Supplementary Methods

Sample collection

Surgically removed fresh samples (primary CRC, lCRC, oCRC and adjacent normal tissues) and PBMC were obtained from 7 patients at the Sixth Affiliated Hospital, Sun Yat-sen University (online supplemental table 1). Sample was divided into two portions. One fresh portion was used for scRNA-seq and another portion embedded in optimal cutting temperature compound (OCT) was used for spatial transcriptomic analysis. CRC diagnosis was confirmed by histopathological examination and the tumor stage was defined according to the 8th edition of the American Joint Committee on Cancer (AJCC). Clinical information about CRC patients was obtained from their medical records. The study was approved by the institutional review board of the SYSUCC and informed consent was obtained from each participant.

Sample preparation and scRNA sequencing

Fresh tissue sample was temporarily placed in RPMI-1640 medium (Corning) with 20% fetal bovine serum (FBS, Cell Technologies) on ice and processed for scRNA-seq following the protocols described previously.[1] Tissue sample was cut into small pieces in a cell-culture dish and digested in a centrifuge tube with a digestion mixture containing Collagenase II (Sigma), Collagenase IV (Sigma) DNase I (Sigma) and Dispase (Corning) for 30 min at 37 °C. The cell suspension was filtered through 40-μm cell strainer (BD Falcon) to obtain single-cell suspension, which was then with blood cell lysis buffer (Boster Biological Technology) on ice for 7 min followed by brief centrifugation. The remaining cells were washed twice with PBST buffer containing 0.1% BSA. Cell viability was measured with Acridine Orange/Propidium Iodide (AO/PI) kit. scRNA-seq libraries were prepared using ChromiumTM Single Cell G Chip (10 × Genomics, 1000120) and Chromium Single Cell 3’ Library & Gel Bead Kit v3.1 (10 × Genomics, 1000121) and sequencing was accomplished on an Illumina NovaSeq6000 System using a paired-end 150 bp.

Single-cell RNA sequencing data analysis

The Cell Ranger Single-Cell toolkit was applied to align reads for each sample based on the human reference genome GRCh38 (https://cf.10xgenomics.com/supp/cell-exp/refdata-gex-GRCh38-2020-A.tar.gz). Seurat [2] R package was used for downstream analyses. Further quality control was applied to cells based on these thresholds: 1) the number of expressed genes larger than 150 and lower than 6,000; 2) the cells with less than 10% mitochondrial RNA content. DoubletFinder [3] R package was applied to remove potential doublets. The filtered gene expression matrix for each sample was normalized and scaled by “NormalizeData” and “ScaleData”

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

functions in Seurat.[2] Harmony [4] R package was used to adjust batch effects between different patients and integrate the gene expression matrices of all samples. Finally, we identified 36,601 genes and detected 230,818 cells from 70 samples. We performed principal component analysis (PCA) on the corrected expression matrix using highly variable genes (HVGs) identified by "FindVariableGenes" function. Next, "RunPCA" function was used to perform the PCA and "FindNeighbors" function was used to construct a K-nearest-neighbor graph. The most representative principal components were used to determine different cell types with "FindCluster" function. We annotated cell types and 10 clusters were identified based on expression of the following marker genes: *CD79A* and *MS4A1* for B cells, *ENG* and *PECAM1* for endothelial cells, *KRT8* and *EPCAM* for epithelial cells, *DCN* and *COL3A1* for fibroblasts; *GPM6B* and *PLP1* for glial cells, *TPSAB1* and *CPA3* for mast cells, *PPBP* and *PF4* for megakaryocytes, *CD14* and *CD68* for myeloid cells, *TNFRSF17* and *JCHAIN* for plasma cells and *CD3D* and *CD3E* for T cells. We performed unsupervised clustering on all T cells. We defined cells as CD4 cells if they had CD4 expression higher than CD8 and CD8 cells if they had CD8 expression higher than CD4. We performing unsupervised clustering on CD4 cells and CD8 cells separately to identify the expression of CD8-negative in all subpopulations of CD4 cells and the expression of CD4-negative in CD8 subpopulations.

Differential expression analysis

To identify differentially expressed genes for each cell subtype, the functions "FindAllMarkers" (multiple condition comparisons) and "FindMarkers" (two condition comparisons) from the Seurat [2] package were used with default parameters. The expression differences with $P < 0.05$ and $\log_2$(fold change, FC) > 0.25 were considered as differentially expressed genes.

Epithelial cell characterization

Fourteen epithelial cell subtypes were identified and annotated based on expression of some markers, including Tuft cells (*PLCG2* and *SH2D6*),[5] proliferation cells (*MKI67* and *PTTG1*),[6] goblet cells (*MUC2* and *ITLN1*),[7] enteroendocrine cells (*PYY* and *CHGA*),[8] enterocytes (*CA2* and *CA7*) [9] and cholangiocytes (*IFI6* and *TM4SF4*). The CytoTRACE algorithm [10] was used for predicting the differentiation status of five malignant cell subtypes. CytoTRACE scores ranged from 0 to 1 with high score indicating low differentiation status and low score indicating high differentiation.

Copy number variation calling

We used the inferCNV R package (inferCNV of the Trinity CTAT Project, provided at

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

https://github.com/broadinstitute/inferCNV) to infer the large-scale chromosomal copy number variations of each cell. Other parameters were set as default. All epithelial cells from adjacent normal tissues served as references. In addition, CopyKAT [11] was used to assume the major genetic distance among the cell subtypes and to classify the single cells into two clusters, i.e., diploid cell cluster and aneuploidy cell cluster.

Tissue preference of each cell subtype

We calculated the ratio of observed to expected cell numbers in each cell subtype (Ro/e) to quantify the preference of each cell subtype across tissues as previously suggested.[12] In brief, the expected cell numbers of each cell subtype in each tissue were obtained from the Chi-square test, and Ro/e > 1 for a cell subtype in a tissue indicated preference of this cell subtype in this tissue.

Genomic instability estimate

To estimate the genomic instability of each malignant cell, we used the genomicInstability R package which used the aREA algorithm to quantify the enrichment of sets of contiguous genes (loci-blocks) on the gene expression profiles and estimate the Genomic Instability Score (GIS) for each analyzed cell.

Stromal cell characterization

Fibroblast cells were separated into 13 distinct subtypes and were annotated as myofibroblasts (marker genes: *ACTA2* and *COL1A1*),[13] pericytes (*RGS5* and *NDUFA4L2*), [13] vascular smooth muscle cells (*PLN* and *RERGL*).[14] Among the 13 subtypes, 9 were enriched in primary CRC or metastatic CRC (Ro/e > 1) and thus were termed as cancer-associated fibroblasts (CAFs). Endothelial cells were separated into 10 distinct cell subtypes and were annotated as vein clusters (*ACKR1* and *VWF*),[15] arterial clusters (*GJA4* and *HEY1*),[16] capillaries (*KDR* and *RGCC*),[16] lymphatics (*PROX1* and *LYVE1*).[17]

Cell developmental trajectory analysis

RNA velocity analysis was conducted using velocyto [18] and scVelo.[19] We used the 10 × velocyto pipeline to count spliced and non-spliced reads for each sample from cellranger-generated BAM files. To predict the root and terminal states of the underlying Markov process, the respective scVelo function was applied. We also used a python package PAGA [20] to verify the pseudotime between each epithelial cell subtype. The single-cell trajectory analysis of myofibroblast cell subtypes was performed with Monocle2 [21] using DDR-Tree and default parameters.

Spatial transcriptomic analysis

OCT embedded tissue samples were cryo-sectioned to slides at 10-μm thick and adhered to the capture regions (6.5 mm × 6.5 mm) of the visium spatial tissue optimization slide (10 × Genomics). After fixation and permeabilization, the released mRNA was captured by the oligonucleotides. cDNA synthesis was performed by using the Master mixture containing reverse transcriptional reagents and fluorescently labeled nucleotides. cDNA was then covalently linked to oligonucleotides and retained in the visium spatial tissue optimization slide when the tissue was removed. The library was built after the above operations and Illumina NovaSeq 6000 System was used to perform the sequencing. Sample information on spatial cohort is shown in online supplemental table 2.

The Spaceranger software from 10 × Genomics was used to process the fastq files and images of spatial transcriptomics for each sample based on the human reference genome GRCh38. UMI count spot matrices, images and spot-image coordinates were imported into R. To infer the spatial organization of certain cell subtypes, we used Seurat [2] R package to integrate spatial and single-cell data. Raw UMI counts were normalized by "SCTransform" function. Dimensionality reduction and clustering were performed as before. Cell subtypes distributions were visualized in spatial context over H&E images.

Single cell gene set enrichment analysis

We conducted the gene set enrichment analysis for select cell subtypes by the irGSEA R package, with the Hallmark or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways being derived. Finally, the "irGSEA.heatmap" and "irGSEA.halfvlnplot" functions were applied to visualize enrichment score.

Transcription factor module analysis

We applied the pySCENIC [22] workflow to detect active transcription factor modules in five malignant cell subtypes, using transcription factor motif scores for hg38 human reference genome from the RcisTarget database.

Identification of master transcription factors

Master transcription factors that regulate differentially expressed genes in P1 and P3 cells were analyzed using the plugin iRegulon [23] in Cytoscape network.[24] iRegulon used > 9,000 known position weight matrices from various sources and different species. Candidate binding TFs were identified using a 'motif2TF' procedure. Predicted master transcription factors were ranked according to the normalized enrichment score.

Li R, *et al*. *Gut* 2024; 73:470–484. doi: 10.1136/gutjnl-2023-330243

Cell-cell interaction analysis

We used CellChat [25] with default parameters to identify significant ligand-receptor pairs within primary CRC samples. We used all categories of ligand-receptor interactions in the database for the analysis, focusing on the differences in predicted cell-cell communications between P1 cells and P6 cells. Finally, the "netVisual_bubble" function was applied to visualize communication probabilities by ligand-receptor pairs in different directions. We also used the intercellular signaling network (iTALK) [26] R package to verify these significant ligand-receptor pairs between stem-like cell subtypes and CAFs.

Estimate of the corresponding cell fractions in TCGA-COAD RNA-seq data

We used a computational R package Estimate the Proportion of Immune and Cancer cells (EPIC) [27] to estimate the cell fractions of P1 stem-like cells, CAFs and 4 endothelial cell subtypes (Art_NOTCH4, Tip_COL4A1, Veins_ACKR1 and Veins_SERPINE1) from TCGA-COAD RNA-seq data. The score of P1-CAFs interaction network was the product of the mRNA levels of ligand-receptor pairs (PDGFA-PDGFRA, DLL4-NOTCH2 and DLL4-NOTCH3) times the cell fractions of corresponding cell subtypes. The score of P1-endothelial cell interaction network was also calculated by integrating the mRNA levels of ligand-receptor pairs (DLL4-NOTCH1, DLL4-NOTCH4, JAG1-NOTCH1 and JAG1-NOTCH4) and the cell fractions of corresponding cell subtypes.

Cell subtype similarity analysis

We used the following steps to evaluate the similarity of cell subtypes: (1) Identify top 1,000 highly variable genes across different cell subtypes, (2) calculate the mean value of the top 1,000 highly variable genes in each cell subtype, and (3) cluster the hierarchical, using the distance defined by (1-Pearson correlation coefficient)/2.

Single-cell flux estimation and cell metabolite prediction

We used scfea [28] tools which utilizes a graph neural network model to estimate cell-wise metabolic flux by using scRNA-seq data. We chose "module gene m168" as moduleGene file, "cmMat c70 m168" as stoichiometry file, and parameter "sc imputation = True".

Cell-cell metabolic communication

We used MEBOCOST,[29] a Python-based computational tool to infer metabolite, mediated cell-cell communication events. The cutoff was set as 0.25 and other parameters were defaulted.

Plasmid and lentiviral constructions and transduction

HT-29 and SW-480 cells with *DLL4*, *ASCL2* or *PTPRO* knockdown were generated using the pLKD-U6-MCS-CMV-Puro vector (Umine Biotechnology). Both shControl vector and recombinant

viruses were produced in 293T cells. The infection was performed in HT-29 and SW-480 cell lines in the presence of polybrene (Sigma-Aldrich) and selected using puromycin. The shRNA sequences are shDLL4-1: 5'-gcaagaagcgcaatgaccactctcgagagtggtcattgcgcttcttgc-3'; shDLL4-2: 5'-gcactccctggcaatgtacttctcgagaagtacattgccagggagtgc-3'. shASCL2-1: 5'-gcgtgaagctggtgaacttctcaagagaaagttcaccagcttcacgc-3'; shASCL2-2: 5'-ccagcaagaagctgagcaattcaagagattgctcagcttcttgctgg-3'. shPTPRO-1: 5'-gcagtgactatgaaactacgtctcgagacgtagtttcatagtcactgc-3'; shPTPRO-2: 5'-gcagcacattcgggatcatgactcgagtcatgatcccgaatgtgctgc-3'. shControl: 5'-ttctccgaacgtgtcacgtttcaagagaacgtgacacgttcggaga-3'.

Quantitative reverse-transcription PCR

Total RNA from MSS CRC cell lines and human ovarian fibroblasts were extracted using TRIzol reagent (Invitrogen) and was reverse-transcribed using the RevertAid First Strand cDNA Synthesis Kit (K1622, Thermo Fisher Scientific) with random primers. Relative mRNA level of *ASCL2*, *PTPRO* and *DLL4* was determined by RT-qPCR on a Light Cycler 480 II using the SYBR Green method. *β-ACTIN* as internal control. The q-PCR primers are *ASCL2*-F: 5'-caaccgcgtgaagctggtgaact-3'; *ASCL2*-R: 5'-tctccaccttgctcagcttcttgc-3'. *PTPRO*-F: 5'-gcagtttgtacacatggtccgac-3'; *PTPRO*-R: 5'-caatgaatgttcctgtccgtccc-3'. *DLL4*-F: 5'-aagagttgcctgagtggaatttc-3'; *DLL4*-R: 5'-agcttgttagggtccttacgg-3'. *β-ACTIN*-F: 5'-cagggcgtgatggtgggcatg-3'; *β-ACTIN*-R: 5'-gtagaaggtgtggtgccagatt-3'.

Cell proliferation and invasion assay

Invasion assay was performed in a 24-well Millicell chamber in triplicate. The 8-μm pore inserts were pre-coated with 30 μg of Matrigel (BD Biosciences). HT-29 and SW-480 cells (4×10^4) in serum-free medium were added to the upper chamber. RPMI-1640 containing 20% FBS was supplied to the lower chambers. After 24 h incubating at 37 °C with 5% CO_2, cells migrated through the filters were fixed with methanol and stained with crystal violet. Cell numbers in 3 random fields were counted and the experiments were performed in triplicate. The migration assay was tested in a similar method without Matrigel. CRC cells (2000 per well) were seeded in 96-well plates with 100 μl of basal or conditional medium collected from the cultivation of ovarian fibroblasts. After treated with vehicle or V9302, cell viability was measured using CCK-8 assays (Dojindo) at defined time of culture. Each experiment was performed with at least three replicates.

Sphere formation assay

HT-29 and SW-480 cells (with or without *PTPRO* knockdown) were prepared as single-cell suspension and plated into a 24-well ultra-low attachment plate (Corning, CLS3473. 1000 cells per

well). Cells were cultured in serum-free DMEM/F12 medium supplemented with 20 ng/mL EGF, 20 ng/mL bFGF and 1% B27 supplement. After 20 days, the sphere pictures were captured by using inverted microscope and sphere area was analyzed by Image J software.

Immunohistochemical staining

Paraffin-embedded tissue samples were used for Immunohistochemical (IHC) analysis. Antibody against DLL4 (Invitrogen, PA5-85931), ETV4 (Proteintech, 10684-1-AP) or ELF3 (Signalway Antibody, 38856) was used. We applied the immune reactive score (IRS) to determine DLL4, ETV4 and ELF3 protein levels. The staining intensity was estimated as negative (0), weak (1), moderate (2), and strong (3). The extent of staining was graded as 1 (≤25%), 2 (26%–50%), 3 (51%–75%), or 4 (>75%). The IRS was calculated by multiplying the score of intensity and extent.

Western blot analysis

Total protein (30 μg) extracted from HT-29 and SW-480 cells was subjected to SDS-PAGE and transferred to PVDF membranes (Millipore). Membranes were then incubated overnight at 4 °C with primary antibody and visualized with a Phototope Horseradish Peroxidase Western Blot Detection kit (WBKLS0100, Millipore). Rabbit anti-DLL4 antibody (Invitrogen, PA5-85931), Rabbit anti-ASCL2 antibody (Signalway Antibody, 34038), mouse anti-PTPRO antibody (Proteintech, 67000-1-Ig) and mouse anti-β-ACTIN antibody (Proteintech, 66009-1-Ig) were used.

Immunofluorescence staining

Paraffin-embedded tissue samples that were performed single-cell sequencing and spatial transcriptomic analyses were used for immunofluorescence analysis of interesting protein levels. Immunofluorescence staining was performed using PANO 7-plex IHC kit (Panovue). For primary CRC tissue slides antibody against ASCL2 (Bioss, bs-12349R), PTPRO or EPCAM (Abcam, ab71916) were used. The slides were washed after incubating each primary antibody, then the secondary antibody and different dyes were applied and incubated at room temperature. After tyramide signal amplification, cell nuclei were counterstained with DAPI. The multispectral images were obtained using the Mantra System (PerkinElmer, Waltham, Massachusetts, US).

CRC cells transendothelial migration assay

HUVEC cells (5×10^4) were seeded on type I collagen coated transwell inserts and grown for 24 h in 5% CO_2 at 37 °C to form monolayer. CRC cells with different treatment were labeled with 10 μM of the lipophilic fluorescent dye DiO (C1038, Beyotime) for 20 min at 37 °C and then suspended in serum free 1640 medium and added to the upper chamber. Medium with serum was added to the lower chamber. CRC cells were cultured 18 h for their migration. After that, we used

cotton swab to remove non-migrating cells on the upper face of the filter, cells on the lower face were then fixed with 4% formaldehyde for 10 min and washed with PBS buffer. The migrated CRC cells were observed using epifluorescence microscope. Cell numbers in 3 random fields were counted.

Glutamine (Gln) detection

Glutamine levels in culture medium and CRC cells were detected following the instruction manual of Glutamine Microplate Assay Kit (AS0160, sabbiotech). Briefly, human ovarian fibroblasts and CRC cells were cultured in glutamine free 1640 medium for 24 h. Then the conditional medium was collected to determine the glutamine levels released from ovarian fibroblasts and CRC cells. For intracellular glutamine detection, CRC cells were cultured with conditional medium from ovarian fibroblasts or control medium. CRC cells were treated with V9302 or vehicle and the cultivation were collected and sonicated. Then the lysate was used for glutamine detection. The results were normalized to cell number.

Establishment of mouse ovarian metastasis model

Five-week-old female BALB/c nude mice were obtained from Beijing Vital River Laboratory Animal Technology and allowed to acclimate to local conditions for 1 week under a 12 h light/12 h dark cycle. Luciferase labeled CRC cells (5×10^5) with *ASCL2*, *PTPRO* or *DLL4* stable knockdown were injected into the tail vein of mice (n = 10). After 40 days, mouse ovarian tissues were collected. Ovarian metastasis was determined by bioluminescence imaging and hematoxylin-eosin staining. Animal handling and experimental procedures were approved by the Institutional Animal Care and Use Committee of Sun Yat-Sen University and performed in accordance with the relevant institutional and national guidelines.

Public datasets used in this study

To increase the statistical power, we downloaded the publicly available scRNA-seq data from previous study (GSE132465 and GSE144735, termed as KUL3 dataset and SMC dataset).[30] We also included the public bulk RNA-seq datasets (GSE50760, GSE75117) from the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/). [31, 32] The dataset GSE50760 (n = 54) included normal colon, primary CRC, and liver metastases generated from 18 CRC patients. The dataset GSE75117 included invasive front (IF), tumor center (TC) and peritoneal metastasis (PM) from 16 CRC patients. Bulk RNA-seq data of EVP and D endothelial cells were obtained from the GEO accession number GSE114528[33] and the ArrayExpress (https://www.ebi.ac.uk/arrayexpress) accession number E-MTAB-7148.[34] Transcriptomic data and clinical information of The Cancer

Genome Atlas (TCGA) COAD cohort were downloaded from the UCSC Xena data portal (https://xenabrowser.net).[35]

Statistical analysis

Comparisons between two groups were performed using two-tailed Student's *t*-test under the normality assumption. One-way ANOVA with Dunnett's T3 multiple-comparison test was used to compare several groups. Spearman's correlation was used to measure the correlation between two continuous variables and r > 0.3 and $P < 0.05$ was considered significant. Log-rank test was used for univariate survival analyses and showed as the Kaplan-Meier plot. Cox proportional hazards model was used to analyze multivariate effects. Receiver operating characteristic (ROC) curve were constructed and areas under the curves (AUC) were used to evaluate the diagnostic value of DLL4. All statistical analyses and visualization were performed using R or GraphPad Prism. The lines in the middle of the box plot are median and the upper and lower lines indicate 25th and 75th percentiles. $P < 0.05$ was considered statistically significant. The number of replicates and statistical tests used in figures were shown in corresponding figure legends.

Supplementary References

[1] Liu Y, He S, Wang XL, *et al*. Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat Commun* 2021;12:741.

[2] Satija R, Farrell JA, Gennert D, *et al*. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;33:495–502.

[3] McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* 2019;8:329–337.e4.

[4] Korsunsky I, Millard N, Fan J, *et al*. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019;16:1289–1296.

[5] Xiong Z, Zhu X, Geng J, *et al*. Intestinal Tuft-2 cells exert antimicrobial immunity via sensing bacterial metabolite N-undecanoylglycine. *Immunity* 2022;55:686–700.e7.

[6] Chen L, Chou CL, Knepper MA. Targeted single-cell RNA-seq identifies minority cell types of kidney distal nephron. *J Am Soc Nephrol* 2021;32:886–896.

[7] Birchenough GM, Nyström EE, Johansson ME, *et al*. A sentinel goblet cell guards the colonic crypt by triggering Nlrp6-dependent Muc2 secretion. *Science* 2016;352:1535–42.

[8] Steinert RE, Feinle-Bisset C, Asarian L, *et al*. Ghrelin, CCK, GLP-1, and PYY(3-36): Secretory

controls and physiological roles in eating and glycemia in health, obesity, and after RYGB. *Physiol Rev* 2017;97:411–463.

[9] Thiagarajah JR, Donowitz M, Verkman AS. Secretory diarrhoea: mechanisms and emerging therapies. *Nat Rev Gastroenterol Hepatol* 2015;12:446–57.

[10] Gulati GS, Sikandar SS, Wesche DJ, *et al*. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 2020;367:405–411.

[11] Gao R, Bai S, Henderson YC, *et al*. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* 2021;39:599–608.

[12] He J, Xiong X, Yang H, *et al*. Defined tumor antigen-specific T cells potentiate personalized TCR-T cell therapy and prediction of immunotherapy response. *Cell Res* 2022;32:530–542.

[13] Zhang W, Conway SJ, Liu Y, *et al*. Heterogeneity of hepatic stellate cells in fibrogenesis of the liver: insights from single-cell transcriptomic analysis in liver injury. *Cells* 2021;10.

[14] Pelka K, Hofree M, Chen JH, *et al*. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* 2021;184:4734–4752.e20.

[15] Pu W, Shi X, Yu P, *et al*. Single-cell transcriptomic analysis of the tumor ecosystems underlying initiation and progression of papillary thyroid carcinoma. *Nat Commun* 2021;12:6058.

[16] Zeng Y, He J, Bai Z, *et al*. Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Res* 2019;29:881–894.

[17] Kalucka J, de Rooij L, Goveia J, *et al*. Single-cell transcriptome atlas of murine endothelial cells. *Cell* 2020;180:764–779.e20.

[18] La Manno G, Soldatov R, Zeisel A, *et al.* RNA velocity of single cells. *Nature* 2018;560:494–498.

[19] Bergen V, Lange M, Peidli S, *et al*. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 2020;38:1408–1414.

[20] Wolf FA, Hamey FK, Plass M, *et al*. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* 2019;20:59.

[21] Qiu X, Mao Q, Tang Y, *et al*. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017;14:979–982.

[22] Aibar S, González-Blas CB, Moerman T, *et al*. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;14:1083–1086.

[23] Janky R, Verfaillie A, Imrichová H, *et al*. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol* 2014;10:e1003731.

[24] Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.

[25] Jin S, Guerrero-Juarez CF, Zhang L, *et al*. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* 2021;12:1088.

[26] Wang Y, Wang R, Zhang S, *et al*. iTALK: an R package to characterize and illustrate intercellular communication. bioRxiv 2019:507871.

[27] Racle J, Gfeller D. EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data. *Methods Mol Biol* 2020;2120:233–248.

[28] Alghamdi N, Chang W, Dang P, *et al*. A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data. *Genome Res* 2021;31:1867–1884.

[29] Zheng R, Zhang Y, Tsuji T, *et al*. MEBOCOST: metabolic cell-cell communication modeling by single cell transcriptome. bioRxiv 2022:2022.05.30.494067.

[dataset] [30] Lee HO, Hong Y, Etlioglu HE, *et al*. Data from: Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. The GEO database, May, 25, 2020. https://doi.org/10.1038/s41588-020-0636-z.

[dataset] [31] Kim SK, Kim SY, Kim JH, *et al*. Data from: A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. The GEO database, Dec, 8, 2014. https://doi.org/10.1016/j.molonc.2014.06.016.

[dataset] [32] Muto Y, Suzuki K, Kato T, *et al*. Data from: Heterogeneous expression of zinc-finger E-box-binding homeobox 1 plays a pivotal role in metastasis via regulation of miR-200c in epithelial-mesenchymal transition. The GEO database, Sep, 2016. https://doi.org/10.3892/ijo.2016.3583.

[dataset] [33] Donovan P, Patel J, Dight J, *et al*. Data from: Endovascular progenitors infiltrate melanomas and differentiate towards a variety of vascular beds promoting tumor metastasis. The GEO database, January, 03, 2019. https://doi.org/10.1038/s41467-018-07961-w.

[dataset] [34] Lukowski SW, Patel J, Andersen SB, *et al*. Data from: Single-cell transcriptional profiling of aortic endothelium identifies a hierarchy from endovascular progenitors to differentiated cells. The ArrayExpress, May, 28, 2019.

https://doi.org/10.1016/j.celrep.2019.04.102.

[dataset] [35] Goldman MJ, Craft B, Hastie M, *et al*. Data from: Visualizing and interpreting cancer genomics data via the Xena platform. The Xena platform, Jun, 2020. https://doi.org/10.1038/s41587-020-0546-8.