.038 (2019).
46. Gao, X-Y. & Wang, X-L. An adoptive T cell immunotherapy targeting cancer stem cells in a colon cancer model. *J. BUON.* **20**, 1456–1463 (2015).
47. Inoda, S. et al. Cytotoxic T lymphocytes efficiently recognize human colon cancer stem-like cells. *Am. J. Pathol.* **178**, 1805–1813. https://doi.org/10.1016/j.ajpath.2011.01.004 (2011).

48. Chan, T. A. et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann. Oncol.* **30**, 44–56. https://doi.org/10.1093/annonc/mdy495 (2019).
49. Wu, L. & Li, B. A case of severe diabetic ketoacidosis associated with pembrolizumab therapy in a patient with metastatic melanoma. *Diabetes Metab. Syndr. Obes.* **14**, 753–757. https://doi.org/10.2147/DMSO.S297709 (2021).
50. Yao, J. et al. Construction of a novel immune response prediction signature to predict the efficacy of immune checkpoint inhibitors in clear cell renal cell carcinoma patients. *Heliyon* **9**, e15925. https://doi.org/10.1016/j.heliyon.2023.e15925 (2023).

Author contributions
JF and HJ wrote and revised the manuscript.MW, BZ and JZ collected data. CX summarized the results. All authors reviewed and approved the final version of the manuscript.

Funding
This work was supported by 900th Hospital of Joint Logistics Support Force of People's Liberation Army(2023GK01).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-33029-z.

Correspondence and requests for materials should be addressed to C.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025