

Single-Cell Transcriptomic Analysis Reveals Two Molecularly and Clinically Distinct Subtypes of Intrahepatic Cholangiocarcinoma

Song Guohe

Department of General Surgery, Zhong shan Hospital, School of Medicine, Fu Dan University

Yang Shi

School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing, China.

Lu Meng

Key Laboratory of Molecular Virology & Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai, China.

Siyuan Huang

Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China.

Jiaqiang Ma

Key Laboratory of Molecular Virology & Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai, China.

Juan Zhang

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University

Yang Yang

First Affiliated Hospital of Zhengzhou University

Jiaxin Li

Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China.

Youpei Lin

Fudan University

Shuaixi Yang

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University

Dongning Rao

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University

Yifei Cheng

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University

Jian Lin

Shanghai Jiao-Tong University

Shuyi Ji

Department of Cancer Center, Jin Shan Hospital, Fudan University, Shanghai, China.

Yuming Liu

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University

Shan Jiang

Key Laboratory of Molecular Virology & Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai, China.

Xiaoliang Wang

Shanghai First People's Hospital, Medical College

Shu Zhang

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University

Ai-Wu Ke

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University

Xiaoying Wang

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Fudan University

Ya Cao

Key Laboratory of Carcinogenesis and Invasion, Chinese Ministry of Education, Department of Otolaryngology Head and Neck Surgery, Xiangya Hospital, Central South University, Changsha 410078

<https://orcid.org/0000-0002-3558-3336>

Yuan Ji

Department of Pathology, Zhongshan Hospital, Fudan University, Shanghai, 200032

<https://orcid.org/0000-0002-2631-9082>

Jian Zhou

Zhongshan Hospital, Fudan University <https://orcid.org/0000-0002-2118-1117>

Jia Fan

Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, Fudan University <https://orcid.org/0000-0001-5158-629X>

Xiaoming Zhang

Institut Pasteur of Shanghai

Ruibin Xi

Peking University <https://orcid.org/0000-0001-7545-7361>

Qiang Gao (✉ gaoqiang@fudan.edu.cn)

Liver Cancer Institute, Zhongshan Hospital, Shanghai Medical College, Fudan University

Article

Keywords: Intrahepatic cholangiocarcinoma (iCCA), cancer, tumor cells

Posted Date: June 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-641386/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on March 28th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-29164-0>.

Abstract

Intrahepatic cholangiocarcinoma (iCCA) is a highly heterogeneous cancer with limited understanding of its classification and tumor microenvironment. Here, we performed single-cell RNA sequencing on 144,878 cells from 14 pairs of iCCA tumors and non-tumor liver tissues. We found that S100P and SPP1 are two reliable markers for iCCA perihilar large duct type (iCCA^{phl}) and peripheral small duct type (iCCA^{pps}). S100P + SPP1- iCCA^{phl} has significantly reduced levels of infiltrating CD3⁺ T cells, CD56⁺ NK cells, and increased CCL18⁺ macrophages compared to S100P-SPP1 + iCCA^{pps}. The transcription factor CREB3L1 is identified to regulate the S100P expression and promote tumor cell invasion. S100P-SPP1 + iCCA^{pps} has significantly more SPP1⁺ macrophage infiltration, less aggressiveness and better survival than S100P + SPP1- iCCA^{phl}. Moreover, S100P-SPP1 + iCCA^{pps} harbors tumor cells at different status of differentiation, such as ALB + hepatocyte differentiation and ID3 + stemness. Our study extends our understanding of the diversity of tumor cells in iCCA and provides clearer understanding of iCCA classification.

Introduction

Intrahepatic cholangiocarcinoma (iCCA) is the second most common primary liver malignancy after hepatocellular carcinoma, with poor outcome and rising incidence globally¹. As a highly heterogeneous disease, iCCA can originate from cholangiocytes located at any point of biliary tree above the second-order bile ducts. Recently, the World Health Organization and European Network for the Study of Cholangiocarcinoma have recognized that iCCA can be classified into two histologically distinct subtypes, including perihilar large duct type (iCCA^{phl}) and peripheral small duct type (iCCA^{pps}), according to the level or size of the affected bile duct^{2,3}. Indeed, emerging evidence has indicated that the two histological subtypes of iCCA harbored distinct cellular origins and pathogenesis⁴.

Generally, iCCA^{phl} is considered to be derived from large intrahepatic bile ducts and mainly composed of mucin-producing cholangiocytes. This subtype of iCCA is characterized by mucus hypersecretion and has higher lymph node metastasis rates and worse survival⁵ compared with iCCA^{pps}. It has been reported that MUC5AC, one of the main components of mucus, is frequently overexpressed in iCCA^{phl} and associated with aggressive tumor behavior⁶. Also, S100P, a member of the S100 family of EF-hand calcium-binding proteins, that highly expresses in various types of cancer and plays crucial roles in tumor progression⁷, is also upregulated in mucin producing iCCAs and suggested to be an important marker^{8,9} for iCCA^{phl}. On the contrary, iCCA^{pps} is commonly believed originating from small intrahepatic bile ducts with no or minimal mucin production. It has been found that iCCA^{pps} expresses CDH2 more frequently than iCCA^{phl} and presents distinctive clinical and molecular features¹⁰. Moreover, NCAM, a marker of hepatic progenitor cells, was also expressed in iCCA^{pps}, as well as cholangiolocellular carcinoma (CLC) which is thought to originate from canals of Hering/bile ductules^{3,5}. Although the two subtypes of iCCA displayed significant differences in mucin production, shape of tumor cells and patient prognosis^{3,4},

there is no consensus and definite panel of markers to distinguish them, and our knowledge on their biological, molecular, and therapeutic difference is still limited.

Single-cell RNA sequencing (scRNA-seq) is a powerful technology for cancer research. Previous scRNA-seq studies have reported the complexity of the tumor microenvironment in iCCAs without taking into consideration of the histological classification, which may not accurately reflect the diversity of this tumor^{11, 12}. Here, we performed a transcriptome analysis of 144,878 single cells from 14 iCCA samples. We identified and independently validated that SPP1, together with S100P, are optimal discriminatory biomarkers for iCCA^{phl} and iCCA^{pps}. As compared with S100P-SPP1+ iCCA^{pps}, S100P+SPP1- iCCA^{phl} had increased CCL18⁺ macrophages infiltration, decreased SPP1⁺ macrophages, aggressive phenotypes and worse prognosis. Our data shed new light on the classification of iCCA, which might guide the development of rational precision therapies to benefit a wide range of patients.

Results

Single-cell profiling of the tumor ecosystem in iCCA.

We applied scRNA-seq and whole-exome sequencing (WES) on tumor and paired adjacent non-tumor liver tissues from fourteen treatment-naïve iCCA patients (**Fig. 1a and Supplementary Fig. 1a**). All tumors were negative for Hep-Par 1 (a sensitive marker of hepatocytes) expression (**Supplementary Fig. 1b**). The patient clinicopathological characteristics are presented in **Supplementary Data 1**. We obtained single-cell transcriptomes for 144,878 cells after quality control. Thirteen main cell clusters with the expression of known marker genes were identified including epithelial cells, monocytes, macrophages, dendritic cells (DC), natural killer (NK) cells, CD4⁺ T cells, regulatory T cells (Treg), CD8⁺ T cells, mucosal-associated invariant T (MAIT) cells, B cells, plasma B cells, fibroblasts, and endothelial cells (**Fig. 1b**). Totally, we identified 23,667 malignant cells by inferring large-scale copy number variations (CNVs) (**Fig. 1c and Supplementary Fig. 1c**). This classification was further confirmed by the significant difference of epithelial marker expression and unique molecular identifiers (UMIs) between malignant and non-malignant cells (**Fig. 1d and Supplementary Fig. 1d**). Consistent with previous findings in other tumors, malignant cells showed strong intertumoral heterogeneity and formed patient-specific clusters^{13, 14} (**Fig. 1e**). Also, infiltrating immune cells were found to be significantly heterogeneous among different patients and between tumor and peri-tumor tissues (**Supplementary Fig. 1e, f**). For examples, macrophages, CD4⁺ T cells and Tregs were highly infiltrated in the tumor, while MAIT cells were mainly distributed in the adjacent liver tissues (**Fig. 1f**).

SPP1 is a representative marker for iCCA^{pps}.

To explore the subtypes of iCCA with different cell origins at the single cell level, we examined the expression of seven previously proposed markers of iCCA^{phl} (*S100P*, *MUC5AC*, *TFF1*, *MUC1*, and *MUC6*) and iCCA^{pps} (*NCAM1* and *CDH2*) in malignant cells^{2, 3} (**Fig. 2a and Supplementary Fig. 2a**). We found that 7 out of 14 iCCAs (P02, P03, P04, P06, P16, P17, and P18) exhibited high expression of iCCA^{phl} markers

such as *S100P* and *MUC5AC*, indicating their origin from large intrahepatic bile ducts. Notably, S100P+ cells accounted for 91.14% of total tumor cells from these seven iCCAs and displayed more representative and extensive expression compared with the other markers (*MUC5AC*: 42.37%, *TFF1*: 77.61%, *MUC1*: 87.21%, and *MUC6*: 22.97%). The 14 iCCAs can be divided into two groups based on *S100P* expression, which was confirmed by immunohistochemistry (**Supplementary Fig 2b, c**). For the remaining seven S100P- iCCAs (P09, P10, P12, P13, P14, P15, and P19), they expressed iCCA^{PPS} markers *NCAM1* and *CDH2*, which were mutually exclusive with the expression of *S100P*, confirming the different origins of these tumor cells. However, the positive cells of *NCAM1* (2.36%) and *CDH2* (31.86%) accounted for a very low proportion of the total tumor cells in these seven S100P- iCCAs. To find more representative markers for iCCA^{PPS}, we searched for genes mutually exclusive with S100P but expressed extensively in iCCA^{PPS}. Gene such as *SPP1* had low expression in S100P+ and high expression in S100P- cells, making it potential biomarkers (**Fig. 2b and Supplementary Data 2**). *SPP1*, also known as osteopontin (OPN), is highly expressed in a variety of tumors and plays important roles in tumor progression¹⁵. We confirmed that the seven S100P- iCCAs showed high expression of *SPP1* both at the cellular (87.17% of S100P- iCCAs' tumor cells) and tissue level (**Fig. 2c and Supplementary Fig. 2d**). Thus, we divided 14 iCCAs into S100P+SPP1- iCCA^{phl} and S100P-SPP1+ iCCA^{PPS} subgroups based on the expression of S100P and SPP1.

According to our scRNA-seq data, most of the tumor cells either expressed *S100P* (23.95%) or *SPP1* (60.05%), while only 10.01% and 5.98% tumor cells showed double negativity or double positivity, respectively (**Supplementary Fig. 2e**). To further explore whether the expression of S100P and SPP1 in iCCA were mutually exclusive in a larger cohort, immunohistochemistry was performed on a tissue microarray (TMA) containing 201 iCCAs. We found that 92.54% iCCAs can be clearly divided into S100P+SPP1- (33.83%, 68 patients) and S100P-SPP1+ (58.71%, 118 patients) iCCAs, while only 5.97% (12 patients) and 1.49% (3 patients) were classified as S100P-SPP1- and S100P+SPP1+ iCCAs, respectively (**Fig. 2d and Supplementary Fig. 2f**). Importantly, S100P+SPP1- iCCAs had a significantly worse prognosis than S100P-SPP1+ iCCAs ($P = 0.008$, **Fig. 2e**), which was further confirmed by the multivariate Cox regression analysis (HR, 1.922; 95% CI, 1.257-2.939; $P = 0.003$, **Supplementary Data 3**). Also, S100P+SPP1- iCCAs significantly correlated with higher CA19-9 ($P < 0.01$), CEA ($P < 0.01$), Ki67 expression ($P = 0.025$), lymph node metastasis ($P = 0.013$) and advanced TNM stage ($P = 0.021$), but negatively correlated with tumor size ($P = 0.019$), HBsAg status ($P < 0.01$) and liver cirrhosis ($P = 0.049$) (**Fig. 2f and Supplementary Data 3**). The higher percentage of HBsAg positive status and liver cirrhosis in S100P-SPP1+ iCCAs further support the notion that iCCA^{PPS} usually develops on a background of chronic liver disease.⁵

We further evaluated the effect of S100P and SPP1 in distinguishing iCCA^{phl} and iCCA^{PPS} in two RNA-seq databases of cholangiocarcinoma. Patients from Jusakul et al.'s dataset¹⁶ includes iCCA (81 patients) and extrahepatic cholangiocarcinoma (ECC, 34 patients), while patients from Sia et al.'s dataset¹⁷ only contains iCCA (109 patients). We found that 81.48% iCCAs can be divided into two independent groups according to the expression of S100P and SPP1 in Jusakul et al.'s dataset (**Fig. 2g and Supplementary**

Data 4). The S100P-SPP1+ samples almost exclusively exist in iCCA instead of ECC, further supporting their distinct origination (**Fig. 2h**). Survival analysis showed that the prognosis of S100P+SPP1- iCCAs were significantly worse than S100P-SPP1+ iCCAs ($P < 0.01$, **Fig. 2i**). Similar results were obtained from Sia et al.'s dataset (**Supplementary Fig. 3a, b**).

Analysis of the WES data found that S100P+SPP1- iCCAs tended to have more *TP53* (4/14), *SYNE1* (3/14), and *EPHA2* (3/14) mutations, while S100P-SPP1+ iCCAs harbored more *BAP1* (3/14) mutations, which was consistent with previous studies^{10, 16} (**Supplementary Fig. 3c, d**). We also found that the DNA methylation level of *S100P* in S100P+SPP1- was significantly lower than that in S100P-SPP1+, while no apparent difference was observed in CNVs¹⁶, indicating potential epigenetic regulation of *S100P* in these two iCCA subtypes (**Supplementary Fig. 3e, f**). Taken together, these results indicate that S100P and SPP1 are two optimal biomarkers for distinguishing iCCA^{phl} and iCCA^{pps}, which can effectively divide the iCCA patients into two subtypes with different cell origins and clinicopathological characteristics.

Molecular profiles and transcription networks of S100P+SPP1- and S100P-SPP1+ iCCAs

The presence of two main subgroups of malignant cells in iCCA prompted us to investigate their unique gene expression profiles. We first evaluated their intratumor heterogeneity (ITH) at the genomic and single-cell transcriptome levels. The results showed no significant difference in genomic ITH, but a significant higher transcriptomic ITH in S100P+SPP1- iCCAs (**Fig. 3a**). This was consistent with previous study that higher transcriptomic ITH predicted poor survival¹¹. Subsequently, we identified 755 differentially expressed genes between these two groups of malignant cells ($|\logFC| > 1.5$ and $P < .01$, **Supplementary Fig. 4a and Supplementary Data 5**). Genes up-regulated in S100P-SPP1+ cells were mainly enriched in regulation of coagulation and complement activation, which were involved in hepatocyte function (**Fig. 3b**). These cells presented high expression of hepatocyte-specific genes such as *SERPINE2*, *APOB*, and *CPB2*, further supporting their hepatocyte-like differentiation (**Supplementary Fig. 4b**). In contrast, genes upregulated in S100P+SPP1- cells were related to mucus secretion, protein localization to endoplasmic reticulum (ER), and epithelial structure maintenance. Remarkably, we found that *PSCA*, which encodes a tumor antigen and up-regulated in prostate¹⁸ and bladder¹⁹ cancers, was highly expressed in S100P+SPP1- iCCAs, making it a promising candidate for immunotherapy of iCCA^{phl} (**Supplementary Fig. 4c**).

We further applied SCENIC analysis to characterize transcription networks between S100P+SPP1- and S100P-SPP1+ cells²⁰. The results showed that transcription factors such as *ATF3*, *CREB5*, *MEIS2*, and *EGR1* were upregulated in S100P-SPP1+ cells, while S100P+SPP1- cells showed up-regulation of transcription factors like *CREB3L1*, *PPARG*, *CDX2*, and *HOXB7* (**Fig. 3c and Supplementary Fig. 4d**). Survival analysis from Jusakul et al.'s dataset showed that transcription factors that highly expressed in iCCA^{phl} (*PPARG*, *MECOM*, *HOXB7*, *IRF7*, *FOXA3*) and iCCA^{pps} (*ONECUT1*, *HNF1B*, *MEIS2*), were associated with worse and better prognosis, respectively¹⁶ (**Supplementary Fig. 4e**). Notably, SCENIC analysis revealed that *CREB3L1*, which is induced by ER stress and contributes to maximal induction of the

unfolded protein response²¹, was a potential transcription factor regulating *S100P*. Also, *CREB3L1* expression strongly and positively correlated with *S100P* expression ($r = 0.58$, $p < 2.2e-16$, **Fig. 3d**). To determine whether *S100P* is a direct target of *CREB3L1*, we performed dual-luciferase report assay and found that the *S100P* promoter activity was markedly increased in a dose-dependent manner after overexpression of *CREB3L1* (**Fig. 3e**). Transwell assays showed that *CREB3L1* knockdown significantly weakened the invasion capacity of HuCCT1 and RBE cells (**Fig. 3f, g**). RNA-seq analysis showed that *CREB3L1* not only modulated the expression of *S100P*, but also affected the expression of various upregulated genes in S100P+SPP1⁻ cells, such as *OASL*, *RCN3*, and *OAS1* (**Fig. 3h**). Pathway analysis indicated that *CREB3L1* was involved in co-translational protein targeting to membrane, establishment of protein localization to ER, and actin filament reorganization (**Fig. 3i**). Together, these results reveal the distinct transcriptional profiles of S100P+SPP1⁻ and S100P-SPP1⁺ cells, identifying CREB3L1 as a potential transcriptor of S100P that promotes invasion of iCCA^{phl}.

Different polarization of infiltrated macrophages in iCCA^{phl} and iCCA^{pps}

Despite studies have profiled the tumor immune microenvironment of iCCA by scRNA-seq^{11,12}, the difference of immune landscape between iCCA^{phl} and iCCA^{pps} remains unclear. First, we evaluated the infiltration of T cells, B cells, NK cells, and macrophages in 186 iCCAs from the TMA cohort by immunostaining. Results showed that more CD3⁺ T cells ($P < 0.01$) and CD56⁺ NK cells ($P < 0.01$) were infiltrated in S100P-SPP1⁺ iCCA^{pps} (118 patients) compared to S100P+SPP1⁻ iCCA^{phl} (68 patients), indicating a more robust antitumor microenvironment in iCCA^{pps} (**Supplementary Fig. 5a**). Although there was no significant difference in CD68⁺ macrophages and CD20⁺ B cells, more CD68⁺CD206⁺ macrophages were found to be infiltrated in S100P+SPP1⁻ iCCA^{phl} ($P < 0.01$) (**Supplementary Fig. 5b, c**). Then, we focused on macrophages to evaluate distinct macrophage subsets infiltrated in the two subtypes of iCCAs.

A total of six clusters present in the myeloid lineage with the expression of specific marker genes, including one monocyte (Mono_FCN1), two macrophages (Macro_c1_SPP1 and Macro_c2_CCL18), and three DCs (DC_c1_CD1C, DC_c2_XCR1, and DC_c3_CD1A) (**Fig. 4a, b and Supplementary Data 6**). Macrophages and CD1a⁺ DCs (DC_c3_CD1A) were significantly enriched in tumors compared with paired non-tumor tissues, while monocytes, CD1c⁺ DCs (DC_c1_CD1C), and cDC1 DCs (DC_c2_XCR1) showed the opposite trend (**Supplementary Fig. 5d**). Indeed, we observed that, SPP1⁺ macrophages, which have been reported in colon cancer and closely interact with cancer-associated fibroblasts (CAFs)²², were more infiltrated in S100P-SPP1⁺ iCCA^{pps}, while CCL18⁺ macrophages, which were abundant in advanced hepatocellular carcinoma²³, were mostly infiltrated in S100P+SPP1⁻ iCCA^{phl} (**Fig. 4c, d**). Though both macrophages subsets have been defined as tumor associated macrophages, they varied in signaling pathways and metabolic features²⁴ (**Supplementary Fig. 6a, b**). Consistently, we found that SPP1⁺ macrophages showed increased level of oxidative phosphorylation and glycine, serine, threonine and tyrosine metabolism, while CCL18⁺ macrophages had elevated cytokine-cytokine receptor interaction,

nitrogen and riboflavin metabolism (**Supplementary Fig. 6c**). By calculating pro-/anti-inflammatory and M1/M2 polarization scores²⁵, we found that SPP1⁺ macrophages were more potent in both pro- and anti-inflammatory responses and skewed toward M1 polarization (**Fig. 4e, f**). In contrast, CCL18⁺ macrophages showed a dominant M2-like phenotype with the high expression of *CD163*, *MARCO*, and *CSF1R*, suggesting their stronger tumor-promoting role than SPP1⁺ macrophages (**Supplementary Fig. 6d**). Immunostaining on TMA cohort further confirmed that SPP1⁺CCL18⁻ macrophages were more abundant in S100P-SPP1⁺ iCCA^{PPS}, while SPP1⁻CCL18⁺ macrophages were mostly enriched in S100P+SPP1⁻ iCCA^{PHI} (**Fig. 4g, h**), which were again validated by the results from Jusakul et al.'s data¹⁶ (**Supplementary Fig. 6e**). Together, these results indicate that iCCA^{PHI} has a unique immune ecosystem, with increased CCL18⁺ macrophages, reduced CD3⁺ T and CD56⁺ NK cells as compared with iCCA^{PPS}.

iCCA^{PPS} contains tumor cells at different status of differentiation

The expression of ALB is generally considered as a marker of hepatocytes. Several studies have demonstrated the expression of ALB in iCCA, but the features of these ALB⁺ tumor cells are still unclear^{8, 26, 27}. Here, we detected a group of ALB expressing tumor cells at the single cell level, most of which (79.4%) were present in the S100P-SPP1⁺ iCCA^{PPS} (**Supplementary Fig. 7a-c**). Due to the different origins of iCCA^{PHI} and iCCA^{PPS}, we here only focused on these seven S100P-SPP1⁺ iCCA^{PPS} to explore their heterogeneity. By comparing the gene expression profiles of ALB⁺ and ALB⁻ cells, we found that ALB⁻ cells highly expressed *ID3*, which negatively regulates the basic helix-loop-helix and involved in cell differentiation, and neoplastic transformation²⁸ (**Fig. 5a, Supplementary Fig. 7d and Supplementary Data 7**). ALB⁺ cells highly expressed hepatocyte specific genes such as *CPB2*, *ASGR1*, *FGA*, as well as cholangiocyte markers *KRT19*, *KRT18*, and *EPCAM*, but did not express *AFP*, a marker of hepatic progenitor cells (**Fig. 5b and Supplementary Fig. 7e**). Genes that highly expressed in ALB⁺ cells were mainly involved in hepatocyte-specific processes, such as complement activation, detoxification, fatty acid catabolic process, and bile acid secretion, suggesting their hepatocyte differentiation (**Supplementary Fig. 7f**). SCENIC analysis showed that genes specifically upregulated in ALB⁺ cells were regulated by *NR5A2*, *BATF*, and *NFIA* (**Supplementary Fig. 7g**). In contrast, ID3⁺ cells highly expressed genes such as *MDK*, *ZEB1*, and *LGR5* that playing important roles in tumor stemness^{29, 30, 31, 32}. SCENIC analysis predicted that transcription factors *SOX11*, *PAX2*, *IRX2*, *IRX3*, *FOXC1*, and *EN2* were responsible for genes upregulated in these cells.

Previous studies have designated ID3⁺ cells as hepatoblasts which could give rise to both hepatocytes and cholangiocytes³³. To reveal the differentiation process³³ in iCCA, we explored the gene expression patterns along this transition by trajectory analysis. Tumor cells from P09 and P10 were selected for this analysis as they contained comparable number of ALB⁺ and ID3⁺ cells (**Fig. 5c and Supplementary Fig. 7h**). We found that ALB⁺ cells were mainly located at the terminal of this trajectory and genes involved in regulation of coagulation, ER lumen and response to ER stress were increased gradually along the trajectory (**Fig. 5d**). Also, the expression of *MKI67* showed the same trend as *ALB*, implying an

increased proliferation capacity of ALB+ cells. ID3+ cells located opposite to ALB+ cells in the trajectory and were enriched for pathways in collagen containing extracellular matrix and negative regulation of cell adhesion. For example, the expression of *COL12A1*, which encodes the alpha chain of type XII collagen and is overexpressed in several cancer types^{34,35}, decreased gradually along transition from ALB- cells to ALB+ cells (**Fig. 5e**).

By evaluating the expression of 16 identified marker genes of ID3+ and ALB+ cells in Jusakul et al.'s dataset¹⁶, we validated that S100P-SPP1+ patients can also be clearly divided into two subclasses with mutually exclusive expression of 16 genes (**Fig. 5f**). In addition to the exclusivity between ALB and ID3, a significantly negative correlation between ID3 and MKI67 expression was also observed, suggesting the slow proliferation of these tumors (**Fig. 5g**). Taken together, these results demonstrate that iCCA^{PPS} is a heterogeneous tumor with tumor cells at various status of differentiation such as hepatocyte differentiation or stemness.

ID3+ tumor cells indicate abundant stroma components and worse prognosis in iCCA^{PPS}.

We next explored the clinical and histological characteristics of ID3+ iCCA^{PPS}. By immunostaining, we find that ID3 was predominantly expressed in the nucleus of tumor cells that located in the tumor center and were surrounded by rich stromal components (**Fig. 6a**). To further explore the relationship between ID3 expression and tumor stroma, we analyzed the correlation between ID3 expression and CAFs in two public databases^{16,17}. Results showed that ID3 expression positively correlated with CAFs' gene signature, such as *PDGFRB*, *COL1A1*, and *PDPN* (**Fig. 6b and Supplementary Fig. 7i**).

Since CAFs play important roles in tumor progression and chemoresistance³⁶, we speculated that ID3 expression was related to iCCA prognosis. We selected 118 S100P-SPP1+ iCCA^{PPS} from our TMA cohort to explore the prognostic values of ID3+ tumor cells and PDGFR β + stromal cells (most of which were CAFs) (**Fig. 6c**). As expected, the proportion of CK19+ID3+ tumor cells positively correlated with the proportion of CK19-PDGFR β + cells ($r = 0.46$, $P < 0.001$), while the proportion of CK19+ID3- tumor cells negatively correlated with the proportion of CK19-PDGFR β + cells ($r = -0.46$, $P < 0.001$) (**Fig. 6d**). Survival curves indicated that the proportion of CK19+ID3+ tumor cells ($P = 0.016$) and CK19-PDGFR β + cells ($P = 0.005$) both significantly correlated with poor prognosis in iCCA^{PPS} (**Fig. 6e**). Thus, these results demonstrate that ID3+ cells commonly correlated with the presence of CAFs and patient survival in iCCA^{PPS}.

Discussion

iCCAs can be divided into different subgroups according to the tumor anatomical location and the origin of tumor cells. In this study, we generated scRNA-seq profiles of 14 primary iCCAs and identified SPP1 as a representative marker for iCCA^{PPS}. We found that 92.5% iCCAs can be classified as iCCA^{phl} and iCCA^{PPS} according to the expression of S100P and SPP1, and there are significant differences in clinicopathological characteristics, gene regulatory networks and immune infiltration between these two

iCCA subtypes. Moreover, we confirmed the presence of tumor cells at various differentiation in iCCA^{PPS} at the single-cell level. This study broadens our understanding of the origin of iCCA and provides valuable markers for patient classification (**Fig. 7**), which may stem for novel diagnostic and therapeutic strategies for iCCA.

Cholangiocarcinoma can be divided into iCCA or ECC according to the tumor location in the biliary tree. It has been found that a part of mucin-producing cholangiocarcinoma presented in iCCA has high invasiveness and high expression of S100P, which is very similar to ECC⁸. We here identified S100P+SPP1- cells highly expressed mucus related genes such as *MUC5AC*, *MUC1*, and *MUC6* at single cell level. Despite that this subtype is often detected in smaller tumor size, it had more lymph node metastases, and high levels of CA19-9, Ki67, and CEA. Mucins synthesis begin in the ER and they are extremely susceptible to misfolding due to their large sizes and structure complexity, which can eventually lead to ER stress³⁷. We indeed observed many genes associated with mucins synthesis or ER stress upregulated in iCCA^{phl}, such as *XBP1*³⁸, *AGR2*³⁹, and *CREB3L1*²¹, which may be involved in the progression of this subtype of iCCA. S100P+SPP1- iCCA^{phl} had less CD3⁺ T and CD56⁺ NK cells, but more CCL18⁺ macrophage infiltration than S100P-SPP1+ iCCA^{PPS}, indicating its dampened anti-tumor immune response that may contribute to the higher invasive potential.

SPP1 is considered to play a cancer-promoting role and is often associated with a worse prognosis in various tumors, but its prognostic significance in iCCA is still controversial^{40, 41}. One important reason for this inconsistency is that the classification of iCCA is not properly considered. iCCA^{PPS} is believed originating from mucin-negative cuboidal cholangiocytes or ductules containing hepatic progenitor cells. It has been reported that *CDH2* and *NCAM1*, are representative markers of these iCCAs^{3, 5}. Based on our results, the expression of *SPP1* is mutually exclusive with *S100P*, showing a better specificity and sensitivity than *CDH2* or *NCAM1* as a marker of iCCA^{PPS}. S100P-SPP1+ iCCA^{PPS} had less lymph node metastasis, larger tumor volume, and better prognosis than S100P+SPP1- iCCA^{phl}. One research has reported that iCCA with cholangiolocellular differentiation highly expressed *CRP* and *CDH2*, while iCCA without cholangiolocellular differentiation highly expressed *TFF1* and *S100P*. The two groups of iCCAs showed significant differences in clinicopathological characteristics and patient outcomes⁹. The results from this study are very similar to the findings of our study. S100P-SPP1+ iCCA^{PPS} showed high expression of *CRP* and *CDH2*, which correspond to the iCCAs with cholangiolocellular differentiation. Studies have revealed that iCCA^{PPS} often occurs in the background of chronic hepatitis or liver cirrhosis⁵. We observed SPP1⁺ macrophages, which has been reported involving in liver inflammation and fibrosis⁴², were highly infiltrated in iCCA^{PPS}, indicating that these macrophages may be involved in the occurrence and development of iCCA^{PPS}.

Heterogeneity in tumor cell differentiation was observed in iCCA because of the complicated cell origin and formation. In the present study, two major subsets of tumor cells, ALB+ and ID3+ tumor cells, were identified in iCCA^{PPS}. The expression of ALB mRNA has been detected by in situ hybridization in about

40% of all iCCAs⁴³, but the specific biology of these ALB+ cells is still not clear. The results of our study showed that these ALB+ cells have the characteristics of hepatocyte differentiation. However, these cells also expressed *EPCAM* and *KRT19*, indicating that these may be hepatocyte-like cells in the early stage of differentiation rather than mature hepatocytes. The stem-like ID3+ cells coexisting in iCCA^{PPS} may be the precursor cells of ALB+ cells. There are several reasons for this conjecture. First, these ID3+ cells highly expressed many stemness-related genes, such as *ID4*, *MDK*, *ZEB1*, and *LGR5*. Of note, it has been reported that the expression of *ID3* and *LGR5* could promote stem cell features in iCCA^{29, 32}. Second, previous study has identified ID3+ cells at the early stages of development in human and mouse fetal livers, which are able to differentiate into both hepatocytes or cholangiocytes³³. Therefore, the presence of ID3+ cells may be one of the reasons for the diversity of iCCA^{PPS}. Additionally, we found that ID3+ cells were generally located in the interior area of the iCCA^{PPS} and positively correlated with the CAF content. The location of these ID3+ cells and the presence of the large amount of CAFs surrounding them may be an important reason for the poor prognosis of this type of iCCA.

A few limitations of the current study should not be ignored. There were 1.49% S100P+SPP1+ and 5.97% S100P-SPP1- iCCAs in our validation cohort. We did not analyze the clinicopathological features of these iCCAs because of their small number. Also, due to the small number of S100P+SPP1+ and S100P-SPP1- cells in scRNA-seq data, we could not evaluate the molecular characteristics of these two types of tumor cells at the single cell level accurately. Therefore, future studies with a larger sample size containing these two iCCA subtypes may help to resolve this issue.

In summary, our study provides two optimal markers to differentiate iCCA from different origins and revealed the clinical, molecular, biological and immunological diversity between iCCA^{phl} and iCCA^{PPS}. The two markers, S100P and SPP1, could serve as the easy-to-use reference to guide iCCA patient stratification and personalized management, and lay the foundation for basic and translational research of iCCAs from different origins.

Methods

Patient samples

Fourteen patients had liver resection and were pathologically diagnosed as iCCA from January 2019 to January 2020 were enrolled for scRNA-seq. None of the patients received chemotherapy, radiotherapy or any other anti-tumor therapy before surgery. Fresh paired tumor and non-tumor liver tissues were obtained during surgical resection. The adjacent normal tissues were at least 3 cm away from the matched tumor tissue. This study was conducted in accordance with the ethical standards of the Research Ethics Committee of Zhongshan Hospital with patients' informed consent.

Tissue microarray and immunohistochemistry

Tissue microarray was constructed as described⁴⁴. Paraffin-embedded tissue samples from 201 iCCA patients who underwent primary and curative resection for their tumor in Liver Cancer Institute, Zhongshan Hospital of Fudan University (Shanghai, China) between 2012 and 2015 were selected. The tissue microarrays were baked at 60 °C for 1 h, dewaxed in xylene, rehydrated through a gradient concentration and blocked the endogenous peroxidase activity by 3% hydrogen peroxide. The sections were incubated with 10% goat serum for 30 min to block non-specific binding sites and then incubated with the primary antibodies including S100P (1:1500 dilution), SPP1 (1:2000 dilution) at 4°C overnight. Detailed information of antibodies was provided in the Supplementary Data S8. After repeated washing, the sections were incubated at room temperature with secondary antibody and visualized by DAB solution and counterstained with haematoxylin. IHC staining score was assessed by two independent pathologists who were blinded to the patients' clinicopathological data. The score for IHC intensity was scaled as 0 for no IHC signal, 1 for weak, 2 for moderate, and 3 for strong. A positive IHC stain was defined by a visible staining pattern (score 1 to 3) compared to the negative control (score 0).

Preparation of single-cell suspensions

Fresh iCCA tumor tissues and adjacent non-tumor liver tissues were obtained immediately following tumor resection and transferred to the 50 mL centrifugal tube filled with RPMI-1640 medium (Gibco) with 10% fetal bovine serum (Gibco) and transported rapidly to the laboratory on ice. Specimens were then washed twice with cold 1× PBS (Gibco) and digested with Miltenyi Tumor Dissociation Kit and the GentleMACS (Miltenyi, Bergisch Gladbach, Germany) following manufacturer's instructions. The dissociated cells were subsequently passed through a 70 µm cell-strainer (BD) to remove clumps and undigested tissue. After centrifugation, the cell pellet was washed twice with MACS buffer (PBS containing 1% FBS, 0.5% EDTA, and 0.05% gentamycin) and then re-suspended in sorting buffer (PBS supplemented with 1% FBS). Single-cell suspensions were stained with DRAQ5 (1:200, 10 min) and DAPI (1:200, 5 min). Finally, DRAQ5+DAPI- cells were sorted into RPMI-1640 media supplemented with 10% FBS by FACSAria (BD Biosciences).

Single-cell RNA-sequencing

Libraries for scRNA-seq were generated using the Chromium Single Cell 3' library and Gel Bead & Multiplex Kit from 10x Genomics. 10xGenomics Chromium barcoding system was used to construct 10xbarcoded cDNA library following the manufacturer's instructions. All libraries were sequenced on Illumina HiSeq4000 until sufficient saturation was reached.

scRNA-seq data processing

Cell Ranger (v3.1.0) was applied for read mapping and gene expression quantification. Cells with less than 1000 UMIs or >20% mitochondria genes were excluded. We also used 3 algorithms (DoubletFinder, DoubletDetection, Scrublet)^{45, 46, 47} to find doublets and remove cells which were identified as a doublet by at least one algorithm. The total number of transcripts in each cell was normalized to 10,000, followed by

log transformation. Then we used Seurat (v3)⁴⁸ to detect highly variable genes, perform PCA, graph-based clustering and t-SNE.

Classification of malignant cells

As malignant cells harbor significantly more copy number variation (CNV) than normal cells, we first estimated CNV from scRNA-seq following the steps described in previous study⁴⁹. In brief, genes were first sorted according to their genomic location at each chromosome, and then a sliding window of 100 genes was applied to calculate the average relative expression values to derive CNVi (CNV of the *i*th window). We defined CNV score of each cell as the mean of squared CNVi across all windows. In addition, we calculated CNV correlation score by computing the Spearman correlation of the CNVi of a cell and the average CNVi of the single-cells with top 3% CNV scores from the same tumor. Malignant cells were then defined as those with CNV signal above 0.04 and CNV correlation above 0.5. Moreover, we determined the epithelial score of each cell using 14 epithelial markers from Puram et al.⁵⁰ and found the epithelial score of malignant cells is significantly higher than that of non-malignant cells, which proves the robustness of our classification.

Classification of non-malignant cells

For all non-malignant cells, we first used SingleR⁵¹ to classify cells into 7 major cell types: myeloid cell, NK cell, CD8⁺ T cell, CD4⁺ T cell, B cell, endothelial cell and fibroblast. Other cell types (e.g. hepatocyte, neutrophil and mast cell) with fewer than 100 cells are excluded. Then we applied the graph-based clustering method implemented in Seurat to group cells into subtypes and each subtype was further annotated according to its marker genes.

Bulk whole exome sequencing and data processing

Whole-exome sequencing was performed as described previously⁴⁴. In brief, DNA was extracted from iCCA tumor and non-tumor liver tissues from these fourteen patients using a DNeasy Blood and Tissue kit (Qiagen), and DNA concentration and purity were determined using a NanoQuant Plate Infinite M200 PRO reader (Tecan Austria GmbH). After enrichment of exonic DNA fragments with a SureSelect Human All Exon Kit (Agilent, 50 Mb V5), sequencing was performed on Illumina HiSeq4000.

Raw sequencing reads were mapped to human genome version 38 (hg38) using BWA-MEM⁵². After removing duplicated reads, SNV and indel were detected using Mutect2 (<https://doi.org/10.1101/861054>) and annotated with Oncotator⁵³. Copy number alteration (CNA) was identified using FACETS⁵⁴.

Tumor heterogeneity analysis

For WES data, cancer cell fraction (CCF) and clonality of each mutation was determined following the process described in Nicholas et al.⁵⁵ Genomic heterogeneity was calculated as the proportion of

subclonal mutations in a tumor. For scRNA-seq data, we estimated transcriptomic heterogeneity according to the method in Ma et al.¹¹

Differential expression and pathway analysis

Differentially expressed genes (fold change > 4 and P value < 0.001) were identified using the QLF model implemented in edgeR (v3.26.3)⁵⁶. Pathway enrichment analysis was performed using clusterprofiler⁵⁷ based on GOBP gene sets from MSigDB.

Gene regulatory network inference

Gene regulatory networks were identified using SCENIC (v1.1.0)²⁰ with default settings. To reduce the computing time, a python implementation in SCENIC (GRNBoost) was used.

Developmental trajectory analysis

Monocle⁵⁸ was applied to infer the developmental trajectory with each tumor. Only top 1000 variable genes identified by differentialGeneTest were selected for constructing the developmental tree.

Public datasets analysis

Mutation, gene expression and methylation data from Jusaka et, al.¹⁶ were retrieved from. Gene expression data from Sia et al.¹⁷ was retrieved from.

Dual luciferase assay

The dual reporter plasmid expressing firefly luciferase under the human S100P promoter and Renilla luciferase under the SV40 promoter was constructed. Different concentrations of expression plasmids were transiently transfected into the HEK-293T cells with Renilla luciferase plasmid. Firefly luciferase activity was measured with a Dual Luciferase Assay Kit (Promega) 24 hours after transfection and normalized with a Renilla luciferase reference plasmid. Results are assessed as the ratio of Firefly luciferase activity to Renilla luciferase activity.

RNAi and transfection

Human CREB3L1 siRNA (si-CREB3L1) lentivirus vectors and non-specific siRNA (si-Ctrl) lentivirus vectors were synthesized by GeneChem Technology (Shanghai, China). The si-CREB3L1 sequences are at nucleotide positions 131 – 149 as reported previously⁵⁹. Non-targeting siRNA was used as the negative control. Lentivirus transfection was performed following the manufacturer's instructions and the efficiency of silencing was confirmed by immunoblotting.

Transwell invasion assay

Cell invasion was determined by Transwell invasion assay. Briefly, transwell inserts were firstly coated with Matrigel (BD, USA). Then, 1×10^5 cells suspended in 0.2 mL serum-free medium were added into inserts and 0.5 mL medium containing 20% FBS was added to the lower compartment as a chemoattractant. After culturing for 48 h, the cells on the upper membrane were carefully removed using a cotton bud and cells on the lower surface were fixed with methanol for 15 min and successively stained with 0.1% crystal violet solution for 10 min. Photographs were then taken and the number of cells that passed through the Matrigel were counted. Assays were performed in duplicate in three independent experiments.

Multiplex immunohistochemistry and quantitative analysis

In brief, 4 μ m FFPE TMAs sections were deparaffinized in xylene and then rehydrated in 100%, 90%, 70% alcohol successively. Antigen unmasking was performed with a preheated epitope retrieval solution, endogenous peroxidase was inactivated by incubation in 3% H₂O₂ for 20 min. Next, the sections were pre-incubated with 10% normal goat serum and then incubated overnight with primary antibodies (panel 1: CK19, S100P, SPP1, CD68, CCL18; panel 2: CK19, ID3, PDGFR β ; panel 3: EPCAM, S100P, PSCA. Detailed information of antibodies was provided in the Supplementary Data S7). Next, sections were incubated with the corresponding HRP-conjugated second antibodies (Vector Lab, CA) for 30 min at room temperature. The antigenic binding sites were visualized using the OPAL dye. Opal -520 (PerkinElmer Inc.), Opal- 570 (PerkinElmer Inc.), Opal -620 (PerkinElmer Inc.), Opal -650 (PerkinElmer Inc.), Opal -690 (PerkinElmer Inc.) were applied to each antibody, respectively.

Data were analyzed as previously described⁶⁰. Images were analyzed and quantified by inForm software (v2.3, PerkinElmer Inc.) based active machine learning algorithm with a pre-visual cutoff followed by single-cell based mean pixel fluorescence intensity to achieve accuracy. Threshold value of each marker was identified and displayed by both FCS Express 6 Plus v6.04.0034 (De Novo Software) with FACS alike density plot and Inform Score that could adjust the cutoff based on the score map and original staining images to improve the accuracy.

Statistical analysis

Statistical analysis was performed with the R (v3.6.1), SPSS (v22, IBM, Armonk, NY) and Prism 6.0 (SanDiego, CA) softwares. Comparisons were performed using χ^2 test and unpaired two-sided Wilcoxon rank-sum test unless specified. The cumulative survival time was estimated by Kaplan–Meier estimator with log-rank test.

Declarations

Data availability

The raw sequence data reported in this paper has been deposited in the Genome Sequence Archive in National Genomics Data Center under the accession number HRA000863, which is publicly accessible at

Acknowledgements

We thank the High Performance Computing Platform of Center for Life Sciences (Peking University) for providing the data analysis platform. We thank the Institut Pasteur of Shanghai for providing instruments. The study was supported by project grants from the National Natural Science Foundation of China (Nos. 81961128025, 91859105, 11971039 and 31800743), Programs of the Science and Technology Commission of Shanghai (No. 20JC1418900 and 19XD1420700), the Strategic Priority Research Program (No. XDPB0303), Frontier Science Key Research Project (No. QYZDB-SSW-SMC036), Chinese Academy of Sciences, Sanming Project of Medicine in Shenzhen (No. SZSM202003009), and National Key Basic Research Project of China (No. 2020YFE0204000).

Author contributions

Q.G., R.X., and X.Z. contributed to study design and supervised the study. G.S. and Q.G. contributed to writing the manuscript. Y.S., S.H., Y.W., and J.L. assisted in the data analysis. G.S. and J.M. performed immunohistochemical staining and image analysis. J.Z., Y.L., J.L., and S.J. assisted in the preparation of the experiments. S.Y., D.R., Y.C. and Y.L. aided in collection of tissue samples. J.S. and X.W. assisted with data collection. S.Z., A.K., X.W., and Y.J. assisted in histopathological analysis. Y.C., J.Z., and J.F. made intellectual contributions.

Competing interests

The authors declare no competing interests.

References

1. Valle JW, Lamarca A, Goyal L, Barriuso J, Zhu AX. New Horizons for Precision Medicine in Biliary Tract Cancers. *Cancer Discov* **7**, 943-962 (2017).
2. Nakanuma Y, Klimstra D, Komuta M, Zen Y. Intrahepatic cholangiocarcinoma. In: WHO Classification of Tumors Editorial Board. *WHO classification of tumours: digestive system tumours* 5th ed: World Health Organization; 2019. p, 254-259 (2019).
3. Banales JM, *et al.* Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol* **17**, 557-588 (2020).
4. Kendall T, *et al.* Anatomical, histomorphological and molecular classification of cholangiocarcinoma. *Liver Int* **39 Suppl 1**, 7-18 (2019).
5. Aishima S, Oda Y. Pathogenesis and classification of intrahepatic cholangiocarcinoma: different characters of perihilar large duct type versus peripheral small duct type. *J Hepatobiliary Pancreat Sci* **22**, 94-100 (2015).

6. Aishima S, *et al.* Gastric mucin phenotype defines tumour progression and prognosis of intrahepatic cholangiocarcinoma: gastric foveolar type is associated with aggressive tumour behaviour. *Histopathology* **49**, 35-44 (2006).
7. Prica F, Radon T, Cheng Y, Crnogorac-Jurcevic T. The life and works of S100P - from conception to cancer. *Am J Cancer Res* **6**, 562-576 (2016).
8. Komuta M, *et al.* Histological diversity in cholangiocellular carcinoma reflects the different cholangiocyte phenotypes. *Hepatology* **55**, 1876-1888 (2012).
9. Rhee H, *et al.* Transcriptomic and histopathological analysis of cholangiolocellular differentiation trait in intrahepatic cholangiocarcinoma. *Liver Int* **38**, 113-124 (2018).
10. Liao JY, Tsai JH, Yuan RH, Chang CN, Lee HJ, Jeng YM. Morphological subclassification of intrahepatic cholangiocarcinoma: etiological, clinicopathological, and molecular features. *Mod Pathol* **27**, 1163-1173 (2014).
11. Ma L, *et al.* Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. *Cancer Cell* **36**, 418-430 e416 (2019).
12. Zhang M, *et al.* Single-cell transcriptomic architecture and intercellular crosstalk of human intrahepatic cholangiocarcinoma. *J Hepatol* **73**, 1118-1130 (2020).
13. Karthaus WR, *et al.* Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science* **368**, 497-505 (2020).
14. Chen Z, *et al.* Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. *Nat Commun* **11**, 5077 (2020).
15. Moorman HR, Poschel D, Klement JD, Lu C, Redd PS, Liu K. Osteopontin: A Key Regulator of Tumor Progression and Immunomodulation. *Cancers (Basel)* **12**, (2020).
16. Jusakul A, *et al.* Whole-Genome and Epigenomic Landscapes of Etiologically Distinct Subtypes of Cholangiocarcinoma. *Cancer Discov* **7**, 1116-1135 (2017).
17. Sia D, *et al.* Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. *Gastroenterology* **144**, 829-840 (2013).
18. Saeki N, Gu J, Yoshida T, Wu X. Prostate stem cell antigen: a Jekyll and Hyde molecule? *Clin Cancer Res* **16**, 3533-3538 (2010).
19. Fu YP, *et al.* Common genetic variants in the PSCA gene influence gene expression and bladder cancer risk. *Proc Natl Acad Sci U S A* **109**, 4974-4979 (2012).

20. Aibar S, *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017).
21. Sampieri L, Di Giusto P, Alvarez C. CREB3 Transcription Factors: ER-Golgi Stress Transducers as Hubs for Cellular Homeostasis. *Front Cell Dev Biol* **7**, 123 (2019).
22. Zhang L, *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell* **181**, 442-459 e429 (2020).
23. Song G, *et al.* Global immune characterization of HBV/HCV-related hepatocellular carcinoma identifies macrophage and T-cell subsets associated with disease progression. *Cell Discov* **6**, 90 (2020).
24. Xiao Z, Dai Z, Locasale JW. Metabolic landscape of the tumor microenvironment at single cell resolution. *Nat Commun* **10**, 3763 (2019).
25. Azizi E, *et al.* Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* **174**, 1293-1308 e1236 (2018).
26. Collins K, Newcomb PH, Cartun RW, Ligato S. Utility and Limitations of Albumin mRNA In Situ Hybridization Detection in the Diagnosis of Hepatobiliary Lesions and Metastatic Carcinoma to the Liver. *Appl Immunohistochem Mol Morphol* **29**, 180-187 (2021).
27. Lin F, *et al.* Detection of Albumin Expression by RNA In Situ Hybridization Is a Sensitive and Specific Method for Identification of Hepatocellular Carcinomas and Intrahepatic Cholangiocarcinomas. *Am J Clin Pathol* **150**, 58-64 (2018).
28. Wang LH, Baker NE. E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease. *Dev Cell* **35**, 269-280 (2015).
29. Huang L, *et al.* ID3 Promotes Stem Cell Features and Predicts Chemotherapeutic Response of Intrahepatic Cholangiocarcinoma. *Hepatology* **69**, 1995-2012 (2019).
30. Filippou PS, Karagiannis GS, Constantinidou A. Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. *Oncogene* **39**, 2040-2054 (2020).
31. Caramel J, Ligier M, Puisieux A. Pleiotropic Roles for ZEB1 in Cancer. *Cancer Res* **78**, 30-35 (2018).
32. Kawasaki K, Kuboki S, Furukawa K, Takayashiki T, Takano S, Ohtsuka M. LGR5 induces beta-catenin activation and augments tumour progression by activating STAT3 in human intrahepatic cholangiocarcinoma. *Liver Int* **41**, 865-881 (2021).
33. Wang X, *et al.* Comparative analysis of cell lineage differentiation during hepatogenesis in humans and mice at the single-cell transcriptome level. *Cell Res* **30**, 1109-1126 (2020).

34. Poschel A, *et al.* Identification of disease-promoting stromal components by comparative proteomic and transcriptomic profiling of canine mammary tumors using laser-capture microdissected FFPE tissue. *Neoplasia* **23**, 400-412 (2021).
35. van Huizen NA, Coebergh van den Braak RRJ, Doukas M, Dekker LJM, JNM IJ, Luider TM. Up-regulation of collagen proteins in colorectal liver metastasis compared with normal liver tissue. *J Biol Chem* **294**, 281-289 (2019).
36. Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. *Nat Rev Drug Discov* **18**, 99-115 (2019).
37. Bansil R, Turner BS. The biology of mucus: Composition, synthesis and organization. *Adv Drug Deliv Rev* **124**, 3-15 (2018).
38. Glimcher LH, Lee AH, Iwakoshi NN. XBP-1 and the unfolded protein response (UPR). *Nat Immunol* **21**, 963-965 (2020).
39. Schroeder BW, *et al.* AGR2 is induced in asthma and promotes allergen-induced mucin overproduction. *Am J Respir Cell Mol Biol* **47**, 178-185 (2012).
40. Zhou KQ, *et al.* Circulating osteopontin per tumor volume as a prognostic biomarker for resectable intrahepatic cholangiocarcinoma. *Hepatobiliary Surg Nutr* **8**, 582-596 (2019).
41. Zheng Y, *et al.* Osteopontin promotes metastasis of intrahepatic cholangiocarcinoma through recruiting MAPK1 and mediating Ser675 phosphorylation of beta-Catenin. *Cell Death Dis* **9**, 179 (2018).
42. Song Z, *et al.* Osteopontin Takes Center Stage in Chronic Liver Disease. *Hepatology* **73**, 1594-1608 (2021).
43. Avadhani V, Cohen C, Siddiqui MT, Krasinskas A. A Subset of Intrahepatic Cholangiocarcinomas Express Albumin RNA as Detected by In Situ Hybridization. *Appl Immunohistochem Mol Morphol* **29**, 175-179 (2021).
44. Gao Q, *et al.* Activating mutations in PTPN3 promote cholangiocarcinoma cell proliferation and migration and are associated with tumor recurrence in patients. *Gastroenterology* **146**, 1397-1407 (2014).
45. DePasquale EAK, *et al.* DoubletDecon: Deconvoluting Doublets from Single-Cell RNA-Sequencing Data. *Cell Rep* **29**, 1718-1727 e1718 (2019).
46. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324 (2019).
47. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289 (2019).

48. Stuart T, *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
49. Tirosh I, *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016).
50. Puram SV, *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624 e1624 (2017).
51. Aran D, *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019).
52. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
53. Ramos AH, *et al.* Oncotator: cancer variant annotation tool. *Hum Mutat* **36**, E2423-2429 (2015).
54. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* **44**, e131 (2016).
55. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* **7**, 283ra254 (2015).
56. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
57. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).
58. Qiu X, *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982 (2017).
59. Denard B, *et al.* The membrane-bound transcription factor CREB3L1 is activated in response to virus infection to inhibit proliferation of virus-infected cells. *Cell Host Microbe* **10**, 65-74 (2011).
60. Ma J, *et al.* PD1(Hi) CD8(+) T cells correlate with exhausted signature and poor clinical outcome in hepatocellular carcinoma. *J Immunother Cancer* **7**, 331 (2019).

Figures

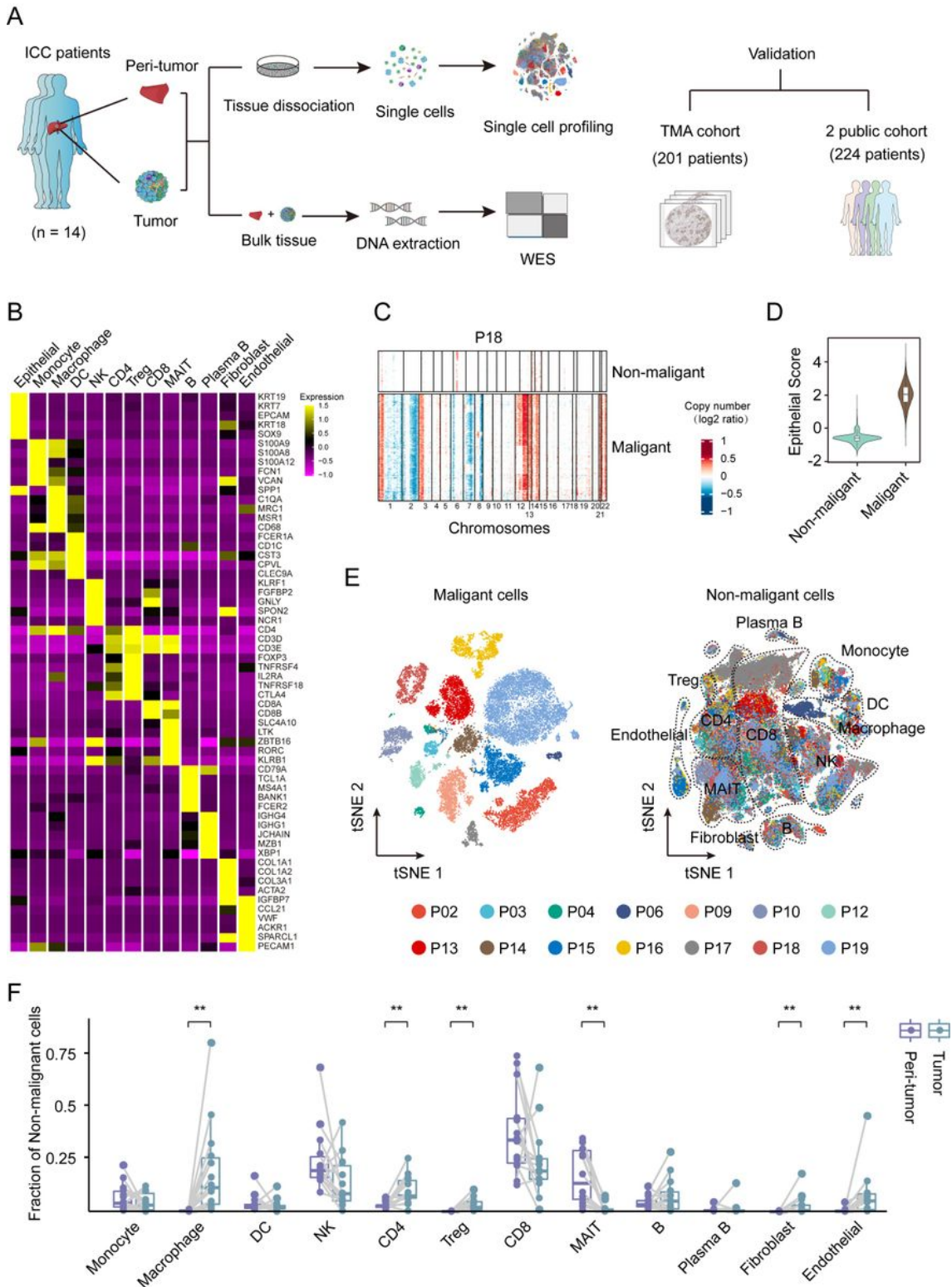


Figure 1

ScRNA-seq profiling of 14 ICCAs. a Schematic representation of the experimental strategy. b Heatmap showing the expression of marker genes in the indicated cell types. c Chromosomal landscape of inferred large-scale CNVs. The P18 tumor is shown with individual cells (rows) and chromosomal positions (columns). Amplifications (red) or deletions (blue) were inferred by averaging expression over 100-gene stretches on the respective chromosomes. d Epithelial score of malignant and non-malignant cells.

Boxplots represent medians and quartiles; violin plots represent overall distribution of the data. e t-SNE plot of malignant and nonmalignant cells from 14 iCCAs. f Boxplot showing the fraction of non-malignant cells in peri-tumor and tumor. **P < 0.01.

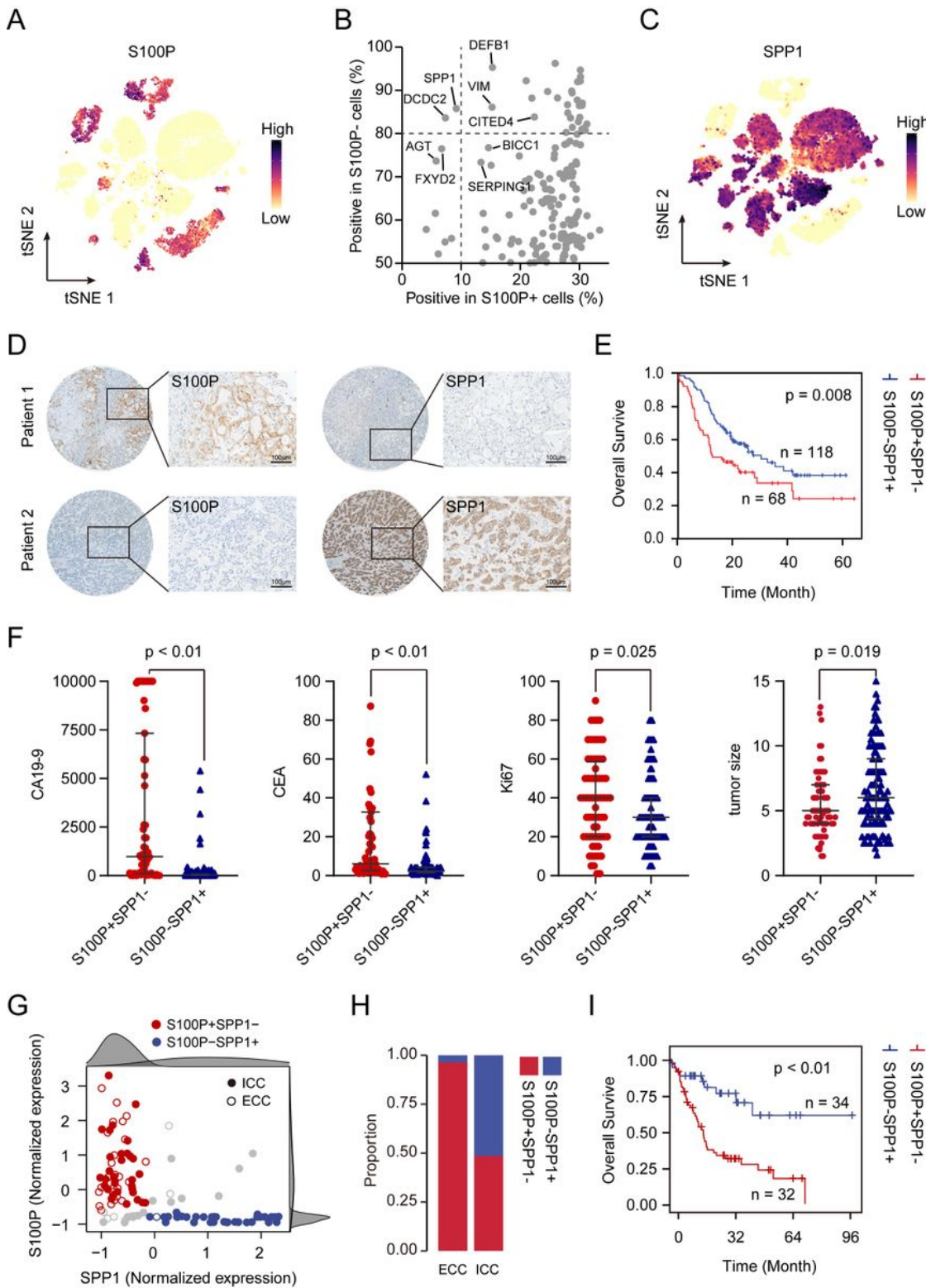


Figure 2

iCCA can be classified into two subtypes according to the expression of S100P and SPP1. a t-SNE plot showing expression level of S100P in malignant cells. b Proportion of positive cells with gene expression

in S100P+ (x-axis) and S100P- cells (y-axis). c t-SNE plot showing expression level of SPP1 in malignant cells. d Representative images of immunohistochemical expression of S100P and SPP1 in iCCAs from TMA cohort. Patient 1: S100P+SPP1-, Patient 2: S100P-SPP1+. Scale bar, 100 μ m. e Kaplan-Meier plot of the S100P+SPP1- and S100P-SPP1+ based on TMA data. f The scatter diagrams showing the differences in CA19-9, CEA, Ki67, and tumor size between the two groups (Wilcox test). g Scatterplot of S100P and SPP1 expression in Jusakul et al. dataset. A Gaussian mixture model with two mixture components was used to identify S100P+/- and SPP1+/- patients (right and top distribution curves). Solid circles represent iCCA and open circles represent ECC. Red represents S100P+SPP1- while blue represents S100P-SPP1+. h Graphical representation of the proportion of S100P+SPP1- and S100P-SPP1+ in iCCA and ECC. i Kaplan-Meier plot of the S100P+SPP1- and S100P-SPP1+ based on Jusakul et al. dataset.

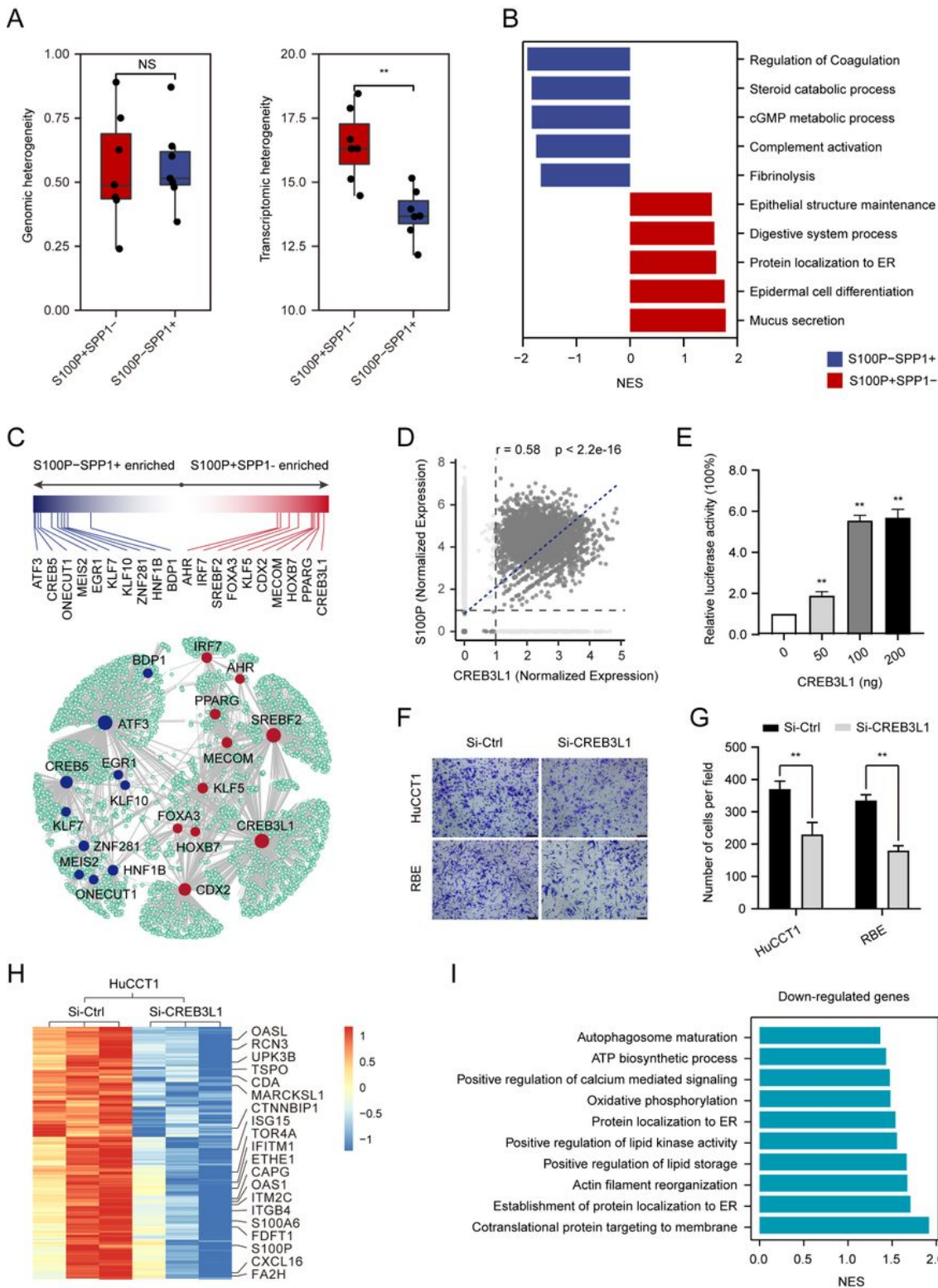


Figure 3

Different gene expression profiles between S100P+SPP1⁻ and S100P-SPP1⁺ cells. a Boxplot of the genomic heterogeneity (left) and transcriptomic heterogeneity (right) of S100P+SPP1⁻ and S100P-SPP1⁺ ICCAs (Wilcox test). b Top enriched pathways for genes with specific expression in S100P+SPP1⁻ and S100P-SPP1⁺ cells. c Network representation of selected differentially expressed transcription factors between S100P+SPP1⁻ and S100P-SPP1⁺ cells, as analyzed by SCENIC. Transcription factors in

S100P+SPP1- are shown in red; transcription factors in S100P-SPP1+ are shown in blue. Bar graph showing the difference score for the selected set of differentially expressed transcription factors in S100P+SPP1- (red) and S100P-SPP1+ (blue). d Scatter plot showing the correlation of CREB3L1 expression (x-axis) with S100P expression (y-axis). Correlation is evaluated by the Spearman correlation coefficient. e The relative luciferase activity in HEK-293T cells following co-transfection with plasmid containing S100P promoter and increasing doses of the CREB3L1 expression vector (Student's t test). f, g Representative images of the Transwell invasion assay f and a statistical histogram g (Student's t test). h Heatmap displaying expression levels of differentially expressed genes in Si-CREB3L1 versus Si-Ctl in HuCCT1 cells. i Top enriched pathways for down-regulated genes in Si-CREB3L1 HuCCT1 cells. **P < 0.01.

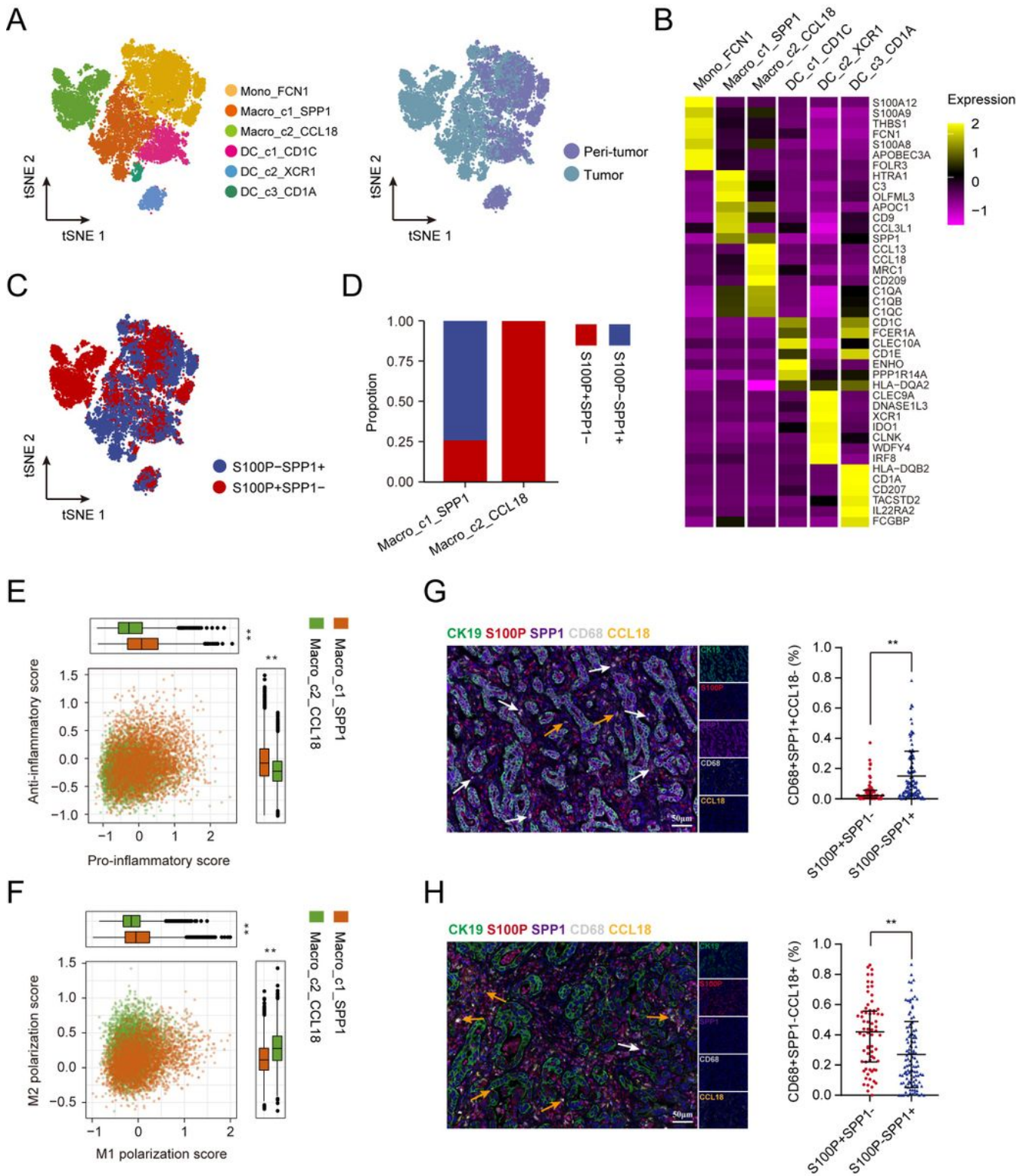


Figure 4

Two different subsets of macrophages infiltrated in iCCApHl and iCCAppS. a The t-SNE plot showing the subtypes of myeloid cells derived from iCCA peri-tumor and tumor. b Heatmap showing the expression of marker genes in each subtype of myeloid cells. c t-SNE plot of myeloid cells from S100P+SPP1- (red dots) and S100P-SPP1+ (blue dots). d Bar plot showing the proportion of macrophage subsets from S100P+SPP1- and S100P-SPP1+. e, f Scatterplots showing pro-/anti-inflammatory scores e and M1/M2

scores *f* for two macrophage subsets (Wilcoxon test). *g*, *h* Representative mIHC images (left) and statistical graphs (right) to show the distribution of CD68+SPP1+CCL18- and CD68+SPP1-CCL18+ macrophages in S100P+SPP1- *g* and S100P-SPP1+ *h*, respectively: CK19 (green), S100P (red), SPP1 (purple), CD68 (white), CCL18 (yellow), and DAPI (blue). White arrows (CD68+SPP1+CCL18-), yellow arrows (CD68+SPP1-CCL18+). Wilcox test, Scale bar, 50µm. ***P* < 0.01.

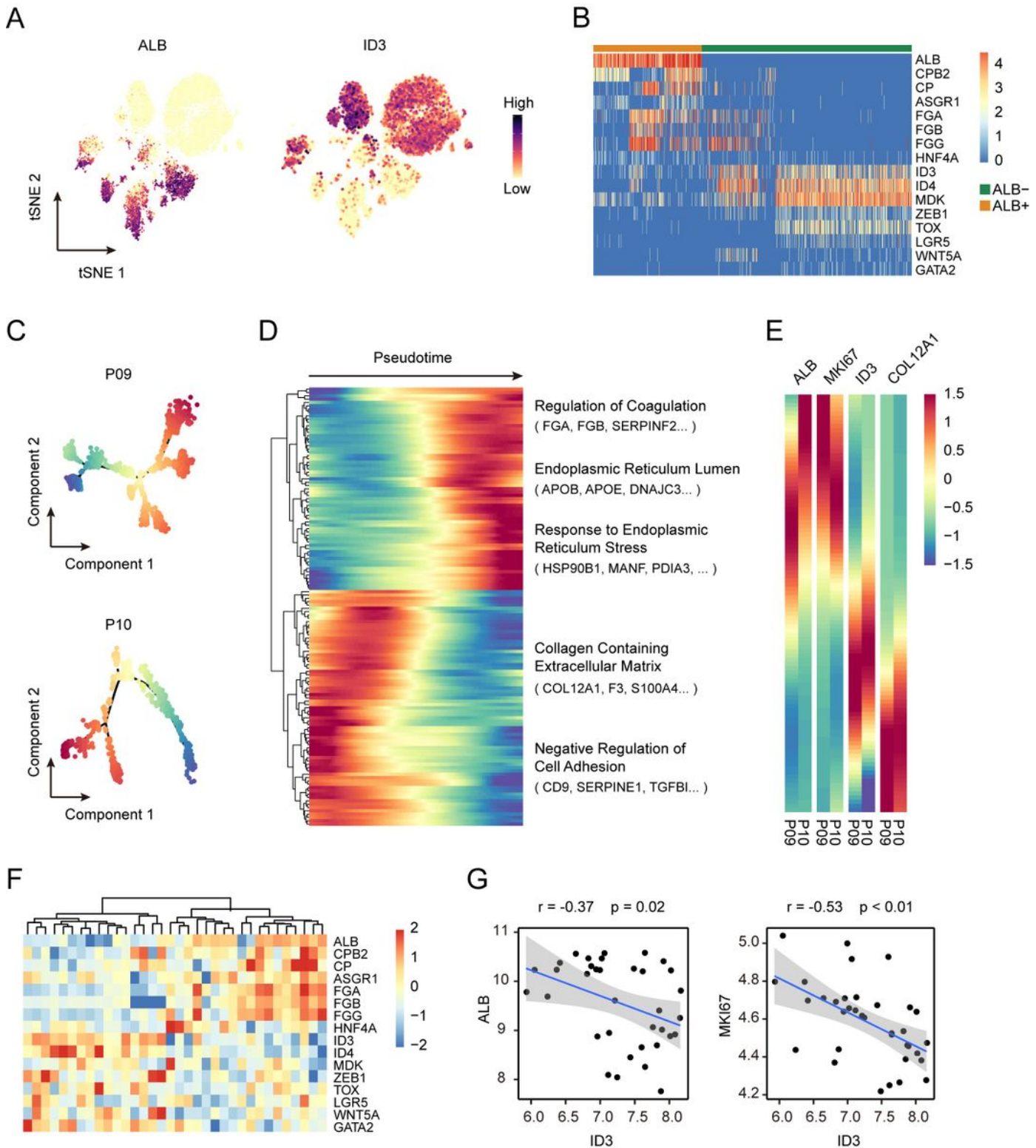


Figure 5

Tumor cells at different status of differentiation exist in S100P-SPP1+ iCCAs. a t-SNE plot showing expression levels of ALB and ID3 in 7 S100P-SPP1+ iCCAs. b Heatmap showing expression levels of differentially expressed genes (rows) between ALB+ and ALB- S100P-SPP1+ tumor cells (columns). c Trajectory of tumor cells from P09 and P10 separately in a two-dimensional state-space defined by Monocle. d Differentially expressed genes along the pseudo-time were clustered hierarchically into two profiles. The representative gene functions and pathways were shown. e Heatmap showing expression of representative genes. Color key from blue to red indicates relative expression levels from low to high. f Heatmap of ALB+ and ALB- specific genes (rows) and hierarchical clustering result in 34 S100P-SPP1+ iCCA (columns) from Jusakul et al. dataset. g Correlation between expression of ID3 and expression of ALB and MKI67. Blue line represents the linear regression curve. Correlation is evaluated by the Spearman correlation coefficient.

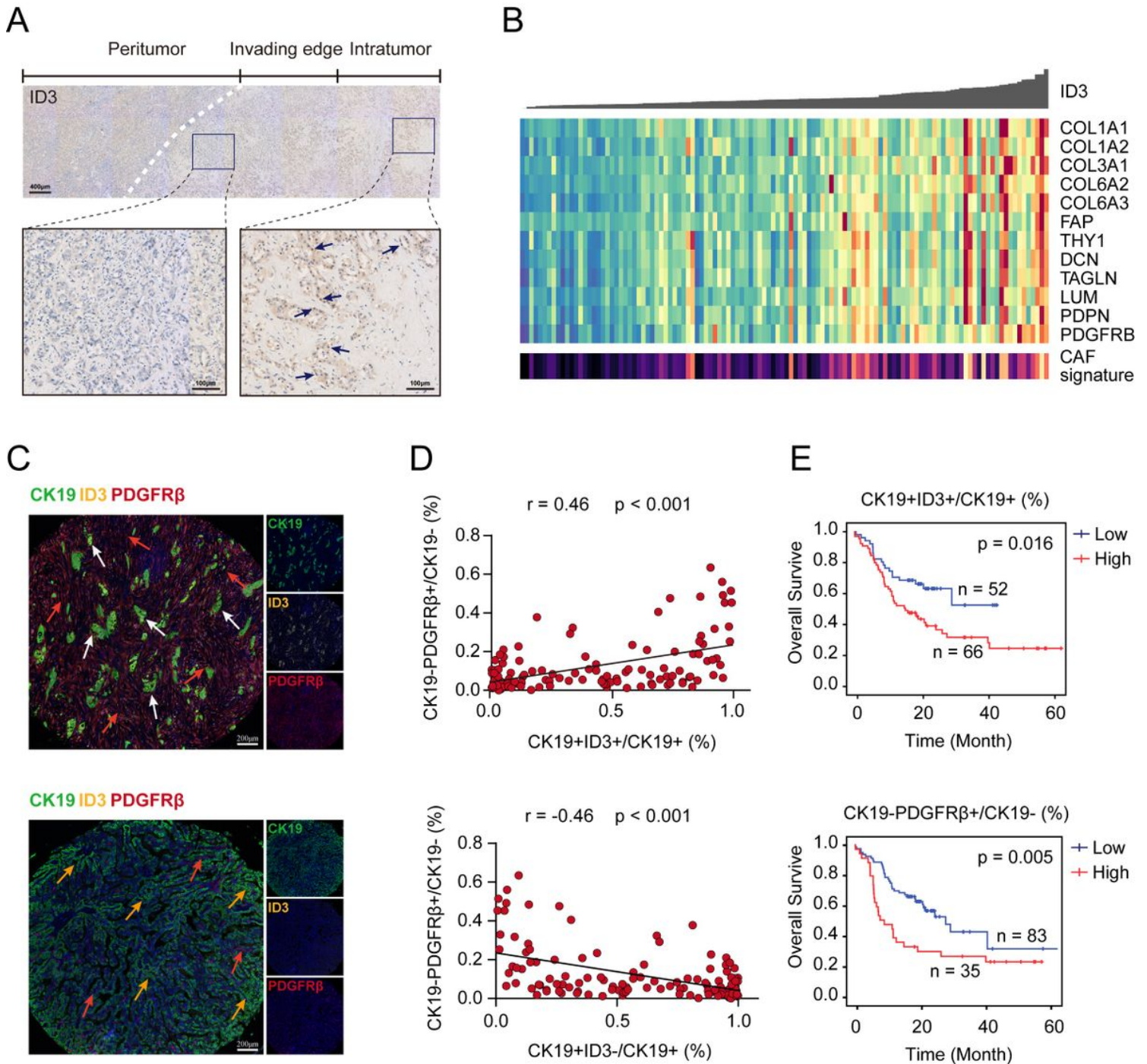


Figure 6

Prognostic significance of CK19+ID3+ tumor cells in S100P-SPP1+ ICCAs. a Representative immunostaining of ID3 in the indicated S100P-SPP1+ ICCAs. ID3+ tumor cells were predominantly located in the intratumor region. Scale bar, 400µm (up) and 100µm (down). b Correlation between ID3 expression and CAFs. iCCA from Jusakul et al. were ordered by their ID3 expression level as shown by bar-plot (top). Heatmap (middle) showing expression levels of selected CAF markers (rows) each tumor (columns). Colored bar (bottom) showing the CAFs score estimated by MCP-Counter of each tumor. c Representative mIHC images showing the distribution of CK19+ID3+, CK19+ID3- tumor cells and CK19-PDGFRβ+ cells in S100P-SPP1+ iCCA from TMA cohort: CK19 (green), ID3 (yellow), PDGFRβ (red), and

DAPI (blue). White arrows (CK19+ID3+), yellow arrows (CK19+ID3-). red arrows (CK19-PDGFR β +) Scale bar, 200 μ m. d Correlation analysis between the proportion of CK19+ID3+ (up) and CK19+ID3- (down) within CK19+ tumor cells and the proportion of CK19-PDGFR β + cells within CK19- cells per core, respectively. (Spearman correlation coefficient). e Kaplan-Meier analysis of overall survival (OS) in S100P-SPP1+ iCCA tumors according to the proportion of CK19+ID3+ within CK19+ tumor cells (up) and CK19-PDGFR β + within CK19- cells (down) in the TMA cohort.

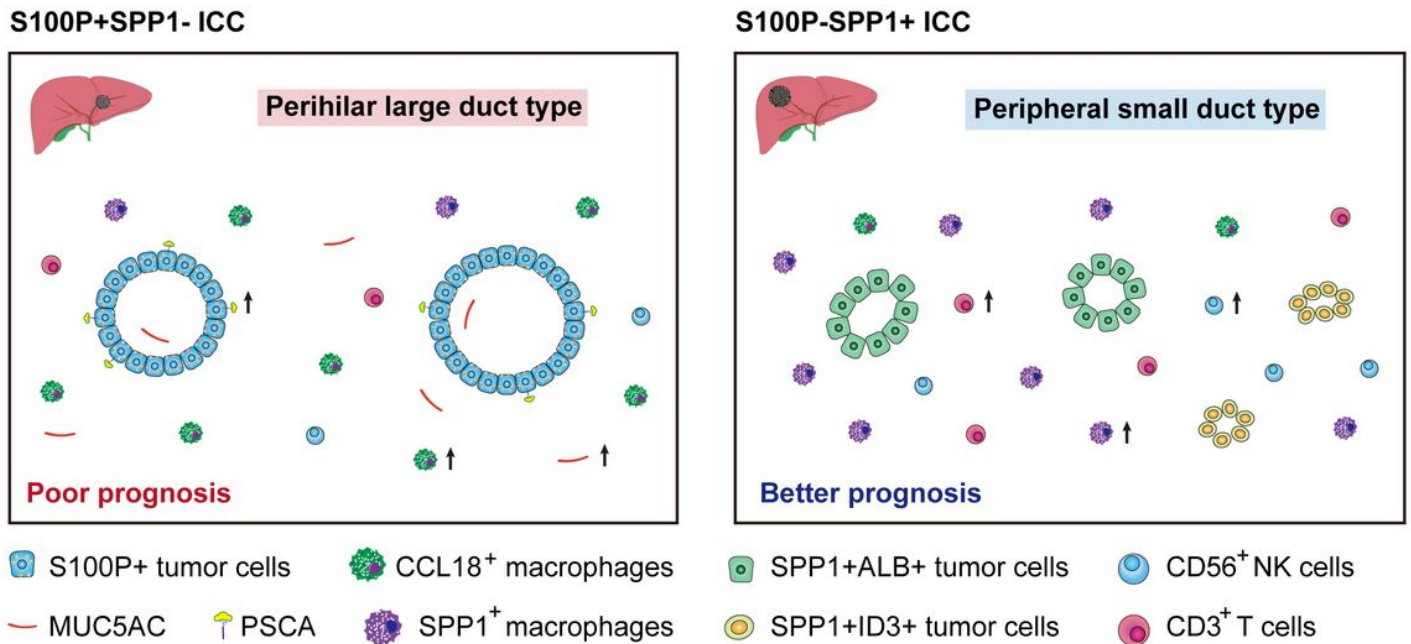


Figure 7

Schematics for the classification of iCCA. Two major subtypes of iCCA were identified in this study. Morphological features, cellular component, immune infiltration and prognosis varied significantly between these two iCCA subtypes.

Supplementary Files

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

- [SupplementaryData1.xlsx](#)
- [SupplementaryData2.xlsx](#)
- [SupplementaryData3.xlsx](#)
- [SupplementaryData4.xlsx](#)
- [SupplementaryData5.xlsx](#)
- [SupplementaryData6.xlsx](#)
- [SupplementaryData7.xlsx](#)

- [SupplementaryData8.xlsx](#)