ARTICLE

Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma

Junya Peng[1], Bao-Fa Sun[2,3,4], Chuan-Yuan Chen[2,3], Jia-Yi Zhou[2,3], Yu-Sheng Chen[2,3], Hao Chen[5], Lulu Liu[1], Dan Huang[1], Jialin Jiang[5], Guan-Shen Cui[2,3], Ying Yang[2,3,4], Wenze Wang[6], Dan Guo[1,7], Menghua Dai[5], Junchao Guo[5], Taiping Zhang[5], Quan Liao[5], Yi Liu[8], Yong-Liang Zhao[2,3,4], Da-Li Han[2,3,4], Yupei Zhao[5,8], Yun-Gui Yang[2,3,4] and Wenming Wu[5]

Pancreatic ductal adenocarcinoma (PDAC) is the most common type of pancreatic cancer featured with high intra-tumoral heterogeneity and poor prognosis. To comprehensively delineate the PDAC intra-tumoral heterogeneity and the underlying mechanism for PDAC progression, we employed single-cell RNA-seq (scRNA-seq) to acquire the transcriptomic atlas of 57,530 individual pancreatic cells from primary PDAC tumors and control pancreases, and identified diverse malignant and stromal cell types, including two ductal subtypes with abnormal and malignant gene expression profiles respectively, in PDAC. We found that the heterogenous malignant subtype was composed of several subpopulations with differential proliferative and migratory potentials. Cell trajectory analysis revealed that components of multiple tumor-related pathways and transcription factors (TFs) were differentially expressed along PDAC progression. Furthermore, we found a subset of ductal cells with unique proliferative features were associated with an inactivation state in tumor-infiltrating T cells, providing novel markers for the prediction of antitumor immune response. Together, our findings provide a valuable resource for deciphering the intra-tumoral heterogeneity in PDAC and uncover a connection between tumor intrinsic transcriptional state and T cell activation, suggesting potential biomarkers for anticancer treatment such as targeted therapy and immunotherapy.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the leading causes of cancer death with only 8% of 5-year survival rate.[1,2] Although surgical resection remains the main option for PDAC treatment, only 10–15% of newly diagnosed patients are eligible.[3] Most patients will eventually die from metastasis,[4] owing to the lack of other treatment modalities for improving the prognosis of PDAC patients.[5]

Genomic and transcriptomic studies have revealed that critical gene mutations or aberrant signaling pathway(s) drive the pathogenesis of PDAC, such as *KRAS* driver mutation (over 90%) and frequent inactivation of *TP53*, *SMAD4*, *CDKN2A* tumor suppressors (over 50%). Other novel recurrent mutations (<10%) have also been identified from unbiased analyses in PDAC.[6] These diverse gene mutations converge on specific pathways and processes, including KRAS, TGF-β, Wnt, Notch, ROBO/SLIT signaling, chromatin remodeling and DNA repair pathways. In addition, alteration of epigenetic pathways is an emerging mechanism of PDAC progression. Inactivating mutations of chromatin modifiers have been identified in PDAC patients. These modifiers include histone modification enzymes (24% of PDAC) and SWI/SNF-mediated chromatin remodeling complexes (14% of PDAC).[7,8] Unfortunately, none of these findings have been translated into clinical use, mainly due to the very limited knowledge about their potential role during PDAC progression, whereas most patients were already at advanced stages at the time of diagnosis.[9]

Although initiation- and metastasis-specific mutations begun to be confirmed,[10,11] dysregulated signal transduction or variation of gene expression within primary tumor cells are also critical for tumor progression.[12] This is further complicated by the signaling cues from the tumor microenvironment and pathways regulating epithelial-to-mesenchymal transition (EMT).[13–15] Meanwhile, intra-tumoral heterogeneity exists between cells within PDAC. In particular, the stroma constitutes over 70% of the tumor mass often embedded with normal pancreatic tissue due to the infiltrative nature of PDAC.[16] This extensive degree of intra-

[1]Department of Medical Research Center, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, 100730 Beijing, China; [2]CAS Key Laboratory of Genomic and Precision Medicine, Collaborative Innovation Center of Genetics and Development, College of Future Technology, Beijing Institute of Genomics, Chinese Academy of Sciences, 100101 Beijing, China; [3]University of Chinese Academy of Sciences, 100049 Beijing, China; [4]Institute of Stem Cell and Regeneration, Chinese Academy of Sciences, 100101 Beijing, China; [5]Department of General Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, 100730 Beijing, China; [6]Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, 100730 Beijing, China; [7]Clinical Biobank, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, 100730 Beijing, China and [8]Tsinghua University-Peking University Joint Center for Life Sciences, School of Medicine, Tsinghua University, 100084 Beijing, China
Correspondence: Yupei Zhao (zhao8028@263.net) or Yun-Gui Yang (ygyang@big.ac.cn) or Wenming Wu (doctorwuu@126.com)
These authors contributed equally: Junya Peng, Bao-Fa Sun, Chuan-Yuan Chen, Jia-Yi Zhou, Yu-Sheng Chen

Received: 19 December 2018 Accepted: 10 June 2019
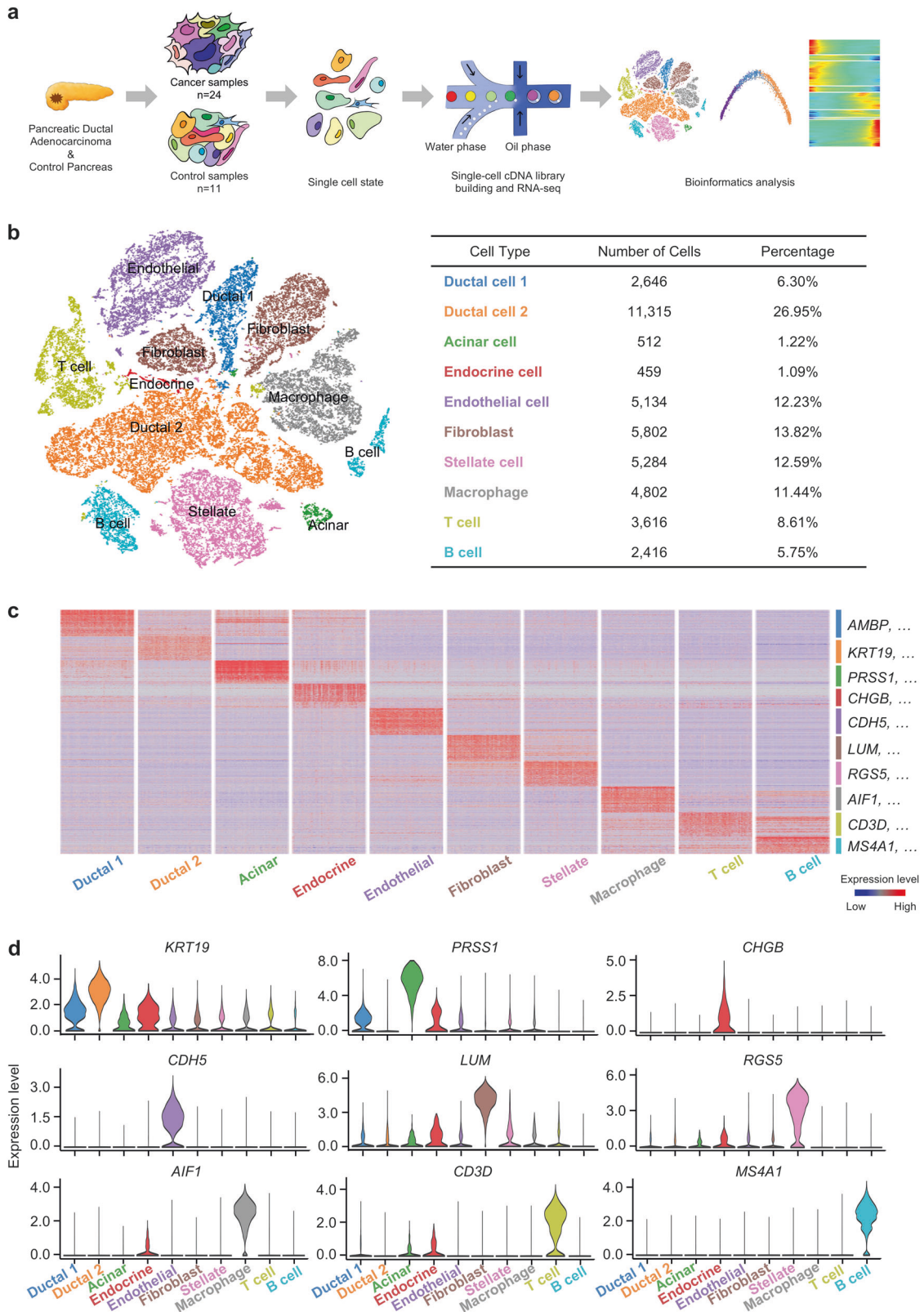Published online: 4 July 2019

a

Pancreatic Ductal Adenocarcinoma & Control Pancreas

Cancer samples n=24

Control samples n=11

Single cell state

Water phase   Oil phase

Single-cell cDNA library building and RNA-seq

Bioinformatics analysis

b

Cell Type	Number of Cells	Percentage
Ductal cell 1	2,646	6.30%
Ductal cell 2	11,315	26.95%
Acinar cell	512	1.22%
Endocrine cell	459	1.09%
Endothelial cell	5,134	12.23%
Fibroblast	5,802	13.82%
Stellate cell	5,284	12.59%
Macrophage	4,802	11.44%
T cell	3,616	8.61%
B cell	2,416	5.75%

c

AMBP, …
KRT19, …
PRSS1, …
CHGB, …
CDH5, …
LUM, …
RGS5, …
AIF1, …
CD3D, …
MS4A1, …

Ductal 1   Ductal 2   Acinar   Endocrine   Endothelial   Fibroblast   Stellate   Macrophage   T cell   B cell

Expression level
Low   High

d

Expression level

Fig. 1 Diverse cell types in PDAC delineated by single cell transcriptomic analysis. **a** Workflow depicting collection and processing of specimens of PDAC tumors and control pancreases for scRNA-seq. **b** The t-distributed stochastic neighbor embedding (t-SNE) plot demonstrates main cell types in PDAC. Cell number and percentage of assigned cell types are summarized in the right panel. **c** Heatmap showing expression levels of specific markers in each cell type. **d** Violin plots displaying the expression of representative well-known markers across the cell types identified in PDAC. The *y* axis shows the normalized read count

tumoral heterogeneity makes it rather challenging to identify genetic variants based on bulk mRNA sequencing. Even though some major treatment breakthroughs have been facilitated in a few tumor types, such as melanoma, by the identification of oncogenic drivers using this approach,[17] the overall progress in identifying actionable diagnostic markers and therapeutic targets is still largely hindered due to the limitation of bulk profiling technologies in capturing intra-tumoral heterogeneity.

Recent advances in single-cell genomics provide powerful tools in exploration of genetic and functional heterogeneity, reconstruction of evolutionary lineages and detection of rare subpopulations.[18,19] In addition, scRNA-seq studies in human tumors revealed new insights into tumor heterogeneity and distinct subpopulations, which are pivotal for dissecting tumor-related mechanism in detail.[20–27] One recent study on head and neck cancer revealed tumor compositions including the subpopulation with partial epithelial-to-mesenchymal transition (p-EMT), shedding new lights into prediction of tumor invasion and metastasis.[24] Apart from the malignant cells, tumor mass also contains macrophages, T cells and fibroblasts, etc., forming tumor microenvironment (TME) supporting tumor progression.[28–36] For instance, in liver cancer, single-cell sequencing had been applied to depict the landscape of 11 subsets of infiltrating T cells in TME, which is potentially valuable in guiding effective immunotherapies.[30]

One recent scRNA-seq study of four intraductal papillary mucinous neoplasias (IPMNs), and two PDACs revealed pathway alterations within epithelial cells, immune cells and fibroblasts during the preneoplastic progression and discovered several biomarkers of early stage pathogenesis.[37] Here, we employed single cell transcriptome approach to dissect PDAC intra-tumoral heterogeneity and associated critical factors in regulating PDAC progression. The transcriptomic profiles of a total of 57,530 cells from 24 primary PDAC tumors and 11 control pancreases were acquired. We found that PDAC tumor mass is highly heterogeneous and composed of diverse malignant and stromal cell types. In addition, malignant ductal subtype could be distinguished by featured gene expression profile and was observed to contain highly proliferative and migratory subpopulations. We further identified a list of novel gene expression changes that affected several known cancerous pathways, and suppressed tumor-related T cell activation that is associated with clinical pathological features. Thus, our findings will improve our current understanding about the mechanism for PDAC progression, and are potentially valuable in providing novel prognosis markers for PDAC.

RESULTS
Single-cell expression atlas and cell typing in PDAC tumors and control samples
To explore the cellular diversity in PDAC, we generated single-cell RNA-seq profiles from 24 PDAC tumor samples and 11 control pancreases without any treatment (Fig. 1a and Supplementary information, Table S1). After initial quality control, we acquired single-cell transcriptomes in a total of 41,986 cells from PDAC samples and 15,544 cells from control pancreases (Supplementary information, Table S2). To explore the cellular composition of tumors, we applied principle component analysis on variably expressed genes across all cells and identified 10 main clusters including type 1 ductal, type 2 ductal, acinar, endocrine, endothelial, fibroblast, stellate, macrophage, T and B cells (Fig. 1b and Supplementary information, Fig. S1a, b). We then generated cluster-specific marker genes by performing differential gene expression analysis to define the identity of each cell cluster (Fig. 1c, d and Supplementary information, Table S3). In most cases, well-known cell type markers were used to identify cell clusters, such as KRT19 for ductal cells,[38] PRSS1 for acinar cells,[38]
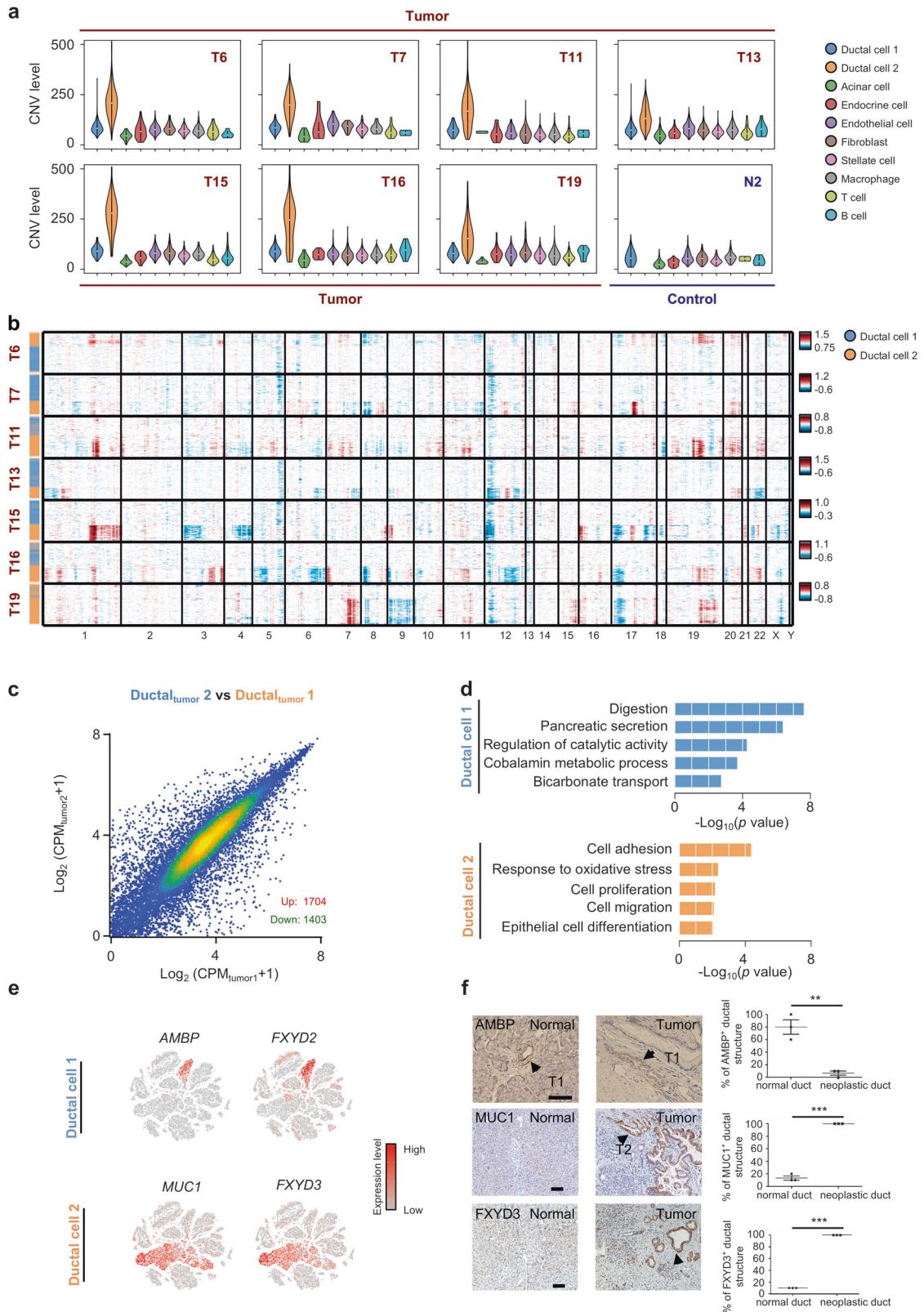
etc. (Fig. 1c, d). We also identified multiple additional markers, such as TFF2 for ductal cells and CELA3A for acinar cells (Supplementary information, Fig. S1c and Table S3). Notably, we identified two different types of ductal cells present in PDAC. Both types of ductal cells showed high level of ductal markers:[38,39] MMP7, TSPAN8, SOX9 and LCN2 (Supplementary information, Fig. S1d). However, type 2 cells had much higher expression of reported poor prognosis PDAC markers, such as CEACAM1/5/6[40] and KRT19[41] (Supplementary information, Fig. S1e), suggesting that type 2 might be a type of malignant ductal cells. We further compared scRNA-seq profiles of 15,544 cells from 11 non-tumor control pancreases, and identified 6 major cell types (Supplementary information, Fig. S1f, g). T cells and B cells are very limited (<1%) in control samples, while about half cells (49.30%) are type 1 ductal cells (Supplementary information, Fig. S1h). Type 2 ductal cells were not detected in control samples. Conversely, type 2 ductal cells showed 4.3-fold more than type 1 (11,315/2,646) in PDAC samples (Fig. 1b), indicating that type 2 ductal cells are significantly expanded in PDACs. In addition, we compared the composition of type 1 and type 2 ductal cells in PDAC patients with or without diabetes (10 vs. 14 cases) (Supplementary information, Fig. S1i). The result showed that the percentage of type 1 and type 2 ductal cells didn't exhibit significant difference (type 1: $p = 0.6605$, Wilcoxon test; type 2, $p = 0.6665$, Wilcoxon test) between these two groups, which excludes the impact of diabetes status on ductal cell population.

CNV landscape distinguishing malignant ductal cells in PDACs
The presence of two different types of ductal cells in PDAC prompted us to investigate their malignant status. To define malignant cells, we calculated large-scale chromosomal copy number variation (CNV) in each cell type based on averaged expression patterns across intervals of the genome.[20] We found that type 2 ductal cells exhibited remarkably higher CNV levels than type 1 ductal cells and other types of cells (Fig. 2a, b and Supplementary information, Fig. S2a, b). In contrast, cells in control samples display low CNV scores (Supplementary information, Fig. S2c).

We then compared type 1 ductal cells between PDAC and control pancreas, and identified 1296 upregulated and 183 downregulated genes (Supplementary information, Fig. S2d). We further found 1102/1296 of these upregulated genes were also highly expressed in type 2 ductal cells, indicating that dysregulation of these genes has already occurred in type 1 ductal cells in PDAC (Supplementary information, Fig. S2e). The functional analysis illustrated that the upregulated genes in type 1 ductal cells in PDAC were significantly enriched for several disease-related terms, such as inflammatory response, cell adhesion and migration, suggesting that type 1 ductal cells in PDAC represent the abnormal state relative to the ones in controls (Supplementary information, Fig. S2f).

We next examined gene expression patterns in two types of ductal cells in PDAC and identified 3107 differentially expressed genes (Fig. 2c and Supplementary information, Fig. S2g). The functional enrichment analyses showed that genes up-regulated in type 2 cells were mainly enriched for cancer-related functions, such as cell proliferation, migration and hypoxia (Fig. 2d), further supporting the malignant state of this subtype. In contrast, genes expressed at higher level in type 1 than in type 2 cells were related to normal pancreatic functions, including digestion, pancreatic secretion and bicarbonate transport (Fig. 2d), supporting that type 1 cells still possess a certain degree of ductal cell functions compared to malignant type 2 cells. We also identified multiple markers, such as AMBP and MUC1, that can be used to distinguish the abnormal and malignant gene expression profiles of ductal cells (Fig. 2e). We then performed Immunohistochemistry (IHC) staining of markers (MUC1, FXYD3 for type 2 ductal cells and AMBP for type 1 ductal cells) in PDAC and control samples to validate these 2 types of ductal cells. We observed that AMBP-

a

b

c Ductal_tumor 2 vs Ductal_tumor 1

d

e

f

positive cells are mainly present in the ductal structure with normal cell features. In contrast, MUC1- or FXYD3-positive cells showed typical neoplastic cell features and were absent in control samples (Fig. 2f). Taken together, these results show that type 2 ductal cells are the major source of malignant cells in PDACs.

Gene expression patterns in ductal cells during PDAC progression
We further analyzed type 1 ductal cells of PDAC samples and detected two different subgroups (Supplementary information, Fig. S3a). We examined *AMBP* and *MUC1* expression signal in these two subgroups and found that subgroup 2 displayed higher *MUC1/*

Fig. 2 CNV and transcriptome landscape of ductal cells in PDAC and control pancreases. **a** Violin plots showing distributions of CNV scores among different cell types from 7 representative PDAC and 1 control samples. **b** Heatmap showing large-scale CNVs of type 1 (blue) and type 2 (orange) ductal cells from 7 representative PDAC samples. The normalized CNV levels were shown, the red color represents high CNV level and blue represents low CNV level. **c** Scatter plots showing gene expression level of type 1 and type 2 ductal cells in PDACs. CPM (Counts per million) values were used to represent the normalized read count for each gene. **d** The enriched gene ontology terms for genes with specific expression in type 1 and 2 ductal cells. **e** Expression levels of representative markers for type 1 and 2 ductal cells are plotted onto the t-SNE map. Color key from gray to red indicates relative expression levels from low to high. The "expression level" was normalized by logNormalize method in Seurat. **f** IHC images of representative control pancreas and PDAC neoplastic tissues stained for type 1 (AMBP) and type 2 ductal cell (MUC1, FXYD3) markers. Scale bar, 100 μm. Ductal structures of fields were assessed and quantified for the presence of AMBP-, MUC1-, FXYD3-positive cells. "Normal duct" indicates ductal structures with normal nuclei, while "neoplastic duct" indicates ductal cells with enlarged nuclei. The *P* value was calculated using Student's *t*-test. Error bars indicate the standard deviation. Three independent experiments were performed. ***$P < 0.001$, **$P < 0.05$

lower *AMBP* expression (Supplementary information, Fig. S3b), suggesting that subgroup 2 cells exhibit some features of type 2 malignant ductal cells relative to subgroup 1, which is supported by the functional analyses showing several enriched tumor-related terms in this subgroup (Supplementary information, Fig. S3c).

It is not very clear yet which pancreatic cell type is responsible for tumor initiation. Some studies support the origin of malignant neoplastic cells from ductal cells,[42,43] while accumulating evidence suggests that PDAC is primarily derived from pancreatic acinar cells.[44,45] However, these findings are mainly based on the observations from transgenic mouse models and not validated in human PDAC. To investigate the origins of neoplastic cells, we performed the trajectory analysis using type 1 (two subgroups), type 2 ductal cells and acinar cells. Our data showed the possibility that both acinar and *MUC1*-low ductal cells could transit to *MUC1*-high ductal cells, and then transform to malignant type 2 ductal cells. (Supplementary information, Fig. S3d). Since only one PDAC sample contains over 100 acinar cells and total number of acinar cells in all PDAC samples is limited (512), we did not perform further analysis on the transition from acinar to malignant ductal cells, but mainly focused on the gene expression patterns along transition from ductal cells with abnormal gene expression profiles to malignant ductal cells by trajectory analysis. Pseudo-time of ductal cells with abnormal (type 1) and malignant (type 2) gene expression profiles was reconstituted (Fig. 3a). Along the trajectory, epithelial cell marker *EPCAM* was expressed in a sustained high level during the transition, while genes previously reported to be involved in tumor progression, such as *MUC1*, gradually increased along PDAC progression (Supplementary information, Fig. S3e).

We further analyzed the gene expression patterns of all genes along the trajectory of PDAC progression and identified 2299 genes with dynamic expression changes (Supplementary information, Table S4). We further clustered these genes into 4 abnormal expression patterns (P1–P4) and 4 malignant patterns (P5-P8) with specific expression patterns (Fig. 3b). Functional enrichment analysis showed that genes with normal pancreatic cell functions, such as digestion, translation, and oxidative phosphorylation, have higher expression levels in ductal cells with abnormal gene expression profiles and down-regulated when transiting to malignant state. In contrast, we identified multiple classical oncogenic pathways, including ErbB and Notch signaling pathways, were activated during PDAC progression. In particular, a large portion of genes in responsible for cell proliferation and migration were remarkably activated at the late stage of tumor progression (Fig. 3b). We also detected the activation of multiple key regulators and TFs that participate in the tumorigenesis of PDACs (Fig. 3c, d and Supplementary information, Fig. S3f), including several genes potentially associated with PDAC progression, such as PI3K-Akt pathway activator *YWHAZ*.[46]

Distinct subgroups in malignant ductal cells
The type 2 ductal cells were further divided into 7 subgroups based on t-SNE analysis (Fig. 3e, Supplementary information, Fig. S1a). By
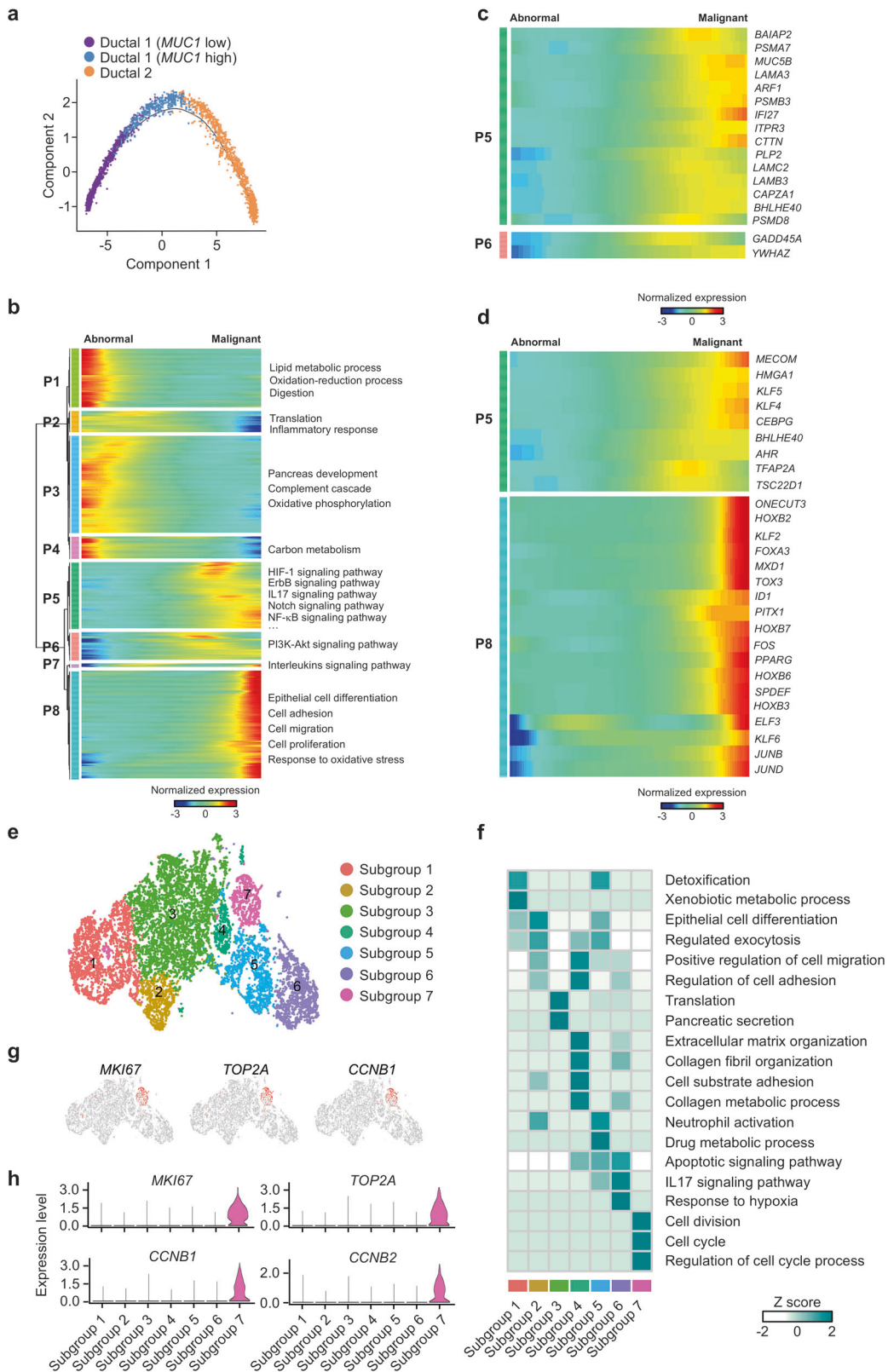
comparing gene expression levels, we found that each subgroup expresses specific gene set that can be used for distinguishing these subgroups (Supplementary Table S5). Notably, we found that subgroup 3 cells are the major population present in most patients (Supplementary information, Fig. S3g), suggesting that this subgroup is shared among PDAC patients. In contrast, subgroups 1, 2, 4, 5 and 6 are exclusively observed in some patients, which reflects the tumor heterogeneity of individual patient. Furthermore, we noticed that subgroup 7 cells with low proportion in ductal 2 malignant cells were commonly observed in 16/24 patients.

Since proliferation was found to be the main feature of PDAC, we then performed functional enrichment analysis for each subgroup and found that the unique functions of subgroup 7 were related to cell cycle and cell proliferation (Fig. 3f), which is further supported by the specific expression of proliferation markers *MKI67*, *TOP2A* and cell cycle markers *CCNB1* and *CCNB2* (Fig. 3g, h). We also observed that functions of genes specific for subgroup 1, 2 or 3 were mainly related to detoxification, epithelial cell differentiation and translation, respectively. Roles of subgroup 5 expressed genes were related to neutrophil activation, indicating a possible association with immune response. Notably, genes for both subgroups 4 and 6 were enriched for migration-related terms. Specifically, genes for subgroup 6 were enriched for GO terms such as response to hypoxia and IL17 signaling pathway, emphasizing their potential functions in migration and metastasis.

Pancreatic intraepithelial neoplasia (PanIN) is believed to be the most common preneoplastic mucinous lesion with ductal morphology.[47] However, PanIN is not well characterized at the molecular level in human PDAC. We performed immunostaining using markers FXYD3 for type 2 ductal cells and Alcian blue dye staining for PanIN. We did find that PDAC tumor sections were both FXYD3 and Alcian blue positive or only FXYD3 positive, validating the presence of PanINs in type 2 ductal cells (Supplementary information, Fig. S3h). We then examined the possible PanIN subgroups by detecting *KRT19* and *MUC5AC* expression[47] and found that subgroups 1/2/3 showed obvious signal for *KRT19 and MUC5AC* (Supplementary information, Fig. S3i), suggesting that these 3 subgroups should represent PanINs state cells. We found that PDAC tumor sections were both FXYD3 and Alcian blue positive or only FXYD3 positive, supporting that type 2 ductal cells are composed of cells in PanIN state or PDAC stage (Supplementary information, Fig. S3i). In addition, we performed trajectory analysis on all 7 subgroups of type 2 ductal cells to explore their transition and found apparently stage-specific progression patterns. Notably, most of the cells in PanIN populations (1/2/3) were present in the early progression stage, while the others appeared in the late stages of progression (Supplementary information, Fig. S3j).

Gene expression pattern analyses in TCGA PAAD samples based on malignant ductal markers
We next analyzed the public invasive pancreatic ductal adenocarcinoma and its variants (PAAD) data from TCGA (https://cancergenome.nih.gov/) to investigate the clinical

value of gene expression patterns in malignant ductal cells, including cancer expression data, clinical information and genomic mutation data. In total, 178 PAAD samples (146 PDAC samples, 32 non-PDAC PAAD samples) were used for the analysis. The malignant ductal cell markers were used to cluster the PAAD tumor samples (Fig. 4a, b and Supplementary information, Table S6). The results indicated that all samples could be assigned to three PDAC groups (cluster 1–3) and one non-PDAC group (cluster 4) based on these markers. Meanwhile, we noted that the proliferative ductal markers were

Fig. 3 Differential gene expression profiles along malignant progression. **a** Pseudo-time of ductal cells with abnormal gene expression profiles and malignant ductal cells inferred by Monocle2. Each point corresponds to a single cell. Clusters information was shown. **b** The differentially expressed genes (rows) along the pseudo-time (columns) is clustered hierarchically into eight profiles. The representative gene functions and pathways of each profile were shown. **c** Heatmap showing expression of representative identified genes potentially associated with PDAC across single cells. Color key from blue to red indicates relative expression levels from low to high. **d** Heatmap showing expression of representative identified TFs across single cells. Color key from blue to red indicates relative expression levels from low to high. **e** t-SNE representation of 7 subgroups generated from sub-clustering malignant ductal cells. **f** Heatmap showing the representative gene ontology and pathway terms enriched in each subgroup. Color key from white to blue indicates z-score of -Log$_{10}$(P value). **g** Expression levels of representative proliferation marker genes in each subgroup are plotted onto the t-SNE map. Color key from gray to red indicates relative expression levels from low to high. **h** Violin plots showing the expression level of representative proliferation marker genes across the subgroups. The y axis shows the normalized read count

clearly separated from other markers. Specifically, proliferative ductal markers were abundant in cluster 2 and 3. In contrast, patients in cluster 4 showed very limited expression level of these ductal markers. Gene mutations of KRAS, TP53 and SMAD4 were demonstrated in over 82% of cluster 1–3 patients, but not in cluster 4 patients (Supplementary information, Fig. S4a). The 5-year overall survival rate in cluster 4 was significantly higher than other 3 clusters (Supplementary information, Fig. S4b). About 89% (8/9) of patients in cluster 4 were diagnosed as pancreatic adenocarcinoma other subtype (PAOS), while 86% (145/169) of patients in cluster 1–3 were PDACs. These findings indicate that the ductal markers identified in our study are specifically expressed in PDACs instead of other types of pancreatic cancers. Taken together, the malignant ductal markers could effectively distinguish PDAC patients from patients with other types of pancreatic cancer.

We then explored the prognostic value of the proliferative ductal markers in PDAC patients. We used unsupervised non-negative matrix factorization (NNMF) clustering and divided PDAC patients into 3 subgroups according to the expression level of proliferative markers (Fig. 4c). We found that patients in group 3 showed a high abundance of proliferative ductal markers, together with a significantly lower survival rate compared to other groups (Fig. 4d). These results indicate that ductal cells with positive proliferative markers are closely associated with PDAC outcome. Next, we sought to identify drugs that can pharmacologically target the proliferative ductal cells and found that CDK1, PLK1 and AURKA can be served as therapeutic targets of PDAC by using their specific drugs (Fig. 4e and Supplementary information, Fig. S4c). We then selected inhibitors for CDK1, PLK1 and AURKA to validate their inhibitory effect on the growth of pancreatic cancer cells. The results demonstrated that these inhibitors can significantly suppress the cell proliferation (Fig. 4f). We also performed CDK1 IHC staining on pancreatic tumor tissues. Ki67 marker was co-immunostained to locate the proliferative ductal cells. We found that CDK1-positive ductal cells are also Ki67 positive (Supplementary information, Fig. S4d, e), further suggesting that CDK1 can be used as the target for proliferative ductal cells.
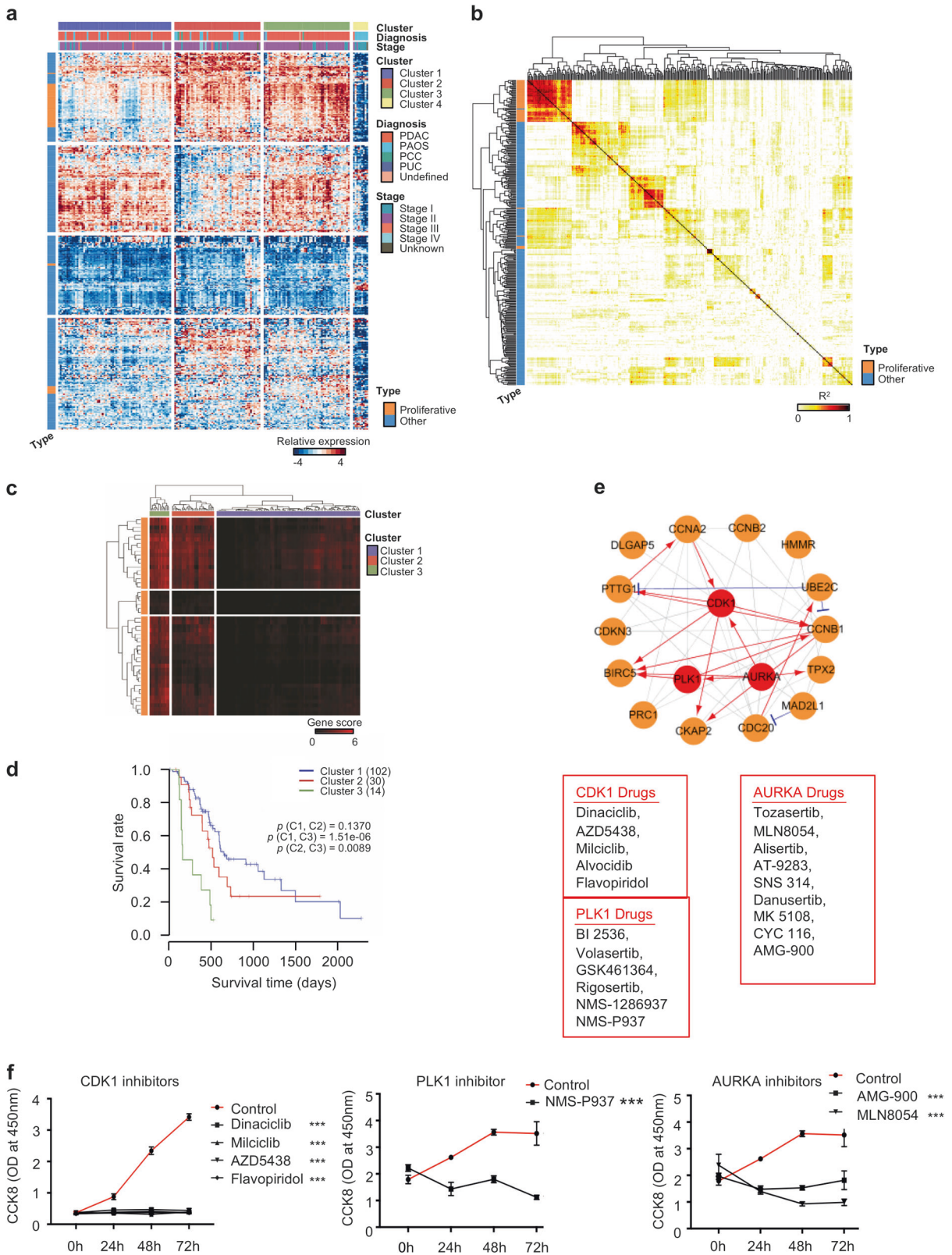
We next compared the TCGA transcriptome data between group 3 and the other two groups to determine the differential expression genes, 759 up-regulated and 946 down-regulated genes were detected in group 3 patients (p.adj ≤ 0.01 and |log$_2$FoldChange| ≥ 0.5) (Fig. 5a and Supplementary table S7). Gene ontology (GO) functional enrichment analysis revealed that genes enriched for terms such as cell cycle, DNA replication and DNA repair were highly up-regulated, whereas the downregulated genes were mainly enriched for T cell selection and lymphocyte activation terms (Fig. 5b). We then analyzed the expression pattern of the above differentially expressed genes in our single-cell data (Fig. 5c). We found that the majority of genes were highly expressed in certain specific cell type. Specifically, many up-regulated genes were expressed in type 2 ductal cells, while down-regulated genes were mainly present in macrophage and T cells (Fig. 5c), suggesting that dysregulated ductal proliferation and immune response occur concurrently in the TME.

Inactivation of T cells in PDAC patients with high abundance of proliferative ductal markers

To evaluate the function of infiltrated T cells in PDACs, we analyzed the T cell activation state within the PDAC stroma using TCGA data. We observed that samples with high level of proliferative ductal markers conversely have low expression level of T cell markers (Fig. 5d). In addition, this inverse relationship was also detected in our PDAC single-cell sequencing data (Supplementary information, Fig. S5a). We then calculated the CD8 T cell activation score[31] and found that the group with high proliferative ductal cell exhibit significantly lower T cell activation score (P value = 0.0081, Fig. 5e). To validate this finding, we co-immunostained Ki67 and CD3D by IHC on 10 PDAC tissues to investigate the relationship between proliferative state of ductal cells and T cell infiltration. For each patient, we randomly selected 5 fields containing tumor cells and quantified the expression level of Ki67 and CD3D. In support, we found an inverse spatial relationship: very limited Ki67$^+$ ductal cells with high T cell infiltration vs. high portion of Ki67$^+$ ductal cells with rare T cell infiltration (Fig. 5f and Supplementary information, Fig. S5b). We divided the patients into two groups based on the mean intensity of Ki67 and then compared their CD3D expression (Supplementary information, Fig. S5c-d). In consistent with the rare T cell infiltration, we found that patients in Ki67-high group show low CD3D expression (Supplementary information, Fig. S5e). Therefore, both the existence of proliferative ductal cells and loss of T cell activation probably contributes to the poor prognosis of PDAC patients.

Previous TCGA bulk RNA-seq analysis reported four subtypes of PDAC samples: squamous, pancreatic progenitor, immunogenic, and aberrantly differentiated endocrine exocrine (ADEX).[8,48] We used signature genes identified from TCGA four subtypes of patients to calculate the scores of each cell type found in our single-cell map. Consistently, progenitor and ADEX subtypes showed higher score in both type 2 ductal cells and acinar cells, while immunogenic subtype showed higher score in T and B cells. Notably, squamous subtype showed relatively higher score in type 2 ductal cells, endothelial, and stellate cells vs. lower score in T and B cells (Supplementary information, Fig. S5f), indicating a complicated immune response in the tumor microenvironment. We then compared these subtype signature scores in the three groups of patients with high, medium and low abundance of proliferative markers (Fig. 4d) and found that the squamous subtype score was highest in group 3 patients and immunogenic subtype score was highest in group 1 patients (Fig. 5g and Supplementary information, Fig. S5g). 50 and 0% of patients in group 3 were classified as squamous and immunogenic subtype (Supplementary information, Fig. S5h). In addition, we compared the proliferative ductal signature scores to squamous and immunogenic subtype scores for each TCGA patients. We found proliferative score had significantly positive correlation with squamous subtype and negative correlation with immunogenic subtype (Supplementary information, Fig. S5i).

Distinct subtypes of immune cells and fibroblasts in PDAC

Recent studies suggest that stromal cells, such as T cells, macrophages, and fibroblasts, are highly heterogeneous,[28–36]

a

b

c

d

e

f

however, the extent of this heterogeneity in PDAC remains poorly characterized. We performed the sub-clustering of T cells, B cells, macrophages and fibroblasts, and detected 5, 6, 5 and 8 subsets, respectively, in these cell types (Supplementary information, Fig. S6a, b). Since T cell and macrophage are the major population

shaping the immune landscape within tumor microenvironment, we delineated the specific of each of these two subtypes. For T cells, the expression of known markers suggested two clusters of CD8$^+$ cells and three clusters of CD4$^+$ cells (Supplementary information, Fig. S6a, b). We noticed that both CD8$^+$ clusters

Fig. 4 TCGA PAAD data analyses based on malignant ductal markers detected in the scRNA-seq data. **a** Heatmap showing the expression patterns of representative malignant ductal markers across the 178 PAAD samples (146 PDAC samples; 4 PCC samples: Pancreas-Colloid Carcinoma; 1 PUC sample: Pancreas Undifferentiated Carcinoma; 26 PAOS samples: Pancreatic Adenorcarconoma Other Subtype; 1 sample was the undefined subtype). Clustering identifies 4 coherent expression patterns across TCGA samples. Rows in the heatmap correspond malignant ductal markers and columns in the heatmap that correspond to TCGA samples. The clinical information including the diagnosis type and stage was also shown. **b** Heatmap depicting pairwise correlations (R^2) on the expression level of the type 2 ductal markers in TCGA PAAD samples. Most proliferative ductal markers were correlated with each other. Correlation coefficients are colored yellow to red to indicate low to high, respectively. **c** Heatmap showing the clustering result for the value of NNMF based on the proliferative ductal markers. PDAC samples were assigned into three groups according to the value of NNMF. The value of NNMF are colored black to red to indicate low to high, respectively. **d** Kaplan–Meier survival analysis of tumor samples grouped in **c**. The sample numbers for each group were shown in brackets. Statistical significance was determined using log-rank test. **e** A simplified scheme showing the functional interaction network of the representative proliferation marker genes and available drugs for the hub genes of *CDK1*, *PLK1* and *AURKA*. The interactions were generated using Ingenuity Pathway Analysis (IPA, Ingenuity Systems). The gray, red and blue lines indicate protein-protein interactions, activated or inhibited regulation, respectively. Circle with red colors indicates genes with available drugs. **f** MIA PaCa-2 cells were treated with CDK1 inhibitors Dinaciclib (10 nM), Milciclib (10 µM), AZD5438 (5 µM), Flavopiridol (500 nM); NMS-P937 (10 µM); AMG-900 (1 µM), MLN-8054 (1 µM) for indicated time. OD at 450 nm were recorded. Three independent experiments were performed. The *P* value was calculated using Student's *t*-test. Error bars indicate the standard deviation. ***$P < 0.001$

shared genes related to effector T cell (C1-CD8 and C5-CD8), marked by the high expression of cytotoxic markers *GZMA* and *GZMH*, while cell-cycle related genes (*CENPA* and *CENPE*) were preferentially expressed in cluster C5-CD8 rather than C1-CD8, suggesting that C5-CD8 T cells still retained their proliferating capacities. In addition, cluster C2-CD4 dominated by homing markers *CCR7* and *SELL* represented naïve CD4 T cells, whereas C4-CD4 expressed both *SELL* and *IL7R*, resembling central memory T cells. Notably, the cluster C3-CD4 was characterized with remarkable higher expression of Treg signature genes such as *FOXP3* and *IL2RA*.

To delineate the diversity of macrophages within tumor microenvironment, we selected cells with known myeloid markers *CD68* and *FCER1G*, and identified five cell clusters (Supplementary information, Fig. S6a, b). We found that genes involved in class II antigen presentation (*HLA* and *CD74*) were present at higher levels in cluster 1. In contrast, genes related to extracellular matrix (ECM) deposition and remodeling (*COL1A*) were preferentially expressed in cluster 2. Since myeloid-derived suppressor cells (MDSC) are recruited and maintained by chemokines and inflammatory cytokines,[49] we noticed that *CCL2* as the dominant chemokine gene for the migration of MDSC was higher in cluster 3, while cluster 5 was specialized in generating proinflammatory cytokines (*S100A8* and *S100A9*). In contrast to the recruitment of myeloid cells, cluster 4 expressed higher levels of chemokine genes responsible for the infiltration of cytotoxic T cells, such as *CXCL9* and *CXCL10*. Considering the complex compositions and functions of macrophages shown in our dataset, we think the traditional M1 vs. M2 classification of macrophage might not be appropriate to reflect the diversity of macrophages and single cell analysis allows us to evaluate the spectrum of dynamic molecular programs in macrophages.

In addition, we performed the pseudo-time analysis for the subsets in T cells and revealed a transition from naïve T cell and memory T cell to Treg. Similarly, a trajectory of B cell populations was ordered from a starting point of naïve B cell to plasmablast (Supplementary information, Fig. S6c).
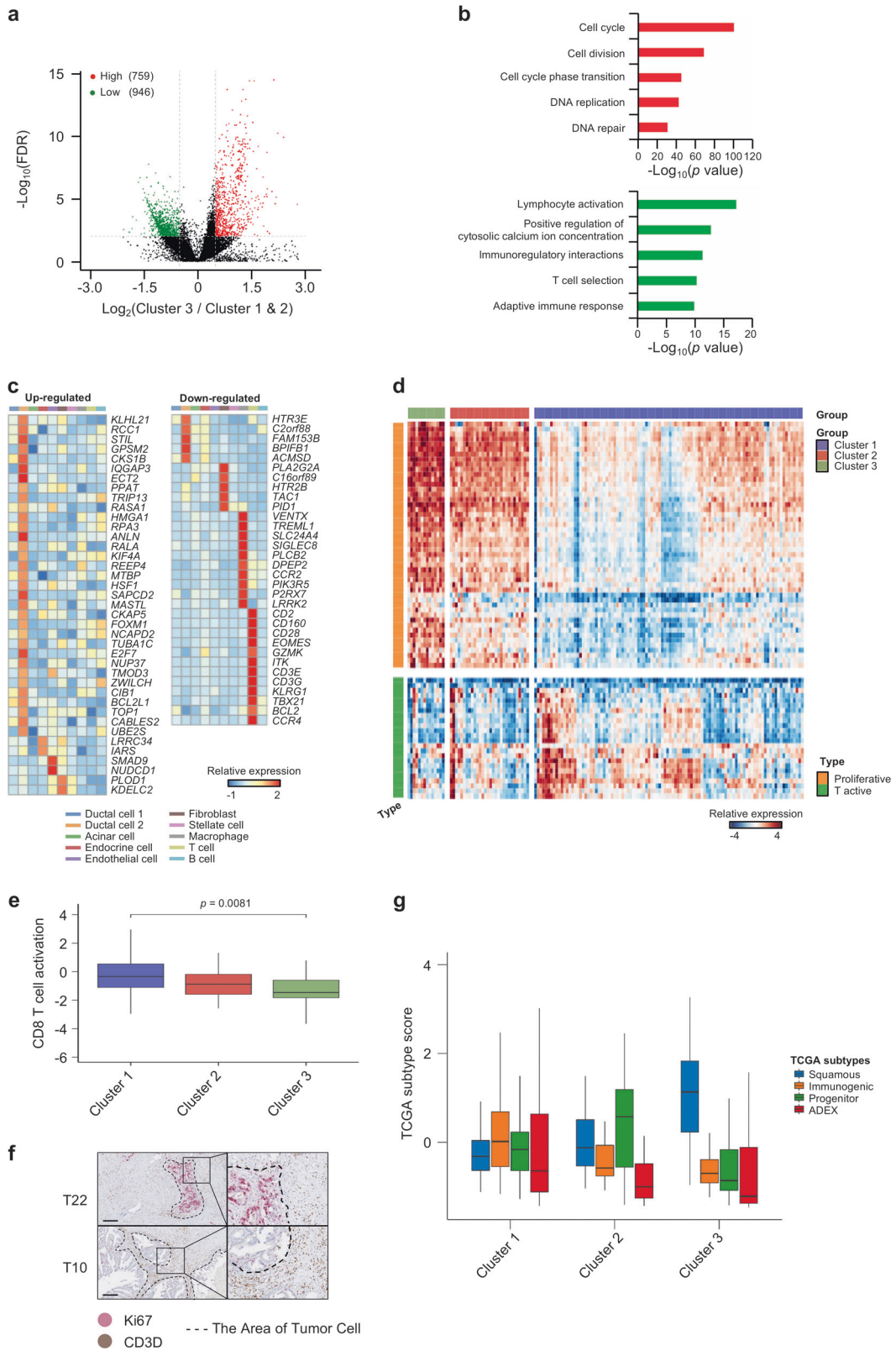
DISCUSSION

PDAC is characterized by a high degree of intra-tumoral heterogeneity that constitutes the main obstacle to effective PDAC treatment. Thus, it is highly desirable to explore the intra-tumoral heterogeneity and the underlying mechanism that are pivotal for PDAC prognostic improvement. In this study, we established the comprehensive gene expression atlas of various cell types of PDAC, and characterized the features of gene expression profiles in each cell type based on scRNA-seq analysis. Notably, two ductal cell types with different transcriptomic

profiles, named type 1 and 2 ductal cells, were identified. Further CNV-based analysis demonstrated that type 1 ductal cells are relative normal and present in both non-cancerous and neoplastic tissues, whereas type 2 ductal cells expanded in PDAC are malignant. Interestingly, type 2 ductal cells also contain highly proliferative subpopulations. Additionally, we observed suppressed T cell activation associated with clinical pathological features.

Multiple signaling pathways were previously implicated in pancreatic tumorigenesis and metastasis, including PI3K, Hedgehog, Wnt and Notch pathways,[50–53] which is also demonstrated by our findings. Notably, we further identified numbers of novel genes participating in PDAC carcinogenesis such as *EGLN3*, *MMP9* and *PLAU*, arising from dysregulated gene expression in malignant ductal cells. We identified multiple previously reported TFs in PDAC progression, such as *HMGA1*, *FOS*, *KLF5*, etc. HMGA1 was found to be tumor progression-related protein associated with advanced tumor grade and decreased survival.[54] Knockdown of its expression in human pancreatic cancer cells blocked tumor growth and metastatic progression.[55] *FOS* gene has been revealed to promote cell growth, cell cycle and migration in the pancreatic cancer cells.[56] *KLF5* was recently shown to induce mouse pancreatic cancer cell proliferation and ductal phenotype.[57] In addition to these known TFs, we also identified the activation of several TFs potentially associated with PDAC. Combined with TCGA PDAC data, we further demonstrated that patients with higher abundant proliferative ductal markers display significantly lower survival rate, suggesting potential prognostic biomarker for PDACs.

T cell infiltration status and T cell characteristics are usually associated with different prognostic outcomes.[58] It has been shown that the degree of T cell infiltration that is associated with disease progression varies in PDAC patients.[8,59–61] PDAC tumor patients with high microsatellite instability (MSI) level showed T cell infiltration and were more sensitive to immunotherapy.[62] These observations underscore the importance of understanding how distinct tumors regulate their immune heterogeneity. PDAC mouse model also revealed that tumor clones with similar oncogenic mutations showed differential capacities in orchestrating a T cell-inflamed tumor microenvironment.[63] Transcriptomic and epigenetic comparison of the T-cell-inflamed and non-T-cell-inflamed tumor clones revealed that tumors with low T cell infiltration showed enhanced *E2F* and *MYC* targeted gene signatures. In PDAC patients, we observed an inverse relationship between tumor intrinsic ductal proliferation transcriptome features and concurrent changes in immune infiltration. Although the causal relationship between proliferating tumor cells and immune infiltrations is still unknown, PDAC patients with lower ductal proliferation score and higher immune signatures showed delayed

Article

e

f

g

tumor progression and better prognosis. Another study in breast cancer found that beside their traditional function, cell cycle inhibitors can enhance tumor immunogenicity and promote T-cell-mediated clearance of tumor cells.[64] Our observation of T cell-inflamed tumors with proliferation ductal signatures suggests

a potential new combination strategy comprising cell-cycle inhibitors and immunotherapies.

Many studies based on bulk sequencing did show the existence of inter-tumoral heterogeneity, and based on different gene expression patterns, PDACs are classified into 4 subtypes: (1)

Fig. 5 Tumor T cell activation states in PDAC patients. **a** Differential expressed genes detected between group 3 and group 1–2 as in Fig. 4d. The *x* axis indicates log$_2$FC of gene expression in group 3 compared to group 1 & 2. **b** Representative enriched GO terms in up- and down-regulated genes (|log$_2$FC| ≥ 0.5 and FDR ≤ 0.01), respectively. **c** Heatmap showing the expression levels of representative up- and down-regulated genes in each cell type of the PDAC scRNA-seq data. **d** Heatmap showing the expression of markers for proliferative ductal cell and T cell activation. Patients were grouped as in Fig. 4c. **e** Boxplot showing distribution of T cell activation score in three PDAC groups as in Fig. 4c. The *P* value was calculated using Wilcoxon rank sum and signed rank test. **f** IHC images of PDAC neoplastic tissues stained for proliferative subgroup cell marker (Ki67) and T cell marker (CD3D). Representative Ki67-high (T22) and Ki67-low (T10) specimens were shown. Dashed line showed region of tumor cells. Scale bars, 200 μm. **g** Boxplot showing the signature scores of 4 known PDAC subtypes (squamous, immunogenic, progenitor and ADEX) in three PDAC groups as in Fig. 4c

squamous; (2) pancreatic progenitor; (3) immunogenic; (4) ADEX.[8] However, bulk sequencing has limitations in clarifying which type of cells has aberrant gene expression, making it difficult to find the potential therapeutic targets. The scRNA-seq approach has successfully been used to characterize the genetic and functional heterogeneity at a cellular resolution.[18,19] Using this technique, we observed different cluster patterns of type 2 ductal cells in each PDAC patient, supporting the existence of inter-tumoral heterogeneity. The 10× genomics used in the single-cell RNA sequencing technology has some limitations, such as only sequencing the 3′ end and relative low coverage. Even though these disadvantages exist, 10× Genomics Chromium has been evaluated to show the best performance in several aspects, including the molecular sensitivity, precision and lowest technical noise, when compared to other ultra-high-throughput single-cell RNA-seq such as inDrop and Drop-seq.[65]

Overall, our findings provide valuable resources for deciphering comprehensive gene expression landscapes of heterogeneous cell types in PDAC and the underlying mechanism for PDAC intra-tumoral heterogeneity. Meanwhile, we revealed the critical signaling pathways that might coordinately regulate the occurrence of intra-tumoral heterogeneity and further tumor progression in PDAC. Thus, these findings are potentially valuable in advancing not only our current understanding of critical genetic network related to PDAC progression, but also the translational use in PDAC prognosis.

MATERIALS AND METHODS
PDAC patient samples
Patients at the department of general surgery of Peking Union Medical College Hospital (PUMCH) had signed the consent forms and the proposed studies were approved by Ethics Committee (JS-1491). Age and gender of patients involved were listed (Supplementary information, Table S1).

Fresh specimens of control pancreases and PDAC were collected at the time of surgical resection under the supervision of a qualified pathologist. The control pancreas samples in our study were harvested from 3 cases of non-pancreatic tumor patients (e.g. bile duct tumors or duodenal tumors) and 8 cases of non-malignant pancreatic tumor patients (e.g. pancreatic cyst) receiving pancreatoduodenectomy or distal pancreatectomy. Control pancreas samples without visible inflammation were collected by a qualified pathologist.

Tissue dissociation and cell purification
Tissues were transported in RPMI 1640 (Gibco, Cat. no. 11875–093) with 1 mM protease inhibitor (Solarbio, Cat. no. P6730) on ice to preserve viability, washed 2–3 times with phosphate buffered saline (PBS; Hyclone, Cat. no. SH30256.01), then minced on ice. We used dissociation enzyme cocktail 1 mg/ml Type VIII Collagenase (Sigma-Aldrich, Cat. no. C2139), 2 mg/ml Dispase II (Sigma-Aldrich, Cat. no.4942078001), 1 mg/ml Trypsin Inhibitor (Sigma-Aldrich, Cat. no. T6522) and 1 unit/ml DNase I (NEB, Cat. no. M0303S) dissolved in PBS with 5% Fetal Bovine Serum (FBS; Gibco, Cat. no. 16000–044) to digest the tissues. Neoplastic tissues were dissociated at 37 °C with a shaking speed of 50 r.p.m for about 40 min. We repeatedly collected the dissociated cells at interval of

20 min to increase cell yield and viability. As for control pancreases, minced tissues were incubated with the same digestion buffer at 37 °C without shaking for about 40 min. We repeatedly collected the dissociated cells at interval of 10 min. Cell suspensions were filtered using a 40 μm nylon cell strainer (Falcon, Cat. no. 352340) and red blood cells (RBC) were removed by RBC lysis buffer (Invitrogen, Cat. no. 1966634) with 1 unit/ml DNase I. Dissociated cells were washed with PBS containing 0.04% Bovine Serum Albumin (BSA; Sigma-Aldrich, Cat. no. B2064) with step by step descending centrifuging speed and increasing time. Cells were stained with 0.4% Trypan blue (Invitrogen, Cat. no. T10282) to check the viability, and then diluted with PBS containing 0.04% BSA to about 1×10^6 cells/ml for single cell sequencing.

10× library preparation and sequencing
The concentration of single cell suspension was counted using Countess (Thermo) and adjusted to 1000 cells/μl. Cells were loaded according to standard protocol of the Chromium single cell 3′ kit in order to capture 5000 cells to 10,000 cells/chip position (V2 chemistry). All the remaining procedures including the library construction were performed according to the standard manufacturer's protocol.

Single cell RNA-seq data processing
Single cell libraries were sequenced on Illumina HiSeqXTen instruments using 150 nt paired-end sequencing. Reads were processed using the Cell Ranger 2.1.0 pipeline with default and recommended parameters. FASTQs generated from Illumina sequencing output were aligned to the human reference genome (hg19) using the STAR algorithm.[66] Next, Gene-Barcode matrices were generated for each individual sample by counting unique molecular identifiers (UMIs) and filtering non-cell associated barcodes. Finally, we generate a gene-barcode matrix containing the barcoded cells and gene expression counts.

This output was then imported into the Seurat (v2.3.0) R toolkit for quality control and downstream analysis of our single cell RNA-seq data.[67] All functions were run with default parameters, unless specified otherwise. Low quality cells (<200 genes/cell, <3 cells/gene and >10% mitochondrial genes) were excluded. Before incorporating a sample into our merged dataset, we individually inspected the cells-by-genes matrix of each as a Seurat object.

Identification of cell types and subtypes by nonlinear dimensional reduction (t-SNE)
The Seurat package implemented in R was applied to identify major cell types. Highly variable genes were generated and used to perform PCA. Significant principle components were determined using JackStraw analysis and visualization of heatmaps focusing on PCs 1 to 20. PCs 1 to 10 were used for graph-based clustering (at res = 0.8 for PDAC samples and 1 for pooled cells from PDACs and controls) to identify distinct groups of cells. These groups were projected onto t-SNE analysis run using previously computed principle components 1 to 10. We characterized the identities of cell types of these groups based on expression of known markers: *AMBP, CFTR, MMP7* (ductal cell 1),[38] *KRT19, KRT7, TSPAN8, SLPI* (ductal cell 2),[38,39,68–71] *PRSS1, CTRB1, CTRB2, REG1B* (acinar),[38,69,71] *CHGB, CHGA, INS, IAPP* (endocrine cell),[38,72,73] *RGS5,*

ACTA2, *PDGFRB*, *ADIRF* (stellate cell),[38,74,75] *LUM*, *DCN*, *COL1A1* (fibroblast),[33,76] *CDH5*, *PLVAP*, *VWF*, *CLDN5* (endothelial cell), *AIF1*, *CD64*, *CD14*, *CD68* (macrophage),[78–80] *CD3D*[33,38,76,77], *CD3E*, *CD4*, *CD8* (T cell),[79,81] *MS4A1*, *CD79A*, *CD79B*, *CD52* (B cell).[82–84] Subclustering of T cells, macrophages, B cells and fibroblasts were further performed with the same approach. In addition, subclusters of T cells without expression of either *CD4* or *CD8A* and subclusters of macrophages without expression of known myeloid markers *CD68* and *FCER1G* were excluded from further analysis.

Cluster markers identification
The cluster-specific marker genes were identified by running the FindAllMarkers function in the Seurat package to the normalized gene expression data. To identify differentially expressed genes between two clusters, we used the 'find.markers' function. We used DAVID (https://david.ncifcrf.gov/home.jsp)[85,86] and Metascape (http://metascape.org)[87] to perform biological process enrichment analysis with the top 100 differentially expressed genes in each cluster or subset.

DEG analysis of pseudo-bulks sequencing
The gene expression level of pseudo-bulks was derived from single-cell sequencing data within random permutation. In this study, we created pseudo-bulks sample by combining every 20 cells in one cell type, and used the average expression level as the expression level for scatter plot. The differential expressed genes (DEG) were identified by edgeR package[88] and DEG were filtered by |fold change| > 2 and FDR < 0.05 (Bonferroni adjust).

CNV estimation
Initial CNVs for each region were estimated by inferCNV R package.[20] The CNV of total cell types were calculated by expression level from single-cell sequencing data for each cell with –cutoff 0 and –noise_filter 0.2. For each sample, gene expression of cells was re-standardized and values were limited as −1 to 1. The CNV score of each cell was calculated as quadratic sum of CNV_{region}.

In order to well study the CNV level in ductal cell 1 and ductal cell 2 for each individual sample, we used cells except ductal cells 1/2 and acinar cells as background to eliminate the individual somatic CNV and re-calculated the CNV level with the same cutoff.

Constructing single cell trajectories in PDAC
The Monocle2 package (v2.8.0)[89] was used to analyze single cell trajectories in order to discover the cell-state transitions. We used top 100 differentially expressed genes in type 1, type 2 ductal cells and acinar cells identified by Seurat to sort cells in pseudo-time order. The ductal state informed us the start point of the pseudo-time in the first round of 'orderCells'. We then set these ductal cells with abnormal gene expression profiles as the root_state argument and called 'orderCells' again. 'DDRTree' was applied to reduce dimensions and the visualization functions 'plot_cell_trajectory' were used to plot the minimum spanning tree on cells. Differentially expressed genes over the Pseudo-time from ductal cells with abnormal gene expression profiles to malignant ductal cells transition were calculated by the "differentialGeneTest" function in Monocle2 (q value $< 10^{-20}$). In the 7 subgroups of type 2 ductal cells trajectory analysis, top 50 differentially expressed genes in each subgroup were used. In the T and B cell trajectory analysis, we used genes meeting the thresholds that mean_expression ≥ 0.5 and dispersion_empirical ≥ 1 * dispersion_fit identified by Monocle2 to sort cells in pseudo-time order.

Identification of malignancy associated TFs
In order to identify malignancy associated transcription factors, we extracted a list of all identified transcription factors from Animal TFDB 2.0.[90] We compared the top 1000 genes differentially expressed as a function of the Pseudo-time with the transcription factor list, then we identified the malignancy transcription factors.

TCGA data analysis
PAAD expression matrix was used to plot the heatmap by pheatmap R package with kmeans clustering results. Mutation information was performed by maftools.[91] For each PDAC tumor samples, non-negative matrix factorization (using rNMF R package with number of factors set to 5 and gamma set to 0.025) was applied to the relative expression matrix, by transforming all negative values to zero. By integrating the sample clinical information, the survival analysis was completed by survival R package. Differentially expressed genes were calculated by DESeq2.[92] Significant differential genes were determined by the threshold of |log$_2$FoldChange| ≥ 0.5 and adjust P value ≤ 0.01. The GO enrichment analysis was completed in Metascape (http://metascape.org).[87] According to the method mentioned by Azizi et al.[31] the related gene set score of each sample was calculated by the mean expression in this gene set. CD8 T cell activation gene set was obtained from Azizi et al.[31] and rlog was used to normalize the expression value. The TCGA four subtype signatures were calculated by comparing a given subtype to each of the other three subtypes (DESeq2 with log$_2$FoldChange ≥ 2 and adjust P value ≤ 0.01). The Pearson correlation of TCGA samples were calculated based on the subtype signatures non-negative matrix factorization value. And the gene-sets scores in cells were calculated by using Jerby-Amon et al.' method.[93]

Immunohistochemistry staining
Tissue sectioning and immunohistochemistry (IHC) staining of formalin fixed paraffin-embedded (FFPE) PDAC and normal pancreatic specimens were performed by following standard protocols. All sections were 4 μm thick and deparaffinized through xylenes and ethanol gradient. Antigen retrieval was performed in a high-pressure heat repair process using citrate buffer at pH 6.0. After endogenous peroxidase was blocked by incubating 10 min in 3% H_2O_2, the slides were incubated with primary antibodies followed by HRP-linked secondary antibodies and diaminobenzidine (DAB; ZhongShan Golden bridge biotechnology Co LTD, Cat No. ZLI-9018) staining. Counterstaining was done with hematoxylin. Slides were dehydrated with sequential ethanol washes for 1 min each starting with 75%, followed by 80% and finishing with a 100% ethanol wash.

Antibodies and reagents used are listed as following: AMBP (Abcam, ab129059), MUC1 (Cell signaling technology, 4538S), FXYD3 (Novus Biologicals, NBP2–01991), Insulin (Cell signaling technology, 3014S), Glucagon (Cell signaling technology, 2760S), CDK1 (Proteintech, 19532-1-AP), Ki67 (ZhongShan Golden bridge biotechnology, ZM0166), CD3D (Abcam, ab109531), Alcian blue (Leagene, DG0041).

Quantification of Immunohistochemical staining of CD3D and Ki67 expression
DAB (3,3′-diaminobenzidine) and PAS (Periodic Acid-Schiff) staining intensity were analyzed by color deconvolution algorithms in Fiji, a derivative software of Image J. Background intensities were deducted and mean DAB and PAS intensities were converted into optical density (OD).

Proliferation assay
MIA PaCa-2 cells cultured at 37 °C with 5% CO_2 were seeded in 96-well plate (5000 cells/well) for 12 h before adding inhibitors (Selleckchem). Concentration of inhibitor was listed as below: Dinaciclib (Cat. S2768, 10 nM), Milciclib (Cat. S2751, 10 μM), AZD5438 (Cat. S2621, 5 μM), Flavopiridol (Cat. S1230, 500 nM), NMS-P937 (Cat. S7255, 10 μM), AMG-900 (Cat. S2719, 1 μM), MLN8054 (Cat. S1100, 1 μM). Cell count kit (CCK-8) reagent was added at 10 μL/well at 0, 24, 48, 72 h after plating. After adding

CCK-8 reagent for 2.5 h, the optical density (OD) at the 450 nm wavelength (OD450) was measured using a microplate reader (Wellscan MK3, Thermo/Labsystems, Finland). OD630 was served as a reference, and the OD of the blank well was used as the base level.

DATA AVAILABILITY
The accession number for the sequencing data reported in this paper is GSA: CRA001160. These data have been deposited in the Genome Sequence Archive under project PRJCA001063.

AUTHORSHIP CONTRIBUTIONS
Y.P.Z., Y.-G.Y. and W.M.W. conceived this project. J.Y.P., H.C., L.L.L., D.H., J.L.J., G.-S.C, Y.Y., W.Z.W., D.G., M.H.D., J.C.G., T.P.Z., Q.L., Y.L., Y.-L.Z. perform experiments under the supervision of W.M.W. and Y.P.Z. B.-F.S., C.-Y.C., J.-Y.Z. and Y.-S.C. perform bioinformatics analysis under the supervision of D.-L.H and Y.-G.Y. D.-L.H., Y.P.Z., Y.-G.Y. and W.M.W. wrote the manuscript.

ADDITIONAL INFORMATION
Supplementary information accompanies this paper at https://doi.org/10.1038/s41422-019-0195-y.

Competing interests: The authors declare no competing interests.

REFERENCES
1. Rahib, L. et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **74**, 2913–2921 (2014).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30 (2018).
3. Winter, J. M. et al. Survival after resection of pancreatic adenocarcinoma: results from a single institution over three decades. *Ann. Surg. Oncol.* **19**, 169–175 (2012).
4. Makohon-Moore, A. & Iacobuzio-Donahue, C. A. Pancreatic cancer biology and genetics from an evolutionary perspective. *Nat. Rev. Cancer* **16**, 553–565 (2016).
5. Castellanos, E., Berlin, J. & Cardin, D. B. Current treatment options for pancreatic carcinoma. *Curr. Oncol. Rep.* **13**, 195–205 (2011).
6. Dunne, R. F. & Hezel, A. F. Genetics and biology of pancreatic ductal adeno-carcinoma. *Hematol. Oncol. Clin. N. Am.* **29**, 595–608 (2015).
7. Waddell, N. et al. Whole genomes redefine the mutational landscape of pan-creatic cancer. *Nature* **518**, 495–501 (2015).
8. Bailey, P. et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
9. Ryan, D. P., Hong, T. S. & Bardeesy, N. Pancreatic adenocarcinoma. *N. Engl. J. Med.* **371**, 1039–1049 (2014).
10. Campbell, P. J. et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
11. Makohon-Moore, A. P. et al. Limited heterogeneity of known driver gene muta-tions among the metastases of individual patients with pancreatic cancer. *Nat. Genet.* **49**, 358–366 (2017).
12. Neureiter, D., Jager, T., Ocker, M. & Kiesslich, T. Epigenetics and pancreatic cancer: pathophysiology and novel treatment aspects. *World J. Gastroenterol.* **20**, 7830–7848 (2014).
13. Kong, S. L. et al. Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. *Mol. Syst. Biol.* **7**, 526 (2011).
14. Öhlund, D., Elyada, E. & Tuveson, D. Fibroblast heterogeneity in the cancer wound. *J. Exp. Med.* **211**, 1503–1523 (2014).
15. Krebs, A. M. et al. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat. Cell Biol.* **19**, 518–529 (2017).
16. Biankin, A. V. & Maitra, A. Subtyping pancreatic cancer. *Cancer Cell* **28**, 411–413 (2015).
17. Chapman, P. B. et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
18. Navin, N. E. The first five years of single-cell cancer genomics and beyond. *Genome Res.* **25**, 1499–1507 (2015).
19. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
20. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
21. Lawson, D. A. et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
22. Tirosh, I. et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
23. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
24. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624 e24 (2017).
25. Venteicher, A. S. et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).
26. Filbin, M. G. et al. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331–335 (2018).
27. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
28. Lavin, Y. et al. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell* **169**, 750–765 e17 (2017).
29. Stubbington, M. J. T., Rozenblatt-Rosen, O., Regev, A. & Teichmann, S. A. Single-cell transcriptomics to explore the immune system in health and disease. *Science* **358**, 58–63 (2017).
30. Zheng, C. et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* **169**, 1342–1356 e16 (2017).
31. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 e36 (2018).
32. Guo, X. et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
33. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).
34. Patil, V. S. et al. Precursors of human CD4$^+$ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci. Immunol.* **3**, eaan8664 (2018).
35. Pace, L. et al. The epigenetic control of stemness in CD8$^+$ T cell fate commitment. *Science* **359**, 177–186 (2018).
36. Slowikowski, K., Wei, K., Brenner, M. B. & Raychaudhuri, S. Functional genomics of stromal cells in chronic inflammatory diseases. *Curr. Opin. Rheumatol.* **30**, 65–71 (2018).
37. Bernard, V. et al. Single cell transcriptomics of pancreatic cancer precursors demonstrates epithelial and microenvironmental heterogeneity as an early event in neoplastic progression. *Clin. Cancer Res.* **25**, 2194–2205 (2018).
38. Segerstolpe, A. et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.* **24**, 593–607 (2016).
39. Reichert, M. & Rustgi, A. K. Pancreatic ductal cells in development, regeneration, and neoplasia. *J. Clin. Invest.* **121**, 4572–4578 (2011).
40. Gebauer, F. et al. Carcinoembryonic antigen-related cell adhesion molecules (CEACAM) 1, 5 and 6 as biomarkers in pancreatic cancer. *PloS One* **9**, e113023 (2014).
41. Yao, H. et al. Glypican-3 and KRT19 are markers associating with metastasis and poor prognosis of pancreatic ductal adenocarcinoma. *Cancer Biomark.* **17**, 397–404 (2016).
42. Pylayeva-Gupta, Y. et al. Oncogenic Kras-induced GM-CSF production promotes the development of pancreatic neoplasia. *Cancer Cell* **21**, 836–847 (2012).
43. von Figura, G. et al. The chromatin regulator Brg1 suppresses formation of intraductal papillary mucinous neoplasm and pancreatic ductal adenocarcinoma. *Nat. Cell Biol.* **16**, 255–267 (2014).
44. Habbe, N. et al. Spontaneous induction of murine pancreatic intraepithelial neoplasia (mPanIN) by acinar cell targeting of oncogenic Kras in adult mice. *Proc. Natl Acad. Sci. USA* **105**, 18913–18918 (2008).

45. Kopp, J. L. et al. Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. *Cancer Cell* **22**, 737–750 (2012).

46. Guo, F. et al. Anticancer effect of YWHAZ silencing via inducing apoptosis and autophagy in gastric cancer cells. *Neoplasma* **65**, 693–700 (2018).

47. Wei, D. et al. KLF4 is essential for induction of cellular identity change and acinar-to-ductal reprogramming during early pancreatic carcinogenesis. *Cancer Cell* **29**, 324–338 (2016).

48. Cancer Genome Atlas Research Network. Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell* **32**, 185–203 (2017).

49. Ostrand-Rosenberg, S. & Sinha, P. Myeloid-derived suppressor cells: linking inflammation and cancer. *J. Immunol.* **182**, 4499–4506 (2009).

50. Morris, J. Pt, Wang, S. C. & Hebrok, M. KRAS, Hedgehog, Wnt and the twisted developmental biology of pancreatic ductal adenocarcinoma. *Nat. Rev. Cancer* **10**, 683–695 (2010).

51. Avila, J. L. & Kissil, J. L. Notch signaling in pancreatic cancer: oncogene or tumor suppressor? *Trends Mol. Med.* **19**, 320–327 (2013).

52. Prabhu, L. et al. Critical role of NF-kappaB in pancreatic cancer. *Oncotarget* **5**, 10969–10975 (2014).

53. Thillai, K., Lam, H., Sarker, D. & Wells, C. M. Deciphering the link between PI3K and PAK: An opportunity to target key pathways in pancreatic cancer? *Oncotarget* **8**, 14173–14191 (2017).

54. Hristov, A. C. et al. HMGA1 correlates with advanced tumor grade and decreased survival in pancreatic ductal adenocarcinoma. *Modern Pathol.* **23**, 98–104 (2010).

55. Liau, S. S., Jazag, A. & Whang, E. E. HMGA1 is a determinant of cellular invasiveness and in vivo metastatic potential in pancreatic adenocarcinoma. *Cancer Res.* **66**, 11613–11622 (2006).

56. You, L. et al. c-Fos/ERK promotes the progression from pancreatic intraepithelial neoplasia to pancreatic ductal adenocarcinoma. *Oncol. Rep.* **36**, 3413–3420 (2016).

57. He, P., Yang, J. W., Yang, V. W. & Bialkowska, A. B. Kruppel-like factor 5, increased in pancreatic ductal adenocarcinoma, promotes proliferation, acinar-to-ductal metaplasia, pancreatic intraepithelial neoplasia, and tumor growth in mice. *Gastroenterology* **154**, 1494–1508 e13 (2018).

58. Sharma, P. & Allison, J. P. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* **161**, 205–214 (2015).

59. Gunderson, A. J. et al. Bruton tyrosine kinase-dependent immune cell cross-talk drives pancreas cancer. *Cancer Discov.* **6**, 270–285 (2016).

60. Balli, D., Rech, A. J., Stanger, B. Z. & Vonderheide, R. H. Immune cytolytic activity stratifies molecular subsets of human pancreatic cancer. *Clin. Cancer Res.* **23**, 3129–3138 (2017).

61. Stromnes, I. M. et al. T-cell localization, activation, and clonal expansion in human pancreatic ductal adenocarcinoma. *Cancer Immunol. Res.* **5**, 978–991 (2017).

62. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).

63. Li, J. et al. Tumor cell-intrinsic factors underlie heterogeneity of immune cell infiltration and response to immunotherapy. *Immunity* **49**, 178–193 e7 (2018).

64. Goel, S. et al. CDK4/6 inhibition triggers anti-tumour immunity. *Nature* **548**, 471–475 (2017).

65. Zhang, X. et al. Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *Mol. Cell* **73**, 130–142 e5 (2019).

66. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

67. Satija, R. et al. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

68. Dorrell, C. et al. Isolation of major pancreatic cell types and long-term culture-initiating cells using novel human surface markers. *Stem Cell Res.* **1**, 183–194 (2008).

69. Dorrell, C. et al. Transcriptomes of the major human pancreatic cell types. *Diabetologia* **54**, 2832–2844 (2011).

70. Dorrell, C. et al. Isolation of mouse pancreatic alpha, beta, duct and acinar populations with cell surface markers. *Mol. Cell. Endocrinol.* **339**, 144–150 (2011).

71. Fang, Z. et al. Single-cell heterogeneity analysis and CRISPR screen identify key beta-cell-specific disease genes. *Cell Rep.* **26**, 3132–3144 e7 (2019).

72. Muraro, M. J. et al. A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385–394 e3 (2016).

73. Xin, Y. et al. RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab.* **24**, 608–615 (2016).

74. Bahrami, A. J. et al. Regulator of G-protein signaling-5 is a marker of hepatic stellate cells and expression mediates response to liver injury. *PloS One* **9**, e108505 (2014).

75. Baron, M. et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* **3**, 346–360 e4 (2016).

76. Vanlandewijck, M. et al. A molecular atlas of cell types and zonation in the brain vasculature. *Nature* **554**, 475–480 (2018).

77. Stevant, I. et al. Deciphering cell lineage specification during male sex determination with single-cell RNA sequencing. *Cell Rep.* **22**, 1589–1599 (2018).

78. Donovan, K. M. et al. Allograft inflammatory factor 1 as an immunohistochemical marker for macrophages in multiple tissues and laboratory animal species. *Comp. Med.* **68**, 341–348 (2018).

79. Gao, S. et al. Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat. Cell Biol.* **20**, 721–734 (2018).

80. Kiss, M., Van Gassen, S., Movahedi, K., Saeys, Y. & Laoui, D. Myeloid cell heterogeneity in cancer: not a single cell alike. *Cell Immunol.* **330**, 188–201 (2018).

81. Noutsias, M. et al. Expression of functional T-cell markers and T-cell receptor Vbeta repertoire in endomyocardial biopsies from patients presenting with acute myocarditis and dilated cardiomyopathy. *Eur. J. Heart Fail.* **13**, 611–618 (2011).

82. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotech.* **36**, 411–420 (2018).

83. Fan, X. et al. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* **28**, 730–745 (2018).

84. Stephenson, W. et al. Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. *Nat. Commun.* **9**, 791 (2018).

85. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucl. Acids Res.* **37**, 1–13 (2009).

86. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

87. Tripathi, S. et al. Meta- and orthogonal integration of influenza "omics" data defines a role for UBR4 in virus budding. *Cell Host Microbe.* **18**, 723–735 (2015).

88. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

89. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

90. Zhang, H. M. et al. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucl. Acids Res.* **43**, D76–D81 (2015).

91. Mayakonda, A. et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).

92. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

93. Jerby-Arnon, L. et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997 e24 (2018).