

Supplementary methods

Human subjects

For the primary cohort, we enrolled 2 healthy donors, 2 patients with treatment-naïve chronic pancreatitis (CP), and 5 patients with PDAC. Additional 138 PDAC patients were involved in the validation cohort (24 patients in validation cohort 1, and 114 patients in validation cohort 2). All the PDAC patients had untreated, resectable, non-metastatic pancreatic tumors that were confirmed to be PDAC according to pathologist assessment, and the patients with infectious diseases, rheumatic diseases, or other malignancies were excluded. In addition, blood samples from 6 healthy donors were collected for *in vitro* immune functional assays. The samples were obtained from Ruijin Hospital, Shanghai Jiaotong University School of Medicine. The study protocol was approved by the Ethics Committee of Ruijin Hospital affiliated to School of Medicine, Shanghai Jiaotong University. All enrolled participants consented to attend this cohort study and signed written informed consent. It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research

Sample collection, processing, and purity assessment

4ml peripheral blood was collected using EDTA anticoagulant tubes prior to surgery, and processed within 15 minutes after collection. Red blood cells were depleted by gravity sedimentation through HetaSep (StemCell). Peripheral blood leukocytes were harvested and labeled with CD66b-phycoerythrin (PE) monoclonal antibody (BioLegend, 305106), and then neutrophils were isolated by anti-PE beads and MACS column (Miltenyi), according to manufacturer's protocol.

For PDAC tumor tissues, following resection in the operating room, small tissue blocks from ≥ 3 different sites of tumor were collected and immersed in tissue storage solution (Miltenyi), and transported to the research facility on ice immediately. On arrival, tissues were rinsed with PBS, and necrotic foci, hemorrhagic foci, and blood vessels were removed. Tissues were then minced into small pieces (< 1 mm in diameter), and then transferred into 2ml digestion medium containing 100ul enzyme H, 50ul enzyme R, and 12.5ul enzyme A (Tumor Dissociation Kit, Miltenyi, 130-095-929) in RPMI-1640. Tissues were enzymatically digested on a shaker at 37°C for 20min. The dissociated cells were collected at the interval of 10min to increase viability, and filtered through a 40 μ m nylon cell strainer to harvest single-cell suspension. Dead cells were removed by dead cell removal kit (Miltenyi). For the primary cohort, CD45⁺ immune cells were isolated using CD45 microbeads (Miltenyi), according to manufacturer's protocol. For validation cohort 1, CD66b⁺ neutrophils were isolated using the same protocol as that for peripheral blood. Throughout the dissociation and isolation procedure, cells were maintained on ice whenever possible.

For the neutrophils isolated from peripheral blood or tumor tissues using CD66b-PE monoclonal antibody and anti-PE beads, the purity was assessed by flow cytometry performed with Beckman CytoFlex S and analyzed using FlowJo software.

Single-cell RNA-sequencing

Single-cell RNA sequencing was performed on the single-cell suspensions with viability $> 70\%$, including CD66b⁺ PMNs from peripheral blood and CD45⁺ immune filtrates from tumor tissues. Single-cell capture was achieved by BD Rhapsody system via a limited dilution approach, in which $\sim 10,000$ cells were randomly distributed across $> 200,000$ microwells. Beads with oligonucleotide barcodes were added to microwells so that each bead was paired with a cell. Cells were lysed in microwells, and poly-adenylated RNA molecules were hybridized to the beads. All the beads were collected from microwells into a single tube. Reverse transcription was performed, and each cDNA molecule was tagged on the 5' end with a unique molecular identifier (UMI) and a cell label barcode. Whole transcriptome libraries were prepared according to the BD Rhapsody single-cell whole-transcriptome amplification workflow. In brief, second strand cDNA was synthesized, and WTA adaptor was ligated for universal amplification. The adaptor-ligated cDNA products were amplified by 18 cycles of PCR. Random priming PCR of the whole-

transcriptome amplification products was performed to enrich the 3' end of the transcripts linked with the UMI and barcode. Sequencing libraries were then quantified by a High Sensitivity DNA chip (Agilent) on a Bioanalyzer 2200 and the Qubit High Sensitivity DNA assay (Thermo Fisher Scientific), and sequenced using HiSeq Xten (Illumina, San Diego, CA) on 150 bp paired-end run.

Single-cell RNA-seq data processing and quality control

Raw data was processed using fastp with default parameter to filter adaptor sequences and remove low-quality reads [1]. Cell barcode whitelist was identified by UMI-tools [2]. The UMI-based clean data was mapped to human reference genome (Ensemble version 91) using STAR algorithm [3] with customized parameter from UMI-tools standard pipeline. UMI count matrices were generated for each sample, and imported into Seurat R toolkit (version 3.2.3) [4]. Low quality cells (<200 genes/cell or >20% mitochondrial genes) were excluded.

Unsupervised clustering, marker identification, and cell type annotation

Seurat package was applied for normalization and scaling of the expression matrix, using default settings. Mitochondrial contamination was regressed out by setting “vars.to.regress” parameter. In the process of subclustering PMNs and TANs, batch effect was removed with ComBat [5]. To reduce the dimensionality of the expression matrix, principal component analysis (PCA) was performed based on 2,000 highly variable genes. Unsupervised cell clusters were acquired by graph-based clustering approach (The top 30 PCs were selected, resolution = 0.5), and visualized by UMAP dimensionality reduction. The clusters were annotated to known biological cell types according to the expression of canonical markers. Marker genes of each cluster were identified by FindAllMarkers function under the following criteria: logFC > 0.25; min.pct > 0.25; adjusted P value < 0.05.

Comparison of single-cell RNA profiles between PMNs from healthy controls, CP patients and PDAC patients, and TANs from PDAC tumor tissues

The averaged expression profiles of PMNs and TANs from each sample was visualized by Barnes-Hut implementation of t-distributed stochastic neighbor embedding (t-SNE) using Rtsne package [6] (1000 iterations, perplexity = 4, trade-off θ = 0.5). To compare the expression profiles of neutrophils, differentially expressed genes were identified using “FindMarkers” function (logFC > 1; min.pct > 0; adjusted P value < 0.05) implemented in the Seurat package, and pathway analysis was performed by gene set enrichment analysis (GSEA) [7].

Definition of cell signatures

To investigate the functions of each neutrophil subcluster, we selected the gene sets associated with neutrophil functions (Table S10), and the expression levels of each gene set were estimated for each individual cell with gene set variation analysis (GSVA) [8] implemented in the GSVA package. Linear models analyzing difference between neutrophils from one cluster and neutrophils from all other clusters were generated for comparison of pathway activities between TAN subclusters. The metabolism signature of each cell was defined as the average normalized expression of the genes involved in hallmark metabolic pathways.

Constructing single-cell trajectories of neutrophils

Pseudotime transitional trajectory of neutrophils was constructed utilizing Monocle2 package (version 2.18.0) with default parameters [9]. Top 2,000 highly variable genes in neutrophils were selected as input, and dimensionality reduction was performed by “DDRTree” method. Differentially expressed genes along the pseudotime trajectory were identified by the “differentialGeneTest” function with a q-value < 10^{-20} , and visualized by “plot_pseudotime_heatmap” function. Pathway enrichment analysis was performed for the differentially expressed genes with Database for Annotation, Visualization and Integrated Discovery (DAVID).

Cell-cell communication analysis

CellPhoneDB (<http://www.cellphonedb.org/>) was applied to analyze cell-cell communication between different cell types in tumor microenvironment [10], based on the normalized expression matrix from Seurat. Mean expression of each receptor-ligand pair was calculated as the mean of the average expression of receptor in one cluster and the average expression of ligand in the other cluster, and the p-value indicates the cell type-specificity of the crosstalk.

SCENIC analysis

To construct gene regulatory networks that predicts the association between transcription factors and target genes in TANs, SCENIC analysis [11] was run with default settings, using the 20-thousand motifs database for RcisTarget and GRNboost. The activities of regulons (transcription factors and target genes) were calculated with AUCell. Unsupervised cluster analysis of regulon activities was performed by graph-based clustering approach (The top 20 PCs were selected, resolution = 0.5), and visualized by t-SNE dimensionality reduction.

Bulk RNA sequencing and data analysis

Isolated PMNs and TANs were lysed with TRIzol (Invitrogen), and total RNA was extracted with RNeasy Mini Kit (QIAGEN, Germany), according to the manufacturer's protocol. RNA sequencing libraries were generated with KAPA RNA library preparation kit (Illumina) and KAPA dual-indexed adapter kit (Illumina), and sequenced on Illumina HiSeq X Ten platform (2 x 150 bp). The raw RNA-Seq reads were aligned to the human reference genome hg38 using STAR (v2.7.0d) [3]. The human reference was downloaded from GENCODE (<https://www.encodegenes.org/>). Name-sorted and indexed BAM files were generated by Samtools (v1.8-47) [12]. To quantify gene expression levels, Salmon (v0.13.1) [13] was used to determine transcripts per kilobase million (TPM) value. Transcript sequences (v34) and the annotation file were downloaded from GENCODE. We used fragments per kilobase million (FPKM) to evaluate expression levels of individual genes. The R package limma [14] was used to calculate differentially expressed genes between PMNs and TANs, and pathway analysis was performed by gene set enrichment analysis (GSEA) [7].

Proteomic assay

Protein extraction and digestion

1x10⁶ CD66b+ cell pellets were resuspended in 150 µL 1% SDS lysis solution (Beyotime, P0013G) for 15 min on ice, vortexed at the highest speed for 10 seconds, then incubated for an additional 15 min, followed by centrifugation for 10 min at 13500g and 4°C. The supernatant was collected and detected protein concentration by BCA assay according to the manufacturer's instructions (ThermoFisher, 23225). Extracted proteins were reduced in 10 mM dithiothreitol (ThermoFisher, 20290) at 56°C for 30 min, then alkylated in 15 mM iodoacetamide (BBI Life Sciences, A600539) at room temperature for 30 min in darkness. The protein was precipitated with acetone at -20°C for 4 hours. Protein samples was digested by trypsin at 37°C for 16 hours. Peptide mixture was desalted by SepPak C18 cartridge (Waters, 186002319) and vacuum-dried by SpeedVac (ThermoFisher).

Nano-LC-MS/MS

EASY-nLC 1200 coupled to Orbitrap QExactive HFX (Thermo Fisher Scientific) were used for proteome analysis. iRT peptides (Biognosys AB) were spiked-in samples according to the manufacturer's instructions (Biognosys AB). Buffer A: 0.1 % Formic acid in water; buffer B: 0.1 % Formic acid in 80 % ACN. Sample peptides were resolved in buffer A, loaded onto a PepMap C18 analytical column (2 µm particle, 250 mm × 75 µm NVFS 1200 bar, ThermoFisher Scientific, 164941) and separated at a flow rate of 300 nL/min with 120 min gradient. The mass spectrometer was operated under data-independent acquisition (DIA) mode with multiple MS1 scans interspersed with 20 MS/MS scans per scan cycle. Precursor ion signals was quantified by full MS scans with ion target value of 3 × 10⁶ charges in the 350–1,250 m/z range, maximum injection time of 60 ms and resolution of 120,000; the MS2 scans with ion target value of 1 × 10⁶ charges for the precursor window, and resolution of 30,000. MS/MS isolation windows were set 9 m/z -25 m/z unit.

Data analysis

All DIA raw files were processed directDIA analysis with Spectronaut software version 14.11.210528.47784 (Biognosys, Zurich, Switzerland). Database selected the human UniProt reference proteome with 20,379 sequences downloaded in May 2020. Both MS1 and MS2 data were used for peptide identification while the parameters of the quantification process were solely derived from the MS1 data. Trypsin/P proteolytic cleavage rule was used, allowing up to two miscleavages and a peptide length of 7–52 amino acids. Carbamidomethylation set as a fixed modification, acetylation of the protein N-terminus and oxidation of methionines as variable modifications. Protein intensities were normalized using the “Global Normalization”. The identification was carried out using a kernel density estimator and false discovery rate (FDR) cut-off 0.01 for precursor and protein levels.

Glycolysis/gluconeogenesis metabolites assay

Sample preparation

1x10⁶ CD66b⁺ cell pellets were mixed with 1 ml ice-cold 50% methanol water solution in 1.5 mL of Eppendorf tube, vortexed for 30 s, incubated in liquid nitrogen and thawed on ice, followed by repetition vortex–freeze–thaw cycle until cells were completely broken. For protein precipitation, the samples was put in the –20 °C refrigerator for 1 h and then centrifuged for 20 min at 16300g and 4°C. The supernatant was collected and lyophilized, then redissolved in methanol/acetonitrile/ water solution (2:2:1). Vortexed until all the precipitates were dissolved. The solutions were centrifuged for 5 min at 16300g and 4 °C, the supernatants were transferred to the loading bottles for LC-MS/MS analysis.

Multiple reaction monitoring (MRM) analysis by HPLC-QTRAP/MS

A HPLC System (SHIMADZU LC-30A, JAPAN) coupled with triple quadrupole mass spectrometer (SCIEX QTRAP6500, USA) were used to establish a new quantitative method for 16 glycolysis/gluconeogenesis metabolites. The glycolysis/gluconeogenesis metabolite library was purchased from Sigma (Sigma, ML0013). 2 µL sample was separated on an Acquity UPLC®BEH HILIC column (2.1 × 100 mm I.D., 1.7 µm) with solvent A (20 mM of ammonium acetate in 95% H₂O/ACN, pH = 9) and solvent B (20 mM of ammonium acetate in 95% ACN/H₂O, pH = 9). The column temperature was kept at 40 °C. The total flow rate was 0.3 mL/min. The gradient was: t = 0.0 min, 90% B solvent; t = 0.5 min, 90% B solvent; t = 7.0 min, 10% B solvent; t = 8 min, 10% B solvent; t = 9 min, 90% B solvent; t = 10.0 min, 90% B solvent. The instrument was operated in MRM mode, the parameters were listed as follows: temperature of electrospray ion source was 400 °C; capillary voltage was 4500 V (positive mode) and -5500 V (negative mode). Declustering Potential (DP), Entrance Potential (EP), Collision Energy (CE), Collision Cell Exit Potential (CXP) were optimized for each metabolite by FIA mode in both positive and negative mode.

Data analysis

LC-MRM MS data were acquired in Analyst 1.6 software (AB SCIEX, Framingham, U.S.A.), quantitative analysis was performed by MultiQuant 3.0.1 (AB SCIEX, Framingham, U.S.A.). The metabolites with RSD below 30% were used for further analysis. Nonparametric Mann–Whitney U test performed in SPSS 18 (SPSS, Chicago, IL) was applied to evaluate the statistic alterations of metabolites' level among two groups with p value set at 0.05.

Immunohistochemistry (IHC) staining of tissue microarrays

PDAC tumor tissues were obtained from surgical resection, and fixed with 4% formalin immediately. Tissue microarrays were constructed from paraffin-embedded tissue specimens (three 1mm cores of different regions from each tissue). IHC staining was performed according to following protocols: Tissue slides were de-paraffinized with xylenes and graded ethanol. Antigen retrieval was performed in 0.01M citrate buffer (pH 6.0), heated in microwave. After blocking endogenous peroxidase with 3% H₂O₂ solution and blocking nonspecific binding with 5% BSA, the slides were incubated with primary antibodies (as listed below), followed by HRP-conjugated secondary antibodies (Dako), and diaminobenzidine (DAB, Dako). The slides were then counterstained with hematoxylin (Dako). The antibodies used are listed as following: CD66b (BioLegend, 305102); VEGFA (Proteintech, 19003-1-AP); NLRP3

(Proteintech, 19771-1-AP); MME (Proteintech, 18008-1-AP); IFIT2 (Proteintech, 12604-1-AP); BHLHE40 (Novus Biologicals, NB100-1800).

We analyzed the percentage of CD66b (+) and TAN subcluster marker (+) double-positive cells (assessed based on serial sections) among total cells in 10 high power fields of each core, and calculated the average proportion of double-positive cells in three tissue cores from each patient. Therefore, the patients were divided equally into two groups based on the infiltration level of each TAN subcluster, and its association with clinical parameters was further investigated.

Immunofluorescence (IF) staining of PDAC tissues

PDAC tissues were embedded in optimal cutting temperature (OCT) medium and immediately frozen at -80°C. Frozen tissues were cut into 6µm thick sections with cryostat, and layered on gelatin-coated slides. The slides were blocked with 10% normal goat serum for 1h at room temperature, and then incubated with primary antibodies at 4°C overnight and washed with PBS, followed by incubation with fluorochrome-conjugated secondary antibody for 2hr at room temperature. After washing with PBS, the slides were mounted in aqueous mounting media with DAPI. The tissues slides were visualized with confocal microscopy (Zeiss LSM 900). The following antibodies were used for IF staining: CD66b (BioLegend, 305102); CD68 (Servicebio, GB113150); p65 (abcam, ab16502); c-JUN (Proteintech, 24909-1-AP); VEGFA (Proteintech, 19003-1-AP); NLRP3 (Proteintech, 19771-1-AP); MME (Proteintech, 18008-1-AP); IFIT2 (Proteintech, 12604-1-AP); BHLHE40 (Novus Biologicals, NB100-1800); goat-anti-rabbit (Servicebio, GB21303); goat-anti-mouse (Servicebio, GB22301).

Flow cytometry analysis of PDAC tissues

Single-cell suspensions from PDAC tissues were obtained as described above. For staining of cell surface markers, the cells were incubated with Zombie NIR™ fixable viability dye (BioLegend, 423105) and Fc block (BioLegend) for 15min before staining with fluorochrome-conjugated antibodies for 30 min at 4°C. For intracellular staining, the cells were fixed and permeabilized with BD Cytofix/Cytoperm™ Fixation/Permeabilization Kit according to manufacturer's protocol, incubated with primary antibody for 30min at 4°C, followed by PE-conjugated secondary antibody for 30min at 4°C. Flow cytometry analysis was performed with Beckman CytoFlex S or Fortessa X-20 flow cytometer. Flow cytometry staining was performed with the following antibodies: anti-CD45 (BD, 563792); anti-CD15 (BioLegend, 323006); anti-CD66b (BD, 564679); anti-CD66b (BioLegend, 305118); anti-LGALS3 (BioLegend, 125418); anti-GLUT1 (abcam, ab115730); anti-HK2 (abcam, ab209847); anti-PFKFB3 (abcam, ab181861); anti-LDHA (Proteintech, 19987-1-AP); and anti-rabbit (abcam, ab72465).

Analysis of public RNA-seq data

Gene expression data and corresponding clinical information of TCGA-PAAD cohort were downloaded from UCSC Xena. Gene expression profiles of the published dataset by Cao et al. were downloaded from Genomic Data Commons. The signature scores of neutrophils/ macrophages/ TAN-1 in each sample were calculated by GSVA, according to the marker genes identified in single-cell sequencing (logFC > 1 and adjusted p value < 0.05). The treatment-naïve patients with standard PDAC histology were divided equally into two groups based on the expression of TAN-1 signature, and Kaplan-Meier survival curves and log-rank tests were used to compare overall survival in different patient groups. Pearson correlation analysis was performed to assess the association between the expression of neutrophil signature and macrophage signature.

Spatial transcriptomics

PDAC tissue was embedded in OCT compound on dry ice and stored at -80°C. PDAC tissue was cut into 6µm thick sections with cryostat, and layered within the frames of capture areas on Visium Spatial slides (10X Genomics). After fixing with methanol and staining with hematoxylin and eosin (H&E), the slide was scanned and visualized using Pannoramic MIDI (3DHISTECH). Optimization of permeabilization time was performed using Visium Spatial Tissue

Optimization Slides & Reagent Kits (10X Genomics), according to manufacturer's instructions. Tissue permeabilization and spatial transcriptomic sequencing were performed with Visium Spatial Gene Expression Slides & Reagent Kits. Briefly, after permeabilization, the slide was incubated in RT Master Mix for 45 min at 53°C, followed by incubation with Second Strand Mix for 15 min at 65°C. Barcoded cDNA was purified and amplified, and then fragmented, A-tailed, ligated with adaptors and index PCR amplified. The concentration and size distribution of Visium library were analyzed with Qubit High Sensitivity DNA assay (Thermo Fisher Scientific) and High Sensitivity DNA chip on a Bioanalyzer 2200 (Agilent). The library was sequenced by Illumina sequencer on a 150 bp paired-end run.

Raw data was processed using fastp with default parameter to filter adaptor sequences and remove low-quality reads [1]. The clean data was mapped to human genome (GRCh38) using SpaceRanger v1.1.0 to obtain feature-barcode matrix. Seurat package was applied for normalization and SCTransform. Neutrophil signatures of each spatial spot were calculated with ssGSEA [15] based on the marker genes of neutrophil cluster identified from single-cell RNA profiles of PDAC tissues, and the spots with top 10% neutrophil signatures (ssGSEA score >0.53) were defined as neutrophil-enriched spots. Similarly, glycolysis signatures of each spatial spot were calculated with ssGSEA [15] according to hallmark glycolysis pathway. Tumor and stromal regions were annotated according to brightfield image of H&E staining. The regions of spots in stromal area adjacent to tumor regions were defined as "adjacent-tumor area".

Cell culture

HL-60, Aspc-1, and PATU-8988 cell lines were obtained from the Cell Bank of Shanghai Institute of Biotechnology, Chinese Academy of Sciences (Shanghai, China). Aspc-1 and PATU-8988 were cultured in DMEM medium (Meilunbio) supplemented with 10% FBS (Gibco) and antibiotics (Beyotime, penicillin, 100U/ml; streptomycin, 0.1mg/ml). HL-60 cells were cultured in RPMI-1640 medium with 10% FBS and antibiotics. Neutrophil-like differentiated HL-60 (dHL-60) cells were induced by supplementing 1% DMSO in culture media of HL-60 for 6 days. For *in vitro* induction of TAN-1 phenotype, dHL-60 cells were stimulated with thapsigargin (Abcam), IL1 β (PreproTech), TNF α (PeproTech), or cultured in hypoxic conditions (1% O₂) for 24h. In co-culture experiments, we used 24mm transwell plates with 0.4 μ m pore polyester membrane insert (Corning), in which dHL-60 cells were cultured in upper chambers, and PDAC cell lines were cultured in lower chambers.

Transient overexpression and knockdown of genes in neutrophil-like differentiated HL-60 (dHL-60) cells

The full-length overexpression plasmids of LDHA and BHLHE40 and knockdown plasmids of BHLHE40 were commercially obtained from Gene Chem (Shanghai, China). The sequence of shRNA of BHLHE40 was as following: 5'-GCAGTGGTTCTTGAACCTTA-3'. The shRNA sequence was ligated into pCLenti-U6-shRNA-CMV-Puro-WPRE vector. Empty vector was also used as negative control. The plasmids were transiently transfected into dHL-60 cells with Lipofectamine 2000 (Invitrogen), according to the manufacturer's protocol.

Real-time RT-PCR (qPCR) analysis

Total RNA was extracted with TRIzol (Invitrogen). 1 mg RNA was used as template for cDNA conversion with HiScript III RT SuperMix (+gDNA wiper) (Vazyme). Real-time RT-PCR analysis of mRNA expression was performed in triplicate with AceQ Universal SYBR qPCR Master Mix (Vazyme). The primers for target genes were listed in Table S11. Gene expression was normalized to β -actin as endogenous control.

Western blotting

Total protein was extracted using RIPA buffer (Solarbio, China) with protease inhibitor cocktail (Roche Applied Science, Switzerland), and quantitated by bicinchoninic acid (BCA) protein assay (Beyotime). Western blotting was performed as described previously [16].

Metabolism assays

For glucose uptake assay, dHL-60 cells were starved overnight, and seeded in 96-well plate (3000 cells/well). Glucose consumption was measured by glucose uptake assay kit (abcam, ab136955), according to manufacturer's protocol. For analysis of lactate production, dHL-60 cells were cultured at the density of 1×10^6 in 24-well plate for 24h, and the concentration of lactate in culture supernatant was measured by L-lactate assay kit (abcam, ab65331).

Cell proliferation assay

Aspc-1 and PATU-8988 were seeded in 96-well plates, and cultured in the presence of supernatants from control and LDHA/BHLHE40 overexpressed dHL-60 cells. The number of viable cells were assessed at each 24h interval using cell counting kit 8 (Dojindo, Japan) according to the manufacturer's protocol. For colony formation assay, Aspc-1 and PATU-8988 cells were seeded in 6-well plates (1000 cells/well), and control and LDHA/BHLHE40 overexpressed dHL-60 cells were cultured in upper chambers in transwell plates. After 2 weeks of co-culture, colonies were stained with crystal violet (Beyotime, C0121).

Migration assay

After co-culturing with control and BHLHE40 overexpressed dHL-60 cells for 3 days, Aspc-1 and PATU-8988 cells were loaded in the upper chamber of transwell plates (8µm pore size; Corning), and incubated at 37°C for 36~48h. Migrated cells were stained with crystal violet, and counted in 3 random high power fields.

In vitro immune functional assays

Peripheral blood mononuclear cells (PBMCs) were isolated from peripheral blood of healthy donors with standard Ficoll-Hypaque procedures. For analysis of cytokine expression, PBMCs were co-cultured with control and LDHA/BHLHE40 overexpressed dHL-60 cells for three days using transwell plates (0.4 µm pore size, Corning), and then stimulated with PMA and ionomycin, in the presence of brefeldin A and monensin (Cell Stimulation Cocktail, Thermo Fisher Scientific) for 4h. The percentage of IFN γ^+ and TNF α^+ cells among stimulated CD8 $^+$ T cells was analyzed by flow cytometry, as described above. The following antibodies were used for flow cytometry staining: anti-CD8 (BioLegend, 344748), anti-IFN γ (BioLegend, 502506), anti-TNF α (BioLegend, 502909). For lymphocyte proliferation assay, PBMCs were labelled with 5µM CellTrace Violet, seeded in 96-well round-bottom plates (1×10^5 cells/well), and co-cultured with control and LDHA/BHLHE40 overexpressed dHL-60 cells at 1:1 ratio. PBMCs were activated with coated anti-CD3 antibody (1µg/ml, clone OKT3, eBioscience) and soluble anti-CD28 antibody (5µg/ml, clone CD28.2, eBioscience) for 4 days. At the end of culture, cells were stained with Zombie NIRTM fixable viability dye and PE-conjugated anti-CD3 (BioLegend, 300441) as described above, and the CellTrace signals of lymphocytes was analyzed with flow cytometry. The proliferation index of lymphocytes was calculated using ModFit software.

Chromatin immunoprecipitation (ChIP)

For each sample, 100 million crosslinked cells were lysed to prepare nuclear extracts. After chromatin shearing by sonication for 10 min, lysates were incubated on a sky wheel overnight at 4 °C with protein A Dynabeads (Invitrogen, USA) coupled with 3-6 µg of BHLHE40 antibody (Novus Biologicals, NB100-1800). After immunoprecipitation, the beads were washed and recovered using magnets. DNA was reverse-cross-linked at 55 °C for 4 hours, followed by purification of the DNA product with a QIAquick PCR Purification Kit (QIAGEN, USA). DNA was quantitated using the Qubit dsDNA HS assay and a Qubit 2.0 Fluorimeter (Invitrogen). ChIP products were used to amplify the PCR products for 40 cycles using the following PCR conditions: 95 °C for 3 min and thermocycling at 95 °C for 15 s, 60 °C for 60 s, followed by 4 °C forever. The ChIP DNA Library was sequenced with Illumina HiSeq X Ten using the paired-end module and with 150 bp reads on each end (Novogene Biotech, China).

Statistical analysis

The association between two variables was assessed by Pearson correlation analysis. Kaplan-Meier survival curves and log-rank tests were used to compare overall survival in different patient groups. Linear models were generated for

comparison of pathway activities between TAN subclusters. Comparisons between two groups were performed with Student's *t* test (for normally distributed data) or Mann-Whitney *U* test (for non-normally distributed data). Benjamini & Hochberg correction was performed for multiple comparison.

References:

- 1 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;**34**:i884-i90.
- 2 Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* 2017;**27**:491-9.
- 3 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15-21.
- 4 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, *et al.* Comprehensive Integration of Single-Cell Data. *Cell* 2019;**177**:1888-902.e21.
- 5 Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;**28**:882-3.
- 6 Krijthe J. Rtsne: T-Distributed Stochastic Neighbor Embedding using Barnes-Hut Implementation. 2015.
- 7 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545-50.
- 8 Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;**14**:7.
- 9 Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;**32**:381-6.
- 10 Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* 2020;**15**:1484-506.
- 11 Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;**14**:1083-6.
- 12 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078-9.
- 13 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017;**14**:417-9.
- 14 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47.
- 15 Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 2009;**462**:108-12.
- 16 Wu J, Li H, Shi M, Zhu Y, Ma Y, Zhong Y, *et al.* TET1-mediated DNA hydroxymethylation activates inhibitors of the Wnt/ β -catenin signaling pathway to suppress EMT in pancreatic tumor cells. *J Exp Clin Cancer Res* 2019;**38**:348.