

Multimodal single cell-resolved spatial proteomics reveal pancreatic tumor heterogeneity

Received: 20 October 2023

Accepted: 12 November 2024

Published online: 21 November 2024



Yanfen Xu^{1,5}, Xi Wang^{1,2,5}, Yuan Li^{1,5}, Yiheng Mao^{1,5}, Yiran Su¹, Yize Mao³, Yun Yang¹, Weina Gao¹, Changying Fu¹, Wendong Chen¹, Xueting Ye¹, Fuchao Liang¹, Panzhu Bai⁴, Ying Sun⁴, Shengping Li³, Ruilian Xu² & Ruijun Tian¹✉

Despite the advances in antibody-guided cell typing and mass spectrometry-based proteomics, their integration is hindered by challenges for processing rare cells in the heterogeneous tissue context. Here, we introduce Spatial and Cell-type Proteomics (SCPro), which combines multiplexed imaging and flow cytometry with ion exchange-based protein aggregation capture technology to characterize spatial proteome heterogeneity with single-cell resolution. The SCPro is employed to explore the pancreatic tumor microenvironment and reveals the spatial alternations of over 5000 proteins by automatically dissecting up to 100 single cells guided by multi-color imaging of centimeter-scale formalin-fixed, paraffin-embedded tissue slide. To enhance cell-type resolution, we characterize the proteome of 14 different cell types by sorting up to 1000 cells from the same tumor, which allows us to deconvolute the spatial distribution of immune cell subtypes and leads to the discovery of subtypes of regulatory T cells. Together, the SCPro provides a multimodal spatial proteomics approach for profiling tissue proteome heterogeneity.

Spatial and cell-type heterogeneity is ubiquitous in the tissue context and plays a vital role in constituting the functional diversity of human diseases¹. While single-cell and spatial transcriptomics have made remarkable achievements in revealing the intra-tumor cell diversity, in-depth profiling of its protein basis at single-cell resolution remains challenging^{1,2}. Antibody recognition-based cell typing technologies by dissociating the tissue specimens into single-cell suspension³ or multiplexed imaging analysis of the tissue microenvironment have been widely used to analyze dozens of proteins at the cellular or even sub-cellular resolution^{4,5}. However, these targeted methods are limited by

the number of available antibodies, thus falling short of comprehensively capturing the intricate cellular proteome. Recently, tremendous progress in mass spectrometry (MS)-based proteomics has rendered it a powerful tool for exploring the proteome in unbiased and global manners⁶.

Spatially resolved proteomics based on various microdissection techniques has made tremendous progress in the profiling of thousands of proteins while preserving spatial information^{7–10}. Guided by hematoxylin-eosin (H&E)- or immunohistochemical (IHC)-stained images, laser microdissection (LMD)-based spatial proteomics has

¹State Key Laboratory of Medical Proteomics and Shenzhen Key Laboratory of Functional Proteomics, Department of Chemistry and Research Center for Chemical Biology and Omics Analysis, School of Science and Guangming Advanced Research Institute, Southern University of Science and Technology, Shenzhen, China. ²Department of Oncology, The Second Clinical Medical College, Jinan University (Shenzhen People's Hospital), Shenzhen 518020, China, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen, China. ³Department of Pancreatobiliary Surgery, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, PR China.

⁴Department of System Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, China. ⁵These authors contributed equally: Yanfen Xu, Xi Wang, Yuan Li, Yiheng Mao. ✉ e-mail: tianrj@sustech.edu.cn

been successfully demonstrated in providing unprecedented insights into the heterogeneous tissue context of lethal disease with cell-type resolution. Starting with either fresh frozen or formalin-fixed, paraffin-embedded (FFPE) tissue slice samples, LMD-based spatial proteomics has been widely applied to ovarian cancer¹¹, colon cancer^{12,13}, tuberculosis¹⁴, pancreatic cancer¹⁵, COVID-19^{16,17}, etc. However, the spatial resolution of these studies largely depends on experienced pathological examination, subjective cell typing, and time-consuming manual annotation of numerous cell contours for LMD. Furthermore, the large-scale cell contours obtained from these studies commonly contain the mixture of multiple cells (hundreds to thousands of cells) and have relatively low cell-type resolution, which inevitably results in an averaging effect and ultimately blurs the spatial and cell-type information.

Efforts have been made to improve the precision and throughput of LMD-based spatial proteomics with spatial and single-cell resolution. Mund et al. recently reported the development of Deep Visual Proteomics (DVP) which combines in-house trained artificial intelligence (AI)-driven algorithms for cell segmentation and a customized automated LMD to enhance cell contour collection efficiency. The platform has been applied to spatial proteomic profiling of IHC-stained melanoma tissue with single-cell resolution¹⁸. Very recently, the DVP platform was also used to study the proteome of immunofluorescence-stained melanoma cells within the epidermal and dermal compartments of primary cutaneous melanoma¹⁹. Despite these pioneered advancements made in image-guided spatial proteomics, the sophisticated process of cell typing, cell contour collection, and sample preparation is not easy to be widely adopted. Therefore, an easily accessible and user-friendly pipeline for

centimeter-scale multi-color IHC (mIHC) image navigation-based spatial proteomic analysis is required to precisely target diverse single-cell types in different tissue microenvironments with high heterogeneity. Technical challenges, such as precise cell typing for generic multiplexed imaging, navigation transfer between high-quality imaging with coverslip and low-quality imaging without coverslip for precisely dissecting multiple cell types in a single tissue section, the no-failure cell collection, as well as highly efficient sample preparation with clean-up steps for processing rare stained tissue cells (e.g., <100 cells to single cells) for spatial proteomic analysis, remain to be addressed.

Here, we introduce the Spatial and Cell-type Proteomics platform, termed SCPro (Fig. 1), which enables the integration of image-guided spatial proteomics and flow cytometry-based cell-type proteomics to uncover cell-type heterogeneity in tissue context. The spatial proteomics aspect of SCPro coordinates accurately defines single-cell contours of centimeter-scale mIHC images based on nuclei and cell membrane identification algorithms of commercially available software without manual cell contour annotation by experts for image training, automated LMD at single-cell resolution with no-failure capture, ion exchange-based protein aggregation capture (iPAC) technology for integrated proteomics sample preparation and highly sensitive proteomics profiling of rare stained FFPE tissue cells (Fig. 1a, c). Furthermore, inspired by deconvolution algorithms that align cell-type information in spatial transcriptomics using single-cell RNA sequencing (scRNA-seq) data as a reference^{20,21}, we seek to extend the cell-type resolution of the SCPro by incorporating flow cytometry-based proteomics data of 14 distinct cell types as a reference map to deconvolute the cell-type composition and proportion in spatial proteome profiles (Fig. 1b–d). Collectively, we apply the SCPro platform to

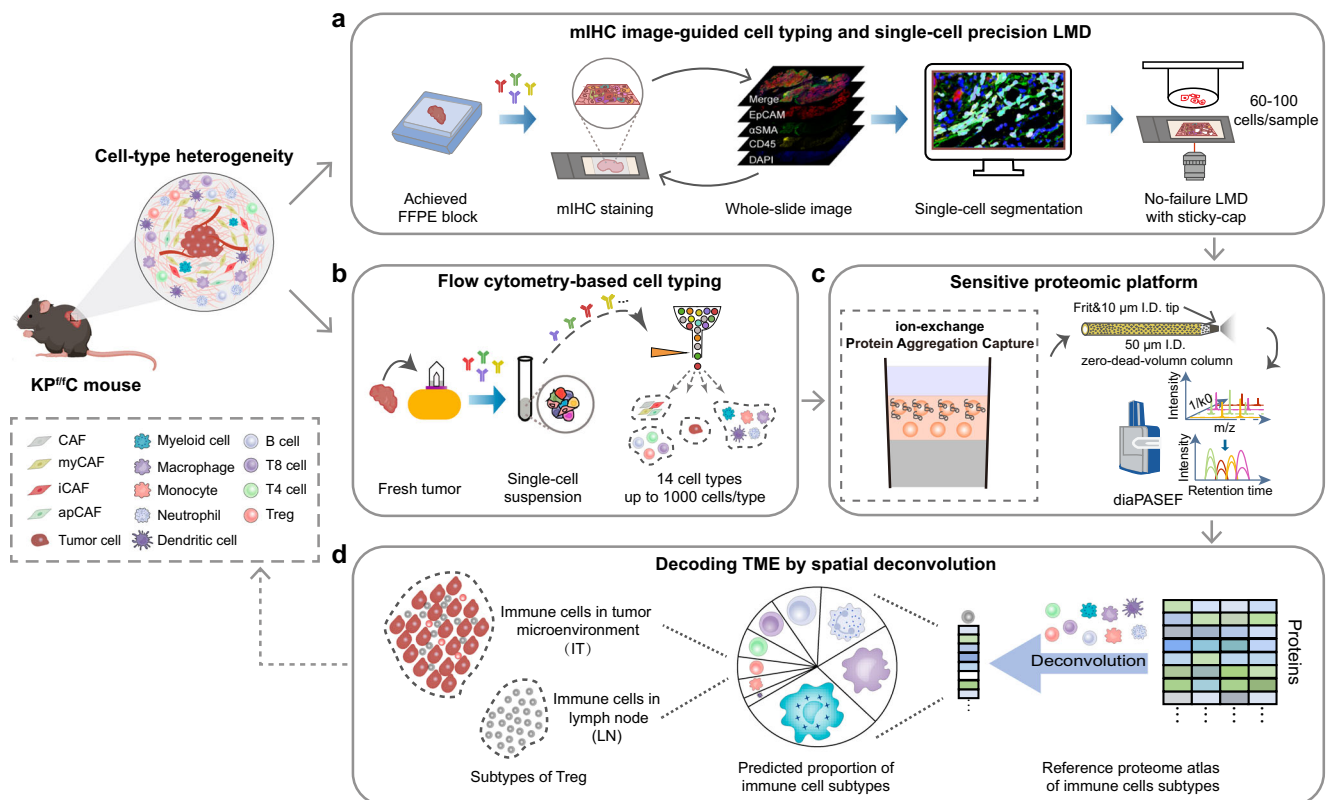


Fig. 1 | Concept and workflow of the SCPro platform. The SCPro platform integrates multiple modules. **a** Antibody-guided cell typing based on high-quality multiplexed imaging and automated LMD with single-cell resolution. **b** Flow cytometry-based cell typing. **c** Ultra-high-sensitivity proteomics platform combining with ion exchange-based protein aggregation capture sample preparation, low-flow chromatography, and high-sensitivity mass spectrometry data acquisition.

d Decoding the pancreatic tumor microenvironment through spatial deconvolution. CAF cancer-associated fibroblast, ICAF inflammatory CAF, myCAF myfibroblastic CAF, apCAF antigen-presenting CAF, Treg regulatory T cell, FFPE formalin-fixed paraffin-embedded, mIHC multiplexed immunohistochemical, LMD laser microdissection, TME tumor microenvironment.

explore the spatial proteome heterogeneity of mouse pancreatic tumor microenvironment (TME) and identify subtypes of regulatory T cells.

Results

iPAC enables in-depth spatial proteomic profiling of <100 cells in FFPE tissue slice

One of the key technical hindrances to integrating antibody recognition-based cell typing technologies and MS-based proteomics to uncover tissue proteome heterogeneity is the manipulation of rare cells. The development of an integrated proteomics sample preparation platform for processing a limited number of cells is therefore crucial for avoiding sample loss and achieving highly sensitive proteome profiling²². Notably, in the case of image navigation-based spatial proteomics, the tissue slides are often FFPE-processed and stained for visualization and cell typing. It is inevitable for hydrophobic chemical dyes and surfactants to be introduced into the tissue lysates as contaminants, which cannot be effectively removed even with the peptide clean-up steps. The accumulation of these contaminants compromises chromatographic performance and significantly influences the high-sensitive MS analysis over time¹³. Here, we introduce the solid-phase extraction (SPE)-based iPAC technology for handling low quantities of cells (e.g., <100 cells to single cells), especially for stained tissue slice samples.

The iPAC device is a spintip packed with strong anion exchange (SAX) and C18 disks in tandem which stems from our previously developed fully integrated SPE-based proteomics sample preparation technology, SISPROT^{22,23}. Notably, we incorporated several key technical advancements to enhance the robustness of the iPAC (Fig. 2a). Firstly, we introduced the “carrier” surfactant N-dodecyl- β -D-maltoside (DDM) to prevent nonspecific adsorption of low nanogram-level proteins during sample processing in the iPAC spintip²⁴. Secondly, we employed ion-exchange disks instead of loosely packed beads that significantly improve protein-capturing efficiency. Thirdly, we implemented *in situ* protein aggregation capture²⁵ by simply introducing an incubation step with pure acetonitrile (ACN) after protein capture and concentration onto the SAX disks at basic pH, which well induces precipitation of proteins and facilitates extended wash to remove contaminants and pH exchange for enzymatic digestion. The resulting peptides were then salt-eluted onto C18 disks at acidic pH for desalting and ultimately transferred to a glass insert for direct injection with negligible sample loss (Supplementary Fig. 1a, b). Last but not least, we significantly improved the sensitivity of the LC-MS system by integrating a homemade 50 μ m I.D. zero-dead-volume (ZDV) column which has a short frit made at the end of the emitter tip with neglectable dead volume and significantly improved ionization efficiency (Fig. 2a)²⁶. The homemade ZDV column running at 100 nL/min, in turn, enables the identification of over 3000 protein groups from 1 ng pre-digested HeLa cell samples in ddaPASEF acquisition mode, without using the match between run (MBR) algorithm (Supplementary Fig. 1c).

We first evaluated the performance of the iPAC by side-by-side comparing the protein recovery rate by either processing 5 ng, 10 ng, and 20 ng of HEK 293T cell lysate or directly injecting the same amounts of pre-digested HEK 293T samples (Fig. 2b). More than 3000, 4000, and 5000 protein groups were identified, respectively, using ddaPASEF mode without the MBR algorithm. Comparing with the direct injection of 5 ng peptides, we observed a recovery rate of over 60% in terms of protein group identification (Fig. 2b). Notably, a higher recovery rate was obtained by increasing the cell input. Next, we assessed the performance of iPAC using 10–1000 flow cytometry-sorted HEK 293T cells (Fig. 2c). Over 2000 and 5000 protein groups were identified from 10 and 100 sorted cells, respectively, while deactivating the MBR algorithm to avoid overestimation of protein identification for 10-cell samples²⁴. The median coefficient of variation

(CV) of maxLFQ intensities within groups was less than 10%, except for the 10-cell group with a median CV of approximately 15% (Supplementary Fig. 1d).

We then went on to validate the performance of the iPAC for analyzing H&E-stained mouse brain tissue slices, which contain much more contaminants as compared with sorted cells. Excitingly, the iPAC identified over 800, 2000, and 3200 protein groups from 20, 50, and 100 μ m-side length square samples, respectively, corresponding to 2.4, 15, and 60 cells in volume (Fig. 2d). Notably, the iPAC demonstrated high quantitative reproducibility with CV values below 15% (Supplementary Fig. 1e). The excellent performance of iPAC is largely attributed to the extended 80% ACN wash of aggregated proteins captured by the SAX disks, which is only available in this double-layer SPE-based sample preparation technology. As shown in Fig. 2e, we identified on average 4445 protein groups and 42,595 unique peptides for processing 200 μ m-side length square samples after adding the extended wash, representing a three-fold increase in the number of protein groups and an eight-fold increase in the number of unique peptides. Moreover, the median CV decreased from 26% to 8.2%, demonstrating greatly improved quantitative reproducibility (Fig. 2e, insert). This result is consistent with a significantly higher number of peptide-spectrum matches (PSMs) across the whole LC gradient for the iPAC-processed samples with the extended wash, although the total ion chromatography (TIC) intensity was similar between samples with (W/) and without (W/O) the extended wash (Fig. 2f, bottom panel).

We hypothesized that this drop in protein identification without the extended wash was primarily due to the inherent nature of stained tissues, rather than sample loss. To investigate this, we evaluated the contamination ratio (CR) which is defined by analyzing the ratio of singly charged and multiple charged precursors in the ion mobility-mass spectrometry (IM-MS) heatmap across the whole LC gradient (Fig. 2f, g and “Methods” section)²⁷. By incorporating the extended wash, the CR significantly decreased, especially in the middle of the LC gradient from 20 to 60 min (Fig. 2f, upper panel). For example, the tissue samples showed 3 and 7 times cleaner at 40 and 48 min of the LC gradient, respectively, which is close to the pre-digested HeLa cell results, but the CR for the samples without extended wash is more than 60% (Fig. 2g). Notably, the iPAC showed excellent performance in processing trace stained tissue samples compared to state-of-the-art sample preparation techniques for spatial proteomics (Supplementary Fig. 1f and Supplementary Table 1).

To further demonstrate the robustness of the iPAC technology in processing complex biological samples, we applied it to analyze the proteome of distinct regions in pancreatic tumor tissue sections. Pancreatic cancer, with its desmoplastic and immunosuppressive tumor microenvironment (TME) characterized by cancer-associated fibroblasts (CAFs) and immune cells surrounding the tumor cells, exhibits high heterogeneity and serves as a prototypical example of solid tumors with a poor prognosis²⁸. We collected acinar, lymph node, and tumor regions by dissecting 100 μ m-side length square tissue (approximately 100 tissue cells) from a fresh frozen tissue slice of a transgenic mouse model K^P/C (Kras^{LSL-G12D/+}; Trp53^{fllox}; Pdx1-Cre) (Fig. 2h and Supplementary Fig. 2a)²⁹. Around 2000 proteins were reliably identified with good quantitative reproducibility (Fig. 2i and Supplementary Fig. 2b). PCA analysis effectively differentiated cell types within the same tissue slice (Fig. 2j). Furthermore, we observed the enrichment of relevant biological processes (GOBP) for distinct cell types (Supplementary Fig. 2c). Importantly, we also applied the iPAC to spatial proteomic analysis of approximately 100 cancer cells and their adjacent stromal cells dissected out based on their cell contour and spatial proximity (Supplementary Fig. 3). The results showed the excellent performance of the iPAC in achieving in-depth spatial proteomic profiling and recapitulating the biological features of <100 cells in stained tissue samples.

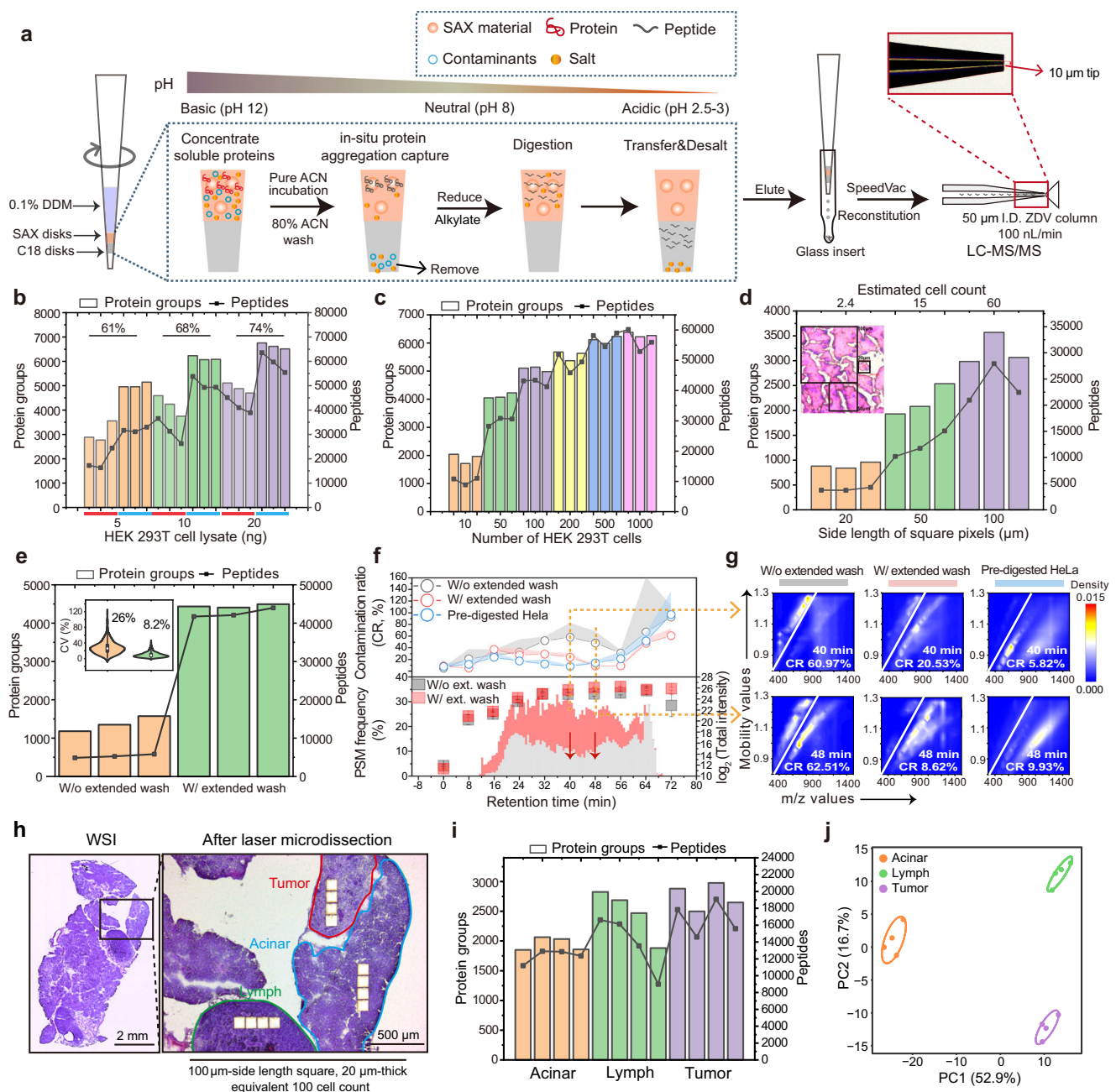


Fig. 2 | Development of iPAC technology for processing rare stained tissue cells. a Workflow and principle of iPAC technology. **b** Identified protein groups and unique peptides from processing nanogram HEK 293T cell lysate or direct injection of an equal number of pre-digested HEK 293T peptides, showed in red and blue lines, respectively. The proportion shown on the graph indicates the protein recovery rate for each group. **c** Identified protein groups and unique peptides from 10 to 1000 flow cytometry-sorted HEK 293T cells. **d** Identified protein groups and unique peptides from 20, 50, and 100 µm-side length squares of 12 µm-thick H&E-stained mouse brain slice. The images shown are representative of 3 independent experiments. **e** Identified protein groups and unique peptides of 200 µm-side length squares from a 12 µm-thick H&E-stained mouse brain without (W/o) or with (W/) extended wash. The insert violin diagram shows the Coefficient of Variation (CV) distributions ($n = 3$ biological replicates). The insert boxplots display the median (white dot), the 25th and 75th percentiles (black box), and the minimum and maximum (whiskers). **f** Upper panel, Contamination Ratios (CR) of

H&E-stained mouse brain samples W/o or W/ extended wash along the LC gradient. The pre-digested HeLa was served as a control. The transparent shades beyond each dot-line indicate the half standard deviation of CR within each group. The bottom panel identified peptide-spectrum matches (PSMs) along with retention time (RT). The total ion chromatogram (TIC) intensity of precursors extracted for generating heatmaps is displayed. **g** Representing heatmap of precursors in the trapped ion mobility (IM)- m/z space at specific RT from one of the three replicates in each group. **h** Workflow showing the acinar, tumor, and lymph regions ($n = 4$ biological replicates from one $KP^{fl/c}$ mouse). The images shown are all of 4 independent experiments. Scale bar, 2 mm and 500 µm for the left and right images, respectively. **i** Identified protein groups and peptides from different regions. **j** Principal Component Analysis (PCA) analysis based on quantified protein groups before data filtering. DDM N-dodecyl- β -D-maltoside, SAX strong anion exchange, ACN acetonitrile, ZDV zero-dead-volume. Source data are provided as a Source Data file.

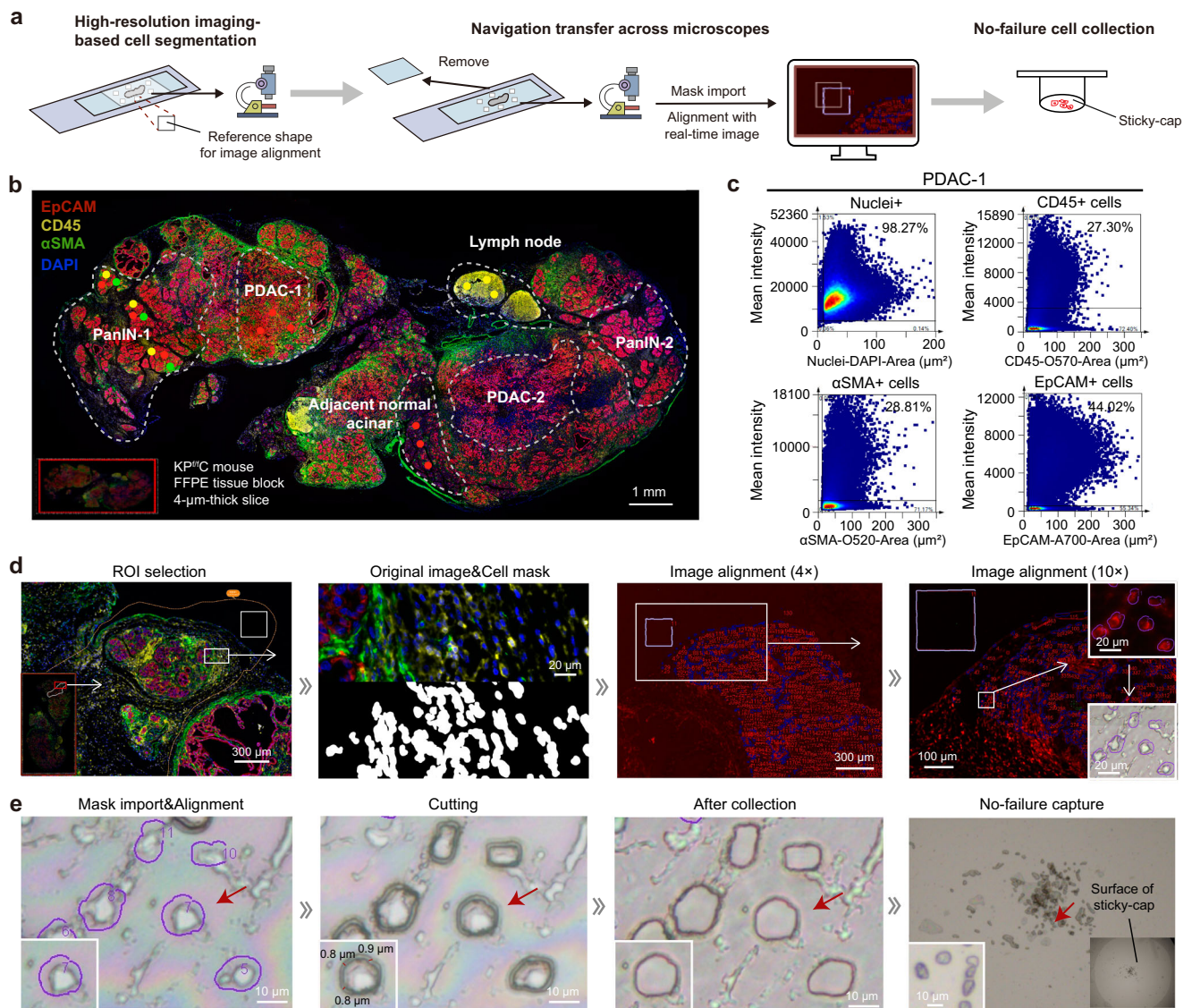


Fig. 3 | mIHC profiling and LMD capture of single cells in the pancreatic TME. **a** Spatial proteomics workflow of the SCPro. **b** Multiplexed immunohistochemical whole-slide image of a 4-μm-thick KP^{fl/c} mouse tissue section. The color dot showing the representative LMD cutting position of distinct cell types ($n = 3$ biological replicates). Scale bar, 1 mm. **c** Representative tissue cytometry charts of the PDAC-1 region. **d** Workflow of the ROI selection, cell typing and cell mask generation, and image alignment based on the real-time image of LMD. Scale bar, 300 μm,

20 μm, 300 μm, 100 μm, and 20 μm for the images from left to right, respectively. **e** Automated LMD with single-cell resolution. The images shown are representative of 3 independent experiments. Scale bar, 10 μm. EpCAM (epithelial cells); CD45 (immune cells); αSMA (fibroblasts); DAPI (nuclei). PanIN pancreatic intraepithelial neoplasia, PDAC pancreatic ductal adenocarcinoma, ROI region of interest.

SCPro captures the pancreatic TME with single-cell resolution

After evaluating the sensitivity of the iPAC for processing rare stained tissue cells, we then focused on accurately defining single-cell boundaries on mIHC-stained tissue slices without laborious and time-consuming manual cell contour annotation by experts and transferring cell typing results into an LMD microscope to guide automated microdissection. Here, we benchmarked the technological development and application of the SCPro utilizing an FFPE tissue block from the KP^{fl/c} mouse, which well recapitulates the malignant transformation process from normal acinar cells to pancreatic intraepithelial neoplasia (PanIN) and ultimately pancreatic ductal adenocarcinoma (PDAC)^{30,31}. After a rigorous pathological screening process to best represent the progression of PDAC on a single tissue section, a 4-μm-thick FFPE tissue section from a 7-week-old KP^{fl/c} mouse was subjected to 4-color mIHC staining. As the commonly used open-faced frame slide in the LMD system without the coverslip generated poor-

quality images (Supplementary Fig. 4a), navigation transfer across microscopes with and without the coverslip is a must for precise cell segmentation. Importantly, square reference shapes for image alignment were marked onto the membrane slide by LMD (Fig. 3a). Then, the high-quality multiplexed whole-slide image was obtained with the coverslip on the TissueFAXS system for accurate cell segmentation. After removing the coverslip from the frame slide, the cellular mask was imported and aligned with the low-resolution real-time image of the LMD microscope system. Subsequently, the annotated single-cell contours were isolated along the borderline of cell contours by automatic LMD and captured by a sticky-cap with no failure for downstream proteomic analysis (Fig. 3a). The centimeter-scale multi-color whole-slide image comprehensively recapitulates the spatial distribution of distinct pathological features and cell types within the pancreatic TME, including the adjacent normal acinar cells, PanIN, and PDAC (EpCAM⁺ cells), as well as the immune cells (CD45⁺ cells) and

CAFs (α SMA⁺ cells) (Fig. 3b and Supplementary Fig. 4b). Furthermore, region-specific distribution of fibroblast-enriched (PanIN-1 and PDAC-1) and fibroblast-deficient areas (PanIN-2 and PDAC-2) were also observed in both PanIN and PDAC stages. Notably, the CD45⁺ immune cells also exhibited a distinct spatial distribution within the TME (IT) and peritumoral lymph nodes (LN). Subsequently, 6 typical regions demonstrating spatial distribution heterogeneity in the pancreatic TME were obtained based on the expression level of surface markers and cellular morphology (Fig. 3b and Supplementary Fig. 4b).

Next, we conducted a quantitative tissue cytometry analysis of the 6 cell types. The gating strategy helps effectively remove weak staining and irregular cell contours in situ with single-cell resolution. (Fig. 3c and Supplementary Fig. 5). The results showed an increasing proportion of CD45⁺ immune cells and α SMA⁺ fibroblasts surrounding EpCAM⁺ tumor cells during tumor progression from acinar cells to PanIN and finally to PDAC (Supplementary Fig. 6a). Additionally, the PanIN-1 and PDAC-1 regions exhibited a higher proportion of infiltrating immune cells and fibroblasts wrapping compared to the PanIN-2 and PDAC-2 regions, illustrating the distinct spatial distribution of different cell types. The spatial distribution of α SMA⁺ CAFs and CD45⁺ immune cells surrounding the tumor cells is a crucial factor in the prognosis of PDAC²⁸. Distance map analysis illustrated a decreasing proportion of CD45⁺ immune cells and α SMA⁺ CAFs as the distance from the tumor increased (Supplementary Figs. 6b, c and 7), indicating the formation of stromal barriers by the aggregation of CAFs and suppressive immune cells, particularly in the advanced stages for contributing to the poor prognosis of pancreatic cancer²⁸.

After quantitative analysis of the multi-color image in situ, we proceeded to address the next challenge of accurately defining single-cell boundaries and transferring cell typing results for precise automated LMD. Specifically, the boundaries of each cell type were delineated by the StrataQuest (SQ) software³² in two straightforward steps based on the multiplexed image. Firstly, the nucleus identification algorithm was employed to determine the location of nuclei based on the morphology of the nucleus and DAPI staining. Then the cell morphology and staining intensity of surface markers were used to determine the cell boundaries using the membrane identification algorithm (“Methods” section). To avoid damage to the cell membrane by the laser of LMD, the offset of the cell contours for about 1 μ m was configured. Finally, a filled mask was built over the corresponding cell types (Fig. 3d and Supplementary Fig. 8). Subsequently, high-purity cell contours with single-cell resolution were generated over the original image (Fig. 3d). The obtained cell contours were imported into the LMD and aligned with the real-time image of the LMD microscope using the reference shape generated at the high-resolution imaging stage, by simply “drag and drop” of the shape to the reference square shape in the fluorescence model (Fig. 3d). Importantly, the well-annotated single cells were automatically dissected and then collected in real-time onto the sticky-cap from the membrane slide with no-failure under brightfield visualization (Fig. 3e and Supplementary Fig. 9). Specifically, the collection efficiency of the sticky-cap was benchmarked and compared with the gravity-based LMD system, which demonstrates the no-failure collection by the sticky-cap (Supplementary Fig. 10 and Supplementary Movie 1 and 2).

SCPro explores spatial proteome heterogeneity of the pancreatic TME

To achieve a balance between sufficient protein identification depth and minimal tissue usage, we isolated 60–100 phenotype-matched cell contours for quantitative proteomic analysis. Benefit from the high sensitivity of the iPAC technology, more than 3000 proteins and nearly 5000 proteins were quantified from only 60 cells and 100 cells, respectively (Fig. 4a). We first conducted spatial proteomic analysis of three cell types (Acinar cell, PanIN, and PDAC) to study the progression of pancreatic cancer. PCA analysis successfully separated these three

neighboring cell types, indicating distinct protein expression profiles (Fig. 4b). Differential expression analysis showed more upregulated proteins in acinar cell than PanIN and PDAC regions (Fig. 4c), highlighting the heterogeneity between acinar cells and the neoplasia cells, which was in line with the fact that acinar cells are normal exocrine cells and the other two cell types are neoplasia cells that exhibit varying degrees of progression^{30,33}. GO analysis revealed the enrichment of tumor-related pathways of wound healing and mitochondrial translation in PDAC, and the digestion and pancreatic juice secretion pathways were enriched in acinar cells, which well reflected their biological functions (Fig. 4d)^{33,34}. Our dataset also revealed many cell-type-specific markers from the proteome of acinar cells, which is consistent with previous studies on the human pancreas (Fig. 4c)^{33,35}. For instance, Cpa1, a known acinar cell marker playing an important role in digestive function, exhibited high expression in acinar cells compared to the other two cell types. Additionally, Reg3g, which promotes pancreatic inflammation and tumor progression from acinar cells to PanIN, showed high expression in acinar cells, implying a poor prognosis for acinar cells adjacent to tumor cells (Fig. 4c and Supplementary Fig. 11a)^{33,35}.

PDAC is a highly malignant solid tumor that is typically diagnosed at advanced stages, underscoring the significance of identifying early detection markers for developing treatment strategies. Excitingly, our spatial proteomics data revealed a progressive increase in the expression level of many proteins during PDAC progression on the same tissue slice (Fig. 4e). Slc4a4 and Anxa2 which are involved in transport and metal-binding, respectively, have been found to contribute to progression and metastasis of PDAC and have been identified as poor prognosis markers in previous studies (Fig. 4e and Supplementary Fig. 11a)^{36,37}. In addition to these two proteins, we identified several others that have previously been recognized as prognostic markers for PDAC or other solid tumors^{38–40}. Many of these proteins are located on the plasma membrane and play important biological functions, such as transporters, cell-adhesion molecules, and receptor tyrosine kinases (RTKs), making them potential therapeutic targets (Fig. 4e). Except for these well-known proteins in PDAC, we also found that Cwh43, a cell membrane protein involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis⁴¹, exhibits differential expression in acinar cells, PanIN, and PDAC and has not been previously reported in relation to PDAC (Supplementary Fig. 11b). Consistently, IHC staining in KP^{fl/c} tissues showed that Cwh43 has an increasing expression level as the tumor progressed (Supplementary Fig. 11c, d).

Expectedly, spatial proteomic analysis on CD45⁺ immune cells within the tumor microenvironment (IT) and peritumoral lymph node (LN) also revealed significant differences between these two regions (Fig. 4f). Differential expression analysis revealed an enrichment of myeloid cell-specific markers in the IT (e.g., Adgre1, Clqb, Clqc, and Siglec1), whereas the LN region exhibited an enrichment of lymphoid cell markers (e.g., Cd3d, Cd4, Cd8b, and Cd79b) (Fig. 4g). Further GO analysis of the differentially expressed proteins in the IT and LN showed that the upregulated proteins in the IT were associated with myeloid leukocyte differentiation and antigen processing and presentation pathways, while the LN exhibited enrichment in T cell- and B cell-associated signaling pathways (Fig. 4h). These findings demonstrate the distinct spatial distribution of myeloid and lymphoid cell lineages within the IT and LN regions, respectively. However, the cell composition information of immune cell subsets in these two regions is limited and needs to be enhanced to further investigate the immune landscape of the pancreatic TME in distinct spatial locations.

SCPro decodes the pancreatic immune TME through spatial deconvolution

There are numerous cell types in the pancreatic TME, and many of them are present in low abundance while performing crucial functions

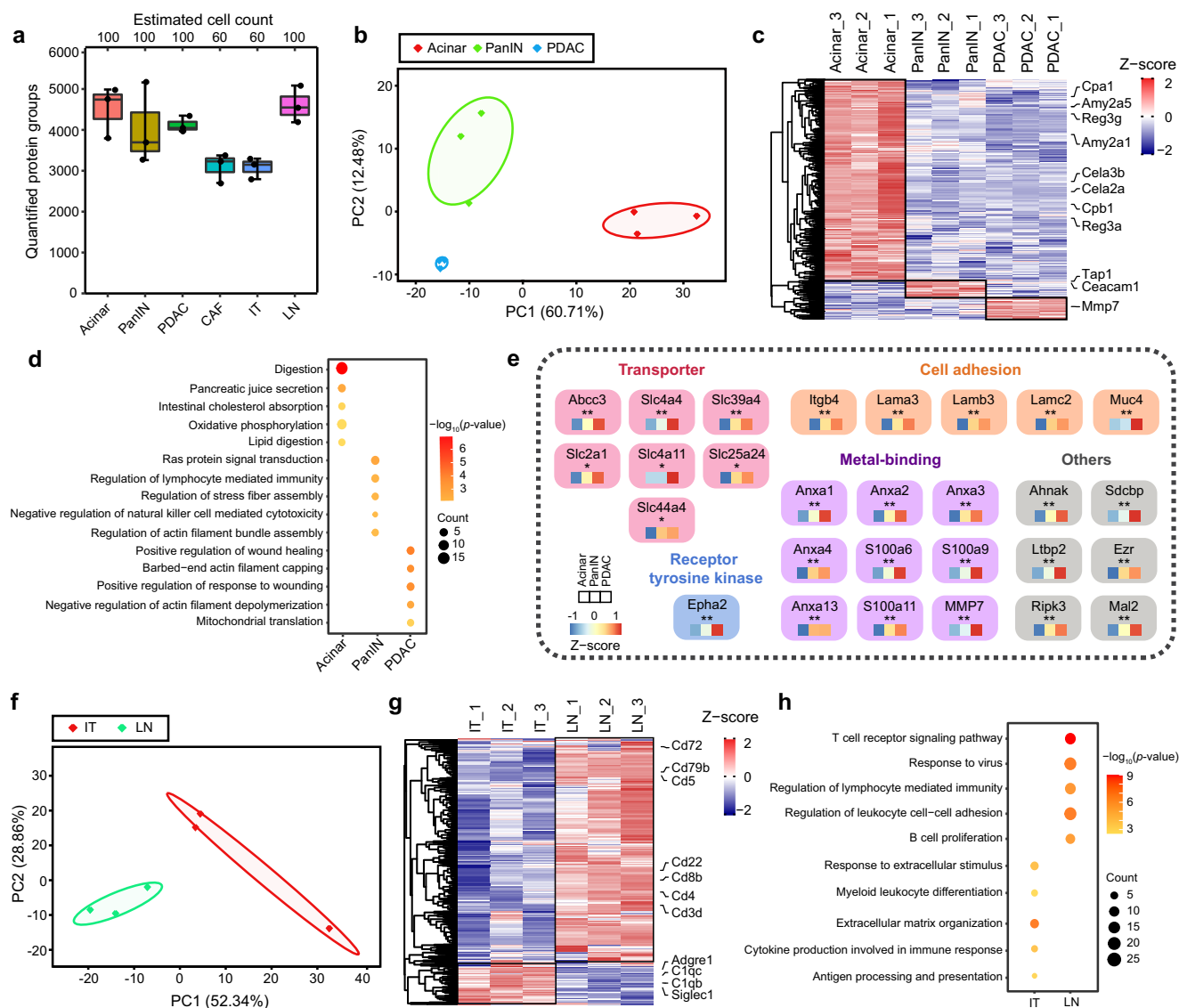


Fig. 4 | Uncovering the spatial proteomic heterogeneity of PDAC TME.

a Quantified protein groups of each cell type. A total area of 120,000 μm^3 cell contours (~60 cells) for the CAF and IT regions, and 200,000 μm^3 cell contours (~100 cells) for the Acinar, PanIN, PDAC, and LN regions were dissected from one KP^{flC} mouse tissue section with 4- μm -thick ($n = 3$ biological replicates from one KP^{flC} mouse). Boxplots display the median (horizontal line), the 25th and 75th percentiles (colored box), and the minimum and maximum (whiskers). **b** PCA analysis based on quantified protein groups before data filtering ($n = 3$ biological replicates from one KP^{flC} mouse). Replicates of each subtype were wrapped in a colored circle. **c** Unsupervised hierarchical clustering of significantly expressed proteins in acinar, PanIN, and PDAC regions ($P\text{-value} < 0.05$, fold change > 2). $P\text{-value}$ was calculated using one-way ANOVA. Known cell-type-specific markers were labeled on the right. Protein expression levels were Z-scored. **d** Dot plot showing the Gene Ontology Biology Process (GOBP) in terms of significantly

differentially expressed proteins. Significance was calculated by one-tailed Fisher's Exact Test ($P\text{-value} < 0.05$). **e** Prognostic markers for pancreatic cancer identified in the dataset and classified by their functions ($P\text{-value} < 0.05$, fold change > 1.2). $P\text{-value}$ was calculated using one-way ANOVA. *, reported in other cancers as a prognostic marker; **, reported in pancreatic cancer as a prognostic marker. **f** PCA analysis based on quantified protein groups before data filtering. Replicates of each region were wrapped in a colored circle ($n = 3$ biological replicates from one KP^{flC} mouse). **g** Unsupervised hierarchical clustering of significantly differential proteins for IT and LN regions (two-tailed Student's $t\text{-test}$, $P\text{-value} < 0.05$, fold change > 2). Known cell-type-specific markers were labeled on the right. Protein expression levels were Z-scored. **h** Dot plot showing the GOBP terms of significantly differentially expressed proteins. Significance was calculated by one-tailed Fisher's Exact Test ($P\text{-value} < 0.05$). No adjustment was made for multiple comparisons in (**c**, **d**, **e** and **h**). Source data are provided as a Source Data file.

(e.g., iCAF, apCAF, and Treg)^{42,43}. However, the spatial proteomic aspect of SCPro has a relatively low cell-type resolution due to the limited abundance of rare cell types and their functional marker signals on a single mIHC-stained slide. To further enhance the resolution of the SCPro for exploring cell subsets in the spatial proteomics data, we seek to generate cell-type-specific proteome expression information from the same tumor. Importantly, we adopted a spatial deconvolution algorithm which has been widely used in the spatial transcriptomics field to systematically explore and correlate the diverse

cell-type composition in different tissue locations within the pancreatic TME.

To build a comprehensive reference map for spatial deconvolution, we first conducted flow cytometry-based proteomic analysis to acquire cell-type-specific proteome data of the main cell types in the pancreatic TME⁴². Herein, 14 distinct cell types, which consisted of CAFs and 3 subtypes of CAFs (myCAF, iCAF, and apCAF), 9 immune cell subpopulations [B cells, CD4⁺ T cells (T4), Tregs, CD8⁺ T cells (T8), myeloid cells (MYE), dendritic cells (DC), macrophages (MAC),

neutrophils (NEU), and monocytes (MO)] and the pancreatic cancer cells (PCCs), were included in our flow cytometry-based proteomic analysis (5 biological replicates, total of 69 samples). To ensure that all the 14 cell types in a viable state could be successfully sorted from the same tumor sample, only up to 1000 cells for each individual cell type were sorted for further proteomic profiling. We successfully collected 1000 cells for all the cell types, except for the typically rare cell types myCAF, iCAF, and apCAF for which we only collected a few hundred cells as expected (Fig. 5a and Supplementary Fig. 12a, b).

Leveraging the sensitivity of the iPAC technology, we identified 4000–6000 protein groups for each cell type and over 7000 protein groups across all 14 cell types with high reproducibility in ddaPASEF mode (Fig. 5b and Supplementary Fig. 12c). PCA analysis revealed distinct distribution patterns of the proteome among the four lineages (Fig. 5c). Lymphoid and myeloid cells showed proximity due to their immune-related characteristics, while non-immune lineages (PCCs and CAFs) exhibited a closer distribution among each other. Notably, apCAF was closely associated with myeloid cells due to their expression of antigen-presenting proteins⁴². The heatmap of the top differentially expressed proteins for each cell type displayed lineage-specific markers with expected abundance, such as PCCs markers (Epcam, Msln, and Krt19), CAFs markers (Fap, Dcn, and Vim), lymphoid cells markers (Cd4, Cd8, and Cd19), and myeloid cells markers (Itgam, S100a8, and Cd74) (Fig. 5d and Supplementary Fig. 12d). Further GO analysis of the differentially expressed proteins revealed distinct functions of the four lineages (Fig. 5e). For instance, the mitochondrial translation pathway was found to be enriched in cancer cells, indicating its important role in cancer development³⁴. Fibroblasts exhibited enrichment in extracellular matrix organization and extracellular structure organization, which well aligns with their critical role in tumor progression through the production of ECM, growth factors, and chemokines⁴⁴. As scRNA-seq is widely used for cell-type exploration, a cross-omics comparison was conducted between the cell-type proteomic dataset and the well-cited scRNA-seq dataset of KPC tumor⁴². We reanalyzed the scRNA-seq dataset by subdividing the 12 main clusters into 19 clusters, including the 14 cell types analyzed in our cell-type proteomics research (Supplementary Fig. 13a). Among the 14 cell types, better correlation can be observed in the proteomics dataset compared to the transcriptomics dataset for the closely related cell types (Supplementary Fig. 13b). The correlation coefficient between RNA and protein with cell-type resolution is close to 0.5 (Supplementary Fig. 13b, numbers marked on the diagonal), which is better than previous proteogenomic study on human PDAC using bulk tissues, where the correlation coefficient was only 0.36⁴⁵. Importantly, the normalized abundance of the well-known cell-type-specific markers between the transcriptome and proteome shows good consistency at both the RNA and protein levels (Supplementary Fig. 13c). These results well demonstrated the reliability of the cell-type proteomics data and the value of multiomics analysis.

To gain deeper insights into the cellular composition and proportion of CD45⁺ immune cells in the spatial proteome data, we went on to utilize the deconvolution algorithm Tangram²⁰ to further decode the proportion of all the 14-flow cytometry-sorted cell subsets in the IT and LN region, respectively (Fig. 5f). To validate the accuracy of the cell-type proportion obtained from Tangram, we compared the relative abundance of myeloid cells and lymphoid cells in the IT and LN regions of the pancreatic TME with the cell composition acquired through mIHC imaging by co-staining Krt19 (cancer cells), CD11b (myeloid cells), CD3 (lymphoid cells), and DAPI (Fig. 5g, h and Supplementary Fig. 13d). Notably, the predicted myeloid cells and lymphoid cells proportion is 77.53% and 22.47% in the IT region and 26.61% and 73.39% in the LN region, respectively, which is generally in accordance with the image results, indicating the reliability of the deconvolution algorithm in processing proteome data. It should be noted that our previous bioinformatic analysis of spatial proteomics

data also indicates the enrichment of myeloid cell and lymphoid cell lineages in the IT and LN, respectively (Fig. 4g, h). In addition, the spatial deconvolution analysis also showed that myeloid cells, including MYE, NEU, and MAC, were the most abundant subtypes of CD45⁺ immune cells in the TME, along with a low fraction of CD8⁺ T cells and DC cells (Fig. 5f). This observation further underscored the immuno-suppressive nature of the pancreatic TME, which is mainly surrounded by the suppressive myeloid cells^{46,47}.

SCPro enables the discovery of cell subtypes in the pancreatic TME

Although the spatial and cell-type resolution of the SCPro is greatly improved by integrating the cell-type information based on the proteome information of previously identified cell subsets, such type of cell composition and proportion predication may limit the discovery of biological cell types. To this end, we went on to discover sub-cell types in the pancreatic TME by analyzing the plasma membrane (PM) proteins in the cell-type proteomic data. PM proteins play vital roles in tumor ecosystems, which commonly serve as surface markers for distinguishing distinct cell types. Notably, most of them also play important biological functions and represent valuable therapeutic targets⁴⁸.

We developed a bioinformatic strategy to explore reliable surface markers for the identification of sub-cell types within the tissue context. The workflow involved two key procedures (Fig. 6a): (1) scoring and ranking cell-type specific PM proteins by fold change and copy number within the measured cell types; (2) categorizing those ranked proteins according to major biological functions. In addition, we curated a mouse PM proteins database by incorporating Uniprot, Phobius⁴⁹, and DeepTMHMM⁵⁰ (“Methods” section). The results showed that close 10% of the identified proteins were annotated as PM proteins, and up to 300 PM proteins were identified for most cell populations (Fig. 6b). To enhance the differentiation of cell types with functional proximity within the same lineage, we employed two strategies: (1) comparing the measured cell lineages (PCC, CAF, lymphoid cell, and myeloid cell) versus the rest; (2) evaluating individual cell types versus the other cell types within the same lineage. This led to the identification of a significant differential surface marker panel, consisting of 206 unique PM proteins across 14 cell types (Fig. 6c and Supplementary Fig. 14a).

Based on normalized scores and major biological functions, we highlighted the top 2 differentially expressed PM proteins representing cell types in the four lineages, including known lineage makers and potentially new surface markers for identifying sub-cell types (Fig. 6d). Interestingly, we found that Tnfrsf18 and Klr1 were significantly over-expressed on Treg, which is a subset of CD4⁺ T cells with tumor-promoting features and has been linked to a poor prognosis for cancer therapy⁵¹. The presence of essential functions in Treg was rationalized, as Tnfrsf18 and Klr1 both function as immune checkpoint molecules^{52,53}. To validate our proteomic data, we conducted flow cytometry analysis and verified the elevated expression levels of Tnfrsf18 and Klr1 on CD25⁺ Treg (Fig. 6e). Nearly all of the Treg cells exhibited the expression of Tnfrsf18, suggesting that Tnfrsf18 may serve as an additional marker for CD25⁺ Treg in murine PDAC. Notably, the t-distributed stochastic neighbor embedding (t-SNE) plot from the flow cytometry analysis showed that CD25⁺ Treg can be subdivided into two subtypes based on the expression of Klr1 (Fig. 6e).

To systematically investigate the role of Klr1 on T cells, we further sorted 8 of the Klr1-associated T cell subtypes (i.e., CD4⁺CD25⁺Klr1⁺, CD4⁺CD25⁺, CD4⁺Klr1⁺, and CD8⁺Klr1⁺ T cells) for further cell-type proteomic analysis (Supplementary Fig. 14b, c). We then investigated the proteome differences between CD4⁺CD25⁺Klr1⁺ Treg (also named Klr1⁺ Treg) and CD4⁺CD25⁺Klr1⁺ Treg (also named Klr1⁺ Treg). These two cell types showed a clear distinction in the PCA plot (Supplementary Fig. 14d). The significantly upregulated proteins

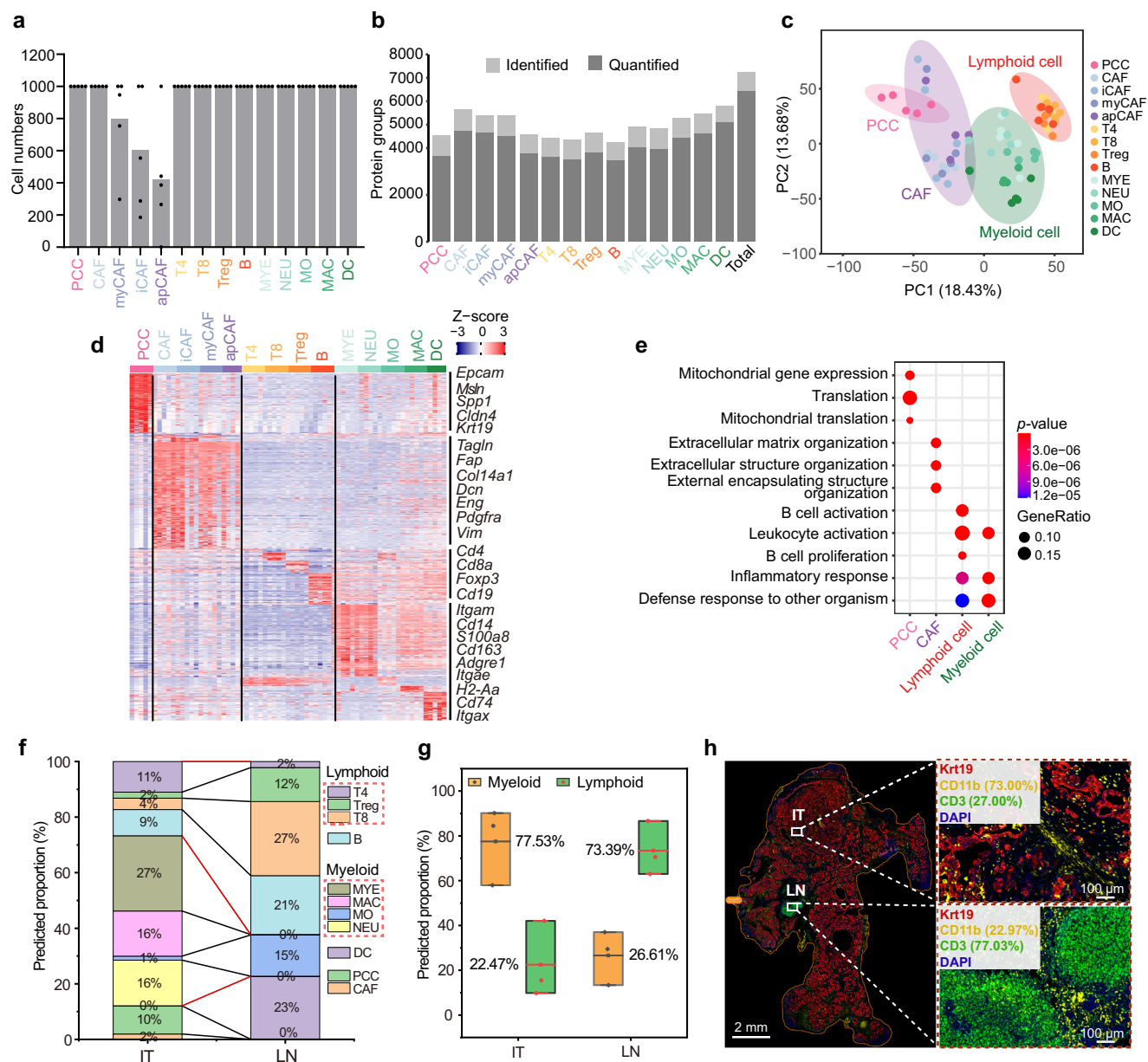


Fig. 5 | SCPro dissects the PDAC TME through spatial deconvolution. a Cell number of each cell type. **b** Proteome coverage of each cell type, the identified and quantified protein groups are indicated in gray and dark, respectively ($n = 4$ and 5 biological replicates for apCAF and the other groups, respectively, from 5 KP^{fl/c} mice). **c** PCA plot. Ellipses cover the four cell lineages indicated by colors. **d** Heatmap showing significantly differentially expressed proteins for 14 cell types with P -value < 0.05 and fold change > 2 . P -values were calculated using the moderated t -statistic (with a two-tailed test) from the LIMMA package, no adjustment was made for multiple comparisons. The well-known lineage markers were labeled on the right. Protein expression levels were Z-scored. **e** Dot plot showing the significantly enriched GOBP terms of cell-lineage-specific expressed proteins ($n = 4$ and 5 biological replicates for apCAF and the other groups, respectively, from 5 KP^{fl/c} mice). Significance was calculated by one-tailed Fisher's Exact Test (P -value < 0.05), and no adjustment was made for multiple comparisons.

f Predicted proportion of the 14-flow cytometry-sorted immune cell subtypes in the IT and LN regions. **g** Boxplots showing the predicted proportion of CD11b⁺ myeloid cells (MYE, MAC, MO, and NEU) and CD3⁺ lymphoid cells (T4, Treg, and T8) in the IT and LN regions ($n = 3$ biological replicates from one KP^{fl/c} mouse tissue section). Boxplots display the mean (horizontal line), and the 25th and 75th percentiles (bounds of box). The sum of the predicted proportions of the two cell types was set to 100% through normalization. **h** mIHC staining of CD11b⁺ myeloid cells and CD3⁺ lymphoid cells. The sum of the proportion of the two cell types was set to 100% through normalization. The images shown are representative of 2 independent experiments. Scale bar, 2 mm and 100 μm for the left and right images, respectively. Krt19 (epithelial cells); CD11b (myeloid cells); CD3 (T cells); DAPI (nuclei). PCC pancreatic cancer cell, MYE myeloid cell, NEU neutrophil, MO monocyte, MAC macrophage, DC dendritic cell. Source data are provided as a Source Data file.

in the Klr1g1⁺ Treg reveal its immunosuppressive and tumor-promoting features (Fig. 6f). For instance, Casp1 plays a crucial role in the process of pyroptosis, which is an inflammatory form of cell death⁵⁴. Kdelr2, localized to the ER-Golgi pathway, is associated with the poor prognosis and tumorigenesis of many cancers⁵⁵. Map4k1 (also known as Hpk1) acts as an immunosuppressive regulatory kinase to inhibit the

function of T cells and DC, resulting in poorer survival outcomes in pancreatic cancer⁵⁶. Additionally, GO biological analysis also indicated that the Klr1g1⁺ Treg plays a more significant role in the activation of myeloid leukocytes (Fig. 6g). To further validate the immunosuppressive function of Klr1g1⁺ Treg, we conducted biological validation by determining the expression levels of two Treg suppressive markers,

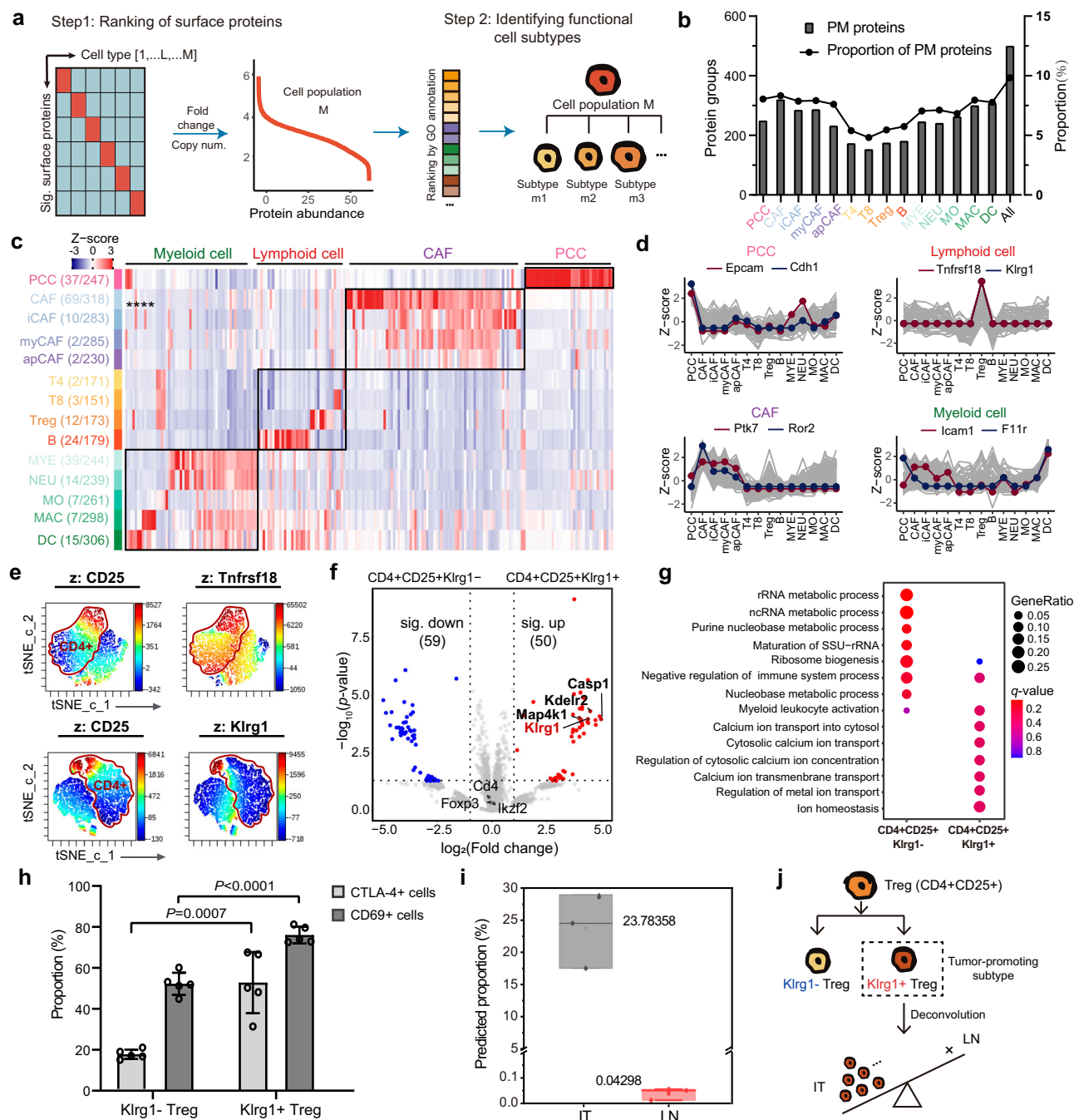


Fig. 6 | Sub-cell types discovery by SCPPro. **a** Workflow for discovering functional sub-cell types. **b** Bar plot showing the number of PM proteins. The line chart showing their proportion in all identified protein groups for individual cell types. **c** Heatmap showing the significantly differentially expressed PM proteins (P-value < 0.05, fold change > 2). P-values were calculated using the moderated *t*-statistic (with a two-tailed test) from the LIMMA package. The numbers on the left and right of brackets represent the number of significantly differentially expressed PM proteins and the total number of PM proteins for each cell type, respectively. Protein expression levels were Z-scored. **d** Line plot showing the scaled expression levels of significant proteins within four representative cell types (PCC, CAF, Treg, and DC), the top 2 of which are colored, respectively. **e** t-SNE plot showing the expression patterns of Tnfrsf18 and Klr1 on CD25⁺ Treg cells. **f** Proteome comparison of Klr1⁻ Treg and Klr1⁺ Treg. The significant proteins were shown in blue and red color, respectively (P-value < 0.05, fold change > 2). P-values were

calculated using the moderated *t*-statistic (with a two-tailed test) from the LIMMA package. **g** Dot plot showing the enriched GOBP terms of the Klr1⁻ Treg and Klr1⁺ Treg (n = 3 biological replicates from 3 KP^{fl/c} mice). Significance was calculated by one-tailed Fisher's Exact Test (P-value < 0.05), and subsequently *q*-values were estimated to control the false discovery rate across multiple comparisons. **h** Bar plot showing the proportion of CTLA-4⁺ and CD69⁺ cells in Klr1⁻ and Klr1⁺ Treg cells, respectively (n = 5 biological replicates from 5 KP^{fl/c} mice). Significance was calculated by two-tailed Student's *t*-test and data are presented as mean ± SD. **i** Predicted proportion of Klr1⁺ Treg among the 8 Klr1-associated cell types in the IT and LN regions, respectively (n = 3 biological replicates from one KP^{fl/c} mouse). Boxplots display the mean (horizontal line), and the 25th and 75th percentiles (bounds of box). **j** Summary of discovering a Treg subtype and predicting its spatial location. Source data are provided as a Source Data file.

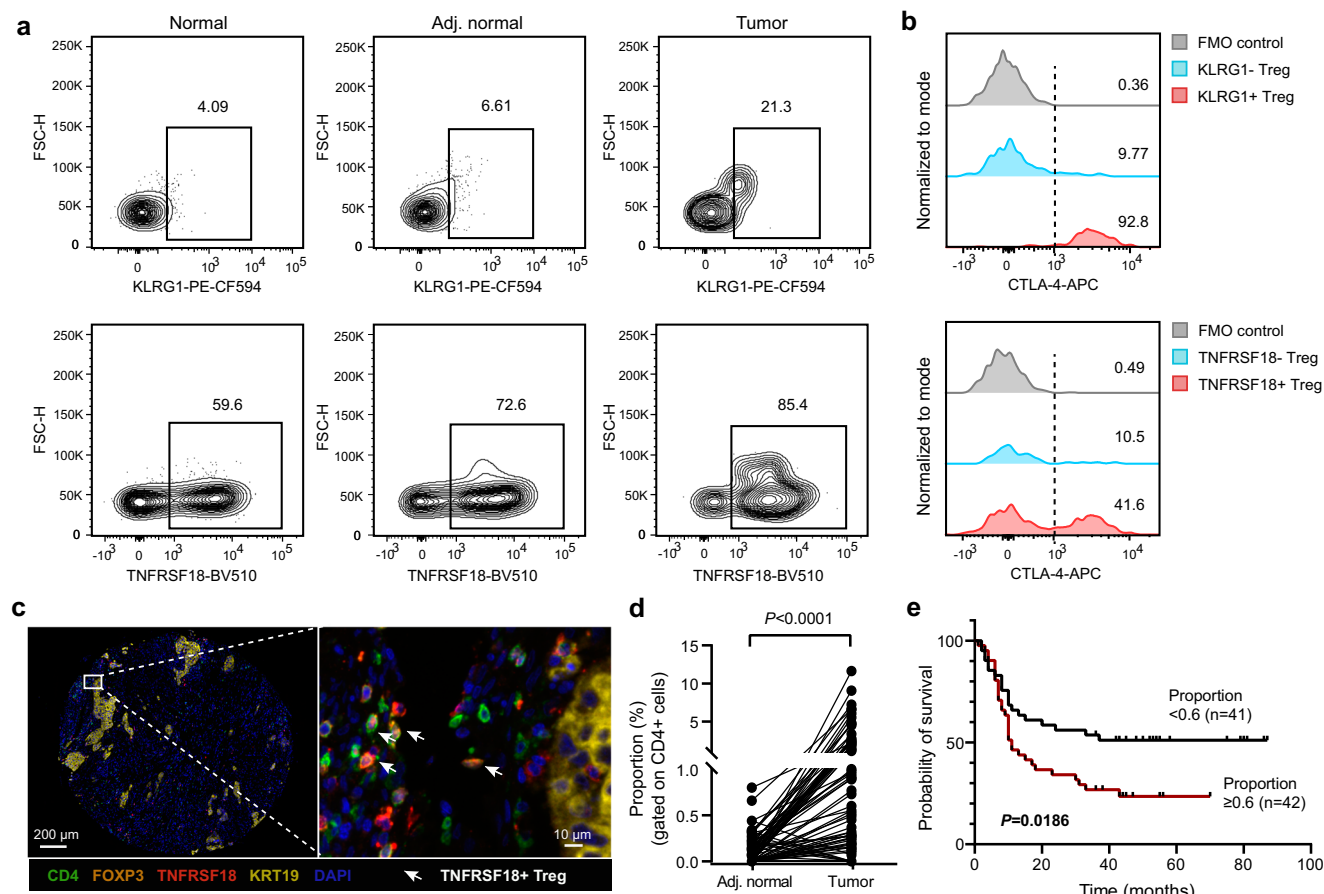


Fig. 7 | Validation of the Treg subtypes in human PDAC samples. **a** Scatter plot showing the expression levels of KLRG1 and TNFRSF18 in the Treg cells of human PDAC samples from the same patient ($n = 1$). Normal, pancreatic tissues located distant from the tumor; Adj. normal, adjacent normal samples near the tumor; Tumor, tumor sample. **b** Histograms with corresponding percentages showing the expression levels of CTLA-4 in the FMO control, KLRG1⁺ Treg, KLRG1⁺ Treg, TNFRSF18⁺ Treg, and TNFRSF18⁺ Treg, respectively. **c** Representative images of TNFRSF18⁺ Treg (CD4⁺FOXP3⁺TNFRSF18⁺ cells) in the tumor sample of human PDAC tissue microarray (TMA) slice using mIHC. Images shown are representative

of 83 pairs of adjacent normal and tumor samples on the TMA slide. Scale bar, 200 μ m and 10 μ m for the left and right images, respectively. **d** Proportion of TNFRSF18⁺ Treg among CD4⁺ T cells in the PDAC TMA slide using mIHC ($n = 81$ pairs of adjacent normal and tumor samples). Significance was calculated by a two-tailed Student's *t*-test. **e** Survival analysis of TNFRSF18⁺ Treg in PDAC TMA cohort ($n = 83$ PDAC tumor samples). The median proportion of TNFRSF18⁺ Treg was used as the cutoff to define the low and high-expression groups. *P*-values were determined by log-rank test. For sample information, see Supplementary Data 5. Source data are provided as a Source Data file.

CTLA-4 and CD69, using flow cytometry analysis. CTLA-4 is a well-known Treg suppressive marker by affecting the potency of antigen-presenting cells to activate other T cells⁵⁷. CD69 is mainly known for its early activation role involved in lymphocyte proliferation. Recent studies have shown the role of CD69 in enhancing the immunosuppressive function of Treg by promoting the production of IL-10⁵⁸. The flow cytometry results from 5 KP^{fl/c} mice show higher expression levels of both CTLA-4 and CD69 in KlrG1⁺ Treg subtype than KlrG1⁻ Treg subtype, which validates our conclusion regarding the immunosuppressive role of KlrG1⁺ Treg in murine PDAC (Fig. 6h and Supplementary Fig. 14e). Last but not the least, to predict the spatial location of KlrG1⁺ Treg, spatial deconvolution was performed utilizing the proteome data of the 8 KlrG1-associated T cell subtypes. The results showed that the KlrG1⁺ Treg cells were mainly enriched in the IT region rather than the LN region (Fig. 6i, j), indicating the recruitment of KlrG1⁺ Treg cells in the TME of PDAC.

Following the validation of the existence of KlrG1⁺ Treg and Tnfrsf18⁺ Treg in murine PDAC, we looked for the existence and biological significance of these cells in human PDAC. Successfully, we identified both KLRG1⁺ Treg and TNFRSF18⁺ Treg in human PDAC (Fig. 7a). Importantly, we also observed that the proportion of KLRG1⁺ Treg and TNFRSF18⁺ Treg increases as the tumor progression,

indicating the recruitment of these two cell types in the TME (Fig. 7a). Flow cytometry analysis also revealed higher expression levels of the Treg suppressive marker CTLA-4 in these two cell types (Fig. 7b). Importantly, the TNFRSF18⁺ Treg subtype was further validated by mIHC analysis using an independent tissue microarray (TMA) cohort ($n = 83$ pairs adjacent normal and tumor samples; Fig. 7c and Supplementary Data 5). A higher proportion of TNFRSF18⁺ Treg was observed in tumor samples than in the pair adjacent normal tissue samples in the cohort, which is consistent with the flow cytometry analysis results (Fig. 7d). We then further explored the association between the proportion of TNFRSF18⁺ Treg subset and patient survival using the TMA cohort data. Interestingly, unfavorable survival outcomes were explored for PDAC patients with a higher proportion of TNFRSF18⁺ Treg (Fig. 7e). The KLRG1⁺ Treg subtype was not further analyzed by mIHC due to the poor specificity of the commercially available antibodies. Collectively, these findings demonstrate the potential immunosuppressive features of these two Treg subtypes and well illustrate the powerfulness of the multimodal spatial proteomic investigation.

Discussion

Spatial proteomics is a rapidly developing field and becoming popular in the toolbox of “spatial multiomics” technologies. Compared with

antibody-based multiplexed protein visualization technologies^{3,4}, MS-based spatial proteomics addresses the increasing demand for systematic identification and quantitative analysis of proteins in the tissue microenvironment while preserving the spatial context information^{11,15}. However, it still suffers from relatively low cell-type resolution when investigating the various cell types in the TME with high heterogeneity. This limitation is primarily due to the lack of a widely available, highly efficient pipeline for accurately delineating and no-failure capturing single-cell contours from multiplex stained tissue slices. Additionally, there is a lack of highly efficient sample preparation for processing rare stained tissue cells with the clean-up steps, as well as a specialized data analysis pipeline to further enhance the cell-type resolution.

In this study, the SCPro platform seamlessly integrates antibody-guided cell typing technologies and high-sensitive proteome profiling for comprehensively uncovering the tissue proteome heterogeneity with single-cell resolution. The essential advances of the spatial proteomics aspect of the SCPro platform mainly include (1) the development of a widely accessible workflow for multi-color imaging-based cell segmentation, which enables accurate spatial proteome profiling of multiple cell types on a single FFPE tissue section without time-consuming manual annotation of cell contours as in previous studies; (2) the formation of reference shapes for aligning cell contours which enables navigation transfer between high-quality multiplexed imaging with coverslip and low-quality real-time imaging of LMD after removing the coverslip to enhance the cell-type resolution of LMD system; (3) the dissected cell contours were collected by the sticky-cap directly depressed on the membrane slide under visualization with no-failure, rather than collected by laser pulse or gravity in other LMD systems that may cause sample loss, especially when collecting cell contours with single-cell resolution. With these advantages, the SCPro significantly enhances the precision and throughput of LMD-based spatial proteomics.

Single-cell-resolved spatial proteomic analysis relies on nanoscale processing of limited stained FFPE tissue slice samples. The development of the iPAC technology in this study ensured the sensitivity and versatility of the SCPro platform. Recently, many integrated proteomics sample preparation methods have been proposed to match the increasing sensitivity of advanced MS instruments^{1,22}. Based on the minimized in-solution digestion concept, most of these methods showed excellent performance in processing flow cytometry-sorted single cells by integrating all sample preparation steps into one pot to minimize sample loss. MS-compatible reagents were used in these methods to avoid interfering MS signal, as they lack subsequent clean-up steps. However, these designs seldom fully considered the removal of chemical dyes and other non-protein containments from stained tissue samples, which are the natural components of this type of sample and the major difference with flow cytometry-sorted single cells. The SPE-based iPAC technology addressed these challenges by incorporating the extended wash after protein capture, resulting in enhanced protein coverage for limited stained tissue samples. Our results indicated the cleanness of peptides is equally important to its recovery rate. Notably, the iPAC technology also showed excellent performance in processing a low number of mLHC-stained cells which are LMD-dissected from the FFPE tissue section. Therefore, the iPAC technology is promised to meet the growing demands of spatial proteomic research.

Last but not least, we also develop a data analysis pipeline by integrating flow cytometry-based cell-type resolved proteome data to infer a more refined cell composition in distinct spatial locations. The deconvolution of spatial transcriptomics data is a routine strategy performed in spatial and single-cell transcriptomics data analysis to enhance cell-type resolution. However, it is still rarely reported for proteomics. The combination analysis of spatial and cell-type proteomic data shows the enhancement of the cell-type resolution of

SCPro and extends the data analysis pipeline for spatial proteomic study. The robustness of the SCPro was well demonstrated by revealing the spatial proteome changes of different types of neoplastic cells and immune cells in the TME of murine PDAC. Importantly, by seamlessly combining with flow cytometry-based cell-type proteomics dataset, subtypes of Treg were discovered and functionally validated in clinical samples.

Collectively, our study presents a streamlined proteomic workflow for advancing our understanding of tissue heterogeneity in the spatial context. The SCPro extends the comprehensiveness of traditional digital histopathology by incorporating the proteomic dimension. Along with future advancements in LC-MS instrumentation and data mining algorithms, the SCPro is expected to become a generic tool for the systematic characterization of spatiotemporal proteomic landscape and cell-cell interaction within TME at nanometer resolution.

Methods

Ethical statement

This study complies with all relevant ethical regulations. Animal work was approved by the Institutional Animal Care and Use Committee at the Southern University of Science and Technology of China. According to the requirements of the ethics committee, the maximal tumor volume allowed was 2000 mm³ and the maximal tumor size in this study was not exceeded. The human PDAC tissue samples for flow cytometry analysis were obtained from the Sun Yat-sen University Cancer Center, Sun Yat-sen University, Guangzhou, China. The Institutional Review Board of Sun Yat-sen University Cancer Center approved this study (No. SL-B2024-534-01). The human PDAC microarray was obtained from the Shanghai Outdo Biotech Company. The use of tissue microarray for research purposes was approved by the Ethics Committee of Shanghai Outdo Biotech Company (No. YBM-05-02). Written informed patient consent was obtained prior to the commencement of the study.

Cell lines

HEK 293T cell line (CRL-11268) was purchased from American Type Culture Collection (ATCC) and cultured in DMEM medium supplemented with 10% Fetal Bovine Serum (FBS) and incubated at 37 °C in 5% CO₂. The HEK 293T cell line was authenticated by the vendor and tested for mycoplasma contamination using the mycoplasma detection kit (Vazyme, D101-01) in accordance with the manufacturer's instructions. Cells were harvested following treatment with trypsin and washed three times with phosphate-buffered saline (PBS). The cells were resuspended by PBS, then sorted using the BD FACSAria SORP flow cytometer (BD Biosciences) and collected into 0.2 mL tubes (Axygen, PCR-02-C) using the one-way sorting and single-cell mode, at 4 °C. In order to reduce potential cell loss, the cells were centrifuged at 400 × g for 5 min at 4 °C after sorting, flash frozen down in liquid nitrogen, and stored at -80 °C freezer for further analysis.

Animal experiment and tissue preparation

Mouse brain was obtained from one male C57BL/6J mouse (8-week-old). The C57BL/6J mouse was purchased from the Jackson Laboratory. The mouse was euthanized by cervical dislocation, mouse brain was obtained and embedded in an Optimal Cutting Temperature (OCT) medium (Sakura Finetek USA) and sliced at -20 °C by using the CM 1900 cryostat (Leica). The sections were mounted onto a frame slide (MMI, 50103) and fixed by ice-cold methanol for 10 min. Then the fixed sections were stained with hematoxylin (Servicebio, G1004) and eosin (Servicebio, G1001), and dehydrated by a series of 70%, 80%, and 90% ethanol, each for 3 min, and then 100% ethanol twice.

Kras^{LSL-G12D/+}; Trp53^{fllox}; Pdx1-Cre mice (denoted as KP^{fl/c} mice) were purchased from the Jackson Laboratory, then bred and raised in the Animal Experiments Center at Southern University of Science and Technology. The genotyping of KP^{fl/c} mice was determined by routine PCR protocol using tail biopsies. PCR primers used for analyzing were listed: (1) Kras: 5'-GCAGGTGAGGACCTAATA-3'; 5'-CTGCATAGTACGCTATACCTGT-3'. (2) Cre: 5'-CCTGGACTACATCTTGAGTTGC-3'; 5'-AGGCAAATTTGGTGACGG-3'. (3) p53: 5'-GGTTAAACCCAGCTTGACCA-3'; 5'-GGAGGCAGAGACAGTTGGAG-3'.

To harvest the KP^{fl/c} tumors, the KP^{fl/c} mice were euthanized by cervical dislocation, the tumors were then removed, washed with ice-cold phosphate-buffered saline (PBS) three times, transferred to a 4% paraformaldehyde solution for fixing for 24–48 h. After fixation, the tissues were paraffin-embedded for further analysis. Two female KP^{fl/c} mice (7-week-old) were used for the benchmark of iPAC, and one male KP^{fl/c} mouse (7-week-old) were used for the benchmark of SCPro.

Sex was not considered in this study, as it focuses on technology development. No statistical methods were used to estimate sample size for animal studies throughout. All mice were housed at strict barrier facilities with macroenvironmental temperature and humidity ranges of 21–26 °C and 40–60%, respectively. Mouse rooms had a 12 h light and 12 h dark cycle. The housing conditions were closely monitored and controlled.

Multiplexed immunohistochemical staining of KPC tissues

Four-μm-thick tissue sections were cut using a microtome (Leica) and mounted onto the frame slides. To ensure optimal staining, the frame slides underwent a deparaffinization process by incubating for 10 min in 100% xylene three times, and rehydrated by a series of 100%, 100%, 90%, 80%, and 70% ethanol, each for 5 min, then the tissue sections were washed with ddH₂O for 5 min. The mIHC staining was conducted using the TSA kit (TissueGnostics, TGF7100) following the manufacturer's instructions. Primary antibodies, Anti-EpCAM (Cell Signaling Technology, clone E6V8Y, dilution 1:500), anti-CD45 (Cell Signaling Technology, clone D3F8Q, dilution 1:1000), anti-αSMA (Cell Signaling Technology, clone D4K9N, dilution 1:500), anti-Krt19 (Cell Signaling Technology, clone D4G2, dilution 1:1000), anti-CD3e (Cell Signaling Technology, clone D4V8L, dilution 1:100), and anti-CD11b (Cell Signaling Technology, clone E6E1M, dilution 1:1000) were used for staining KP^{fl/c} mouse tumor sections. Finally, the samples were mounted with DAPI Fluoromount-G® antifade mountant (SouthernBiotech, 0100-20) and coverslips to obtain high-quality images for further analysis.

Multiplexed immunohistochemical image acquisition and analysis of KPC tissues

Before image acquisition, the square reference shapes for image alignment were marked onto the membrane slide by LMD. The whole-slide image was first acquired by the TissueFAXS Spectra Systems (TissueGnostics) using a 5× objective to identify the location of the tissue section. Then the multiplexed whole-slide images were acquired at 40× high magnification, and grayscale images of high magnification were extracted for each dye channel for further analysis. The StrataQuest software version 7.1 (TissueGnostics) was used for the quantitative analysis and cell typing of the high-magnification images. For the cell typing of the KP^{fl/c} tumor tissue section, the nuclei identification algorithm was used to identify the nucleus. The parameter for nuclei size was set at 10 pixels. Then, the cell membrane identification algorithm was utilized to identify the cell boundaries of corresponding cell types. The parameters for cell membrane identification were set at −0.32 μm interior radius, 0.63 μm exterior radius, and 4 μm maximum growing step. After identifying the nuclei and cell membrane, a filled mask was generated over the original image of the corresponding cell type. Then, the cutting path was offset for about 1 μm. The square reference shapes

were then exported with the cell contours after cell typing to guide automated LMD.

Laser microdissection

The CellCut system (MMI) was used to collect cell contours. The mask files exported from the StrataQuest software were imported into the CellCut system to guided automated LMD. The clearly recognizable square reference shape was used to ensure the alignment of the cell contours generated from the StrataQuest software with the real-time image of the LMD under fluorescence mode. The cell contours were cut at 40× objective in brightfield mode. The cell contours were collected using the IsolationCap (MMI) and stored at −20 °C for further analysis. The LMD7 system (Leica) was used for the benchmark of cell contour collection efficiency by gravity-based collection.

Immunohistochemical staining and image analysis of KPC tissues

Four-μm-thick KP^{fl/c} tissue sections were cut and deparaffinized as described in multiplex immunohistochemical staining. For validating the expression level of Cwh43, anti-Cwh43 (Novus, NBP2-30438, 1:100) was used to stain KP^{fl/c} mouse tissue sections. The images were acquired by NanoZoomer S60 (Hamamatsu) and viewed using the NDP.Viewer platform (version 2.9.25). The average optical density (AOD) of Cwh43 in KP^{fl/c} tissues was calculated by the ImageJ software. The threshold for each image was set to (0.1033, 2.7076). Three regions were randomly selected from each KP^{fl/c} tissue slice to calculate the AOD for three different cell types.

Tumor dissociation and cell sorting of KPC tissues

To harvest the KP^{fl/c} tumors, the KP^{fl/c} mice were euthanized by cervical dislocation. The tumor resected from the KP^{fl/c} mouse was washed with ice-cold PBS to remove redundant fat and vessels. Then the tumor was minced into 2–4 mm pieces, and transferred into the gentleMACS C tube (Miltenyi, 130-096-335) with the enzyme mix solution from the tumor dissociation kit (Miltenyi, 130-096-730) prepared following the manufacturer's instructions. The tumor was dissociated using the gentleMACS™ dissociator (Miltenyi). The cell suspension was then filtered through a 70 μm cell strainer (Corning, 431751) and washed twice with ice-cold RPMI-1640 (Corning, 10-040-CMR) to obtain a single-cell suspension. The single-cell suspension was centrifuged at 300 × g for 5 min at 4 °C and the supernatant was completely aspirated. One milliliter of stain buffer (BD Pharmingen, 554657) was used to resuspend the cell precipitation. The cell number was counted and divided into three panels for cell staining to reduce sample loss. The detailed information about the antibodies used in flow cytometry and the panel design is shown in Supplementary Data 1 and 2. The cells were stained according to the manufacturer's instructions and sorted using the BD FACSAria SORP flow cytometer (BD Biosciences). The cells were collected into 1.5 mL Protein LoBind tubes (Eppendorf, 022431081) using the 4-way sorting and single-cell mode, at 4 °C. In order to reduce potential cell loss, the cells were centrifuged at 400 × g for 5 min at 4 °C after sorting, flash frozen down in liquid nitrogen, and stored at −80 °C freezer for further analysis. Two female KP^{fl/c} mice (9-week-old) and three male KP^{fl/c} mice (9-week-old) were used for the establishment of cell-type proteomics dataset of 14 cell types, three female KP^{fl/c} mice (9-week-old) were used for the cell-type proteomics analysis of 8 KlrG1-associated T cell subtypes, three female KP^{fl/c} mice (9-week-old) and two male KP^{fl/c} mice (9-week-old) were used for the flow cytometry analysis of the expression levels of CTLA-4 and CD69 in KlrG1⁺ and KlrG1[−] subtypes.

Patient sample dissociation and flow cytometry analysis

The normal, adjacent normal, and tumor tissue specimens were acquired from one treatment-naïve PDAC patient. The tissues were

minced into pieces and dissociated by tumor dissociation kit (Miltenyi, 130-095-929), then processed as we described in KPC tissues above to obtain the single-cell suspension for staining and flow cytometry analysis. The detailed information about the antibodies used in flow cytometry analysis and the patient information is shown in Supplementary Data 3 and 4, respectively.

Multiplexed immunohistochemical staining and analysis of PDAC TMA

The PDAC TMA ($n = 83$ pairs adjacent normal and tumor samples, 52 males and 31 females; diameter: 1.5 mm; sample information, see Supplementary Data 5) was purchased from Outdo Biotech (HPan-Ade180Sur-01). We performed the multiplexed immunohistochemical staining for Treg cells (CD4⁺FOXP3⁺ cells), TNFRSF18⁺ Treg cells (CD4⁺FOXP3⁺TNFRSF18⁺ cells), and epithelial cells (KRT19⁺ cells) using anti-CD4 (Cell Signaling Technology, clone EP204, dilution 1:100), anti-FOXP3 (STARTER, clone SDT-R064, dilution 1:1000), anti-TNFRSF18 (Cell Signaling Technology, clone D5V7P, dilution 1:100), and anti-KRT19 (Cell Signaling Technology, clone D4G2, dilution 1:1000) antibodies. The multiplexed images were acquired at 20× magnification. The procedures for image acquisition and analysis were consistent with those used for the KPC tissues, which have been described in detail above. No sex and gender analysis were performed in this study.

Sample preparation by iPAC

For tissue and sorted cell samples, 20 μ L of lysis buffer composed of 1% (w/v) DDM, 10 mM HEPES (pH 7.4), 150 mM NaCl, 600 mM guanidine HCl, and a protease inhibitor mixture (Roche) was used. The collected tissue slices were sonicated in the lysis buffer with the non-contact sonication using Bioruptor (Diagenode) for 20 cycles (30 s-on, 30 s-off) at 4 °C. Two plugs of C18 disks (3M Empore) and two plugs of SAX disks (3M Empore) were inserted into 200 μ L pipette tips to fabricate the iPAC spintips. Before sample loading, the tips were equilibrated with DDM coating buffer (0.1% DDM in NH₄OH) with a brief centrifuge, then the equal volume of tissue or cell lysates was mixed and loaded with the DDM coating buffer. Afterward, protein aggregation on the SAX disks was induced by loading and incubating in pure ACN for 10 min. The samples were then subjected to extended wash with 80% (v/v) ACN twice and the initial version of iPAC without the 80% (v/v) ACN wash step. The proteins were reduced using 50 mM dithiothreitol (DTT) in 20 mM ammonium bicarbonate (ABC) and incubated for 30 min at 37 °C. Specifically, 4 μ L of digestion buffer containing 20 ng/ μ L sequencing-grade trypsin (Promega), 20 ng/ μ L sequencing-grade Lys-C (Wako), and 10 mM iodoacetamide (IAA) in 20 mM ABC were added to the tips and incubated in darkness for 3 h at 37 °C for digestion. The digested peptides were then transferred to the C18 layer of the iPAC tip through 60 μ L 1 M NaCl in 1% (v/v) formic acid (FA). After desalting with 60 μ L 1% (v/v) FA twice, the resulting clean peptides were eluted by 60 μ L 80% (v/v) ACN into a glass insert. Peptides were lyophilized to dryness for MS analysis. The number of samples was $n = 3$ technical replicates for 5, 10, 20 ng HEK 293 T cell lysate, $n = 3$ biological replicates for 10–1000 HEK 293T cells, $n = 3$ biological replicates for 20, 50, 100 μ m-side length square tissue samples of 12 μ m-thick H&E-stained mouse brain, $n = 3$ biological replicates for 12 μ m-thick, 200 μ m-side length square H&E-stained mouse brain samples W/o or with W/ extended wash, $n = 4$ biological replicates for acinar, tumor, and lymph regions dissected from KP^{fl/c} mouse tissue section, $n = 4$ biological replicates for cancer and stroma cells dissected from KP^{fl/c} mouse tissue section, $n = 3$ biological replicates for acinar, PanIN, PDAC, CAF, IT, LN in spatial proteomics analysis, $n = 4$ biological replicates for the cell-type proteomics analysis of apCAF and $n = 5$ biological replicates for the other 13 cell types, $n = 3$ biological replicates for the 8 Klrg1-associated T cell subtypes.

High-pH reversed-phase chromatography fractionation

We conducted high-pH reversed-phase chromatography fractionation to generate a deep proteome library of KP^{fl/c} mouse tissues for data-independent analysis (DIA). Around 100 μ g of peptides from 5 KP^{fl/c} mice tumor sections were fractionated on an XBridge peptide BEH C18 column (130 Å, 5 μ m, 2.1 mm \times 150 mm) using a 60 min-gradient and concatenated into 24 fractions on a microflow HPLC (Agilent 1260). The peptide samples were vacuum dried in a SpeedVac (Thermo Fischer), then reconstituted in 0.1% FA spiked with iRT peptides (Biognosys) for LC–MS/MS analysis.

Liquid chromatography

The lyophilized peptides were reconstituted in 2.5 μ L of 0.1% (v/v) FA. Only 2 μ L of the redissolved peptides were injected for the single-shot LC–MS/MS analysis. A homemade 50 μ m I.D. \times 20 cm separation column with an integrated fritted tip was used by packing with 1.9 μ m C18 beads (Dr. Maisch) and coupling to a nanoElute liquid chromatography system (Bruker Daltonics). The temperature of the separation column was maintained at 50 °C using an integrated column oven. Mobile phases A and B consisted of 0.1% FA and ACN, respectively. A segmented 80-min gradient was used for LC–MS analysis. The gradient was set as follows: from 2 to 22% (v/v) buffer B in 50 min, from 22 to 35% (v/v) buffer B in 10 min, from 35 to 80% (v/v) buffer B in 10 min, holding at 80% (v/v) buffer B for the last 10 min.

Data acquisition in DDA mode and DIA mode

The timsTOF Pro (Bruker Daltonics) was used to analyze the eluted peptides. For DDA acquisition, the scan range was set to m/z 300–1500 in the positive mode. The ramp time was 200 milliseconds, and the total cycle time was 1.03 s with one MS scan and 4 parallel accumulation-serial fragmentation (PASEF) scans. The ion mobility ($1/K_0$) was scanned from 0.75 to 1.30 Vs/cm². For DIA acquisition, the dia-PASEF method was optimized using the py_diAID software⁵⁹ with the m/z range of 300–1500, the ion mobility range was set to 0.75–1.30 Vs/cm², and the ramp time was 200 milliseconds. Each dia-PASEF scan with variable isolation window widths that were adjusted according to the precursor densities. The optimized method includes two ion mobility windows and 12 dia-PASEF scans⁶⁰.

Raw data analysis

All the MS raw files for the iPAC benchmark were acquired in DDA mode and defaulted LFQ-MBR workflow of MSFragger software (version 3.7) integrated into the Fragpipe platform (version 19.0). The MS raw files were searched against the reviewed mouse UniProt FASTA database (21,984 entries). The spatial proteomic data of KP^{fl/c} mice were acquired in DIA mode and searched by Spectronaut (version 17.4). The MS raw files were searched against the reviewed mouse UniProt FASTA database (21,984 entries). All the raw files of flow cytometry-based proteomic data of KP^{fl/c} mice were acquired in DDA mode and searched by the MSFragger software (version 3.5) integrated into the Fragpipe platform (version 17.0). The MS raw files were searched against the reviewed mouse UniProt FASTA database (17,101 entries). Cysteine carbamidomethylation was set as a fixed modification, while N-terminal acetylation and deamidation at NQ were set as dynamic modifications. Trypsin and Lys-C were set as digestion enzymes and a maximum of two missed cleavages were allowed for all the raw files. Protein and peptide false discovery rate (FDR) levels were set to 1%.

All flow cytometry raw data were saved as .fcs files and processed by the FlowJo software (version 10, BD Biosciences) for further analysis. For the t-SNE plot analysis, the .fcs files were uploaded to the Cytobank (<https://www.cytobank.cn/>). The t-SNE-CUDA algorithm was used in the dimensionality reduction analysis. An equal number of 50,000 cells were randomly selected for the analysis of each sample. The perplexity parameter was set according to the default factory setting.

Bioinformatics and statistical analysis

The contamination ratios of the heatmaps related to Fig. 2f, g were defined as the ratio between the summed intensity of the up-left and bottom-right precursors in the heatmap. The white lines in the figures pass through points of (350 m/z, 0.81M) and (950 m/z, 1.31M). Up-left precursors were often regarded as +1 contaminants²⁷. Precursors were extracted by AlphaTims software (version 1.0.7)⁶¹. This calculation was performed using a custom R programming script. Furthermore, the total MS intensity at specific retention times and the distribution of PSM were calculated based on the CSV files exported from AlphaTims.

Proteomics data analysis and visualization were performed using Perseus software (version 2.0.9.0)⁶² and R (version 4.1.0). In this study, three spatial proteomics datasets and two cell-type proteomics datasets have been generated. For spatial proteomics dataset 1 and 2, especially for dataset 1 related to Fig. 2h–j and Supplementary Fig. 2 ($n = 4$ biological replicates for acinar, lymph, and tumor regions) and dataset 2 related to Supplementary Fig. 3 ($n = 4$ biological replicates for PCC and stroma cells), we applied a rigorous filtering criterion to retain only the quantified protein groups with at least three valid values in at least one group, resulting in 2168 proteins for dataset 1 and 2045 proteins for dataset 2. Missing values were imputed by sampling from a normal distribution tailored to the proteome abundance of each sample (width = 0.3; downshift = 1.8) using Perseus (version 2.0.9.0)⁶². Significance was calculated by one-way ANOVA for dataset 1 followed by permutation-based FDR for multiple hypothesis testing (FDR < 0.05). Significance was calculated by a two-tailed Student's *t*-test for dataset 2, average protein's abundance of each group was used to calculate the difference.

For spatial proteomics dataset 3 related to Fig. 4 and Supplementary Fig. 11a, b ($n = 3$ biological replicates for Acinar, PanIN, PDAC, CAF, IT, and LN regions), we filtered the quantified protein groups for at least two valid values in at least one group, resulting in 5822 proteins. Missing values were imputed the same as spatial proteomics dataset 1 and dataset 2. The significance of EpCAM⁺ cells ($n = 3$ biological replicates for Acinar, PanIN, and PDAC cells) was calculated by one-way ANOVA, and the median protein's abundance of one versus the rest group was used to calculate the difference. The significance of IT and LN ($n = 3$ biological replicates) was calculated by a two-tailed Student's *t*-test, median protein's abundance of each group was used to calculate the difference.

For cell-type proteomics dataset 1 related to Figs. 5b–g, 6b–e, Supplementary Fig. 12c, d, Supplementary Fig. 13b, c and Supplementary Fig. 14a ($n = 4$ biological replicates for apCAF, $n = 5$ biological replicates for the remaining 13 cell types) and dataset 2 related to Fig. 6f, g and Supplementary Fig. 14c, d ($n = 3$ biological replicates for 8 sub-cell types), we filtered the quantified proteins for at least two valid values in at least one cell type, resulting in 5900 proteins and 3777 proteins, respectively. Missing values were imputed by 0.1 of the minimum value of each protein. Significance and difference were calculated by LIMMA package (version 3.48.0) in R. To find enriched proteins of dataset 1, we used the following comparison strategies. PCC was compared with the other major cell lineage (CAF, T4, T8, B, and MYE); CAF was compared with the other major cell lineage (PCC, T4, T8, B, and MYE); sub-CAFs (iCAF, apCAF, and myCAF) were compared with each other; Lymphoid cells (T4, T8, Treg, and B) were compared with each other; Treg cells were ignored when compare T4 versus other cells; MYE cells were compared with the other major cell lineages (PCC, CAF, T4, T8, and B cells); MYE cells (MO, NEU, MAC, and DC) were compared with each other. We filtered the significantly enriched proteins with *P*-value < 0.05 and fold change > 2, resulting in 1259 proteins. Copy number of proteins for each cell type was estimated by “proteomic ruler” based on protein's intensity⁶³. GO Enrichment analysis for all datasets in this study was performed with clusterProfiler⁶⁴ (version 4.0.5) in R, and significance was calculated by hypergeometric distribution (*P*-value < 0.05).

To construct the membrane protein databases, we initiated by retrieving the reviewed mouse proteins sequence from Uniprot (17,119 entities), and retained 4487 proteins annotated with “transmembrane”. Then, Phobius⁴⁹ and DeepTMHMM⁵⁰ were individually used to predict transmembrane proteins with previously downloaded protein sequences. Proteins predicted as “transmembrane” by less than two approaches were discarded. Combined with reported surfaceome⁶⁵, the membrane protein databases encompass 4518 high-quality proteins. For cell-type proteomics dataset 1, a total of 1718 membrane proteins were identified. Specially, among these, 712 were classified as plasma membrane (PM) proteins, and 206 of these PM proteins exhibited significant alterations in their expression levels across 14 cell types. (related to Fig. 6b–d and Supplementary Fig. 14a).

To decipher the cell-type composition within the spatial proteomics data of pancreatic cancer as depicted in Figs. 5f, g and 6i, j, we applied the Tangram²⁰ (version 1.0.4) for deconvolution. This process involved using spatial proteomics dataset 3 as the input and cell-type proteomics datasets 1 and 2 as reference datasets. Before proceeding with the deconvolution, we retained the top 1000 highly variable proteins from log2-normalized cell-type proteomics datasets, without imputing missing values. The remaining parameters were left at their default settings during the deconvolution process to deduce the cell-type composition within the spatial proteomics data.

scRNA-seq dataset of KP^{trf}C mice related to Supplementary Fig. 13a–c ($n = 4$ biological replicates) was downloaded from the Gene Expression Omnibus (GSE129455)⁴². Ensemble IDs were converted to official gene symbols using the AnnotationDbi package (version 1.54.1), resulting in 13,813 genes for 11,261 cells. FindVariableFeatures function in R package Seurat (version 4.1.0)⁶⁶ was used to define 2000 highly variable genes. After calculating a cell cycle difference using the CellCycleScoring function, each feature was regressed against cell cycle difference and counts individually, and the resulting residuals were then scaled and centered with default parameters. Dimensionality reduction was processed using the RunPCA function with previously identified variable features. Harmony (version 0.1.0)⁶⁷, an algorithm iteratively corrects PCA embeddings, was used to correct each scRNA-seq sample. FindNeighbors and FindClusters functions were used with resolution = 0.5, resulting in 19 clusters. FindAllMarkers function was used to construct a marker genes panel with adjusting *P*-value < 0.05 and fold change > 2, which were used for annotation of known markers of cell-type labels⁴². The AverageExpression function was employed to derive the average RNA expression levels for each cell type. Subsequently, these average expression values were utilized to calculate the Pearson Correlation Coefficient with the corresponding copy number data from the cell-type proteomics dataset 1, focusing specifically on the 14 shared cell types between the two datasets.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The mass spectrometry proteomics data generated in this study have been deposited in ProteomeXchange via the iProX⁶⁸ partner repository with the dataset identifier [IPX0005555000](https://doi.org/10.26434/chemrxiv-2024-00000) and the dataset identifier [PXD038882](https://doi.org/10.26434/chemrxiv-2024-00000). The flow cytometry data generated in this study have been deposited in the Mendeley Data platform [<https://data.mendeley.com/datasets/7n6t858tf4/1>]. The raw mIHC imaging data of human PDAC tissue microarray generated in this study can be accessed at the BioStudies⁶⁹ database with the accession code [S-BIAD1424](https://www.ebi.ac.uk/biostudies/arrayexpress/studies/GSE129455). The publicly available scRNA-seq data was obtained from the Gene Expression Omnibus (GEO) database under the accession code [GSE129455](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129455). The processed scRNA-seq data was integrated and uploaded to the public repository⁷⁰. The remaining data are available within the Article,

Supplementary Information, or Source Data file. Source data are provided with this paper.

Code availability

Custom code used for figure generation is available on GitHub [<https://github.com/liyuan-bioinfo/SCPro>] and Zenodo [<https://doi.org/10.5281/zenodo.13978420>]⁷¹.

References

- Mund, A., Brunner, A. D. & Mann, M. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol. Cell* **82**, 2335–2349 (2022).
- Mao, Y., Wang, X., Huang, P. & Tian, R. Spatial proteomics for understanding the tissue microenvironment. *Analyst* **146**, 3777–3798 (2021).
- Wagner, J. et al. A single-cell atlas of the tumor and immune ecosystem of human breast cancer. *Cell* **177**, 1330–1345.e1318 (2019).
- Giesen, C. et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
- Schurch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* **182**, 1341–1359.e1319 (2020).
- Kelly, R. T. Single-cell proteomics: progress and prospects. *Mol. Cell. Proteom.* **19**, 1739–1748 (2020).
- Coscia, F. et al. A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J. Pathol.* **251**, 100–112 (2020).
- Li, L. et al. Spatially resolved proteomics via tissue expansion. *Nat. Commun.* **13**, 7242 (2022).
- Ma, M. et al. In-depth mapping of protein localizations in whole tissue by micro-scaffold assisted spatial proteomics (MASP). *Nat. Commun.* **13**, 7736 (2022).
- Piehowski, P. D. et al. Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100-um spatial resolution. *Nat. Commun.* **11**, 8 (2020).
- Eckert, M. A. et al. Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723–728 (2019).
- Wisniewski, J. R., Ostasiewicz, P. & Mann, M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J. Proteome Res.* **10**, 3040–3049 (2011).
- Xu, R. L. et al. Spatial-resolution cell type proteome profiling of cancer tissue by fully integrated proteomics technology. *Anal. Chem.* **90**, 5879–5886 (2018).
- Marakalala, M. J. et al. Inflammatory signaling in human tuberculosis granulomas is spatially organized. *Nat. Med.* **22**, 531–538 (2016).
- Grünwald, B. T. et al. Spatially confined sub-tumor microenvironments in pancreatic cancer. *Cell* **184**, 5577–5592.e5518 (2021).
- Leng, L. et al. Spatial region-resolved proteome map reveals mechanism of COVID-19-associated heart injury. *Cell Rep.* **39**, 110955 (2022).
- Mao, Y. et al. Deep spatial proteomics reveals region-specific features of severe COVID-19-related pulmonary injury. *Cell Rep.* **43**, 113689 (2024).
- Mund, A. et al. Deep visual proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231 (2022).
- Thielert, M. et al. Robust dimethyl-based multiplex-DIA doubles single-cell proteome depth via a reference channel. *Mol. Syst. Biol.* **19**, e11503 (2023).
- Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
- Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
- Ye, X. et al. Integrated proteomics sample preparation and fractionation: method development and applications. *Trends Anal. Chem.* **120**, 115667 (2019).
- Chen, W. et al. Simple and integrated spintip-based technology applied for deep proteome profiling. *Anal. Chem.* **88**, 4864–4871 (2016).
- Martin, K. et al. Facile one-pot nanoproteomics for label-free proteome profiling of 50–1000 mammalian cells. *J. Proteome Res.* **20**, 4452–4461 (2021).
- Batth, T. S. et al. Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Mol. Cell. Proteom.* **18**, 1027–1035 (2019).
- Yang, Y. et al. Fritted tip capillary column with negligible dead volume facilitated ultrasensitive and deep proteomics. *Anal. Chim. Acta* **1201**, 339615 (2022).
- Meier, F. et al. Online Parallel Accumulation–Serial Fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteom.* **17**, 2534–2545 (2018).
- Halbrook, C. J., Lyssiotis, C. A., Pasca di Magliano, M. & Maitra, A. Pancreatic cancer: advances and challenges. *Cell* **186**, 1729–1754 (2023).
- Shi, Y. et al. Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. *Nature* **569**, 131–135 (2019).
- Gopinathan, A., Morton, J. P., Jodrell, D. I. & Sansom, O. J. GEMMs as preclinical models for testing pancreatic cancer therapies. *Dis. Model Mech.* **8**, 1185–1200 (2015).
- Hingorani, S. R. et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
- Blenman, K. R. M. & Bosenberg, M. W. Immune cell and cell cluster phenotyping, quantitation, and visualization using in silico multiplexed images and tissue cytometry. *Cytom. A* **95**, 399–410 (2019).
- Tosti, L. et al. Single-nucleus and in situ RNA-sequencing reveal cell topographies in the human pancreas. *Gastroenterology* **160**, 1330–1344.e1311 (2021).
- Wallace, D. C. Mitochondria and cancer. *Nat. Rev. Cancer* **12**, 685–698 (2012).
- Hwang, W. L. et al. Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. *Nat. Genet.* **54**, 1178–1191 (2022).
- Cappellesso, F. et al. Targeting the bicarbonate transporter SLC4A4 overcomes immunosuppression and immunotherapy resistance in pancreatic cancer. *Nat. Cancer* **3**, 1464–1483 (2022).
- Jurcak, N. R. et al. Axon guidance molecules promote perineural invasion and metastasis of orthotopic pancreatic tumors in mice. *Gastroenterology* **157**, 838–850.e836 (2019).
- Chen, R. & Chen, L. Solute carrier transporters: emerging central players in tumour immunotherapy. *Trends Cell Biol.* **32**, 186–201 (2022).
- Zhuang, H. et al. Characterization of the prognostic and oncologic values of ITGB superfamily members in pancreatic cancer. *J. Cell. Mol. Med.* **24**, 13481–13493 (2020).
- Ji, Y. F., Huang, H., Jiang, F., Ni, R. Z. & Xiao, M. B. S100 family signaling network and related proteins in pancreatic cancer (Review). *Int. J. Mol. Med.* **33**, 769–776 (2014).
- Lee, C. C. et al. CWH43 is a novel tumor suppressor gene with negative regulation of TTK in colorectal Cancer. *Int. J. Mol. Sci.* **24**, 15262 (2023).
- Elyada, E. et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019).

43. Zhang, Y. et al. Regulatory T-cell depletion alters the tumor microenvironment and accelerates pancreatic carcinogenesis. *Cancer Discov.* **10**, 422–439 (2020).
44. Biffi, G. & Tuveson, D. A. Diversity and biology of cancer-associated fibroblasts. *Physiol. Rev.* **101**, 147–176 (2021).
45. Cao, L. et al. Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell* **184**, 5031–5052.e5026 (2021).
46. Hegde, S. et al. Dendritic cell paucity leads to dysfunctional immune surveillance in pancreatic cancer. *Cancer Cell* **37**, 289–307.e289 (2020).
47. Nielsen, S. R. et al. Suppression of tumor-associated neutrophils by lorlatinib attenuates pancreatic cancer growth and improves treatment with immune checkpoint blockade. *Nat. Commun.* **12**, 3414 (2021).
48. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).
49. Kall, L., Krogh, A. & Sonnhammer, E. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**, W429–W432 (2007).
50. Hallgren, J. et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. Preprint at *bioRxiv*, <https://doi.org/10.1101/2022.1104.1108.487609> (2022).
51. Plitas, G. & Rudensky, A. Y. Regulatory T cells in cancer. *Annu. Rev. Cancer Biol.* **4**, 459–477 (2020).
52. Beyersdorf, N., Ding, X., Tietze, J. K. & Hanke, T. Characterization of mouse CD4 T cell subsets defined by expression of KLRG1. *Eur. J. Immunol.* **37**, 3445–3454 (2007).
53. Shimizu, J., Yamazaki, S., Takahashi, T., Ishida, Y. & Sakaguchi, S. Stimulation of CD25(+)CD4(+) regulatory T cells through GITR breaks immunological self-tolerance. *Nat. Immunol.* **3**, 135–142 (2002).
54. von Moltke, J., Ayres, J. S., Kofoed, E. M., Chavarria-Smith, J. & Vance, R. E. Recognition of bacteria by inflammasomes. *Annu. Rev. Immunol.* **31**, 73–106 (2013).
55. Wei, H. et al. KDEL2 promotes breast cancer proliferation via HDAC3-mediated cell cycle progression. *Cancer Commun.* **41**, 904–920 (2021).
56. Si, J. et al. Hematopoietic progenitor kinase1 (HPK1) mediates T cell dysfunction and is a druggable target for T cell-based immunotherapies. *Cancer Cell* **38**, 551–566.e511 (2020).
57. Wing, K. et al. S. CTLA-4 control over Foxp3+ regulatory T cell function. *Science* **322**, 271–275 (2008).
58. Yu, L. et al. CD69 enhances immunosuppressive function of regulatory T-cells and attenuates colitis by prompting IL-10 production. *Cell Death Dis.* **9**, 905 (2018).
59. Skowronek, P. et al. Rapid and in-depth coverage of the (phospho-) proteome with deep libraries and optimal window design for diaPASEF. *Mol. Cell. Proteom.* **21**, 100279 (2022).
60. Meier, F. et al. diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).
61. Willems, S., Voytik, E., Skowronek, P., Strauss, M. T. & Mann, M. AlphaTims: indexing trapped ion mobility spectrometry–TOF data for fast and easy accession and visualization. *Mol. Cell. Proteom.* **20**, 100149 (2021).
62. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
63. Wisniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A. proteomic ruler” for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteom.* **13**, 3497–3506 (2014).
64. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
65. Bausch-Fluck, D. et al. The in silico human surfaceome. *Proc. Natl Acad. Sci. USA* **115**, E10988–E10997 (2018).
66. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
67. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
68. Chen, T. et al. iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.* **50**, D1522–D1527 (2022).
69. Sarkans, U. et al. The BioStudies database—one stop shop for all data supporting a life sciences study. *Nucleic Acids Res.* **46**, D1266–D1270 (2018).
70. Xu, Y. et al. Multimodal single cell-resolved spatial proteomics reveal pancreatic tumor heterogeneity. *Zenodo* <https://doi.org/10.5281/zenodo.13981062> (2024).
71. Xu, Y. et al. Multimodal single cell-resolved spatial proteomics reveal pancreatic tumor heterogeneity. *Zenodo* <https://doi.org/10.5281/zenodo.13978420> (2024).

Acknowledgements

R.T. is approved by the funding from the China State Key Basic Research Program Grants (2021YFA1301602, 2022YFC3401104, 2021YFA1301601, 2021YFA1302603, and 2020YFE0202200), the National Natural Science Foundation of China (92253304, 22125403, 32201218, and 22104047), the Shenzhen Innovation of Science and Technology Commission (JSGGZD20220822095200001, JCYJ20200109141212325, JCYJ20210324120210029, JCYJ20200109140814408, and KJZD20230923114220041), Shenzhen Science and Technology Program (ZDSYS20230626090803004) and Guangdong province (2019B151502050).

Author contributions

Y.X. investigated, performed experiments, analyzed data, and wrote the paper. X.W. performed experiments, analyzed data, and wrote the paper. Y.L. conducted the bioinformatic analysis and wrote the paper. Yiheng.M. performed the experiments, analyzed data, and wrote the paper. Yiran.S., Yize.M. and Y.Y. experiment design and performed experiments. W.G. provided advice for the data analysis and data organization. C.F. provided the necessary reagents and gave advice for the experiment design. W.C. gave advice for the experiment design. X.Y., F.L. and P.B. gave advice for the experiment design and technical support. Ying.S. and R.X. gave advice for the experiment design. S.L. provided experiment materials. R.T. conceived and supervised the project, designed the experiment, and wrote the paper.

Competing interests

R.T. is the founder of BayOmics, Inc. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54438-0>.

Correspondence and requests for materials should be addressed to Ruijun Tian.

Peer review information *Nature Communications* thanks Maarten Bijlsma, Oliver Crook, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2024