



## Original Articles

## Single-cell transcriptome profiling of primary tumors and paired organoids of pancreatobiliary cancer



Kai Chen<sup>a,1</sup>, Yongsu Ma<sup>a,1</sup>, Xiejian Zhong<sup>a</sup>, Jianqiang Lan<sup>b</sup>, Di Long<sup>a</sup>, Xiaodong Tian<sup>a,\*</sup>, Yanlian Yang<sup>c,d,\*\*</sup>, Yinmo Yang<sup>a,\*\*\*</sup>

<sup>a</sup> Department of General Surgery, Peking University First Hospital, Beijing, China

<sup>b</sup> Guangdong Research Center of Organoid Engineering and Technology, Accurate International Biotechnology Co Ltd, Guangzhou, China

<sup>c</sup> CAS Key Laboratory of Biological Effects of Nanomaterials and Nanosafety, CAS Key Laboratory of Standardization and Measurement for Nanotechnology, CAS Center for Excellence in Nanoscience, National Center for Nanoscience and Technology, Beijing, China

<sup>d</sup> University of Chinese Academy of Sciences, Beijing, China

## ARTICLE INFO

## ABSTRACT

## Keywords:

Organoid  
Pancreatobiliary cancer  
ScRNA-seq  
Molecular subtypes  
Heterogeneity

Single-cell RNA-seq (scRNA-seq) and cancer organoid model have shown promise in investigating tumor microenvironment heterogeneity and facilitating chemotherapeutic drug testing to inform treatment selection. It is still unknown whether the scRNA-seq results based on organoid can faithfully reflect the heterogeneity of primary pancreatobiliary cancer. To reveal the similarities and differences between primary tumors and their matched organoids at transcriptome level, we conducted scRNA-seq for paired primary tumors and organoids from one cholangiocarcinoma (CCA) and two pancreatic ductal adenocarcinoma (PDAC) patients. We identified inter-patient and intra-tumor heterogeneity and found that the organoids retained copy number variation (CNV) patterns of primary tumors. There was no significant difference in cancer stem cell (CSC) properties between the primary tumors and the organoids, whereas organoid from one PDAC case had increased mesenchymal-score and decreased epithelial-score compared with the primary tumors. All organoids showed a transition tendency from the classical subtype to the basal-like subtype in the transcriptional level. Organoids and primary tumors differed in metabolic and unfolded protein response (UPR) signatures. In addition, we revealed the heterogeneity of cancer associated fibroblasts (CAFs) and T cells, and explored the developmental trajectory of T cells. Our findings facilitate further understanding of organoid model and confirm its application prospects in pancreatobiliary cancer.

## 1. Introduction

Pancreatic cancer is a lethal gastrointestinal malignancy with an overall 5-year survival rate of 12% [1]. Pancreatic ductal adenocarcinoma (PDAC) is the most common histological type of pancreatic cancer. Currently, radical surgery is the only potential curative treatment for PDAC. Nevertheless, approximately half of PDAC patients are diagnosed at advanced stages and have lost the opportunity for surgery. Even after radical surgery, the majority of patients have local or distant recurrence within two years [2]. Multidisciplinary management and integrated supportive care are recommended for the patients with PDAC

[3]. In spite of the continuous introduction of new regimens, there is still a lack of drugs that can be used to significantly improve the prognosis of patients with PDAC. Remarkable heterogeneity of PDAC poses a challenge for personalized therapy. The prognosis and treatment of cholangiocarcinoma (CCA) involved in the pancreas are similar to those of PDAC. It is imperative to develop an appropriate patient-derived model to help personalized drug testing.

Tumor heterogeneity has been widely identified [4]. The traditional 2D-cultured cell lines cannot faithfully represent the heterogeneous characteristics of *in vivo* tumors and recapitulate cell-to-cell interactions in tumor microenvironment (TME), therefore are not suitable as

\* Corresponding author. Department of General Surgery, Peking University First Hospital, 8th Xishiku Street, Beijing, 100034, China.

\*\* Corresponding author. National Center for Nanoscience and Technology, Beijing, 100190, China.

\*\*\* Corresponding author. Department of General Surgery, Peking University First Hospital, 8th Xishiku Street, Beijing, 100034, China.

E-mail addresses: [tianxiaodong@pkufh.com](mailto:tianxiaodong@pkufh.com) (X. Tian), [yangyl@nanoctr.cn](mailto:yangyl@nanoctr.cn) (Y. Yang), [YangyinmoSCI@bjmu.edu.cn](mailto:YangyinmoSCI@bjmu.edu.cn) (Y. Yang).

<sup>1</sup> These authors contributed equally.

preclinical cancer model. Patient-derived xenograft (PDX) models, as an ideal patient-derived model, can stimulate the TME and retain genetic features of original tumors. However, they are very time-consuming, and therefore not good as a tool for screening drugs for patients with PDAC [5]. The in vitro tumor organoid model is a big technological breakthrough, which has been established as a significant tool in deciphering the mechanisms of tumor progression and screening personalized treatments [6]. Recently, single-cell RNA sequencing (scRNA-seq) has paved a way for uncovering the cell landscapes and identifying tumor heterogeneity of various tumor tissues [7,8]. The transcriptional profile of each cell was generated by scRNA-seq analysis, consisting of single cell capture, cDNA library preparation, RNA sequencing, and data mining [9]. However, whether tumor organoids could faithfully represent the matched original tumors at single-cell transcriptome level remains to be investigated. The findings obtained by comparing primary tumors and their organoids in PDAC and CCA will provide new insight into drug screening and personalized treatments.

Here, we applied scRNA-seq technology to dissect the similarities and differences between primary tumors and their matched organoids. The paired primary tumors and organoids from one cholangiocarcinoma (CCA) and two pancreatic ductal adenocarcinoma (PDAC) patients were recruited for scRNA-seq. The cell compositions of primary tumors and their matched organoids were explored. We found that the organoids were made of almost all epithelial cells. The inter-patient and intra-tumor heterogeneity of primary tumors and organoids were identified. We also compared primary tumors and their matched organoids, regarding copy number variation (CNV) patterns, epithelial-mesenchymal transition (EMT) and cancer stem cell (CSC) properties, classical and basal-like subtypes, and metabolic and unfolded protein response (UPR) signatures. In addition, we revealed the heterogeneity of cancer associated fibroblasts (CAFs) and T cell, and explore the developmental trajectory of T cells. The receptor-ligand interactions between tumor cells and stromal cells were investigated. Overall, our findings provide insight into the similarities and differences between primary tumors and their matched organoids by scRNA-seq analysis. This study is also a valuable resource for facilitating further understanding of organoid model.

## 2. Materials and methods

### 2.1. Human specimens

Human primary CCA and PDAC resection specimens were obtained from the Department of General Surgery of Peking University First Hospital. No patients received any treatments before pancreatic resections. This study was reviewed and approved by Ethics Committee of Peking University First Hospital (Approval No. 2020-352) and was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained prior to the acquisition of specimens from all patients. Pathological diagnosis was confirmed based on pathologist assessment.

### 2.2. Tumor organoids culture

Fresh resection specimens were preserved in tissue preservation solution (Miltenyi, Cat. no. 130-100-008) on ice and transported to Guangdong Research Center of Organoid Engineering and Technology. Fresh tissues were washed three times with Hanks' balanced salt solution (Servicebio, Cat. no. G4203) containing penicillin/streptomycin and 10  $\mu$ M Y-27632 (Selleck, Cat. no. S1049). All tissues were minced into pieces, and digested at 37 °C for 1.5 h using tissue pre-treatment kit (Accurate Int., Guangzhou, China). The digested mixture was filtered with a 100  $\mu$ m cell strainer, then washed with advanced DMEM/F12 (Gibco, Cat. no. 12634010), and resuspended in advanced DMEM/F12/growth factor reduced Matrigel with the ratio of 1:1.5. The complex was inoculated on pre-warmed culture plate and allowed to solidify at 37 °C

for 15 min. After the solidification of the Matrigel domes, the organoids culture medium was supplemented and changed every 2–4 days. The medium comprised advanced DMEM/F12, 1x GlutaMax, 1x B27, 10 mM Nicotinamine, 1.25 mM N-Acetyl Cysteine, 20 mM HEPES, 10 nM Gastrin, 5 ng/mL FGF-basic, 15 ng/mL FGF-10, 50 ng/mL EGF, 100 ng/mL Wnt3a, 200 ng/mL Noggin, 250 ng/mL R-spondin-1, 500 nM A83-01, 10  $\mu$ M SB202190, and 10  $\mu$ M Y-27632. After 7–8 days culture, organoids were digested into small pieces using TrypLE Express (Gibco, Cat. no. 12604013) for further single cell sequencing and embedded into paraffin for histological analysis. The bright field images of organoids were captured on day 0, 3, 4, and 6.

### 2.3. Tissue dissociation and cell purification

Fresh resection specimens were preserved in the tissue preservation solution (Miltenyi, Cat. no. 130-100-008) on ice and transported to CapitalBio Technology within 2 h. Multiple tissue blocks were sampled to increase the success rate of tissue dissociation. All tissues were cut into around 1-mm pieces and incubated with a proper digestive solution including enzyme cocktail, consisting of the Type VIII Collagenase (Sigma-Aldrich, Cat. no. C2139), Dispase II (neutral protease, grade II) (Sigma-Aldrich, Cat. no. 4942078001), trypsin inhibitor (Sigma-Aldrich, Cat. no. T6522), and DNase I (Sigma-Aldrich, Cat. no. D5025). The cell suspension was filtered with a 40- $\mu$ m cell strainer (BD, Cat. no. 352340), followed by incubation with red blood cell lysis buffer (Roche, Cat. no. 11814389001) on ice for 10 min.

### 2.4. cDNA library preparation

The concentration of single cell suspension was determined using the Count Star instrument and adjusted to 1000 cells/ $\mu$ l. Using single cell 3' Library and Gel Bead Kit v3.1 (10x Genomics, Cat. No. 1000268) and Chromium Next GEM Chip G Single Cell Kit (10x Genomics, Cat. No. 1000120), the cell suspension was loaded onto the Chromium single cell controller (10x Genomics) to generate single-cell gel beads in the emulsion according to the manufacturer's instructions. In brief, single cells were suspended in PBS containing 0.04 % BSA. About 5000 cells were added to each channel, and the target cell will be recovered. Captured cells were lysed and the released RNA were barcoded through reverse transcription in individual GEMs. Standard reverse transcription was performed in 200  $\mu$ l tubes (NEST Biotechnology, Cat. No. 401001) on a S1000TM Touch Thermal Cycler (Bio Rad) at 53 °C for 45 min, then at 85 °C for 5 min, and hold at 4 °C. The cDNA libraries were generated, amplified, and quality assessed using the Agilent 4200. Finally, the cDNA libraries were sequenced using an Illumina Novaseq6000 sequencer with a sequencing depth of at least 100,000 reads per cell with paired-end 150 bp (PE150) reading strategy.

### 2.5. ScRNA-seq data mining

Raw data (BCL files) from Illumina Novaseq6000 platform was converted to fastq files using Illumina-implemented software bcl2fastq (v2.19.0.316). cDNA reads were aligned to human reference genome (GRCh38). Low-quality cells and genes filtering, barcode and UMI counting were performed with the cellranger software (v6.1.2) to generate the filtered gene-cell matrixes. Then, gene-cell matrixes were imported into R software to further filtered out low-quality cells (<500 genes/cell, >30 % mitochondria genes, <1000 transcripts/cell) and genes (<10 cells/gene) using 'Seurat' R package (v3.2.3). Gene expression levels were normalized (LogNormalize) with "NormalizedData" function. A total of 2000 highly variable genes were selected and used to conduct PCA reduction dimension. The t-distributed stochastic neighbor embedding (t-SNE) was performed. The 'Soupx' R package was applied to reduce the ambient mRNA contamination. Doublets were identified using the 'DoubletFinder' R package (v2.0.3), assuming that it was around 5 % doublet formation rate to the loaded

cells per sample in a droplet channel. Furthermore, the ‘Harmony’ R package (<https://github.com/immunogenomics/harmony>) was used to integrate gene-cell matrixes derived from different samples. Each cell cluster was identified by matching the cluster-specific genes with known signatures of cell populations reported in previous studies and Cell-Marker database (<http://bio-bigdata.hrbmu.edu.cn/CellMarker/>).

## 2.6. Single-cell copy number variation (CNV) inference

Somatic large-scale chromosomal copy number variation was inferred using the ‘inferCNV’ R package (v1.6.0). Briefly, a new gene-cell matrix of epithelial cells, annotation data, and gene/chromosome position files were prepared. Both T cells and endothelial cells were used as reference cells because they are considered to have no CNV. The CNV score of each cluster was calculated as quadratic sum of CNV region.

## 2.7. Single-cell regulatory network inference and clustering (SCENIC)

SCENIC analysis was conducted using the ‘SCENIC’ R package (v1.1.3) with default parameters [10]. In brief, we randomly picked out 200–400 cells from each subcluster to construct a new gene-cell matrix. Regions for transcription factor (TF) searching were restricted to 10 k distance centered the transcriptional start site (TSS) or 500 bp upstream of the TSSs. Transcription factor binding motifs (TFBS) overrepresented on a gene list and networks inferring were done using the ‘Rcis-Target’ R package (v1.6.0) and the ‘GENIE3’ R package (v1.8.0), with the 20-thousand motifs database. The cluster-specific TFs were defined as the top 10 or 15 highly enriched TFs according to a decrease in fold change compared with all the other subcluster using a Wilcoxon rank-sum test. Next, visualization of TFs network was performed.

## 2.8. GO and KEGG analysis

The differentially expressed genes (DEGs) and cluster-specific genes were identified using the ‘FindMarkers’ function in ‘Seurat’ R package. The online tool Metascape (<http://metascape.org/gp/index.html#/main/step1>) was used to conduct GO and KEGG analysis for top 200 subcluster-specific genes.

## 2.9. Single-cell receptor-ligand interactions

The gene-cell matrixes of tumor cells, Tregs, and complement-secreting CAFs (csCAFs) were merged and imported into R environment. The receptor-ligand pairs were mapped using the ‘iTALK’ R package (v0.1.0) with default parameters (<https://github.com/Coolgenome/iTALK>). Considering the test efficiency and computational burden, we randomly selected 200–400 cells of each subcluster for analysis. The visualization of ligand-receptor pairs in ‘iTALK’ was divided into four groups, consisted of cytokines, growth factors, immune checkpoints, and others.

## 2.10. Developmental trajectory of T cells

The cellular trajectory of T cells was inferred using the ‘Monocle 3’ R package (v1.2.9), with the assumption that one-dimensional ‘time’ can describe the high-dimensional expression values, so called pseudo-time analysis of single cells (<http://code-trapnell-lab.github.io/monocle-release/monocle3>). General pipeline was implemented. Cell trajectory and position were overlaid on the Uniform manifold approximation and projection (u-MAP) plot. The cell-state direction was shown using the heatmap.

## 2.11. Hematoxylin-eosin (H&E) staining

HE staining was performed on freshly prepared 4 % paraformaldehyde fixed paraffin-embedded tissue sections. The 5-μm

sections were cut and mounted on slides. Slides were deparaffinized in xylene, rehydrated in decreasing concentration of ethanol, then washed with PBS twice. Next, the slides were subjected to staining using the Hematoxylin and Eosin Staining Kit (Beyotime, Cat. no. C0105 M). The representative images were obtained for each section.

## 2.12. Multiplex immunohistochemistry

All sections were from the Department of General Surgery of Peking University First Hospital and subjected to multiplex immunohistochemistry (mIHC) using PANO Multiplex IHC kit (Panovue, Cat. No. 10234100100) according to manufacturer’s instructions. Briefly, sections were incubated in hot-air oven at 65 °C overnight, deparaffinized in fresh xylene for 10 min three times, rehydrated in graded ethanol (100 %–95 % – 70 %), then washed three times with PBS. The antigen retrieval was carried out with microwave heating method and cooled down for at least 15 min in the ice water bath. After blocking in blocking solution (Panovue, Cat. No. 0018001120) for 15 min at room temperature, sections were incubated with primary antibody for 30 min, secondary antibody for 10 min, and TSA Opal fluorophores for 10 min. Repeated antigen retrieval, blocking, primary and antibody incubation, TSA Opal fluorophores staining for each marker. Finally, sections were counterstained with DAPI (Sigma-Aldrich, Cat. No. D9542) for 5 min and mounted. The following primary antibodies were used: CK19 (Abcam, Cat. no. 52625, 1:200), GATA6 (CST, Cat. no. 5851T, 1:100), CK5 (CST, Cat. no. 71536T, 1:100), E-Cadherin (CST, Cat. no. 3195T, 1:100), and Vimentin (CST, Cat. no. 5741T, 1:200).

## 2.13. Statistical analysis

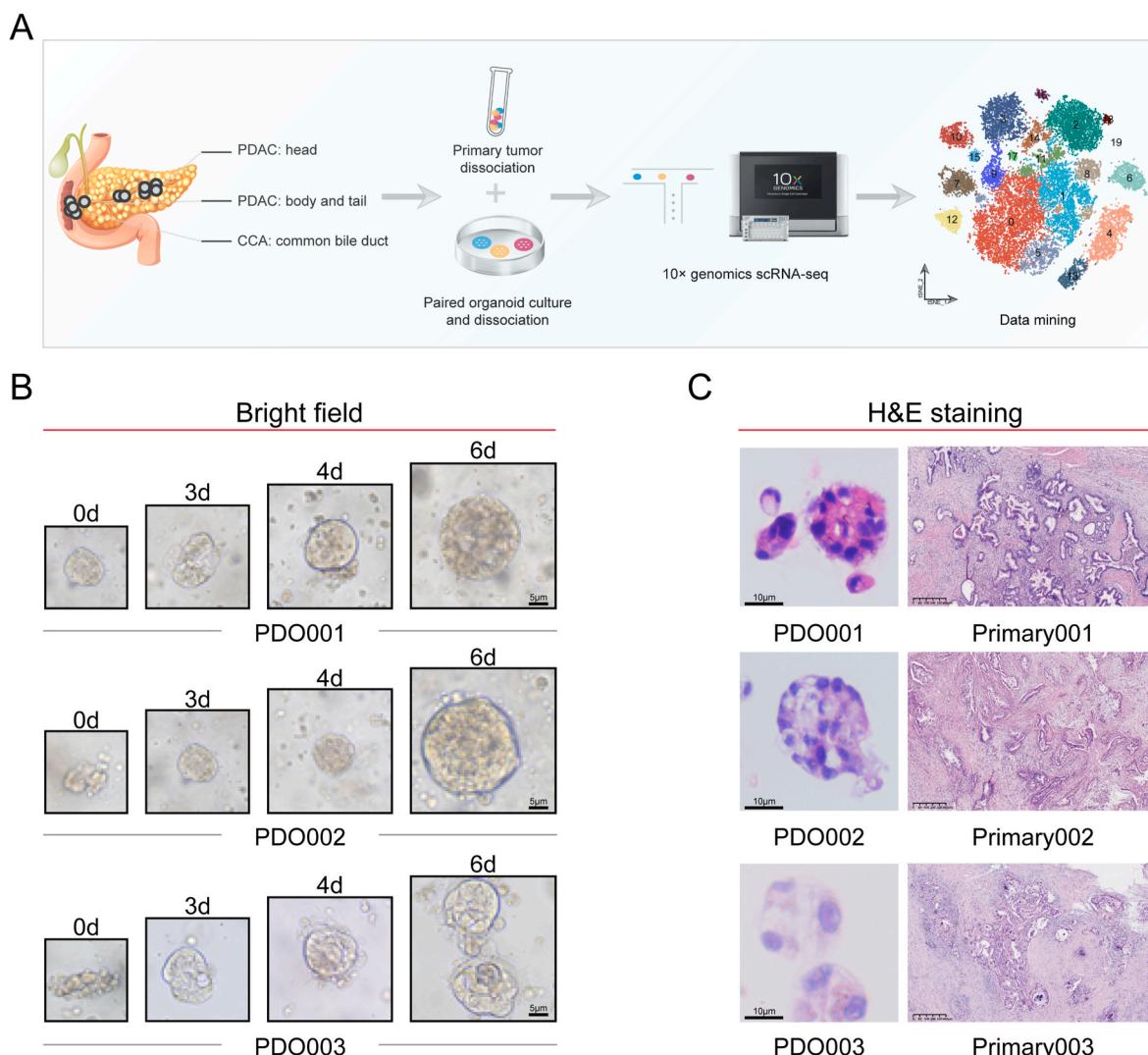
All statistical analyses were performed using the R software. For continuous variable, the independent-samples *t*-test and Mann-Whitney *U* test were performed to compare means between two groups. For categorical variable, the chi-square test or rank sum test was performed. The KM method and the corresponding log-rank test were performed to evaluate the prognostic value of selected markers. Statistical significance was defined as \**P* < 0.05, \*\**P* < 0.01, and \*\*\**P* < 0.001.

## 3. Results

### 3.1. ScRNA-seq reveals cell landscape of primary tumors and organoids

To compare primary tumors and organoids of PDAC at single-cell transcriptome level, patient-derived organoids were established. The general characteristics of the patients were shown in **Supplementary Table S1**. All patients did not receive any treatments prior to pancreatic surgery. We conducted scRNA-seq for three primary tumors and three paired organoids (**Fig. 1A**). The typical morphology of organoids was observed by bright-field microscopy and H&E staining (**Fig. 1B** and C). The pathological diagnosis confirmed that one of the patients was diagnosed with CCA and the other two were diagnosed with PDAC (**Fig. 1C**). In fact, the patient with CCA was predicted to have PDAC based on preoperative and intraoperative surgeons’ judgment.

A large gene-cell matrix, consisting of 8646 cells and 21,442 genes from the CCA and 30,162 cells and 24,395 genes from PDAC, was generated after stringent quality control using the CellRanger and Seurat tools. Cells were segregated into various clusters in two dimensions according to their transcriptional profiles using the t-distributed stochastic neighbor embedding (t-SNE). A total of 17 and 20 original clusters were identified in the CCA and PDAC groups, respectively (**Fig. 2A** and B). By cross-referencing the cluster-specific genes with known markers of cell populations [11–13], these clusters were identified as known cell types, including cholangiocytes (EPCAM), fibroblasts (LUM), T cells (CD3D), endothelial cells (CDH5), macrophages (AIF1), pericytes (RGS5), mast cells (CPA3), neutrophils (G0S2), ductal cells (EPCAM), acinar cells (PRSS1), B cells (MS4A1), Schwann cells (GAP43)



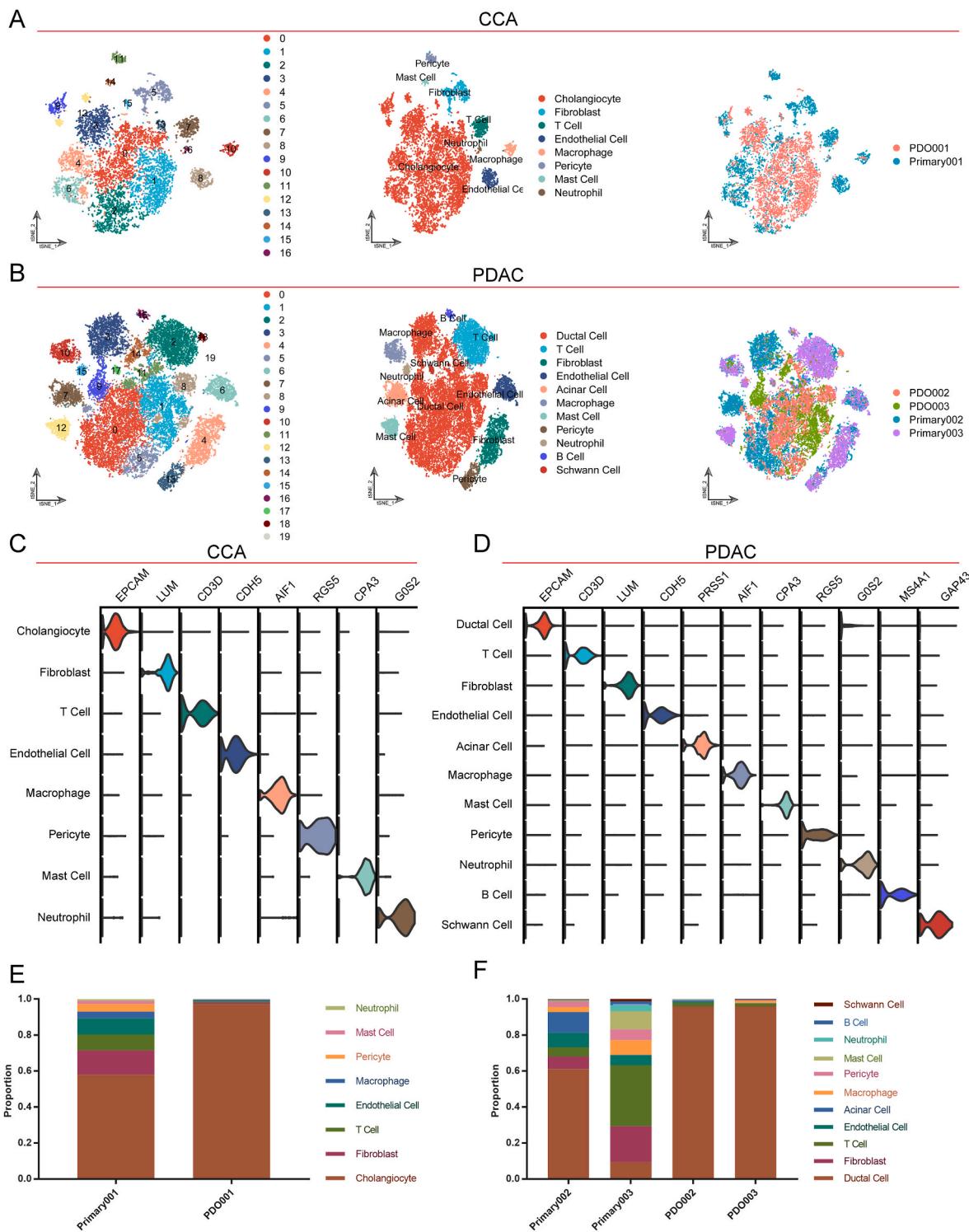
**Fig. 1.** Establishment of tumor organoids for scRNA-seq. (A) Graphical scheme describing the workflow of this study. (B) Representative bright field images of organoids from one CCA (PDO001) and two PDAC (PDO002/3) samples. The clinical characteristics of patients were described in Table S1 (C) Representative H&E staining images of organoids and their matched primary tumors. Scale bars were shown on the lower left corner of the images.

(Fig. 2A–D, and Supplementary Table S2). The top five cell-specific genes for each cell type were shown in Supplementary Figs. S1A–B. The cell composition of each sample was analyzed (Fig. 2E and F). We found that both CCA and PDAC primary tumors had complex tumor microenvironment, consisting of various epithelial and stromal cell types, whereas their matched organoids were made of almost all epithelial cells. In addition, we found different cell compositions between the two PDAC primary samples. One of them had a higher proportion of T cells and a lower proportion of epithelial cells (Primary003), indicating anti-tumor immune response.

### 3.2. Patient-derived organoids recapitulate CNV patterns of primary tumors

To uncover the similarities and differences in CNV profiles between the primary tumors and their matched organoids, epithelial cells from CCA and PDAC were isolated and integrated into a new gene-cell matrix. We conducted a separated clustering analysis, and identified 21 original epithelial subclusters (Fig. 3A). Remarkably, subclusters derived from different samples clustered separately, indicating the inter-patient heterogeneity. Multiple subclusters with different transcriptional profile were identified in the same sample, indicating intra-tumor

heterogeneity (Fig. 3A and B). All subclusters expressed EPCAM, verifying their epithelial identities (Fig. 3C). We inferred somatic large-scale chromosomal CNV and calculated CNV scores of each epithelial subcluster using the “inferCNV” package (Fig. 3D and E). Both T cells and endothelial cells were used as reference cells, as they are considered to be free of CNV. Notably, primary tumors and their matched organoids had the same CNV profiles. There were significant differences in CNV regions between two PDAC patients. Primary/PDO002 (subcluster 0/1/6/7/18/19/20) had copy number deletions on chromosome 5/7/8/10, while primary/PDO003 (subcluster 4/9/14) had copy number amplifications on chromosome 19, indicating inter-patient heterogeneity in PDAC. In addition, the amplifications of chromosome 12 were observed in primary/PDO001 from CCA, but deletions in primary/PDO002 from PDAC (Fig. 3D). We found that all epithelial subclusters, except for subcluster 11/15/17/19, had significantly higher CNV scores than reference cells (Fig. 3E). Therefore, we defined subcluster 11/15/17/19 as the normal ductal cells (Ductal 1–4), and others as the malignant ductal cells (Tumor 1–17) (Fig. 3A). Then, we calculated the sample composition of each subcluster and found most tumor clusters were derived solely from one specimen (Fig. 3F).

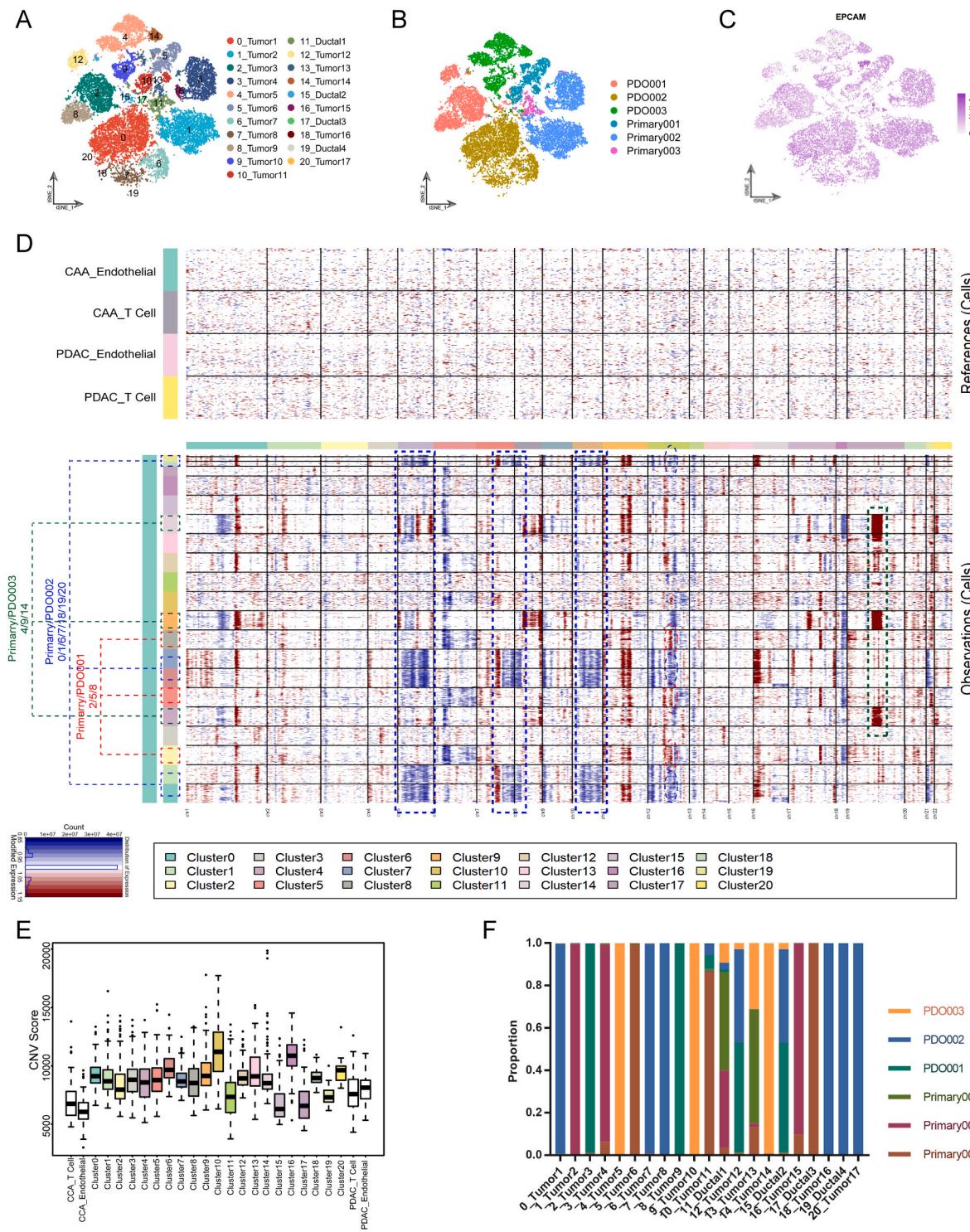


**Fig. 2.** ScRNA-seq reveals the cell landscape of the organoids and primary tumors. (A–B) t-SNE plot showing original cell clusters, renamed cell types, and samples information in CCA and PDAC. (C–D) Violin plots showing the normalized expression levels of known signature genes of distinct cell types. (E–F) Box plots showing the proportion of cell types among distinct samples from CCA and PDAC.

### 3.3. Cell cycle inference

Cell cycle phase score of each cell was calculated using the “Cell-CycleScoring” function implemented in the Seurat package. The results showed that subcluster 6/7 had significantly higher proportion of S and G2M phases (Fig. 4A and B). There were lower proportion of S and G2M phases in the PDAC primary tumors than in their matched organoids

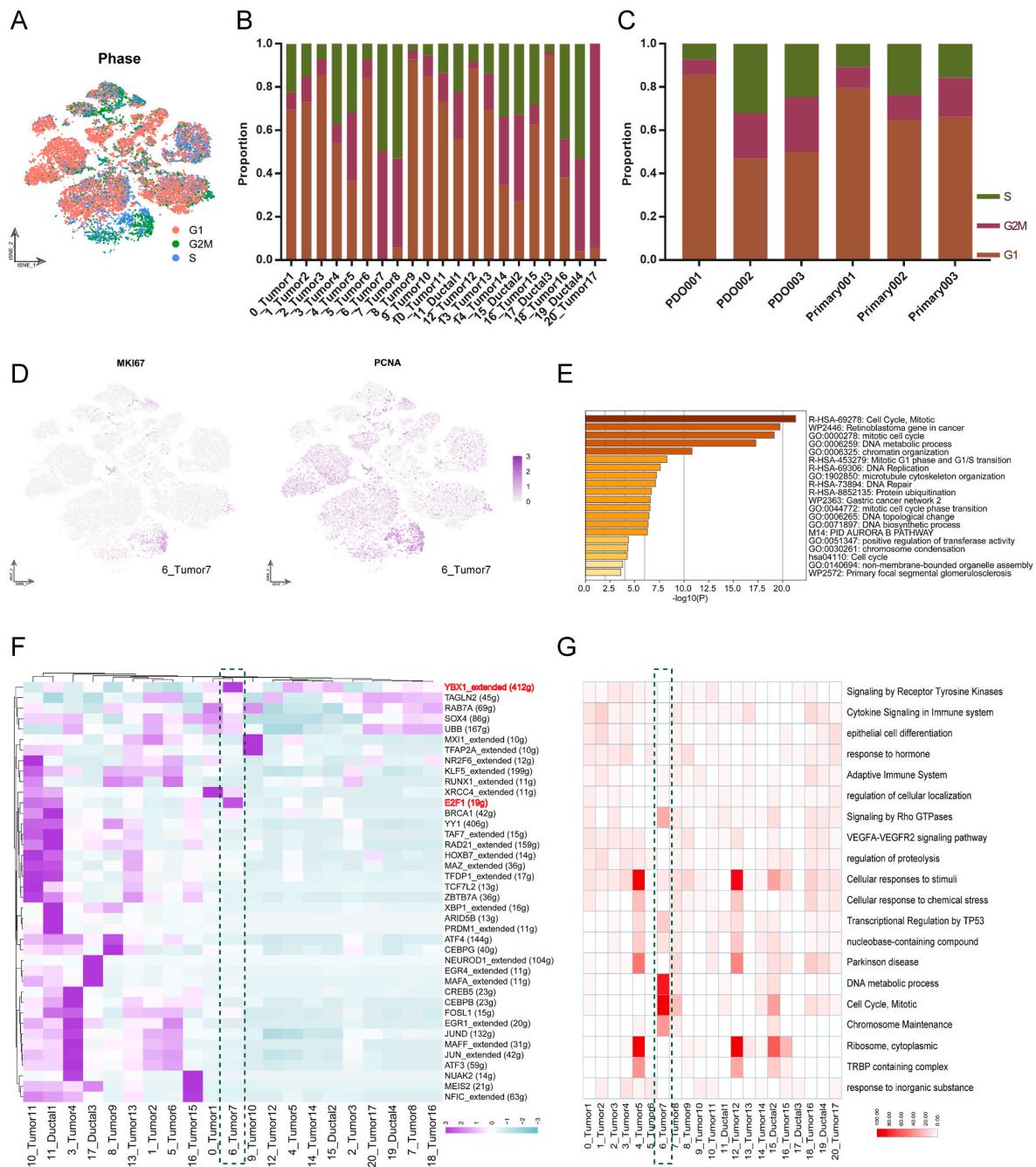
(Fig. 4C). The expression levels of MKI67 and PCNA were overlaid on the t-SNE plot, which indicates that subcluster 6 (6\_Tumor7) had enriched expression of these two proliferation-related markers (Fig. 4D). The subcluster 6 was solely derived from one sample (PDO002), and their specific genes were identified using the “FindMakers” function. GO and KEGG annotation analysis was performed (Fig. 4E). These genes were enriched in cell cycle, mitotic, DNA replication, and DNA repair,



**Fig. 3.** Organoids retain the CNV patterns of the matched primary tumors. (A–B) t-SNE plot showing epithelial cell subclusters and samples information. (C) The expression of EPCAM is overlaid on the t-SNE plot. The intensity of color (from white to purple) indicated the average expression level of EPCAM. (D) Heatmap showing large-scale CNV profile of each epithelial cell subcluster. Blue and red colors represented low and high CNV levels, respectively. Both T cells and endothelial cells were used as reference cells. The interested CNV regions are labeled with different colored frames. (E) Box plots are used to compare the CNV scores of distinct epithelial cell subclusters. White box represented reference cell clusters. (F) Box plots showing the sample composition of each epithelial cell subcluster.

suggesting the potential for proliferation of this tumor subcluster. Transcription factors (TFs) have been demonstrated to be the important regulators in cell fate and have the ability to shape different cell subpopulations. The cell cluster-specific TFs were generated using the single-cell regulatory network inference and clustering (SCENIC) method. Each epithelial subcluster had its' specific TFs pattern (Fig. 4F).

The subcluster 6 had enriched expression of YBX1 and E2F1 (Fig. 4F). YBX1 encodes a highly conserved DNA and RNA binding protein, engaged in DNA repair and transcription regulation [14]. It is now believed that the E2F family plays a crucial role in the control of cell cycle [15]. The potential biological function of each epithelial subcluster was investigated. The subcluster 6 was found to be enriched in signaling



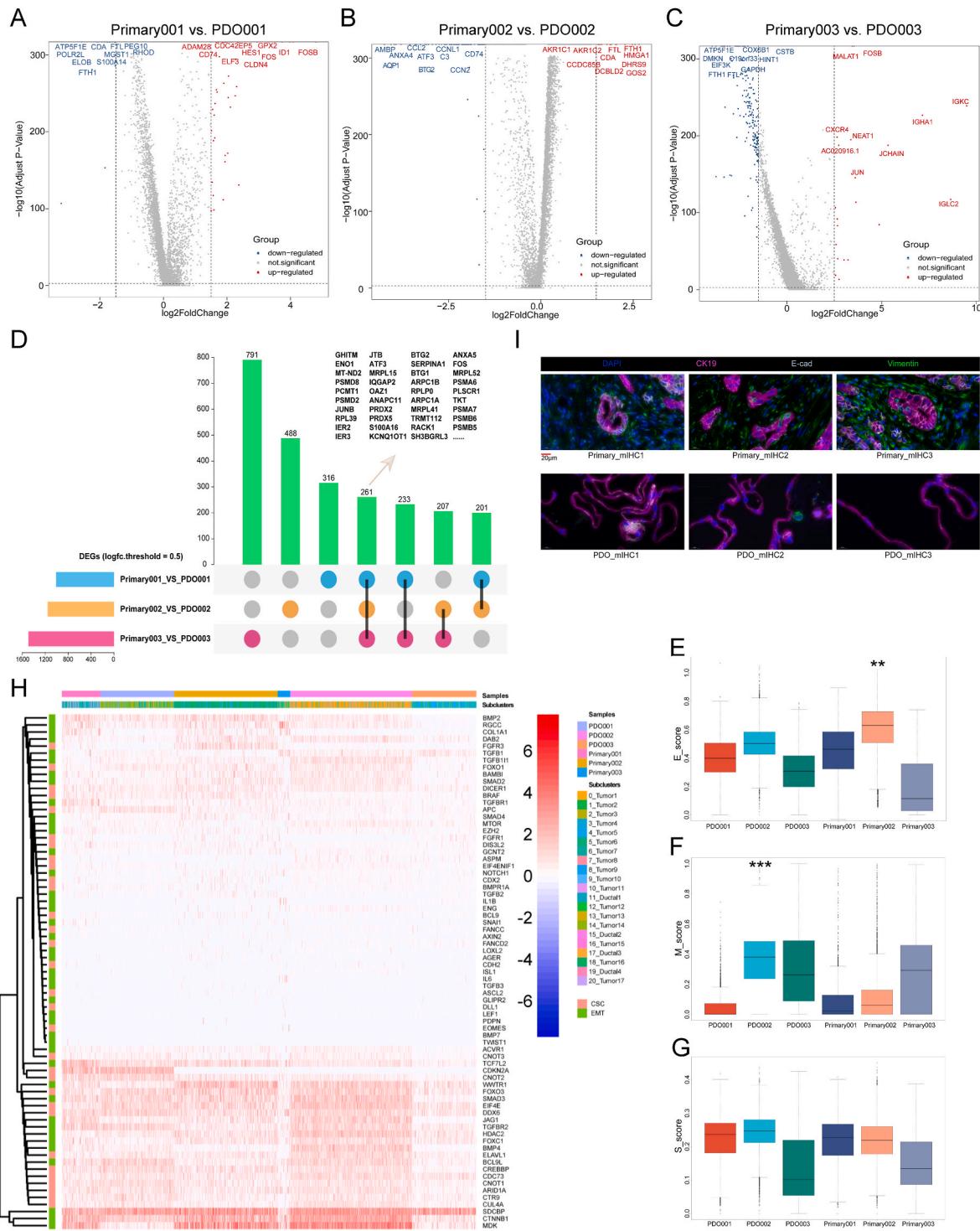
**Fig. 4.** Cell cycle state of epithelial cells. (A) Cell cycle phase is marked with different colors and overlaid on the t-SNE plot. (B–C) Box plots showing the proportion of cell cycle phase among the epithelial cell subclusters and samples. (D) The expressions of proliferation-related markers are overlaid on the t-SNE plot. The intensity of color (from white to purple) indicated the average expression levels of MKI67 and PCNA. (E) Gene annotation analysis of 6\_Tumor7-specific genes. Top GO terms were shown. Bar plots are colored according to their  $-\log_{10}(P)$ -values. (F) Heatmaps of the activation scores of each epithelial cell subcluster for expression regulated by transcription factors (TFs) using the SENIC method. Cell subcluster are indicated on bottom. (G) Gene Ontology enrichment analysis of epithelial cell subclusters, which are colored according to their  $-\log_{10}(P)$ -values.

by Rho GTPases, DNA metabolic process, cell cycle, mitotic, and chromosome maintenance. These findings suggested that organoid culture might alter the proliferation status of tumor cells in PDAC (Fig. 4G).

#### 3.4. Epithelial-mesenchymal transition (EMT) and cancer stem cell (CSC) properties

We compared epithelial cells from primary tumors and their matched organoids. Many DEGs were found between the two groups (Fig. 5A–C). The common DEGs were obtained taking the intersection of different

samples (`logfc.threshold = 0.5`), which were enriched in cellular responses to stimuli (Fig. 5D). Overall, organoid culture has a significant impact on transcriptional profiles of epithelial cells. Next, EMT and CSC scores of each cell were calculated using the “AddModuleScore” function according to a set of gene sets (Supplementary Table S3). The *E\_score* (Epithelial\_score) and *M\_score* (Mesenchymal\_score) of most epithelial subclusters showed an opposite trend. However, subcluster 7 and 9 from PDO002 and PDO003 respectively, scored relatively high in both *E\_score* and *M\_score*, therefore were considered as an intermediate state (Supplementary Figs. S2A–C). There were no significant differences in



**Fig. 5.** Epithelial-mesenchymal transition (EMT) and cancer stem cell (CSC) properties. (A–C) Volcano plot showing the differential expression genes (DEGs) between the organoids and their matched primary tumors. Red and blue dots represent the upregulated and downregulated genes, respectively. (D) The upset plots showing common DEGs across different samples, which labeled with different colors. The number of DEGs is shown with bar plots. The common DEGs are summarized in box. (E–G) Box plots are used to compare the organoids and their matched primary tumors regarding E\_score (Epithelial score), M\_score (Mesenchymal score), and S\_score (cancer stem properties score). (H) Heatmap showing the expression levels of markers of epithelial, mesenchymal, and cancer stem cell properties across distinct epithelial cell subclusters and samples. (I) Representative miHC images of the primary tumors and the organoids. Scale, 20  $\mu$ m.

E\_score and M\_score between primary tumors and the matched organoids (Primary001/003 vs. PDO001/003), whereas, the PDO002 significantly decreased the E\_score but increased the M\_score compared with the Primary002 (Fig. 5E and F). No significant difference was found in S\_score (cancer stem properties\_score) between these two groups (Fig. 5G). The expression of EMT and CSC markers among all subclusters

and samples were shown in Fig. 5H. Because the scRNA-seq lacks spatial context, we conducted miHC to further investigate the EMT status of tumor cells in the primary tumors and organoids. We found that most of tumor cells expressed E-cadherin (Epithelial\_marker). There were a few tumor cells expressing high level of Vimentin (Mesenchymal\_marker) in both primary tumors and organoids, indicating EMT (Fig. 5I).

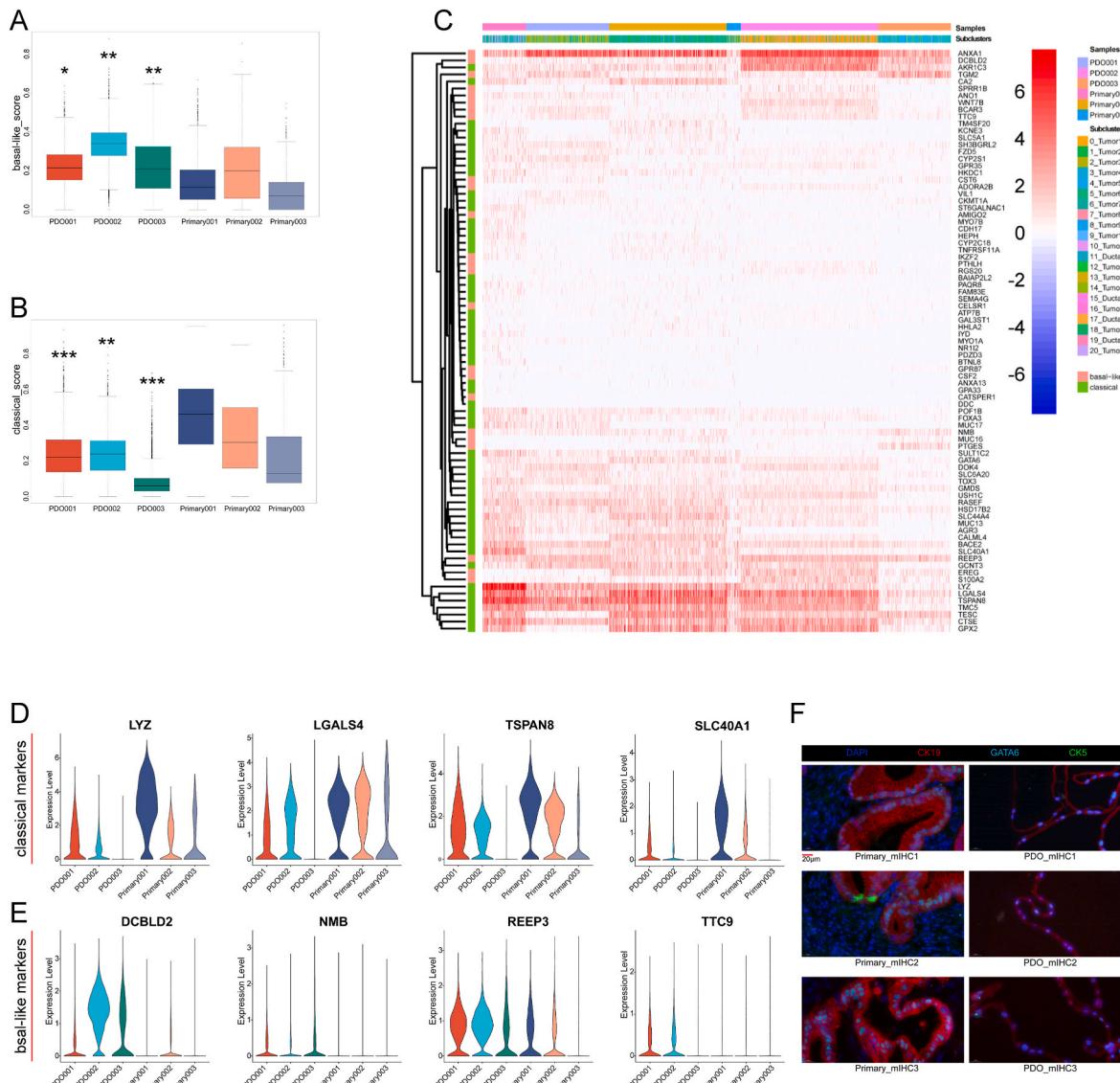
### 3.5. Molecular subtype transition

A growing body of study has reported and validated the molecular subtypes of PDAC [16,17]. Different subtypes of tumors have different prognosis and expression profiles. Two major subtypes of PDAC are thought to exist, including basal-like and classical subtypes. The basal-like and classical scores of each cell were calculated according to the corresponding marker genes (Fig. 6A and B, and Supplementary Figs. 2D–E and Table S4). Compared to primary tumors, the matched organoids had significantly higher basal-like\_score but lower classical\_score (Primary001/002/003 vs. PDO001/002/003; CCA and PDAC) (Fig. 6A and B). The expression of markers used for molecular subtype among all subclusters and samples were shown in Fig. 6C. Cells in the organoid group had lower expression of classical markers, such as LYZ, LGALS4, TSPAN8, and SLC40A1, in comparison with those in the primary tumor group (Fig. 6D). However, the expression of basal-like markers, such as DCBLD2, NMB, REEP3, and TTC9, showed the opposite trend (Fig. 6E). Among them, DCBLD2, encoding a cell surface protein involved in negative regulation of cell growth, and REEP3,

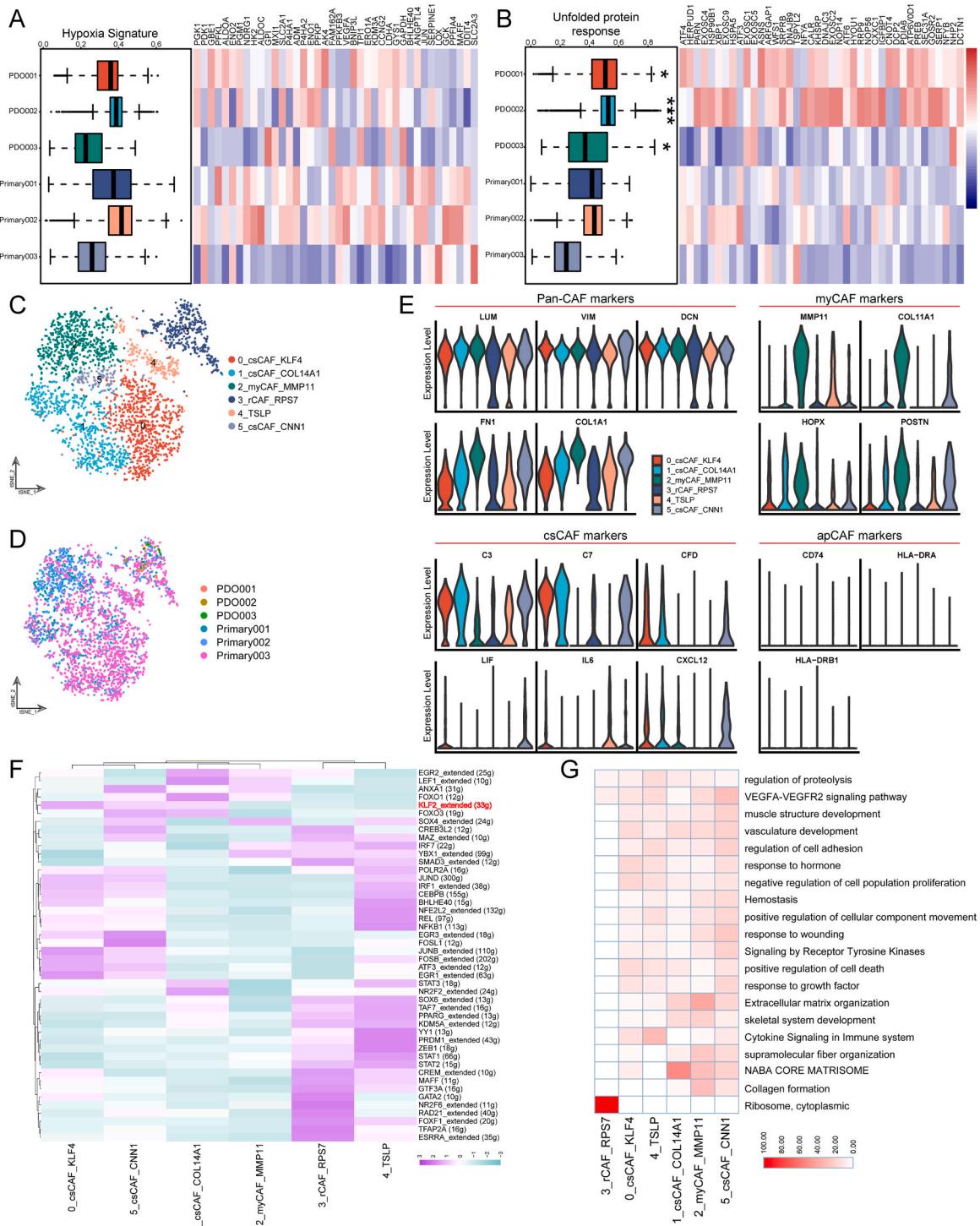
receptor expression enhancing protein, are two of the strongly related basal-like markers. Both of them were the promising prognostic biomarkers for the overall survival (OS) and disease-free survival (DFS) of PDAC (Supplementary Fig. S2F). We found that there existed both GATA6 (classical\_marker) positive- and CK5 (basal-like\_marker) positive-tumor cells in the same malignant duct in a primary tumor (Fig. 6F). However, we did not detect the CK5 positive-tumor cells in the organoids, which was inconsistent with scRNA-seq analysis (Fig. 6F). Taken together, these observations indicated that organoid culture might cause the transition of tumor cells to basal-like subtype in the transcriptional level.

### 3.6. Hallmark gene set signatures

Next, we compared the differences in hallmark gene set signatures between primary tumors and the matched organoids. The inter-patient variation in metabolic and unfolded protein response (UPR) signatures was observed (Fig. 7A and B, and Supplementary Figs. S3A–B and Table S5). There was no significant difference in the average degree of



**Fig. 6.** Molecular subtype transition. (A–B) Box plots are used to compare the organoids and their matched primary tumors regarding basal-like\_score and classical\_score. (C) Heatmap showing the expression levels of markers of basal-like and classical subtypes across distinct epithelial cell subclusters and samples. (D–E) Violin plots showing the normalized expression levels of selected marker genes across distinct samples. (F) Representative miHC images of the primary tumors and the organoids. Scale, 20  $\mu$ m.



**Fig. 7.** Full spectrum of cancer associated fibroblasts (CAFs). (A–B) Hypoxia- (A) and Unfolded protein response - (B) pathways-related scores using the AddModuleScore function. Left, box plots of expression of hallmark pathways-related signatures (defined as the mean normalized expression of genes in the signatures) across distinct samples. Right, heatmap of Z-scored mean expression of genes in signatures. (C–D) t-SNE plot showing CAFs subclusters and samples information. (E) Violin plots showing the normalized expression levels of selected marker genes of distinct CAFs subpopulations. (F) Heatmaps of the activation scores of each CAFs subcluster for expression regulated by transcription factors (TFs) using the SENIC method. Cell subcluster are indicated on bottom. Interested TF is marked with red color. (G) Gene Ontology enrichment analysis of CAFs subclusters, which are colored according to their -log10P-values.

hypoxia signatures between the samples from the primary tumors and the matched organoids (Fig. 7A). It is note that all samples from the organoids had significantly higher average degree of UPR signatures than from the primary tumors (Fig. 7B). In addition, a significantly higher degree of fatty acid metabolism and glycolysis signatures was found in the samples from organoids of PDAC (PDO002/003) compared

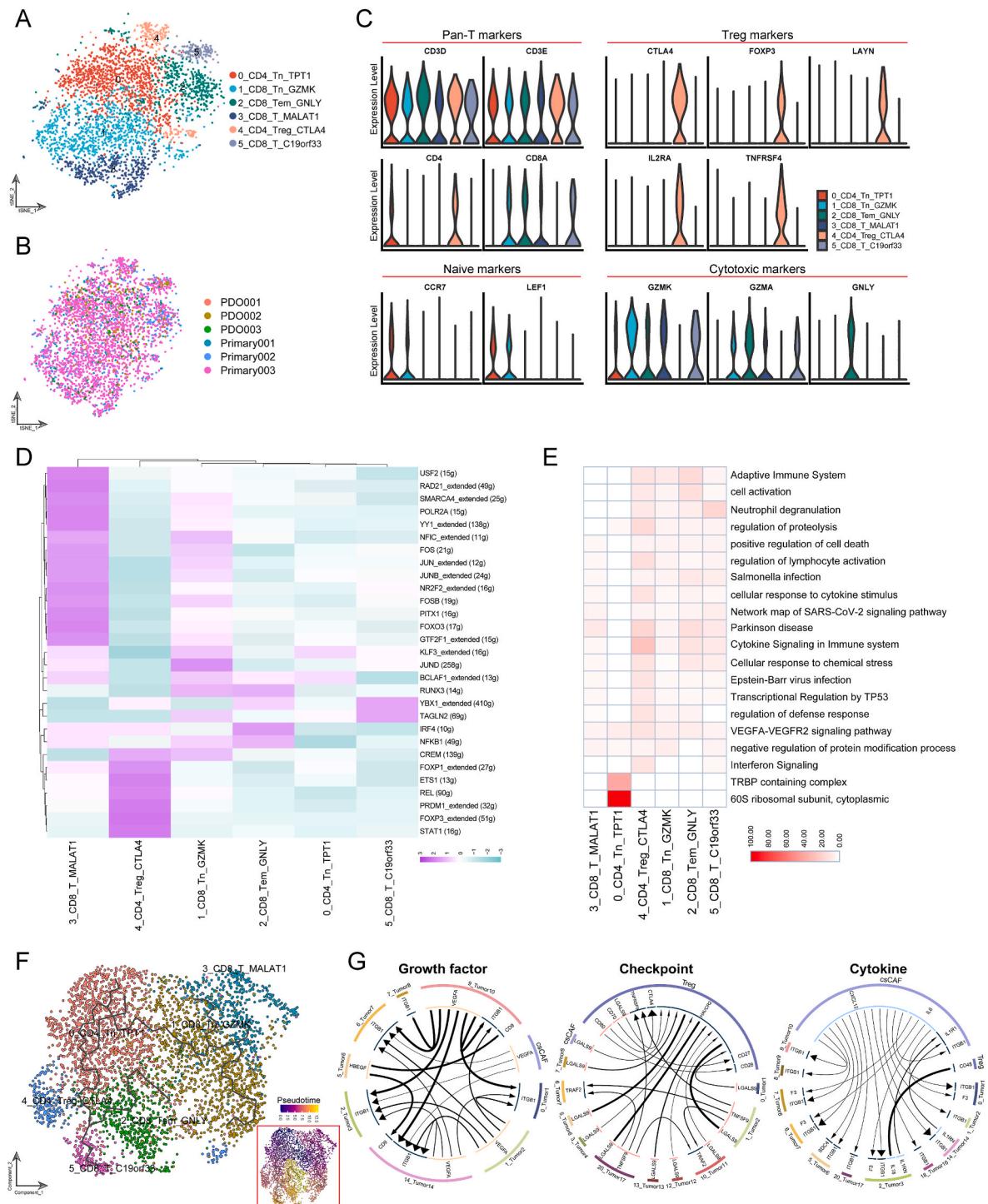
with their counterpart (Supplementary Figs. S3A–B).

### 3.7. ScRNA-seq explores the full spectrum of cancer associated fibroblasts (CAFs)

Both CCA and PDAC are characterized by complex tumor

microenvironment, consisting of a variety of stromal cell subpopulations. Among them, CAFs play a vital role in regulating tumor cell progression by various cell-to-cell communications [18]. To further identify the heterogeneity of CAFs, we isolated CAFs and conducted a separated clustering analysis. A total of six CAFs subclusters were identified (Fig. 7C). A very small number of CAFs were found to be

derived from organoids and all clustered in subcluster3 (3\_rCAF\_RPS7) (Fig. 7D). All subclusters expressed the common markers of fibroblast, including LUM, VIM, DCN, FN1, and COL1A1, confirming their CAFs identities (Fig. 7E). In accordance with our previous complement-secreting CAFs (csCAFs) characterization [11], subcluster 0/1/5 had enriched expression of C3, C7, and CFD, and was therefore



**Fig. 8.** The heterogeneity and developmental trajectory of T cells. (A–B) t-SNE plot showing T cells subclusters and samples information. (C) Violin plots showing the normalized expression levels of selected marker genes of distinct T cells subpopulations. (D) Heatmaps of the activation scores of each T cells subcluster for expression regulated by transcription factors (TFs) using the SENIC method. Cell subcluster are indicated on bottom. (E) Gene Ontology enrichment analysis of T cells subclusters, which are colored according to their -log10P-values. (F) Pseudo-time analysis trajectory analysis of T cells. Each dot represents one cell and is colored based on their subcluster. In the left bottom, pseudo-time score from dark blue to yellow indicates trajectory direction. (G) Chord diagram showing the receptor-ligand pairs among tumor cells, Tregs, and csCAFs, which were classified into three categories, including growth factors, cytokines, immune checkpoints.

identified as csCAF<sub>s</sub>. Notably, subcluster 0/5 also expressed low levels of inflammation-related factors, including LIF, IL6, and CXCL12 (Fig. 7E). Subcluster 2 was identified as myofibroblastic CAF<sub>s</sub> (myCAF<sub>s</sub>) because it expressed high levels of MMP11, COL11A1, HOPX, and POSTN [19]. However, we did not detect antigen-presenting CAF<sub>s</sub> (apCAF<sub>s</sub>) reported in previous study [19], which specifically expresses CD74, HLA-DRA, and HLA-DRB1. Then, CAF<sub>s</sub> subcluster-specific TFs were generated using the SCENIC method. Each subcluster had its' specific TFs pattern (Fig. 7F). The KLF2, a kruppel family member, was specifically enriched in csCAF<sub>s</sub> subpopulations, including subcluster 0/1/5. The effects of KLF2 expressed in csCAF<sub>s</sub> on tumor progression need to be further investigated. GO and KEGG annotation analysis indicated that subcluster 2 was mainly involved in extracellular matrix organization and collagen formation, whereas subcluster 0/4/5 was related to cytokine signaling in immune system and response to growth factor.

### 3.8. T cell heterogeneity and developmental trajectory

A large number of T cells were also identified in our scRNA-seq dataset. Thus, we further conducted a separated clustering analysis to explore T cell heterogeneity. A total of six distinct T cell subclusters were identified. A small percentage of T cells were derived from organoids and scattered in different subclusters (Fig. 8A and B). All subclusters expressed CD3D and CD3E, confirming their T cell identities (Fig. 8C). Subcluster 0/4 expressed CD4, while other subclusters expressed CD8, therefore regarded as CD4<sup>+</sup> and CD8<sup>+</sup> T cell, respectively. Subcluster 0/1 had enriched expression of CCR7 and LEF1, and was identified as naïve T cells (Tn). Subcluster 2 expressed cytotoxic markers and identified as effector memory T cells (Tem). Subcluster 4 was identified as regulatory T cells (Treg) because it expressed high levels of co-inhibitory molecules, including CTLA4, FOXP3, LAYN, IL2RA, and TNFRSF4 (Fig. 8C). Furthermore, we performed the SCENIC analysis to identify cell cluster-specific TFs. We identified well-defined TFs for Tregs, such as ETS1, REF, PRDM1, FOXP3, and STAT1 (Fig. 8D). This T cell subpopulation was enriched in adaptive immune system, regulation of proteolysis, and interferon signaling (Fig. 8E). To better understand the developmental state of T cells, we performed pseudo-time analysis using the Monocle 3. The trajectory tree in two dimensions rooted from subcluster 0/1 (Tn), sprouting into subcluster 4 (Treg) and subcluster 2 (Tem), ended with subcluster 5 (5\_CD8\_T\_C19orf33) (Fig. 8F).

It is widely accepted that there are complex cell-to-cell communications among various cell subpopulations in TME, promoting tumor progression together. To gain insight into the crosstalk between Tregs or csCAF<sub>s</sub> and tumor cells, we inferred the receptor-ligand interactions between them using the iTALK method (Fig. 8G, and Supplementary Fig. S3C). We observed a strong connection between tumor cells and Tregs or csCAF<sub>s</sub>. There were the VEGFA-ITGB1, CXCL12-ITGB1, IL6-F3, TIMP1-CD63 receptor-ligand pairs between csCAF<sub>s</sub> and tumor cells. Tregs could interact with tumor cells by various immune checkpoint molecules, such as TNFRSF9, CTLA4, and HAVCR2.

## 4. Discussion

The majority of solid tumors, including CCA and PDAC, are characterized by the complex TME. Sophisticated cell-to-cell communications play a pivotal role in facilitating tumor progression and chemotherapy resistance [20,21]. In this study, the cellular compositions of primary tumors and their matched organoids were investigated using scRNA-seq analysis. We found that the organoids were made up almost entirely of epithelial cells, with only a very small number of stromal cells (Fig. 2E and F). Thus, organoid could not simulate the real TME, and not conducive to studying crosstalk between tumor cells and other cell types. However, we observed that organoids retained the CNV patterns of primary tumors (Fig. 3D). In addition to stable genetic background and easy culture in vitro [22], organoids had a higher tumor purity than primary tumors, in which tumor cells only account for a small

percentage. These features make the organoid model more suitable for anti-tumor drug testing to inform treatment selection. The epithelial cell subclustering analysis indicated that tumor subclusters derived from different samples clustered separately. Even for the same sample, multiple tumor subclusters were identified (Fig. 3A and B). These findings suggested the inter-patient and intra-tumor heterogeneity. It is now believed that obvious tumor heterogeneity poses a huge challenge for the precision medicine [23]. We also identified a tumor subpopulation (6\_Tumor7) that had significantly enriched in proliferation-related markers, PCNA and MKI67, and cell cycle signaling pathway. This tumor subpopulation was derived from organoid sample (PDO002), thus we speculated that culture environment of organoid might promote the proliferation of tumor cells at a certain stage.

We compared the primary tumors and the matched organoids using the “FindMakers” function and found a large number of DEGs (Fig. 5A–C), indicating that organoid culture brings about the transcriptional profile change of tumor cells. Organoid sample (PDO002) significantly decreased E\_score and increased M\_score compared to the primary tumor. No significant difference in S\_score between the primary tumors and the matched organoids was detected. Therefore, EMT rather than CSC properties are affected by organoid culture. Currently, two major PDAC subtypes, basal-like and classical, have been well established based on a set of markers [17]. The classical subtype is closely related to higher level of differentiation and immune infiltration, while basal-like subtype is associated with a poorer prognosis [24]. In this study, we found that all organoid samples, including CCA and PDAC, had significantly higher basal-like score and expression levels of basal-like-related markers than their matched primary tumors, whereas classical\_score and classical-related markers showed an opposite trend by scRNA-seq analysis, indicating a molecular subtype transition in transcriptional level. However, we did not find that tumor cells in the organoids had higher expression level of CK5 (basal-like\_marker) than those in the primary tumors. In addition, we compared hallmark gene set signature between the primary tumors and the matched organoids, and observed that UPR signature was significantly upregulated in organoid samples. The UPR is a pro-survival mechanism triggered by accumulation of unfolded or misfolded proteins in the endoplasmic reticulum (ER) [25]. Therefore, tumor cells in organoid exploit intrinsic adaptive mechanisms, such as UPR, to overcome environmental stresses during culture.

The heterogeneity of CAF<sub>s</sub> has been attracting a lot of attention. Different CAF<sub>s</sub> subpopulations perform tumor-promoting or tumor-suppressing functions. Recently, a variety of CAF<sub>s</sub> subpopulations were identified using scRNA-seq analysis. In accordance with our previous study [11], a CAF subpopulation was identified, which specifically expressed many complement components, such as C3, C7, and CFD, therefore named as csCAF<sub>s</sub>. The SCENIC analysis indicated csCAF<sub>s</sub> had high activity of KLF2, and the underlying role of this TF in csCAF<sub>s</sub> remains to be explored. In addition to the expression of complement components, some csCAF<sub>s</sub> subclusters had enriched expression of LIF, IL6, and CXCL12, which are similar to inflammatory CAF<sub>s</sub> (iCAF<sub>s</sub>) reported by previous study [19]. However, we did not identify the apCAF<sub>s</sub>, expressing CD74, HLA-DRA, and HLA-DRB1, which might be attributed to limited sample size in this study. Recently, cross-species scRNA-seq analysis were conducted and we identified csCAF<sub>s</sub> in multiple human and mouse PDAC scRNA-seq datasets, and there was correlation between csCAF<sub>s</sub> and clinical characteristics (data not shown). These findings demonstrated the existence of csCAF<sub>s</sub> in PDAC. It is necessary to isolate CAF<sub>s</sub> using the fluorescence activating cell sorter (FACS) and explore its role in regulating tumor immune microenvironment. The heterogeneity of T cell was also uncovered in this study. Consistent with previous study [17], naïve T cells could develop towards effector memory T cells and Tregs, which deepens our understanding of T cell developmental trajectory and will facilitate the development of immunotherapy strategy.

This study has several limitations. First, the limited sample size was

used to compare primary tumors and their matched organoids by scRNA-seq analysis. Second, the duration of organoid culture was 7–8 days in this study. Therefore, the transcriptional profile changes of organoid under longer culture condition remained to be determined. Finally, we did not perform *in vitro* validation for the findings of scRNA-seq analysis, such as molecular subtype transition.

In summary, our work compares primary tumors and their matched organoids by scRNA-seq analysis. Organoid model retains the CNV pattern of the primary tumor, but had significant changes at transcriptome level, regarding EMT, molecular subtype, and UPR signature. Our findings facilitate further understanding of organoid model and confirm its application prospects in pancreaticobiliary cancer.

## Funding

This study was supported by the National Natural Science Foundation of China (NO. 82171722, 82271764, and 81871954), Peking University Medicine Seed Fund for Interdisciplinary Research (NO. 34302), and Innovation Fund for Outstanding Doctoral Candidates of Peking University Health Science Center.

## Data availability

The scRNA-seq dataset for the primary tumors and organoids has been deposited in Gene Expression Omnibus (GEO) repository with accession number: GSE214295 (PRJNA885258). All related codes and data analysis scripts will be provided upon request to first author Kai Chen ([Drchenkai@pku.edu.cn](mailto:Drchenkai@pku.edu.cn)).

## CRediT authorship contribution statement

**Kai Chen:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Yongsu Ma:** Validation, Resources, Methodology, Data curation. **Xiejian Zhong:** Validation, Software. **Jianqiang Lan:** Validation. **Di Long:** Validation. **Xiaodong Tian:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Conceptualization. **Yanlian Yang:** Validation, Resources, Project administration. **Yinmo Yang:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition, Conceptualization.

## Declaration of competing interest

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

## Acknowledgements

We would like to thank Jianming Zeng for writing many online tutorials about scRNA-seq analysis and Shuai He for his expert guidance.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.canlet.2023.216586>.

## References

- [1] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Cancer statistics Jemal, CA Cancer J Clin 73 (2023) 14–48, <https://doi.org/10.3322/caac.21763>, 2023.
- [2] W. Park, A. Chawla, E.M. O'Reilly, Pancreatic cancer: a review, JAMA 326 (2021) 851–862, <https://doi.org/10.1001/jama.2021.13027>.
- [3] S. Tsai, D.B. Evans, Updates on the management of pancreatic cancer, Surg. Oncol. Clin. 30 (2021) xvii–xviii, <https://doi.org/10.1016/j.soc.2021.07.001>.
- [4] L. Gonzalez-Silva, L. Quevedo, I. Varela, Tumor functional heterogeneity unraveled by scRNA-seq technologies: (trends in cancer 6, 13–19, 2020), Trends Cancer 7 (2021) 265, <https://doi.org/10.1016/j.trecan.2021.02.001>.
- [5] G.J. Yoshida, Applications of patient-derived tumor xenograft models and tumor organoids, J. Hematol. Oncol. 13 (2020) 4, <https://doi.org/10.1186/s13045-019-0829-z>.
- [6] A. Fatehullah, S.H. Tan, N. Barker, Organoids as an *in vitro* model of human development and disease, Nat. Cell Biol. 18 (2016) 246–254, <https://doi.org/10.1038/ncb3312>.
- [7] J. Vallejo, C. Cochain, A. Zernecke, K. Ley, Heterogeneity of immune cells in human atherosclerosis revealed by scRNA-Seq, Cardiovasc. Res. 117 (2021) 2537–2543, <https://doi.org/10.1093/cvr/cvab260>.
- [8] S. Ding, X. Chen, K. Shen, Single-cell RNA sequencing in breast cancer: understanding tumor heterogeneity and paving roads to individualized therapy, Cancer Commun. 40 (2020) 329–344, <https://doi.org/10.1002/cac2.12078>.
- [9] C. Ziegenhain, et al., Comparative analysis of single-cell RNA sequencing methods, Mol Cell 65 (2017) 631–643 e634, <https://doi.org/10.1016/j.molcel.2017.01.023>.
- [10] S. Aibar, et al., SCENIC: single-cell regulatory network inference and clustering, Nat. Methods 14 (2017) 1083–1086, <https://doi.org/10.1038/nmeth.4463>.
- [11] K. Chen, et al., Single-cell RNA-seq reveals dynamic change in tumor microenvironment during pancreatic ductal adenocarcinoma malignant progression, EBioMedicine 66 (2021), 103315, <https://doi.org/10.1016/j.ebiom.2021.103315>.
- [12] J. Peng, et al., Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma, Cell Res. 29 (2019) 725–738, <https://doi.org/10.1038/s41422-019-0195-y>.
- [13] M. Zhang, et al., Single-cell transcriptomic architecture and intercellular crosstalk of human intrahepatic cholangiocarcinoma, J. Hepatol. 73 (2020) 1118–1130, <https://doi.org/10.1016/j.jhep.2020.05.039>.
- [14] A.K. Jayavelu, et al., Splicing factor YBX1 mediates persistence of JAK2-mutated neoplasms, Nature 588 (2020) 157–163, <https://doi.org/10.1038/s41586-020-2968-3>.
- [15] H.A. Segern, et al., Excessive E2F transcription in single cancer cells precludes transient cell-cycle exit after DNA damage, Cell Rep. 33 (2020), 108449, <https://doi.org/10.1016/j.celrep.2020.108449>.
- [16] P. Bailey, et al., Genomic analyses identify molecular subtypes of pancreatic cancer, Nature 531 (2016) 47–52, <https://doi.org/10.1038/nature16965>.
- [17] R.A. Moffitt, et al., Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma, Nat. Genet. 47 (2015) 1168–1178, <https://doi.org/10.1038/ng.3398>.
- [18] X. Mao, et al., Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives, Mol. Cancer 20 (2021) 131, <https://doi.org/10.1186/s12943-021-01428-1>.
- [19] E. Elyada, et al., Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts, Cancer Discov. 9 (2019) 1102–1123, <https://doi.org/10.1158/2159-8290.CD-19-0094>.
- [20] J. Kang, et al., Tumor microenvironment mechanisms and bone metastatic disease progression of prostate cancer, Cancer Lett. 530 (2022) 156–169, <https://doi.org/10.1016/j.canlet.2022.01.015>.
- [21] Y. Du, et al., Single-cell RNA sequencing unveils the communications between malignant T and myeloid cells contributing to tumor growth and immunosuppression in cutaneous T-cell lymphoma, Cancer Lett. 551 (2022), 215972, <https://doi.org/10.1016/j.canlet.2022.215972>.
- [22] H. Tiriac, et al., Organoid profiling identifies common responders to chemotherapy in pancreatic cancer, Cancer Discov. 8 (2018) 1112–1129, <https://doi.org/10.1158/2159-8290.CD-18-0349>.
- [23] I. Dagogo-Jack, A.T. Shaw, Tumour heterogeneity and resistance to cancer therapies, Nat. Rev. Clin. Oncol. 15 (2018) 81–94, <https://doi.org/10.1038/nrclinonc.2017.166>.
- [24] N. Juiz, et al., Basal-like and classical cells coexist in pancreatic cancer revealed by single-cell analysis on biopsy-derived pancreatic cancer organoids from the classical subtype, Faseb. J. 34 (2020) 12214–12228, <https://doi.org/10.1096/fj.202000363RR>.
- [25] E. Madden, S.E. Logue, S.J. Healy, S. Manie, A. Samali, The role of the unfolded protein response in cancer progression: from oncogenesis to chemoresistance, Biol Cell 111 (2019) 1–17, <https://doi.org/10.1111/boc.201800050>.