Single-cell pseudotime and cell communication analysis of pancreatic cancer

Chengming Ni[1,2,A,D], Xiaohang Wang[3,B], Zhensheng Cai[1,2,B], Yang Chen[1,2,C], Huan Wang[1,2,C], Qianqian Wang[1,2,B,C], Hao Lin[4,B], Yunting Zhou[5,B], Yang Yuan[1,2,B], Bo Sun[6,C], Zilin Sun[1,2,F]

[1] Department of Endocrinology, Zhongda Hospital, Institute of Diabetes, School of Medicine, Southeast University, Nanjing, China

[2] Institute of Diabetes, School of Medicine, Southeast University, Nanjing, China

[3] Institute of Translational Medicine, Jiangsu Key Laboratory of Integrated Traditional Chinese and Western Medicine for Prevention and Treatment of Senile Diseases, Medical College, Yangzhou University, China

[4] Department of Clinical Science and Research, Zhongda Hospital, School of Medicine, Southeast University, Nanjing, China

[5] Department of Endocrinology, Nanjing First Hospital, Nanjing Medical University, China

[6] State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, China

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;
D – writing the article; E – critical revision of the article; F – final approval of the article

Abstract

Background. Pancreatic cancer (PC) is among the most aggressive and lethal malignancies, characterized by development within a complex tumor microenvironment (TME) that includes a desmoplastic stroma composed of extracellular matrix (ECM) and various cellular components.

Objectives. This study aims to elucidate the cellular and molecular mechanisms regulating PC progression through an integrated analysis of single-cell pseudotime trajectories and intercellular communication.

Materials and methods. We constructed pseudotime trajectories using single-cell RNA sequencing (scRNA-seq) data from PC tissues to trace the developmental progression of cancer cells. Transitional cell states and critical genes involved in the shift from early-to-advanced disease stages were identified. Through a comprehensive analysis, we pinpointed key transcription factors and signaling pathways implicated in tumor progression. Expression of stemness-associated genes in pancreatic stellate cells (PSC) was validated using immunofluorescence and transmission electron microscopy (TEM). Additionally, cell–cell communication analysis was performed to examine interactions within the TME, with particular emphasis on ligand–receptor pairings.

Results. Our analysis identified key transcription factors and signaling pathways that drive the cellular transitions associated with cancer progression. The findings revealed extensive intercellular crosstalk between cancer cells, stromal fibroblasts, and diverse immune cell subpopulations. Notably, the study underscored the distinct functional contributions of these cell populations to tumor development, immune evasion and metastatic dissemination.

Conclusions. The study uncovers the complex cellular diversity and intercellular crosstalk in PC, providing novel avenues for therapeutic interventions and early predictive markers in diagnosis. These findings support the potential for more targeted, personalized treatment strategies in combating PC.

Key words: pancreatic cancer, core genes, scRNA-seq, cell communication

Highlights

- Single-cell RNA sequencing (scRNA-seq) exposes fine-scale cellular heterogeneity and lineage routes in pancreatic cancer (PC).
- Pseudotime trajectory analysis spots the genes that drive PC progression from early lesions to invasive disease.
- High-resolution tumor-microenvironment mapping reveals cross-talk among malignant, stromal and immune cells, highlighting druggable signaling hubs.
- Immunofluorescence and TEM confirm stemness markers in pancreatic stellate cells (PSC), validating a stem-like niche within the tumor stroma.
- The resulting single-cell atlas uncovers novel biomarkers and therapeutic targets for earlier diagnosis and precision treatment of PC.

Background

Pancreatic cancer (PC) is associated with one of the highest mortality rates due to its aggressive nature and poor overall prognosis. The 5-year survival rate for patients with PC remains approx. 9%, with minimal improvement over recent decades. This is partly attributed to late detection and a tumor microenvironment (TME) that supports cancer progression and therapy resistance. Understanding the cellular and molecular events associated with the emergence of PC is crucial for developing more effective diagnostic and therapeutic strategies.[1–3]

Emerging technologies, such as Single-cell RNA sequencing (scRNA-seq), have dramatically transformed our ability to uncover the complexity of cancer in detail. Unlike bulk RNA sequencing, which averages gene expression profiles over a large population of cells, scRNA-seq allows for the analysis of gene expression at the individual cell level. This technology has revealed the diversity of cellular composition and states within tumors, highlighting the functional roles that various cell types play throughout cancer progression.[4–7] Clinically, these insights are vital for tailoring personalized treatment approaches and enhancing early detection methods. By understanding the unique cellular interactions and resistance mechanisms in PC, researchers and clinicians can develop targeted therapies that more effectively disrupt tumor growth and dissemination. This could potentially lead to improved survival rates and enhanced quality of life for patients.

This catalog aims to provide a molecular atlas for the cellular heterogeneity and intercellular communication that characterize PC. To achieve this, scRNA-seq approaches were employed to analyze PC tissues at single-cell resolution to define developmental trajectories in this study, enhancing understanding of how various cell populations within the TME interact. By utilizing pseudotime trajectories, the evolution of individual cancer cells from early to late stages of the disease was traced. This method identified key transitional states and the genes driving these transitions, elucidating the hierarchical organization and temporal dynamics of tumor evolution.[8–11] In addition, a cell communication analysis was conducted to examine interactions between cancer cells and stromal and immune cells. Various cell types are known to affect tumor progression and inhibition through immune evasion, metastasis and therapy resistance in the TME. The analysis of ligand–receptor interactions uncovered widespread intercellular crosstalk and emphasized the intricate signaling networks that maintain the TME.[12–15]

To further support our single-cell analysis, we combined immunofluorescence and transmission electron microscopy (TEM) to identify the expression of stemness genes in pancreatic stellate cells (PSC). These methods confirmed and spatially contextualized our findings, highlighting the role of stemness in PC pathogenesis.

These discoveries offer additional therapeutic targets and biomarkers for early detection and treatment. These discoveries emphasize the critical role that single-cell technologies can play in advancing our understanding of cancer biology and driving precision oncology.

Objectives

The study aims to elucidate the cellular and molecular features regulating PC progression through an integrated analysis of single-cell pseudotime and cell communication.

Materials and methods

Patients

The study included PC patients sourced from public databases (such as The Cancer Genome Atlas (TCGA) https://www.cancer.gov/ccg/research/genome-sequencing/tcga, Gene Expression Omnibus (GEO) https://www.ncbi.nlm.nih.gov/geo, GSM8541021, GSM8541020). Inclusion criteria comprised confirmed cases of PC with documented patient characteristics, including age, sex, geographic location, and stage at diagnosis. Exclusion criteria encompassed patients with incomplete clinical or demographic data.[16–18]

Specimen characteristics

Gene expression data from PC patients were collected from the GEO database. Control samples were obtained from the same datasets, consisting of non-cancerous tissue from PC patients. The datasets used were publicly available and underwent quality control in their respective studies. Data storage complied with open-source data-sharing standards.

Gene expression analysis

RNA-seq and microarray data were utilized for expression profiling. Data were processed with normalization techniques to account for batch effects and technical variability. The analysis focused on differentially expressed genes (DEGs) related to PC. All assays were conducted blinded to clinical endpoints, ensuring unbiased interpretation.

Study design

The study was retrospective and utilized data from the GEO database. The analysis concentrated on immune infiltration and intercellular communication in PC. The primary endpoint focused on exploring gene expression differences in the TME using single-cell sequencing data. The secondary endpoint centered on intercellular communication pathways. Variables initially considered included gene expression levels, selected based on their relevance in PC prognosis. A large dataset from the GEO database was used to ensure adequate statistical power.

Statistical analyses

The clustering and annotation process for single-cell data was performed as follows: First, the Seurat package (v. 4.4.0; https://github.com/satijalab/seurat) was used to filter, perform quality control, normalize, reduce dimensionality, and cluster the single-cell data from GSM854102 and GSM8541020. The steps involved are outlined below:

1. Data preprocessing: The CreateSeuratObject function was employed to filter the single-cell sequencing data, retaining genes expressed in at least 3 cells and cells detecting more than 350 genes (parameters set to min.cell = 3 and min.features = 350).

2. Mitochondrial and ribosomal gene scoring: The PercentageFeatureSet function was utilized to calculate the mitochondrial and ribosomal gene scores based on the expression data of mitochondrial and ribosomal genes, respectively. A robust gene expression window was established, excluding cells with fewer than 200 or more than 2,500 unique genes. This approach mitigated the risk of including cell doublets and excluded low-quality cells that could compromise subsequent analyses. Furthermore, mitochondrial gene expression was assessed, and cells with

more than 5% mitochondrial transcript expression were systematically removed. This step helped identify and exclude cells potentially undergoing stress or apoptosis, which could distort analytical outcomes. Only cells with mitochondrial and ribosomal gene scores below 20% were retained for analysis.

3. Data normalization and variable gene identification: The SCTransform function was used to normalize the data and identify the top 3,000 highly variable genes (variable. features.n = 3000).

4. Principal component analysis and dimensionality reduction: Principal component analysis (PCA) was performed, and an elbow plot was generated to determine the optimal number of dimensions. Principal components up to the inflection point were selected for further analysis. The first 15 principal components were chosen, and batch effects were removed using the Canonical Correlation Analysis (CCA) method in Seurat. The FindNeighbors and FindClusters functions were then used for unsupervised cluster analysis of the data, following batch removal. The optimal resolution parameter was determined using the "clustree" function to assess cluster stability across different resolutions. In this study, the optimal resolution was set to 0.7, resulting in the identification of 26 clusters. Uniform manifold approximation and projection (UMAP) visualization was employed for cluster representation. Cells were annotated based on reference marker genes provided by CellMarker,[13] and the proportions of annotated cells were computed and visualized using ggplot2.[19–22]

5. Cell communication analysis: To assess cell communication, the interaction between ligands and receptors, along with the activation of specific cell signaling pathways, was analyzed. The CellChat package was used to predict differences in cell communication between groups with high and low Müller ratings. Ligand-receptor pairs were selected for analysis, with a threshold p-value of less than 0.05 considered statistically significant. Only membrane-bound ligand-receptor pairs were analyzed, and the expression of these pairs in different cell types was evaluated.

6. Immunofluorescence: Cells were fixed using either pre-warmed 4% paraformaldehyde (PFA) or cold methanol for 20 min. The PFA-fixed cells were then permeabilized with 0.1% Triton X-100 in phosphate-buffered saline (PBS) for 8 min. A pre-boiled ethylenediaminetetraacetic acid (EDTA) solution (1 mM EDTA) was added to the PFA-fixed cells in a LabTek 8-well chambered coverglass and incubated for 10 min at room temperature. This procedure was repeated 3 times. Following fixation and permeabilization, cells were blocked with 3% bovine serum albumin (BSA) in PBS for 30 min. Primary antibodies were incubated at 4°C overnight, diluted 1:500 in 3% BSA/PBS. After washing, secondary antibodies (also diluted 1:500) were incubated for 1 h at room temperature. Nuclear DNA was stained using DAPI (4′,6-diamidino-2-phenylindole;

1 μg/mL). Samples were then analyzed using an LSM800 GaAsP laser scanning confocal microscope with Zen software (Carl Zeiss AG, Jena, Germany).[23–25]

All statistical analyses were performed using R v. 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). Differences were considered statistically significant when the p-value was less than 0.05, unless otherwise specified.

Results

Analysis of single-cell RNA sequencing data comparing normal and tumor cells

Figure 1 presents a comprehensive analysis of scRNA-seq data comparing normal and tumor cells. Panels A–D display density plots of various metrics, whereas panels E, F present results from PCA. The density plot illustrates the distribution of total RNA counts per cell for normal (blue) and tumor (red) cells, revealing a broader and higher distribution of RNA counts in tumor cells, indicative of increased transcriptional activity (Fig. 1A). This plot demonstrates the distribution of the number of detected genes per cell, with tumor cells exhibiting a higher number of detected genes, suggesting increased gene expression diversity (Fig. 1B). The density plot depicting mitochondrial gene expression percentage shows a similar distribution for both normal and tumor cells, although tumor cells exhibit a slight shift towards higher mitochondrial gene expression, potentially indicating altered metabolic states (Fig. 1C). The plot representing hemoglobin gene expression percentage shows overlapping distributions for normal and tumor cells, with tumor cells displaying slightly higher hemoglobin gene expression (Fig. 1D). The scree plot illustrates the variance explained by each principal component, with the first few principal components accounting for the majority of variance in the data. Although p-values are provided for each principal component, they indicate statistical significance rather than effect magnitude. The relative importance of principal compotents is better assessed by the proportion of variance they explain, as visualized in the scree plot (Fig. 1E). The scatter
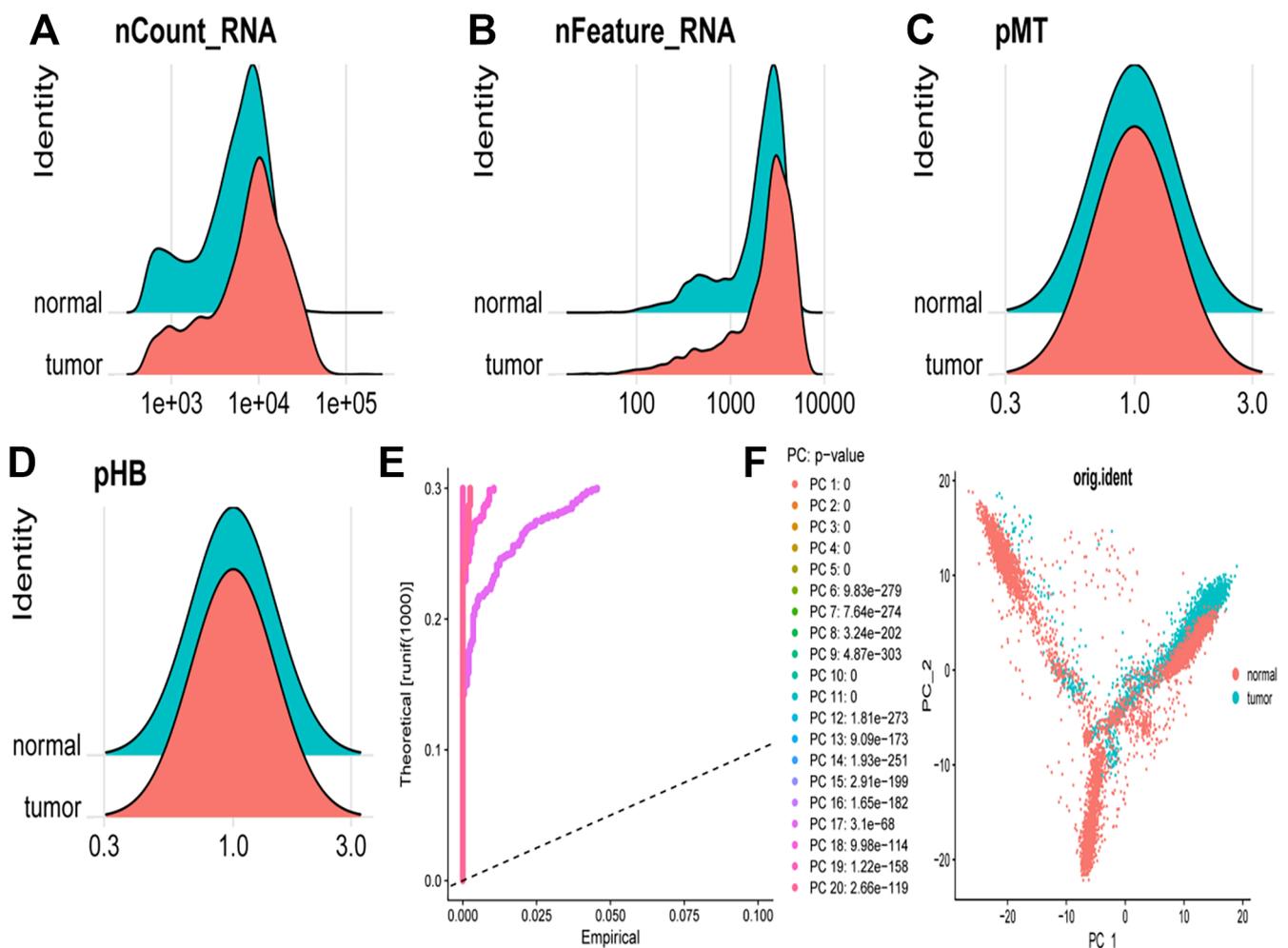

Fig. 1. Analysis of single-cell RNA sequencing (scRNA-seq) data comparing normal and tumor cells. A, B. Tumor cells exhibit higher RNA content and transcriptional diversity compared to normal cells; C. A slight increase in mitochondrial gene expression in tumor cells; D. Marginally higher hemoglobin gene expression in normal cells; E. Significant principal components that capture variance in the data; F. Transcriptional profiles of normal and tumor cells, emphasizing the separation between these 2 groups
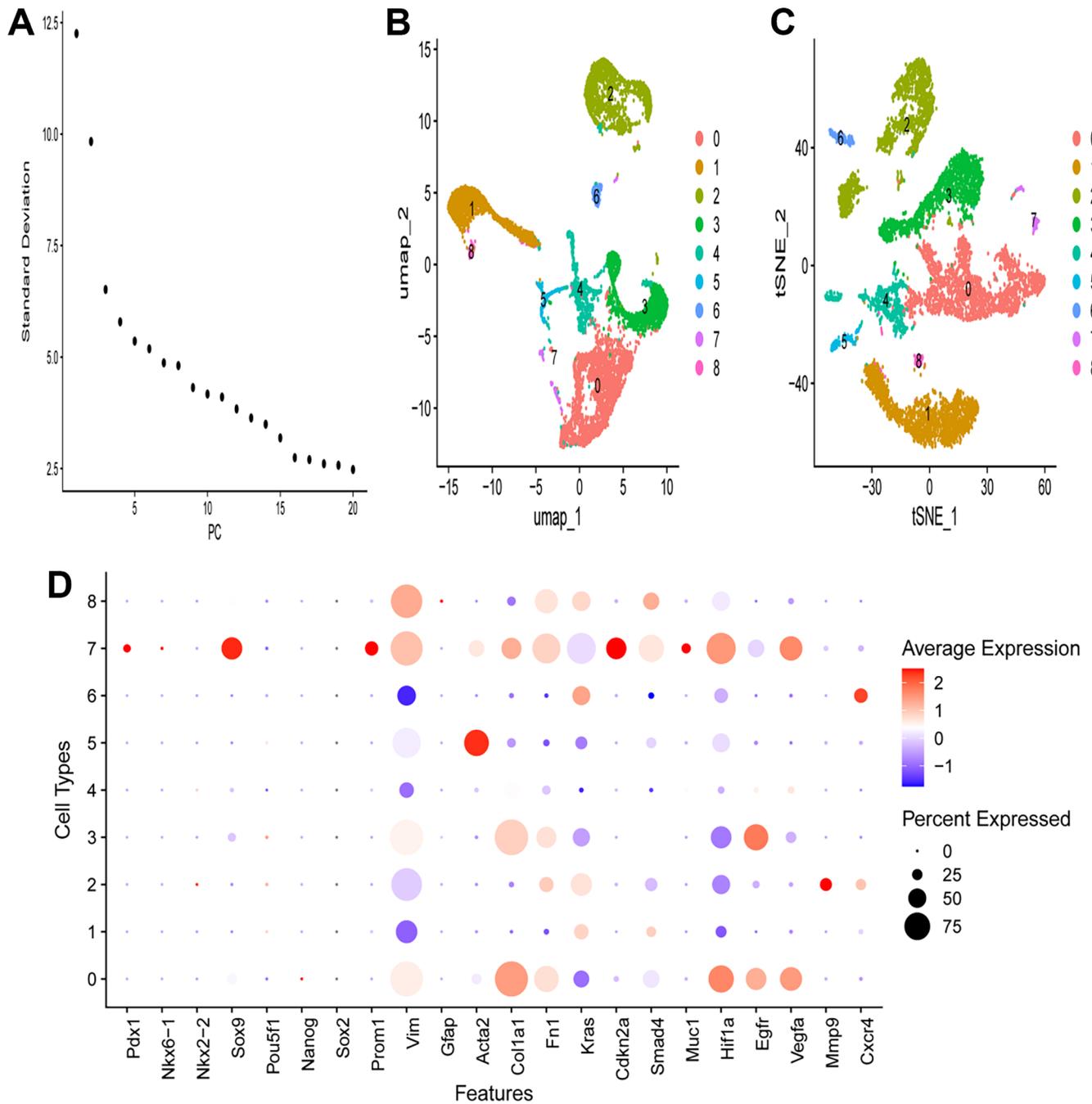
Fig. 2. Principal component analysis and dimensionality reduction techniques. A. The principal component analysis (PCA) scree plot illustrates that the initial principal components account for the majority of the variance in the data, indicating that these components are adequate for subsequent analysis; B. The uniform manifold approximation and projection (UMAP) plot displays a clear separation of cells into varied clusters, revealing the heterogeneity and different cellular states or types within the dataset; C. The t-SNE plot supports the UMAP findings, offering an additional visualization of the cell clusters and their interrelations. The dot plot conveys the expression patterns of key genes across different cell types. Genes such as *Pdgfr*β, *Sox9*, *Nanog*, and *Vim* demonstrate varied expression across the clusters, highlighting their roles in various cellular states or types

plot visualizes the PCA results, with cells plotted according to their scores on the first 2 principal components. Normal (blue) and tumor (red) cells formed distinct clusters, indicating substantial differences in their gene expression profiles. The separation along principal compotent 1 and principal component 2 suggests that these principal components capture the major sources of variation between normal and tumor cells (Fig. 1F).

Principal component analysis and dimensionality reduction techniques

The scree plot (Fig. 2A) illustrates the standard deviation (SD) of the principal components, indicating that the first few principal components capture the most variance. As the number of principal components increases, the SD decreases, suggesting that later principal components
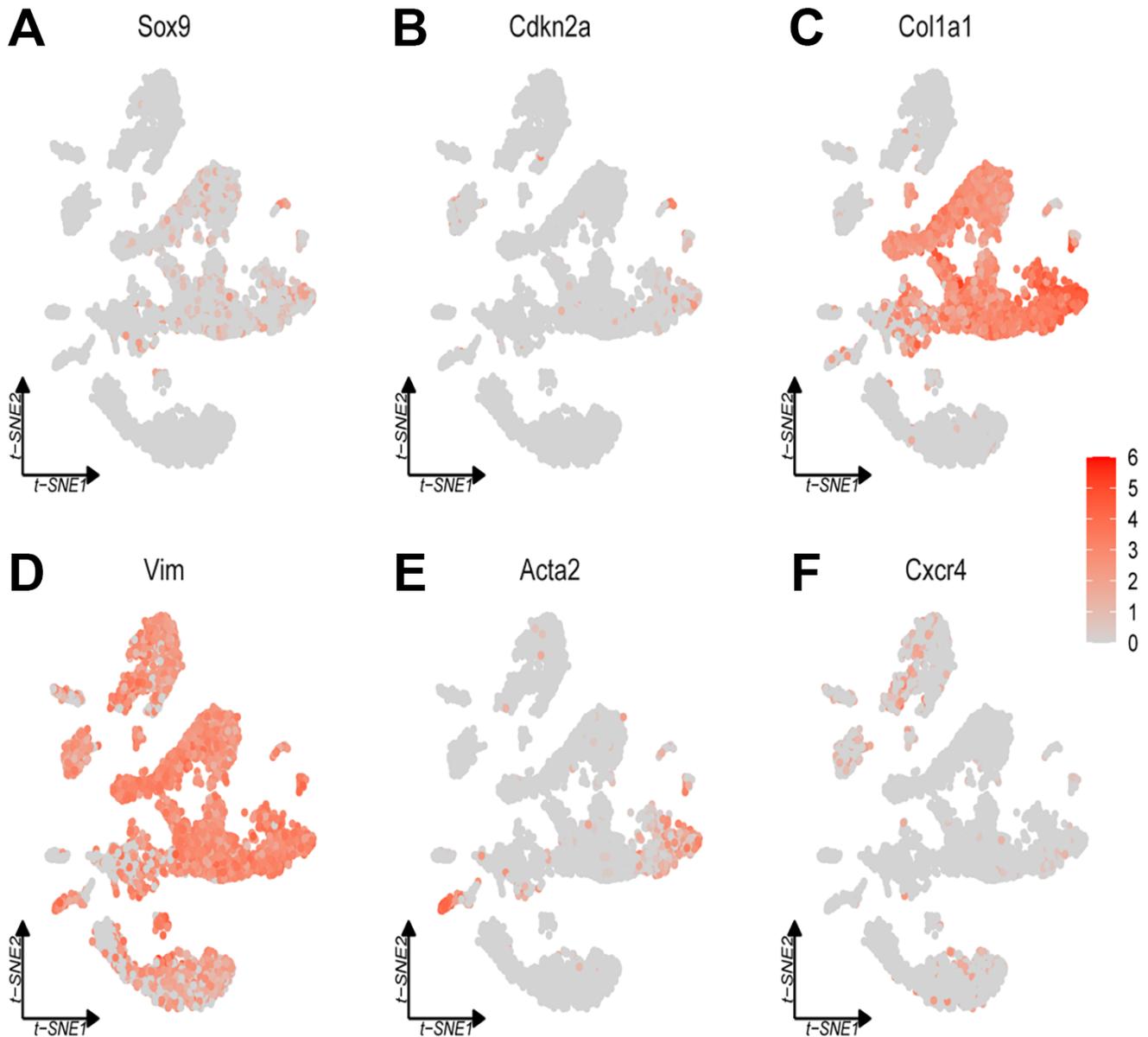
Fig. 3. The expression of specific genes across different cell populations in single-cell RNA sequencing. A. *Sox9* reveals distinct clusters of high expression, signifying its role in stem cell maintenance and differentiation; B. *Cdkn2a* exhibits scattered high expression, emphasizing its role in cell cycle regulation; C. *Col1a1* demonstrates strong expression in specific clusters, highlighting its significance in extracellular matrix formation; D. *Vim* is broadly expressed in mesenchymal cells, indicating its role in cellular integrity and motility; E. *Acta2* shows high expression in clusters associated with muscle contraction; F. *Cxcr4* displays scattered high expression, suggesting its role in cell migration and signaling

contribute less to overall variance (Fig. 2A). The UMAP plot visualizes the scRNA-seq data in a 2-dimensional space, where each dot represents a single cell colored based on assigned clusters (0–8). The distinct clusters indicate different cell populations, with spatial separation suggesting differences in gene expression profiles (Fig. 2B). The t-distributed stochastic neighbor embedding (t-SNE) plot also visualizes the scRNA-seq data in 2 dimensions. Similar to the UMAP plot, cells are colored by clusters (0–8), further confirming the separation of cell populations and highlighting heterogeneity within the dataset (Fig. 2C). The dot plot displays expression levels of selected genes across different cell types, where the size of each dot represents the percentage of cells expressing the gene, and the color indicates the average

expression level. Notable observations show that certain genes are highly expressed in specific cell types, as indicated by larger and darker dots (Fig. 2D). For example, Sox2 and Prom1 show high expression in cell type 7, whereas Egfr and Vegfa are broadly expressed across multiple cell types. The differential expression patterns provide insights into the functional roles of these genes in various cell populations.

Figure 3 presents t-SNE plots illustrating the expression of specific genes across various cell populations in a scRNA-seq dataset. The t-SNE plot for *Sox9* gene expression reveals that Sox9 is present in multiple, yet discrete clusters. The staining intensities vary among clusters, with some clusters exhibiting higher levels of *Sox9*

expression (Fig. 3A). The visualization of Cdkn2a expression in the t-SNE plot indicates a more cluster-specific expression pattern, suggesting a cell-type-dependent expression profile. Each major group of genes (*MGA-D*, *MGB-E* and *2GGs*) is ordered by peak expression time within each cluster, with shades of gray representing expression levels. Gene Ontology (GO) term analysis was performed for each functional map, consistently applying a threshold or flag-based criteria at the boundaries of the last exons.

Figure 3C displays the expression of *Col1a1*, which is concentrated in specific regions of the dataset, indicating that *Col1a1* is expressed at lower levels throughout much of the brain, suggesting that *Col1a* promoter-driven Cre recombinase primarily exists in specific cell populations, such as those producing extracellular matrix (ECM; Fig. 3C). The t-SNE plot for *Vim* demonstrates broad expression across various clusters, with several regions exhibiting higher expression levels. The lower cluster is enriched with cell types similar to mesenchymal cells, characterized by the expression of their derived marker, *Vim* (Fig. 3D). The Acta2 expression plot indicates that *Acta2* is concentrated in specific clusters, suggesting predominant expression in certain cell populations, potentially smooth muscle cells or myofibroblasts, due to its role in actin filament formation (Fig. 3E). The expression of *Cxcr4* is depicted in the t-SNE plot, showing scattered and relatively low expression across the clusters, indicating that this gene is expressed at lower levels or in fewer cell populations compared to the other genes illustrated (Fig. 3F).

The proportions of different cell types in each group

The t-SNE plot (Fig. 4) visualizes the distribution of normal cells within a 2-dimensional space, with each dot representing a single cell. The clustering pattern illustrates the presence of distinct cell populations within the normal tissue, with spatial separation suggesting heterogeneity among normal cells (Fig. 4A). Another t-SNE plot displays the distribution of tumor cells, where each dot again represents a single cell. The clustering pattern indicates the presence of various cell populations within the tumor tissue. In comparison to normal cells, tumor cells appear more densely packed and less dispersed, indicating alterations in cellular composition and gene expression profiles (Fig. 4B). A bar plot compares the proportions of different cell types across the total, normal, and tumor groups. The cell types are color-coded as follows: activated stellate cells (orange), cancer-associated fibroblasts (CAF) (green), embryonic stem cell-like cells (cyan), endothelial cells (blue), immune cells (purple), other mesenchymal cells (brown), PC cells (dark pink), pancreatic progenitor cells (light blue), and PSC (pink). In the normal group, PSC and other cell types are well-represented. Conversely, in the tumor group, PC cells dominate, accompanied by a noticeable increase in CAF and a decrease in other cell types such as PSC.

Using pseudotime trajectories to explore cellular differentiation and transitions

Figure 5 presents visualizations of cell type distributions along pseudotime and comparisons between normal and tumor cells. This plot displays the pseudotime trajectory of cells, color-coded by cell types. The trajectory is represented in a 2-dimensional space, with components 1 and 2 indicating the PC. The cell types are color-coded: activated stellate cells (orange), CAF (green), embryonic stem cell-like cells (cyan), endothelial cells (blue), immune cells (purple), other mesenchymal cells (brown), PC cells (dark pink), pancreatic progenitor cells (light blue), and PSC (pink). The trajectory suggests a progression from one cell type to another, implying a potential differentiation pathway or cellular transition process (Fig. 5A).

A subsequent plot shows the same pseudotime trajectory as in panel A, with cells colored according to their pseudotime values, ranging from 0 to 50 (light to dark blue). Key transition points (1, 2 and 3) are marked along the trajectory, indicating significant stages in the progression of cellular states (Fig. 5B). Additional plots compare the pseudotime trajectories of normal and tumor cells, with cells colored by their respective types. The left plot illustrates the trajectory for normal cells, whereas the right plot depicts the trajectory for tumor cells. Observed differences in the trajectories suggest varying differentiation pathways or cellular transitions between normal and tumor cells. Notably, tumor cells exhibit a more pronounced clustering of PC cells and CAF (Fig. 5C). This combined plot presents pseudotime trajectories for both normal and tumor cells in a single view, with cells colored by their types. The trajectories are aligned to emphasize the differences and similarities in the progression of cell states between normal and tumor conditions. The transitions and clustering patterns provide insights into how cellular states evolve differently within the TME compared to normal tissue (Fig. 5D).

Using pseudotime analysis to understand the progression and differentiation of various cell types

The UMAP plot (Fig. 6A) illustrates the distribution of cells in 2-dimensional space, color-coded by cell type. The pseudotime trajectory is overlaid, demonstrating the progression of cells through various states. The cell types are color-coded as follows: activated stellate cells (orange), CAF (green), embryonic stem cell-like cells (cyan), endothelial cells (blue), immune cells (purple), other mesenchymal cells (brown), PC cells (dark pink), pancreatic progenitor cells (light blue), and PSC (pink). The trajectory lines connect different clusters, suggesting pathways of cellular differentiation or transition. This plot also depicts the density of various cell types along the pseudotime axis, with each cell type represented by a different color,

A Normal Cells

B Tumor Cells

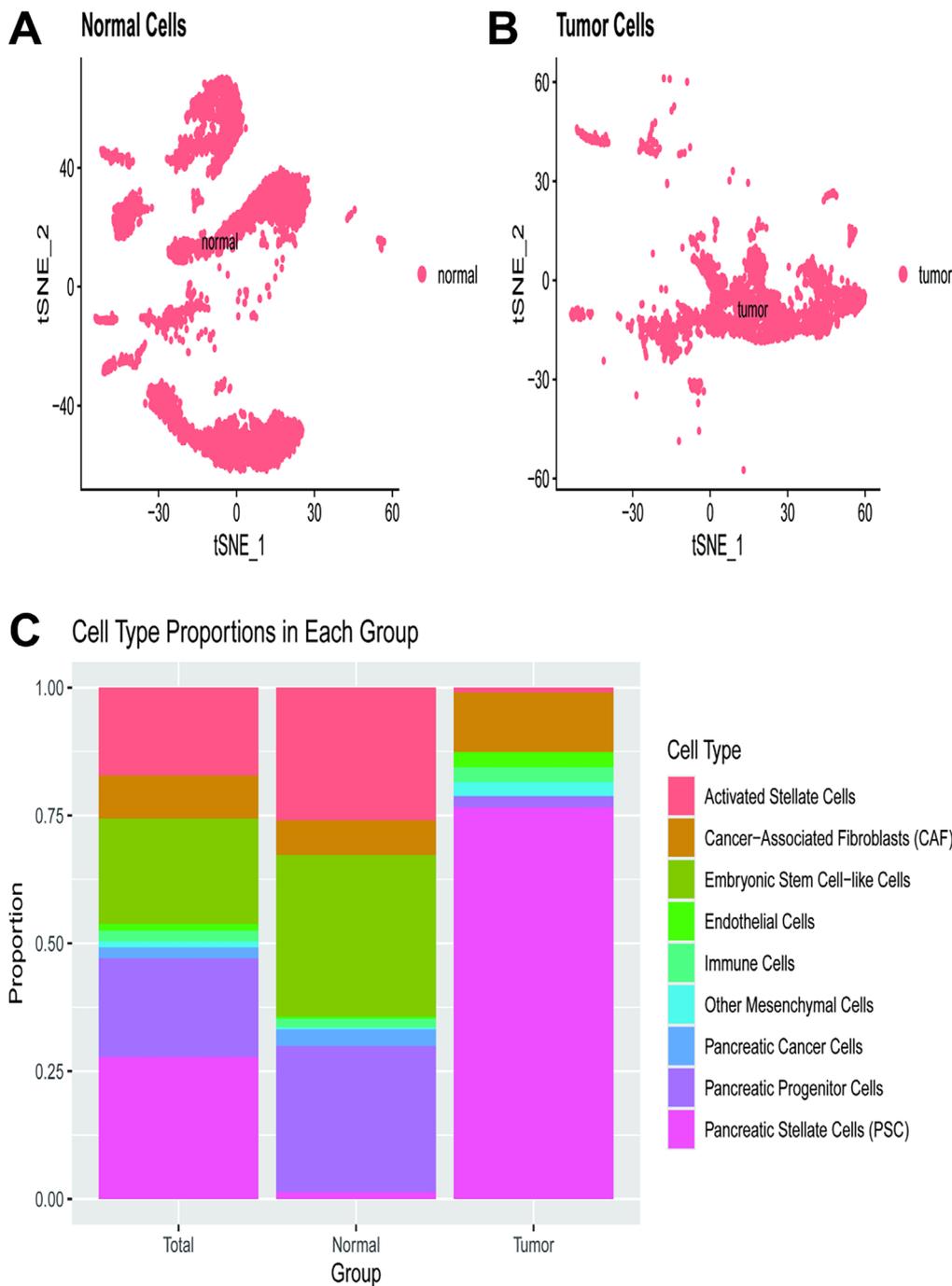**C** Cell Type Proportions in Each Group

Fig. 4. The proportions of different cell types in each group. A. The t-distributed stochastic neighbor embedding (t-SNE) plot of normal cells reveals distinct clusters, indicating the presence of various subpopulations within the normal cell population; B. The t-SNE plot of tumor cells also shows distinct clusters, albeit with different patterns compared to normal cells. This suggests significant transcriptional heterogeneity within the tumor cell population, reflecting the complexity of tumor biology; C. The bar plot of cell type proportions reveals differences in the composition of cell types among the total, normal and tumor groups

consistent with those in Fig. 6A. The density peaks indicate where specific cell types are most prevalent along the pseudotime trajectory. For instance, PC cells (dark pink) and cancer-associated fibroblasts (CAFs; green) exhibit distinct expression peaks, suggesting critical transition points in their progression. These plots illustrate the expression levels of 4 genes (*Lypla1*, *Mrpl15*, *Sox17*, and *Tcea1*) over pseudotime, with each dot representing a single cell colored by its state (1–7) (Fig. 6B). The black line represents the smoothed trend of gene expression over pseudotime. Key observations include: *Lypla1* shows high expression early in pseudotime, decreasing as pseudotime progresses; *Mrpl15* exhibits peak expression around the midpoint

of the pseudotime trajectory; *Sox17* displays a distinct peak in expression later in pseudotime, indicating its role in a specific state; *Tcea1* has high expression early in pseudotime, followed by a gradual decrease (Fig. 6C).

UMAP plots illustrating the expression of specific genes across different cell populations in a single-cell RNA sequencing dataset

As shown in Fig. 7, each UMAP plot demonstrates the distribution of cells in 2-dimensional space, with gene expression levels indicated by color intensity. The color gradient
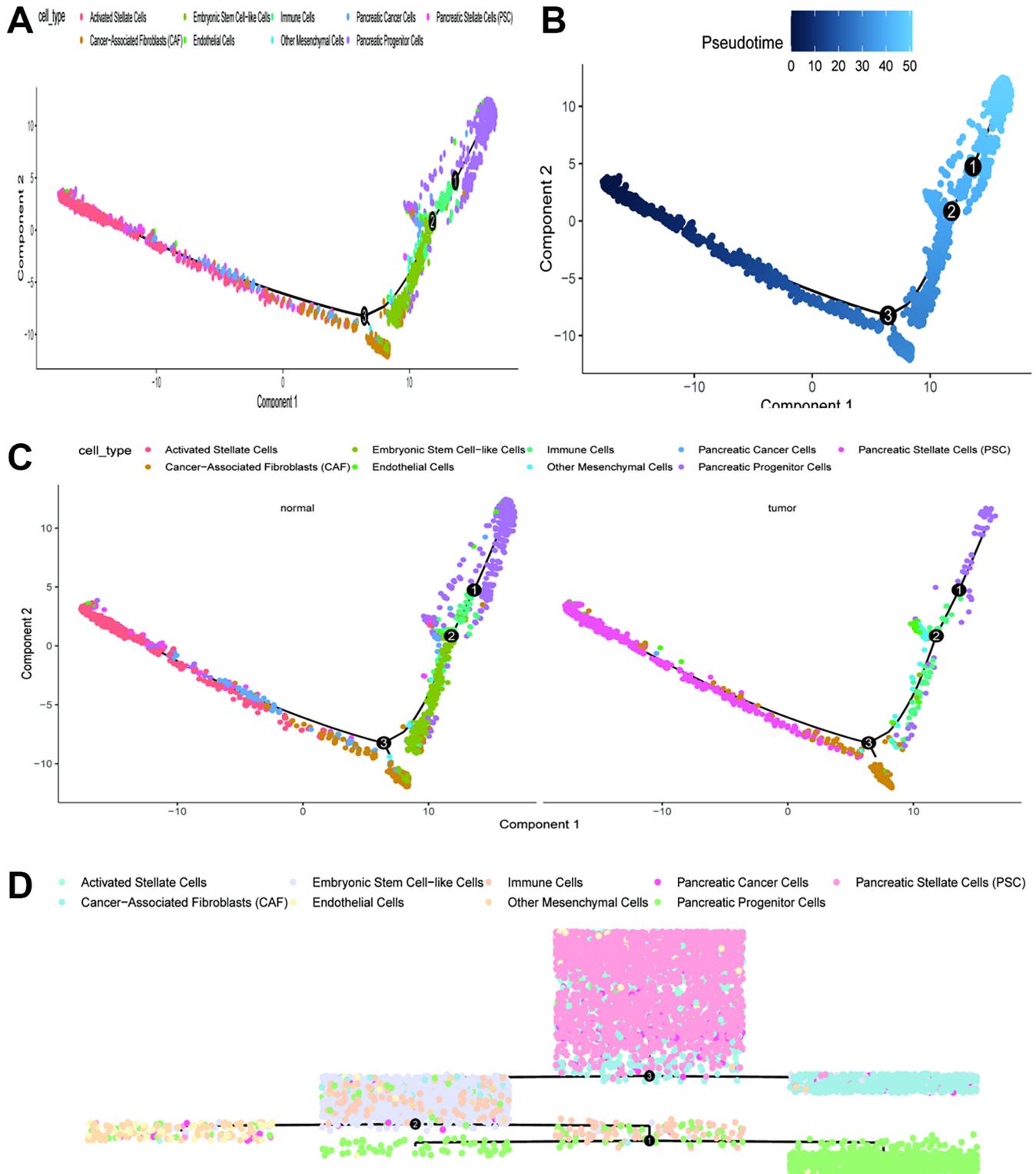
Fig. 5. Pseudotime trajectories used to explore cellular differentiation and transitions. A. The trajectory plot illustrates the differentiation path of cells, with distinct clusters representing different cell types; B. The pseudotime plot highlights the temporal progression of cells along the differentiation path, with key transition points marked; C. The separate trajectory plots for normal and tumor cells reveal differences in their differentiation paths, reflecting the altered progression in tumor cells; D. The linear trajectory plot simplifies the visualization of the differentiation process, showing the sequential progression of cells along the trajectory

ranges from purple (low expression) to green (high expression), representing the percentage of maximum expression for each gene. Areas of high expression (green) exhibit a punctate pattern, implying roles in specific cell types.

Col3a1 (collagen type III alpha 1 chain) shows expression in the same clusters as its paralog in certain regions (75), indicating an association with the ECM and connective tissue. Decorin (Dcn) exhibits scattered expression across
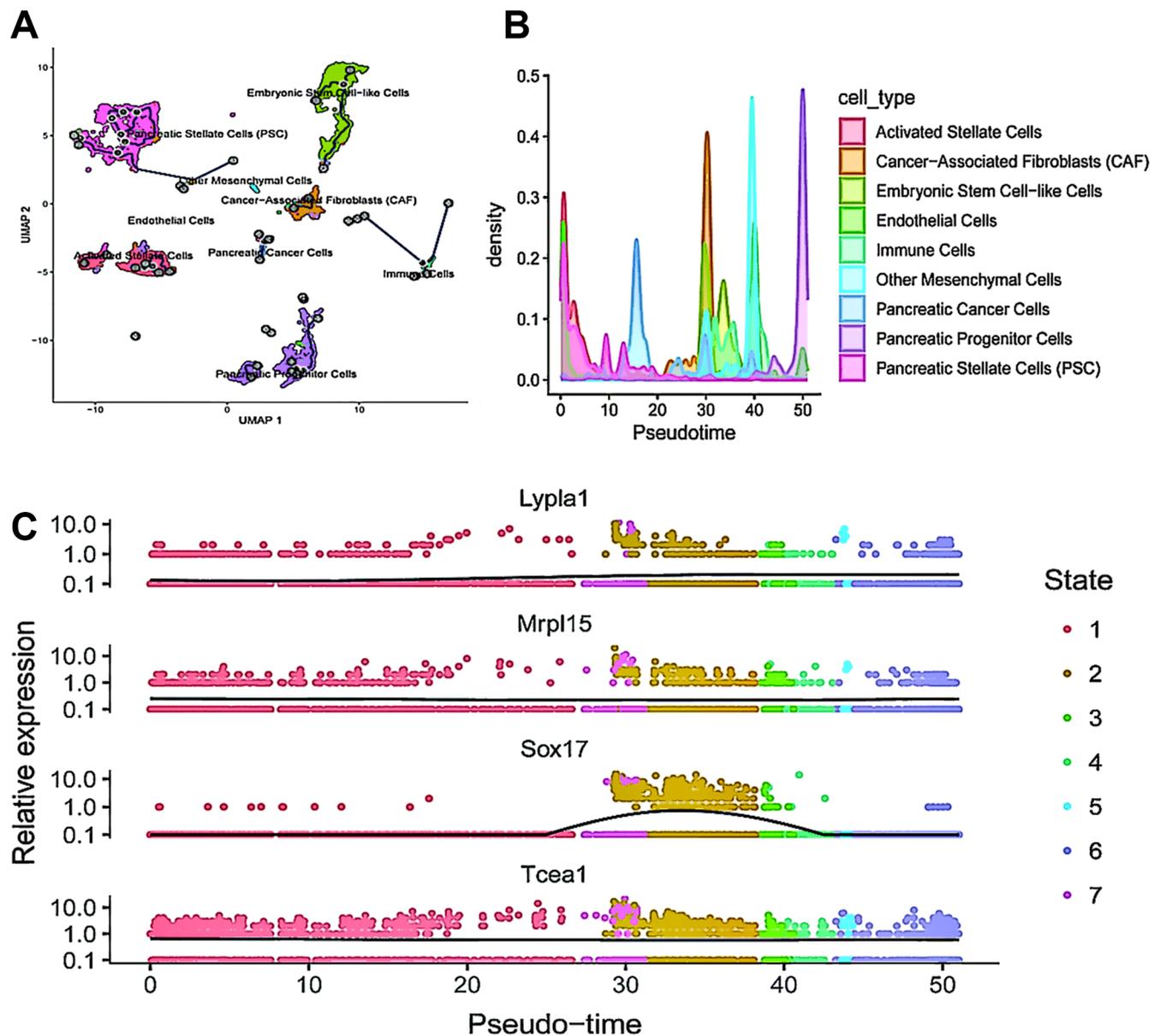
Fig. 6. Pseudotime analysis used to understand the progression and differentiation of various cell types. A. The uniform manifold approximation and projection (UMAP) plot with trajectory inference displays the differentiation path of cells, with distinct clusters representing different cell types. The trajectory line indicates the progression and transitions between these cell types; B. The density plot reveals the distribution of cell types along pseudotime, highlighting key transition points where certain cell types are more prevalent; C. The gene expression trends over pseudotime illustrate how selected genes are regulated during the differentiation process. Each gene exhibits unique expression patterns, reflecting their specific roles at different stages of cell differentiation

multiple clusters, with some regions showing higher levels, highlighting its role in collagen fibrillogenesis and ECM organization. Serping1 (decreased expression in cluster L10 but similar levels to P7) displays a highly restricted expression pattern, with high fractional spike-ins limiting its detection. *Csf1r* (colony stimulating factor 1 receptor) is expressed in existing clusters, indicating its association with specific cell types likely associated with macrophage differentiation and function. *C1qc* (complement C1q C chain) is uniquely expressed, with clear enrichment in clusters associated with immune response and the complement system.

Surface markers of PSC and its differentiation characteristics

Figure 8 presents a comprehensive list of marker expression in PSC and their differentiation capacity. The top immunofluorescence images depict specific markers indicating co-expression in PSC (Vimentin and GFAP), corresponding to embryonic stem cell markers (*Oct4*, *Nanog* and *CD133*) and pancreatic progenitor cell markers (*Ngn3* and *Sox9*) (Fig. 8). The extent to which PSC differentiate into osteogenic, adipogenic, hepatic, and chondrogenic lineages is illustrated through phase-contrast microscopy
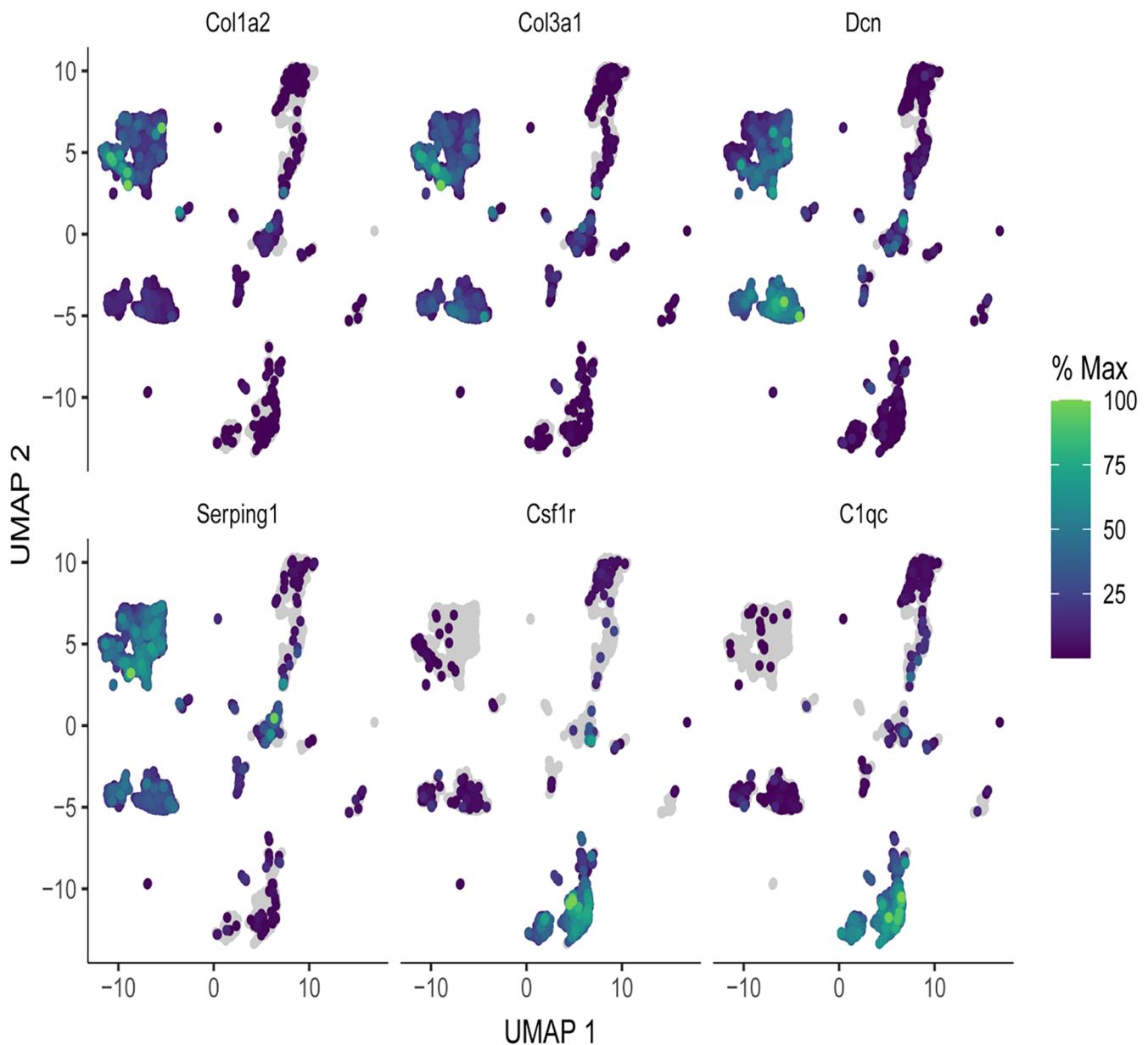
Fig. 7. The uniform manifold approximation and projection (UMAP) plots illustrating the expression of specific genes across different cell populations in a scRNA-seq dataset. The UMAP plots provide a comprehensive visualization of the expression patterns of key genes across various cell populations in the scRNA-seq dataset. Each gene exhibits a unique distribution, reflecting its specific roles and functions within the tissue

images, which show the morphological changes associated with each differentiated state (Fig. 8). Collectively, these analyses provide valuable insights into the cellular heterogeneity within PSC and their differentiation potential.

Discussion

Pancreatic cancer is among the most lethal forms of cancer, characterized by a dismal prognosis and aggressive behavior. Despite some advancements in medical research, the 5-year relative survival rate for individuals diagnosed with PC remains only 7%.[26–28]

The scRNA-seq of pancreatic ductal adenocarcinomas was conducted to provide an extensive analysis of cellular heterogeneity and intercellular communications within the TME. The application of scRNA-seq is rapidly yielding unprecedented insights into cellular dynamics, revealing potential advancements and druggable targets in PC. The construction of pseudotime trajectories has been essential for tracing the developmental pathways of individual cancer cells, identifying critical intermediate states and key regulatory genes that drive these transitions. This has provided a profound understanding of the temporal dynamics of tumor evolution – an essential concept for comprehending the progression of PC from early stages to advanced malignancy. Immunofluorescence studies

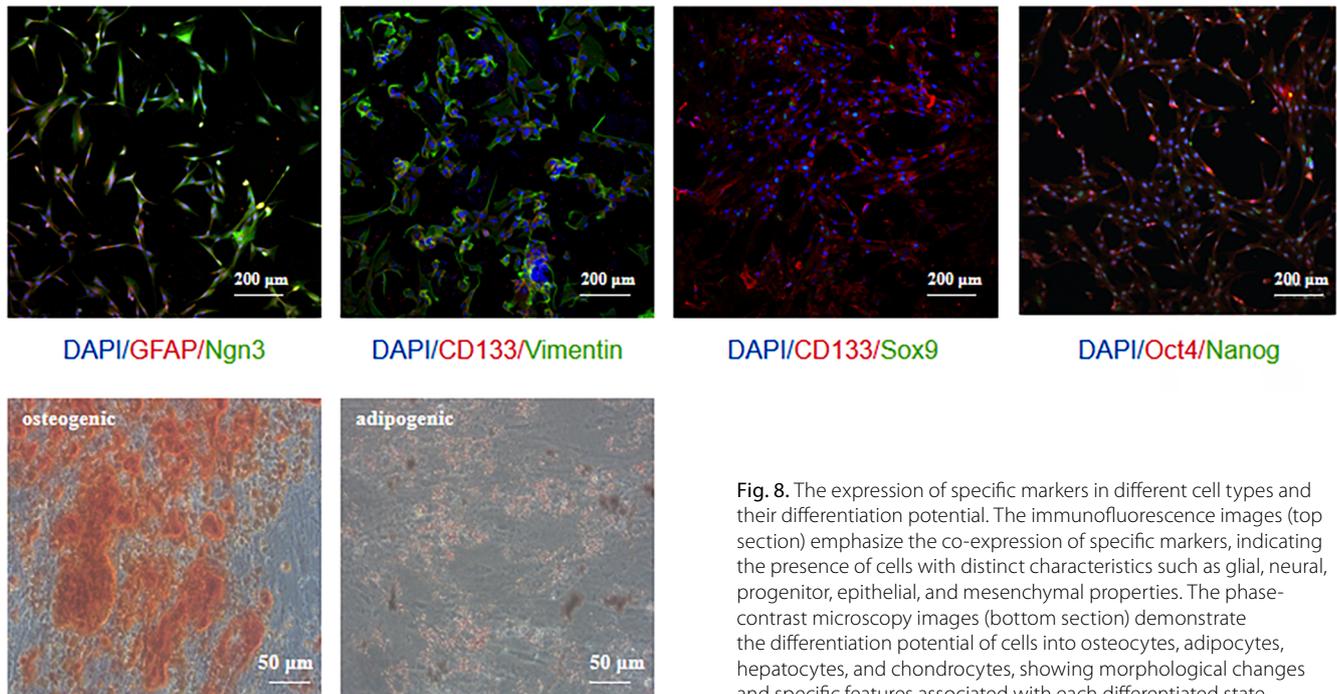DAPI/GFAP/Ngn3   DAPI/CD133/Vimentin   DAPI/CD133/Sox9   DAPI/Oct4/Nanog

Fig. 8. The expression of specific markers in different cell types and their differentiation potential. The immunofluorescence images (top section) emphasize the co-expression of specific markers, indicating the presence of cells with distinct characteristics such as glial, neural, progenitor, epithelial, and mesenchymal properties. The phase-contrast microscopy images (bottom section) demonstrate the differentiation potential of cells into osteocytes, adipocytes, hepatocytes, and chondrocytes, showing morphological changes and specific features associated with each differentiated state

corroborate the spatial context of scRNA-seq data, particularly regarding the examination of stemness genes. The co-expression of vimentin and GFAP, markers characteristic of PSC, with embryonic stem cell markers (*Oct4*, *Nanog* and *CD133*) and pancreatic progenitor markers (*Ngn3* and *Sox9*), suggests an unrecognized pluripotent potential in these cells. These findings highlight the role of stemness in tumor activity, with implications for various aspects such as cancer recurrence and therapeutic resistance. Pseudotime analysis is pivotal in scRNA-seq studies, as it illuminates the developmental trajectories of cancer cells, identifies key intermediate states, and reveals essential genes that govern these transitions. This analytical framework provides unprecedented insights into the temporal dynamics of tumor evolution, which is crucial for understanding the progression of PC from early disease to advanced malignant states. The selection of parameter settings is critical for validating the results of pseudotime analysis. Several factors must be considered, including biological variability, technical biases and intersample variability. For example, methods based on PCA, such as scran and Seurat, perform competitively with default parameters, whereas more sophisticated models like ZinbWave, DCA, and scVI demonstrate enhanced performance with parameter optimization. This indicates that parameter adjustment is a crucial step in optimizing results for various analytical approaches.

CellChat offers ligand-receptor databases for humans and mice, containing extensive validated molecular interaction data essential for understanding and predicting intercellular communication. By comparing ligand and receptor expression across different cell populations or conditions, key molecules related to specific biological processes can be identified. This differential analysis may reveal changes in cell states and variations in intercellular communication.

Our investigation of cell communication has elucidated complex interactions among cancer cells, the stromal compartment, and immune cells within the TME. The findings have illuminated the roles of various cell populations that contribute to tumorigenesis, immune suppression and metastasis by identifying significant signaling crosstalk through ligand-receptor interactions. Understanding the functions of these diverse cell types is critical for developing targeted therapies that address the complexities of the TME. By dissecting cellular heterogeneity, researchers can pinpoint specific cellular interactions and pathways that drive cancer progression, leading to more effective and personalized treatment strategies. Although mutations in genes such as *KRAS*, *TP53* and *SMAD4* are prevalent in PC, the specific mutation patterns and their interactions can differ significantly from those in other cancers.

The identification of critical transcription factors and signaling pathways controlling cellular transitions offers potential therapeutic targets. Furthermore, our research has identified disease markers that may facilitate early detection of PC, thereby significantly enhancing patient survival. The findings highlight the efficacy of single-cell technologies in elucidating cancer characteristics and their contribution to precision oncology. These technologies provide a more detailed understanding of the TME, which is essential for personalized treatment approaches.

The integration of these findings into clinical practice will rely on validation in larger cohorts and through clinical trials. A prerequisite for employing this concept,

even in its preliminary form, will likely require defining specific genes or pathways modified by these critical drivers (and perhaps several additional alterations) so that targeted agents can be developed. The identification of biomarkers and therapeutic targets has the potential to enable precision diagnostics, prognostics and treatment strategies that could significantly improve patient care. Ultimately, such advances may contribute to increased survival rates and enhanced quality of life for individuals affected by PC.

Limitations

Despite the intriguing results of the study, several limitations exist. Single-cell transcriptome analyses yield highly detailed data but may not capture the dynamic and time-resolved nature of cellular interactions within the TME. Although we identified individual cells as representing fresh and aged states, more dynamic studies must incorporate longitudinal sample collections to map transitions between cellular states.

Conclusions

This study contributes to the growing body of knowledge on PC and demonstrates the potential of single-cell technologies in elucidating, with remarkable clarity, the features that underlie its biology. Immunofluorescence data confirmed the stem-like properties of PSC, revealing their pluripotent potential in tumor dynamics. Analyzing key cellular transitions and intercellular communication provides a foundation for developing innovative therapeutic strategies and underscores the need for further studies to translate these insights into clinically relevant outcomes.. Continued research is essential to validate findings related to personalized treatment and prognostic biomarkers, with the goal of advancing clinical applications and improving patient selection, particularly in the effective management of cancer. This ongoing research is essential for transforming scientific discoveries into effective clinical solutions.

Data Availability Statement

The datasets supporting the findings of this study are openly available in Figshare at https://figshare.com/s/d51989569cd0bf8a0fce.

Consent for publication

Not applicable.

Use of AI and AI-assisted technologies

Not applicable.

ORCID iDs

Chengming Ni ● https://orcid.org/0009-0006-7277-2108
Xiaohang Wang ● https://orcid.org/0000-0003-2111-9601
Zhensheng Cai ● https://orcid.org/0000-0003-2964-0565
Yang Chen ● https://orcid.org/0009-0000-5258-6662
Huan Wang ● https://orcid.org/0009-0009-7343-0478
Qianqian Wang ● https://orcid.org/0009-0001-9098-1062
Hao Lin ● https://orcid.org/0000-0001-7981-7707
Yunting Zhou ● https://orcid.org/0000-0002-1385-3841
Yang Yuan ● https://orcid.org/0000-0003-2873-1262
Bo Sun ● https://orcid.org/0009-0001-5483-0910
Zilin Sun ● https://orcid.org/0000-0001-7936-0196

References

1. Jiang J, Krishna SG. Early detection of concomitant pancreatic cancer during intraductal papillary mucinous neoplasms surveillance. *Gland Surg*. 2024;13(9):1659–1661. doi:10.21037/gs-24-158
2. Kinoshita S, Terai T, Nagai M, et al. Clinical significance and therapeutic implication of CD200 in pancreatic cancer. *Pancreatology*. 2024;24(8):1280–1287. doi:10.1016/j.pan.2024.10.007
3. Li R, Liu R, Xu Y, et al. Suppressing pancreatic cancer survival and immune escape via nanoparticle-modulated STING/STAT3 axis regulation. *Bioconjugate Chem*. 2024;35(11):1815–1822. doi:10.1021/acs.bioconjchem.4c00379
4. Ding S, Yi X, Gao J, et al. Prognostic risk model of LIHC T-cells based on scRNA-seq and RNA-seq and the regulation of the tumor immune microenvironment. *Discov Oncol*. 2024;15(1):540. doi:10.1007/s12672-024-01424-z
5. Lou Z, Wei X, Hu Y, Hu S, Wu Y, Tian Z. Clustering scRNA-seq data with the cross-view collaborative information fusion strategy. *Brief Bioinform*. 2024;25(6):bbae511. doi:10.1093/bib/bbae511
6. Tang M, Xiong L, Cai J, et al. Comprehensive analysis of scRNA-Seq and bulk RNA-Seq reveals transcriptional signatures of macrophages in intrahepatic cholestasis of pregnancy. *J Inflamm Res*. 2024;17:6863–6874. doi:10.2147/JIR.S471374
7. Wang H, Torous W, Gong B, Purdom E. Visualizing scRNA-Seq data at population scale with GloScope. *Genome Biol*. 2024;25(1):259. doi:10.1186/s13059-024-03398-1
8. Cao L, Zhang S, Yao D, et al. Comparative analyses of the prognosis, tumor immune microenvironment, and drug treatment response between left-sided and right-sided colon cancer by integrating scRNA-seq and bulk RNA-seq data. *Aging*. 2023;15(14):7098–7123. doi:10.18632/aging.204894
9. Guo G, Zhou Z, Chen S, et al. Characterization of the prognosis and tumor microenvironment of cellular senescence-related genes through scRNA-seq and bulk RNA-seq analysisin GC. *Recent Pat Anticancer Drug Discov*. 2024;19(4):530–542. doi:10.2174/0115748928255417230924191157
10. Lin Z, Li X, Shi H, et al. Decoding the tumor microenvironment and molecular mechanism: Unraveling cervical cancer subpopulations and prognostic signatures through scRNA-Seq and bulk RNA-seq analyses. *Front Immunol*. 2024;15:1351287. doi:10.3389/fimmu.2024.1351287
11. Zhong J, Sun Y, Wu S, et al. The impact of the Hedgehog signal pathway on the tumor immune microenvironment of gastric adenocarcinoma by integrated analysis of scRNA-seq and RNA-seq datasets. *Funct Integr Genomics*. 2023;23(3):258. doi:10.1007/s10142-023-01187-w
12. Jiang X, Chen N, Wei Q, et al. Single-cell RNA sequencing and cell–cell communication analysis reveal tumor microenvironment associated with chemotherapy responsiveness in ovarian cancer. *Clin Transl Oncol*. 2024;27(3):1000–1012. doi:10.1007/s12094-024-03655-6
13. Lyu N, Wu J, Dai Y, et al. Identification of feature genes and molecular mechanisms involved in cell communication in uveal melanoma through analysis of single-cell sequencing data. *Oncol Lett*. 2024;28(5):503. doi:10.3892/ol.2024.14636
14. Xie S, Xu J, Chen L, Qi Y, Yang H, Tan B. Single-cell transcriptomic analysis revealed the cell population changes and cell–cell communication in the liver of a carnivorous fish in response to high-carbohydrate diet. *J Nutr*. 2024;154(8):2381–2395. doi:10.1016/j.tjnut.2024.06.016

332

C. Ni et al. Pancreatic cancer single-cell study

15. Zhao Z, Wu Y, Geng X, Yuan C, Yang G. Single-cell analysis reveals histone deacetylation factor guide intercellular communication of tumor microenvironment that contribute to colorectal cancer progression and immunotherapy [published online as ahead of print on April 18, 2024]. *Biochem Genet*. 2024. doi:10.1007/s10528-024-10730-8

16. Shen C, Geng R, Zhu S, et al. Characterization of tumor suppressors and oncogenes evaluated from TCGA cancers. *Am J Clin Exp Immunol*. 2024;13(4):187–194. doi:10.62347/XMZW6604

17. Cheng X, Lin J, Wang B, Huang S, Liu M, Yang J. Clinical characteristics and influencing factors of anti-PD-1/PD-L1-related severe cardiac adverse event: Based on FAERS and TCGA databases. *Sci Rep*. 2024;14(1):22199. doi:10.1038/s41598-024-72864-4

18. Lin B, Wang K, Yuan Y, et al. A novel approach to the analysis of overall survival (OS) as response with progression-free interval (PFI) as condition based on the RNA-seq expression data in The Cancer Genome Atlas (TCGA). *BMC Bioinformatics*. 2024;25(1):300. doi:10.1186/s12859-024-05897-1

19. Ding Q, Yang W, Xue G, et al. Dimension reduction, cell clustering, and cell–cell communication inference for single-cell transcriptomics with DcjComm. *Genome Biol*. 2024;25(1):241. doi:10.1186/s13059-024-03385-6

20. Pagnotta SM. Unsupervised single-cell clustering with asymmetric within-sample transformation and per-cluster supervised features selection. *Methods Mol Biol*. 2024;2812:155–168. doi:10.1007/978-1-0716-3886-6_8

21. Sun Y, Kong L, Huang J, et al. A comprehensive survey of dimensionality reduction and clustering methods for single-cell and spatial transcriptomics data. *Brief Funct Genomics*. 2024;23(6):733–744. doi:10.1093/bfgp/elae023

22. Yao Z, Li B, Lu Y, Yau ST. Single-cell analysis via manifold fitting: A framework for RNA clustering and beyond. *Proc Natl Acad Sci U S A*. 2024;121(37):e2400002121. doi:10.1073/pnas.2400002121

23. Hsieh Y, Chen S. Visualization and analysis of neuromuscular junctions using immunofluorescence. *Bio Protoc*. 2024;14(1354):e5076. doi:10.21769/BioProtoc.5076

24. Lin CCJ, Jaafar N, Tanzi RE. Protocol to characterize mouse dural mast cells by flow cytometry and immunofluorescence. *STAR Protoc*. 2024;5(4):103364. doi:10.1016/j.xpro.2024.103364

25. Wang Y, Bai W, Wang X. Progress on three-dimensional visualizing skin architecture with multiple immunofluorescence staining and tissue-clearing approaches [published online as ahead of print on September 16, 2024]. *Histol Histopathol*. 2024. doi:10.14670/HH-18-815

26. Wang Q, Wang J, Xu K, Luo Z. Targeting the CSF1/CSF1R signaling pathway: An innovative strategy for ultrasound combined with macrophage exhaustion in pancreatic cancer therapy. *Front Immunol*. 2024;15:1481247. doi:10.3389/fimmu.2024.1481247

27. Yang DH, Park SH, Yoon S. Differential diagnosis of pancreatic cancer and its mimicking lesions. *J Korean Soc Radiol*. 2024;85(5):902. doi:10.3348/jksr.2023.0100

28. Yee KX, Lee YC, Nguyen HD, et al. Uncovering the role of *FXYD3* as a potential oncogene and early biomarker in pancreatic cancer. *Am J Cancer Res*. 2024;14(9):4353–4366. doi:10.62347/LUDE7524