


RESEARCH

Open Access



A dynamic molecular landscape in colorectal cancer progression at single-cell resolution

Jianwen Sheng^{1,2†}, Nianshuang Li^{3†}, Shuai Li^{1,4}, Yuman Ye⁵, Yuanhang Liao^{1,2}, Yupeng Zhang^{1,2}, Huan Chen^{1,2}, Mengxian Tu^{1,2}, Guping Zhong^{1,2}, Chaoliang Xiong^{1,2}, Pan Zheng³, Yuting Lei³, Zekun He^{1,2}, Xingxing He³, Yao Zhang^{1,4}, Lei Wang⁶, Xiaolin Gao⁷, Yin Zhu³, Jianping Liu^{3,8*}  and Huizhen Fan^{1,2*}

Abstract

Background Precursor lesions like polyps and adenomas in the colon commonly precede colorectal cancer (CRC), advancing through the “normal-polyp-adenoma-carcinoma” sequence towards malignancy. Yet, the cellular heterogeneity and molecular mechanisms involved in CRC development remain inadequately characterized.

Methods To understand the molecular mechanisms driving the onset and progression of CRC, we conducted a comprehensive analysis of ten clinical colorectal samples representing sequential pathological stages using single-cell RNA sequencing (scRNA-seq). Validation was performed through immunofluorescence and immunohistochemistry analyses in a separate human colorectal tissue cohort. Additional verification was carried out using bioinformatics analyses of public TCGA and GEO datasets.

Results Our comprehensive analyses not only reveal the cellular diversity and transcriptomic differences throughout disease progression but also highlight the importance of leveraging ligand–receptor gene expression to distinguish various cell subtypes. Subsequent examination and validation with a larger sample cohort uncover the specific involvement of ligand–receptor genes, transcription factors, immunoglobulin genes, and heat shock genes in regulating immune responses and microenvironment changes during colorectal tumorigenesis.

Conclusions Our extensive transcriptome dataset provides valuable insights and acts as a fundamental resource to deepen our understanding of the complex molecular landscape in CRC. This dataset facilitates improved diagnostic accuracy and the creation of more personalized therapeutic approaches.

Keywords Single-cell RNA-seq, Ligand–receptor genes, Colorectal cancer, Transcription factors, Heat shock proteins

[†]Jianwen Sheng and Nianshuang Li contributed equally to this work.

*Correspondence:

Jianping Liu

jianping.liu.2@ki.se; ndyfy10307@ncu.edu.cn

Huizhen Fan

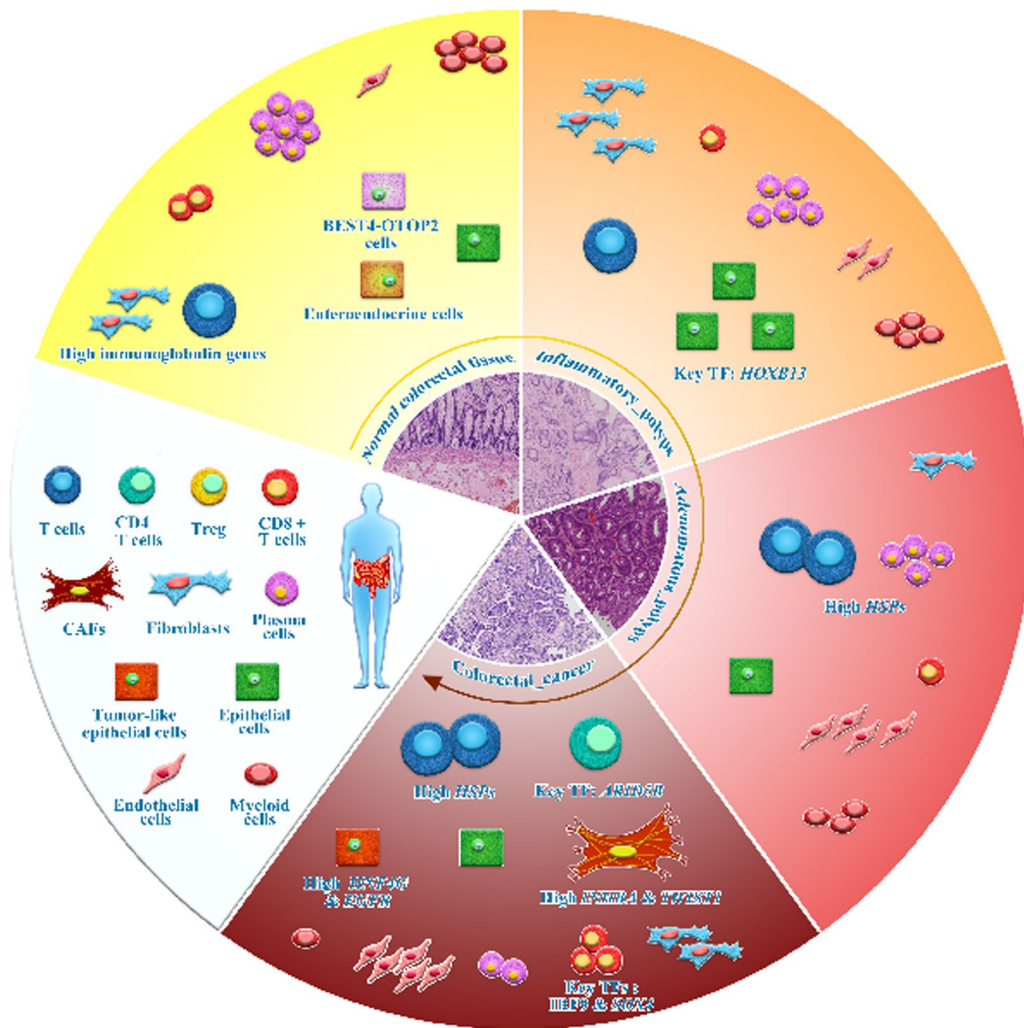
fanfanyc@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **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/>.

Graphical Abstract



Introduction

Colorectal cancer (CRC) ranks as the third most prevalent cancer and a contributor to global cancer mortality [1]. Projections estimate a rise to 3.2 million annual diagnoses and 1.6 million deaths by 2040 [2]. Late-stage diagnosis, incomplete understanding of molecular pathological progression, and constrained treatment options for advanced stages collectively drive poor outcomes. Addressing these challenges necessitates three critical advances: identification of early-stage biomarkers, mechanistic dissection of tumor evolution, and development of targeted therapies to enhance clinical management.

CRC originates from mucosal polyps—benign growths that protrude into the intestinal lumen. While most polyps remain asymptomatic and non-neoplastic, 70–90% of

CRC cases arise from adenomatous polyps (adenomas) via stepwise progression over 5–15 years [3, 4]. Post-inflammatory polyps further elevate colorectal neoplasia risk, including dysplasia and invasive carcinoma [5]. Therefore, systematic characterization of stage-specific molecular landscapes—through single-cell transcriptomic profiling of polyp-to-CRC transitions, is critical to define actionable biomarkers and therapeutic vulnerabilities, enabling early interception and precision management of CRC evolution.

Single-cell RNA sequencing (scRNA-seq) enables high-resolution dissection of tumor ecosystems by delineating cellular heterogeneity, novel phenotypic states, and therapeutic vulnerabilities [6–11]. The scRNA-seq technique has been extensively employed in CRC research [12–14],

particularly in investigating microenvironments [15–19]. For instance, a detailed analysis of immune and stromal cell populations in CRC patients identified specific macrophage and conventional dendritic cell subsets as key players in cellular interactions within the tumor microenvironment, shedding light on their role in response to myeloid-targeted immunotherapy [20]. However, systematic interrogation of dynamic molecular trajectories across CRC tumorigenesis—from premalignant polyps to metastatic lesions—remains limited. Resolving these spatiotemporal transitions at single-cell resolution is critical to uncover stage-specific drivers and actionable therapeutic windows.

While a seminal scRNA-seq study profiled paired adenoma-carcinoma tissues to model CRC progression [12], its limited sample size ($n=1$) warrants validation. Here, we performed a comprehensive single-cell profiling (10× Genomics) across four CRC pathological stages—normal mucosa, inflammatory polyps, adenomatous polyps, and invasive carcinoma ($n=10$)—coupled with multi-omics and experimental validation (immunofluorescence, TCGA, Kaplan–Meier analyses and functional assays). This revealed a dynamic stage-specific molecular landscape, highlighting the crucial involvement of ligand–receptor genes, transcription factors, immunoglobulin genes, and heat shock proteins in orchestrating colorectal carcinogenesis. Our findings decode molecular drivers of CRC evolution and nominate actionable targets for stage-adapted therapeutic strategies.

Materials and methods

Collection of human tissue samples

A cohort of ten human colorectal tissues—comprising normal mucosa ($n=3$), inflammatory polyps ($n=2$), adenomatous polyps ($n=2$), and invasive carcinoma ($n=3$)—was procured via endoscopic or surgical resection at Yichun People's Hospital, China. All specimens underwent 10× Genomics Chromium single-cell RNA sequencing, with histopathological confirmation (H&E-stained sections reviewed by board-certified pathologists) and clinical characteristics (Table S1). Parallel validation employed independent matched FFPE samples for immunofluorescence and immunohistochemical profiling. Ethical approval was obtained from the Ethics Committee of Yichun People's Hospital (protocol number 2021-009), with written informed consent prior to tissue acquisition.

Dissociation of human colorectal tissues and preparation of single-cell suspension

Excised tissues were immediately transferred to DMEM (Cat#11995065, Gibco) supplemented with 10% FBS (Cat#10091148, Gibco). Tissues were minced into

approximately 0.5 mm³ fragments and enzymatically digested at 37 °C with agitation at 500 rpm for 30–40 min using a cocktail of trypsin (Cat#15090046, Gibco) and collagenase I/II/IV (Cat. #17018029, #17101015 and #17104019, Gibco). Suspensions were mechanically dissociated every 10 min, filtered sequentially through 70 µm and 40 µm strainers (Corning), and centrifuged (300g, 5 min, 4 °C). Erythrocytes were lysed with a red blood cell elimination reagent (Cat#130094183, MACS), followed by centrifugation at 300g for 5 min and resuspension in 100 µL of medium. Cell viability and density were quantified using a Luna Cell Counter (Logos Biosystems, South Korea).

Single-cell RNA sequencing and raw data processing

Single-cell suspensions (700–1200 cells/µL) were processed into barcoded libraries using the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Cat#1000268, 10× genomics) as per manufacturer instructions. Libraries were sequenced (PE150) on an Illumina NovaSeq 6000 platform. OE Biotech Co., Ltd. (Shanghai, China) performed the sequencing and subsequent bioinformatics analyses.

FASTQ files were aligned to the GRCh38 human reference genome using Cell Ranger software (v 7.0.1) to generate a unique molecular identifier (UMI) counts matrix, which was analyzed utilizing the Seurat (version 4.0.0) R package [21]. Using DoubletFinder package (version 2.0.3) [22], potential captures of multiple cells and low-quality cells were filtered (genes < 200, UMIs < 1000, log10GenesPerUMI < 0.7, genes expressed in < 3 cells, mitochondrial UMIs > 25%). To address dropout events, MAGIC (Markov affinity-based graph imputation) [23] to recover dropout-affected genes, particularly for low-abundance transcripts. To correct batch effects, datasets were aligned across samples using Harmony to remove batch-driven variance while preserving biological heterogeneity [24]. Library size normalization was performed using the NormalizeData function. Subsequently, the global-scaling normalization method “LogNormalize” was utilized to standardize gene expression measurements for each cell by total expression, scaled by a default factor (10,000), and then logarithmically transformed.

The top 4000 highly variable genes (HVGs) were determined using the Seurat function FindVariableGenes (mean.function = FastExpMean, dispersion.function = FastLogVMR). Principal-component analysis (PCA) was then employed to reduce the dimensionality through the RunPCA function. Subsequently, Graph-based clustering was conducted to cluster cells based on their gene expression profiles using the FindClusters function. Cell visualization was achieved using a two-dimensional Uniform Manifold Approximation and

Projection (UMAP) algorithm utilizing the RunUMAP function. To identify marker genes for each cluster, the FindAllMarkers function (test.use=presto) was utilized. Differentially expressed genes (DEGs) were selected by employing the function FindMarkers (test.use=presto). A significance threshold of P value < 0.05 and $|\log_2 \text{fold-change}| > 0.58$ (corresponding to 1.5 for fold change) was set for differential expression analysis. (Sub-) cell types were annotated using canonical markers. Violin plots, dot plots, and bar plots were generated using the VlnPlot, and DotPlot functions from the Seurat package, and the geom_bar function in ggplots, respectively. Furthermore, correlation analyses between clusters or (sub-) cell types determined from HVGs expression and expressed ligand–receptor genes were conducted using the Pearson algorithm.

Pseudotime analysis

Pseudotime trajectories across colorectal carcinogenesis stages were reconstructed using Monocle2 package (v2.9.0) [25]. Initially, raw counts were transformed from Seurat objects to CellDataSet objects using the importCDS function in Monocle. The differentialGeneTest function was then employed to identify ordering genes (qval < 0.01) informative to guide trajectory ordering. Subsequently, dimensional reduction clustering analysis and trajectory inference were conducted with the reduceDimension function and by utilizing the orderCells function, respectively. Temporal gene expression dynamics were visualized via plot_genes_in_pseudotime function.

SCENIC analysis

Single-Cell rEgulatory Network Inference and Clustering (SCENIC) analysis [26] was performed to construct gene regulatory networks and explore transcription factors (TFs) and cis-regulatory sequences. Using RcisTarget package (v1.10.0), TF binding motifs enriched in target gene sets were identified, while GRNBoost (v1.2.4) inferred co-expression networks. Regulon activity per cell was scored via AUCell package (v1.12.0). To evaluate cell type specificity, the regulon specificity score (RSS) was calculated using Jensen–Shannon divergence (JSD), measuring the similarity between two probability distributions. Specifically, the JSD between each vector of binary regulon activity overlaps with the assignment of cells to a particular cell type [27] was computed. Furthermore, the connection specificity index (CSI) for all regulons was determined using the scFunctions package (<https://github.com/FloWuenne/scFunctions/>) to assess regulon associations.

Copy number variation analysis

Large-scale chromosomal copy number variation (CNV) was inferred from scRNA-seq data across ten colorectal tissues using inferCNV (version 1.0.4) [28] R package. CNV for all cell types was determined based on their expression levels, with a specified cutoff at 0.1. Genes were then ordered by chromosomal location, and a moving average of gene expression was computed using a window size of 101 genes. After mean-centering expression values, epithelial cells were classified as malignant, while others served as normal references. Final CNV profiles were generated after de-noising.

Functional pathway annotation

Functional annotation of cell subtypes was conducted in R (v4.5.0) using the clusterProfiler package (v4.15.2), systematically analyzing Gene Ontology (GO) terms across biological processes, molecular functions, and cellular components, while Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping delineated key metabolic and signaling cascades, and gene set enrichment analysis (GSEA) assessed coordinated expression patterns via MSigDB molecular signatures. All analyses were performed on normalized count matrices derived from single-cell RNA sequencing data.

TCGA database analysis

Validation of scRNA-seq findings was performed using bulk RNA-seq datasets (downloaded on February 16th, 2023) from The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) and The Cancer Genome Atlas Rectum Adenocarcinoma (TCGA-READ). After merging and quality control, gene expression profiles in tumor versus adjacent normal tissues were visualized via box plots using the ggplot2 (v3.4.0) R package (v4.2.0). Statistical significance of the differences between two groups was evaluated using t-tests.

CellPhoneDB analysis

Cell–cell communication analysis was performed using CellPhoneDB (v4.1.0) [29] to identify biologically relevant ligand–receptor (LR) interactions from scRNA-seq data. A ligand or a receptor was considered “expressed” in a specific cell type if detected in $\geq 10\%$ cells. Statistical significance was assessed through 1000 permutations of cell cluster labels, generating null distributions for each LR pair. P-values were derived from the normal distribution of permuted interaction scores. Communication networks were constructed between any two cell types expressing complementary ligands and receptors, visualized using igraph and Circlize R packages.

CellChat analysis

Cell communication analysis was performed using the CellChat (v1.1.3) [30] R package. Initially, the normalized expression matrix was imported to create the Cellchat object using the createCellChat function. Subsequently, the data underwent preprocessing steps utilizing functions such as identifyOverExpressedGenes, identifyOverExpressedInteractions, and projectData. Following this, the computeCommunProb, filterCommunication (min.cells=10), and computeCommunProbPathway functions were utilized to reveal potential ligand–receptor interactions. Finally, the cell communication network was consolidated using the aggregateNet function. All steps were performed using default parameters.

Generation of a searchable colorectal single-cell transcriptome database

To enable interactive exploration of our dataset, we developed an open-access database (<https://www.colon-singlecell.cn>) using R Shiny package (v1.3.2) based on the codes available on <https://github.com/SGDDNB/ShinyCell> and insights from the previous reports [21, 31]. In brief, we employed functions like renderPlot to display feature plots, UMAP visualizations, and bar plots, which were created by FeaturePlot, DimPlot, and geom_bar functions from the Seurat package (v3.2.0), respectively. The renderDataTable function was employed to present a statistical table showing the cell numbers corresponding to each sub-cell type across the four types of colorectal samples.

Hematoxylin & Eosin (H&E) staining

The H&E staining was performed according to a standard protocol. First, the FFPE colorectal tissue slides were deparaffinized and rehydrated with distilled water, followed by staining in Mayers Hematoxylin for 1 min. Second, the slides were washed with tap water for 4–5 times, and counterstained with DAPI in 1× PBS for 1 min. After rinsing with distilled water for three times, the slides were stained with alcoholic eosin for 1 min without rinsing. Then, the slides were dehydrated through graded ethanol (95% and 100%, two times for each reagent and 1 min for each time). The slides were cleared in xylene for three times with 1 min incubation between two changes, and finally subjected to mounting and image acquisition at a Nikon Eclipse microscope.

Immunofluorescence staining

Immunofluorescence staining was performed on deparaffinized and rehydrated tissue sections following antigen retrieval in citrate buffer (pH 6.0). Sections were permeabilized with 0.3% Triton X-100/PBS for 15 min and blocked with 3% BSA/TBST for 1 h. Primary antibodies (Table 1) were applied overnight at 4 °C, followed by PBS washes and incubation with Alexa Fluor-conjugated secondary antibodies (Table 1) at RT for 1 h. Nuclei were counterstained with Hoechst 33342 (Solarbio, Cat# C0030) before mounting, and images were acquired using a LEICA Stellaris 5 confocal microscope.

Immunohistochemistry staining

Immunohistochemistry (IHC) staining was performed using standard protocols to localize target proteins in tissue sections. Following deparaffinization and antigen

Table 1 List of antibodies used in this study

Target protein	Antibody name	Vendor	Cat. No	Dilution
BEST4	Anti-BEST4 antibody produced in rabbit	Sigma	HPA058564	1:100
CEACAM5	Anti-CEA (CD66e) CEACAM5 Rabbit Monoclonal Antibody	Boster	M00356-1	1:100
CD4	CD4 (RPA-T4) Mouse mAb (FITC Conjugate)	Cell Signaling	48705	1:800
CLAUDIN4	Claudin 4 Monoclonal Antibody (3E2C1)	Invitrogen	32-9400	1:100
COL1A1	COL1A1 (E3E1X) Mouse mAb	Cell Signaling	66948	1:100
EGFR	EGF Receptor (E746-A750del Specific) (D6B6) XP® Rabbit mAb	Cell Signaling	2085	1:100
EPCAM	EpCAM (VU1D9) Mouse mAb	Cell Signaling	2929	1:400
FOXP3	FoxP3 (D6O8R) Rabbit mAb	Cell Signaling	12653	1:400
GAPDH	GAPDH (D4C6R) Mouse mAb	Cell Signaling	97166	1:1000
HNF4G	HNF4G Rabbit Polyclonal antibody	Proteintech	25801-1-AP	1:1000
IGHA1	Human IgA Heavy Chain Polyclonal antibody	Proteintech	11449-1-AP	1:800
HSP90AA1	HSP90 alpha/beta rabbit mAb	ZenBio	R24635	1:50
INHBA	Inhibin Beta A Polyclonal antibody	Proteintech	10651-1-AP	1:100
Secondary antibodies	Alexa FourTM 488 goat anti-mouse IG(H+L)	Invitrogen	2610355	1:500
	Alexa FourTM 555 goat anti-rabbit IG(H+L)	Invitrogen	2633537	1:500

retrieval, sections were permeabilized, blocked to reduce non-specific binding, and incubated with primary antibodies (Table 1). After washing, enzyme-conjugated secondary antibodies were applied for detection. Counterstaining was used to visualize cellular morphology before mounting and microscopic analysis of protein expression patterns.

Cell culture and transfection

Colorectal cancer cell line HT29 was cultured according to the standard protocol. The siRNA targeting HNF4A (si-1 sense: 5'-GGCCAAUCGUGUUCUAGA U-dTdT-3' and antisense: 5'-AUCUAGAACACGAUUGGCC-dTdT-3'; si-2 sense: 5'-GGCUAAGCGAUC CAGUAAA-dTdT-3' and antisense: 5'-UUUACUGGA UCGCUUAGCC-dTdT-3') were synthesized by Shanghai Genechem Co., Ltd. HT29 cells were transfected with the relevant siRNAs using Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The impact on target protein expression was assessed by Western blotting 48 h post-transfection.

Western blotting

Western blotting was performed as previously described [32]. Proteins were extracted using RIPA lysis buffer containing protease inhibitors (Roche), quantified by BCA assay (Thermo Scientific), and separated by SDS-PAGE. After transfer to nitrocellulose membranes, blocking with 5% milk, and incubation with anti-HNF4G primary antibody (Table 1) at 4 °C for overnight, membranes were probed with HRP-conjugated secondary antibodies (Invitrogen) at room temperature for 1 h. Protein bands were detected using SuperSignal West Pico substrate (Thermo Scientific) on an iBright imaging system, with GAPDH as the loading control.

CCK-8 and colony formation assays

We utilized the Cell Counting Kit-8 (CCK-8) assay to assess cell viability.

Cells were transfected with HNF4G siRNAs, along with the control siRNA, and subsequently seeded into 96-well plates at a density of 4×10^3 cells/200 μ l per well. Cell viability was then evaluated using a CCK-8 (Solarbio, CA1210) assay at a wavelength of 450 nm. The colony formation assay was also used to detect cell proliferation, with 2000 transfected cells from each group being plated in 6-well plates and cultured. After 1 week, cell colonies in the plates were fixed with 4% paraformaldehyde (Servicebio, G1101) and stained with 1% crystal violet (Servicebio, G1014). The number of cell colonies was counted using Image J.

Statistical analysis

All statistical analyses and graphical representations were performed using SPSS software (V 21.0) and GraphPad Prism 9. A t-test was employed to assess the correlation with parametric variables between two independent samples. Statistical significance was determined at $p < 0.05$, indicated as follows: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Results

Clinical characteristics of ten colorectal tissues and generation of single-cell transcriptome data

Ten colorectal specimens spanning normal mucosa (N1, N2, N3), inflammatory polyps (P1, P12), adenomatous polyps (P3, P4), and CRC (T2, T3, T5) were analyzed using 10 \times Genomics single-cell RNA sequencing to elucidate transcriptomic changes during CRC progression (Fig. 1A). Histopathological analysis (Fig. 1B) revealed progressive architectural alterations: normal tissues exhibited intact crypts and vasculature; inflammatory polyps showed pedunculated morphology; adenomatous polyps displayed glandular hyperplasia; and CRC specimens demonstrated invasive glands, epithelial stratification, and increased mitotic activity.

Quality-controlled single-cell RNA sequencing yielded 51,819 transcriptomes from ten samples (sheet 'Cell_Number' in Table S1), with malignant tissues showing elevated gene detection (average 1624 genes/cell versus 1357, 1353 and 1624 genes, respectively, in non-malignant states), indicating increased gene transcription activity in malignant colorectal tissues, particularly in cancerous lesions. Unsupervised clustering using Seurat pipeline [33] of 4000 highly variable genes (HVGs) identified 23 clusters (Fig. S1A) representing 10 cell types (Fig. 1C). Excluding rare populations (smooth muscle, neural, and mast cells), analysis revealed distinct pathological associations (Fig. S1B, Fig. 2A, and sheet 'Percentage_Cell_Type' in Table S1): adenomatous polyps were T cell-enriched (75.66%), while plasma cells decreased progressively from normal (36.42%) to CRC (8.77%), suggesting tumor-protective functions. In contrast, inflammatory polyps showed elevated B cells (11.16%), fibroblasts (17.04%), epithelial cells (20.46%), and endothelial cells (8.30%). Cell identity was validated through marker gene heatmaps (Fig. 2B), dot plots (Fig. S1C), and the top 25 specific markers (Table S2).

The expression of ligand–receptor genes can be utilized to identify the (sub-) cell types present in colorectal tissues

Analysis of Fig. S2B and Table S2 revealed a substantial representation of ligand–receptor genes among the identified markers. From a curated database of 2593 ligand–receptor pairs (sheets 'Receptor_ligand_interaction' and

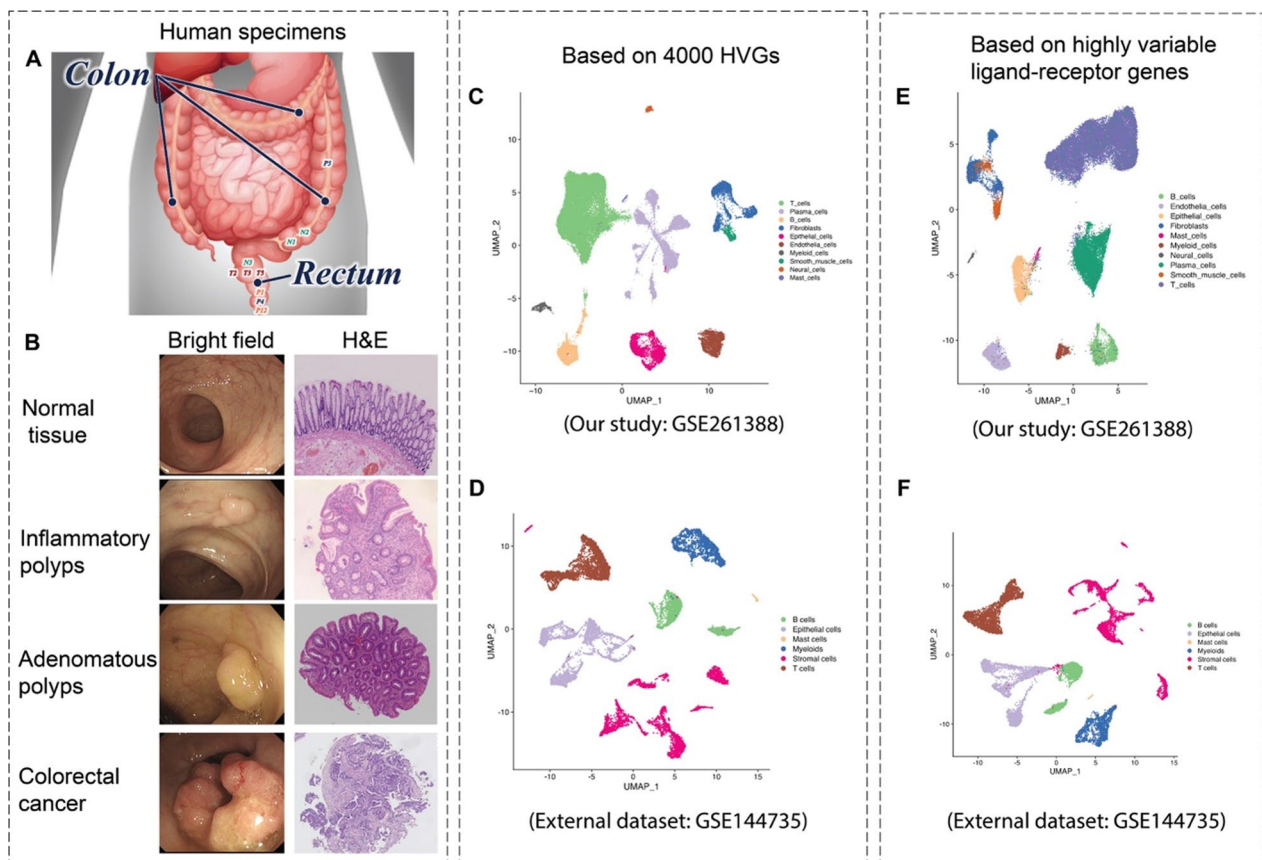


Fig. 1 Characterization of human colorectal samples and cell type identification based on highly variable genes (HVGs) and highly variable ligand/receptor genes. **A** Ten human specimens were collected from the colon or rectum under informed consent for single-cell RNA sequencing. These included three samples from normal colon mucosa (N1, N2, and N3), two samples representing inflammatory polyps (P1 and P12), two samples representing adenomatous polyps (P3 and P4), and three samples representing CRC (T2, T3, and T5). While the CRC tissues were surgically obtained, the remaining seven samples were collected via endoscopic examination. **B** Brightfield and H&E images of the four types of human colorectal specimens. NA: not applicable. H&E: Hematoxylin and eosin staining. **C, E** UMAP plots illustrating cell type annotation with corresponding color codes from our study (GSE261388) based on the 4000 HVGs (**C**) and the 1062 highly variable ligand/receptor genes (**E**), respectively. **D, F** Classification of six cell types from an external CRC scRNA-seq dataset (GSE144735) based on the 4000 HVGs (**D**) and the 1128 highly variable ligand/receptor genes (**F**), respectively

'2593_genes' in Table S3), we found 1062 overlapped with our 4000 HVGs (Fig. S3A, sheet '4000_genes' in Table S3). Strikingly, 40% of the top 25 cell-type-specific markers (Table S2) were ligand–receptor genes (Fig. S3B). Clustering analysis using only the 1062 ligand–receptor genes successfully recapitulated the cellular composition obtained with all 4000 HVGs, generating consistent UMAP projections (Fig. 1E) and maintaining equivalent cell type distributions (Fig. 1C vs E). Marker gene expression patterns (Fig. S2C, D) showed strong concordance with the original analyses (Figs. S1C & S2B). The ligand–receptor approach effectively resolved major cell types (T cells, B cells, plasma cells, epithelial cells, fibroblasts, endothelial cells, myeloid cells) into distinct subpopulations (Fig. S3C, E, G, I, K & M), with clustering

correlations between methods being robust for all cell types except heterogeneous T cell subsets (Fig. S3D, F, H, J, L & N), as further detailed in later analyses. Validation analyses confirmed the specificity of the 1062 ligand–receptor genes for cell type identification, as alternative gene sets showed minimal overlap with the 4000 HVGs (Fig. S4). Notably, clustering using 87 GPCR genes failed to recapitulate the cellular heterogeneity observed with the full 4000 HVGs (Fig. 1C vs Fig. S5), demonstrating their limited contribution to cell type distinction.

We then validated the generalizability of ligand–receptor (LR) gene set for cell typing by analyzing an independent CRC scRNA-seq dataset (GSE144735) [34] comprising matched colorectal cancer (CRC) and normal mucosa samples from three patients (Tumor:

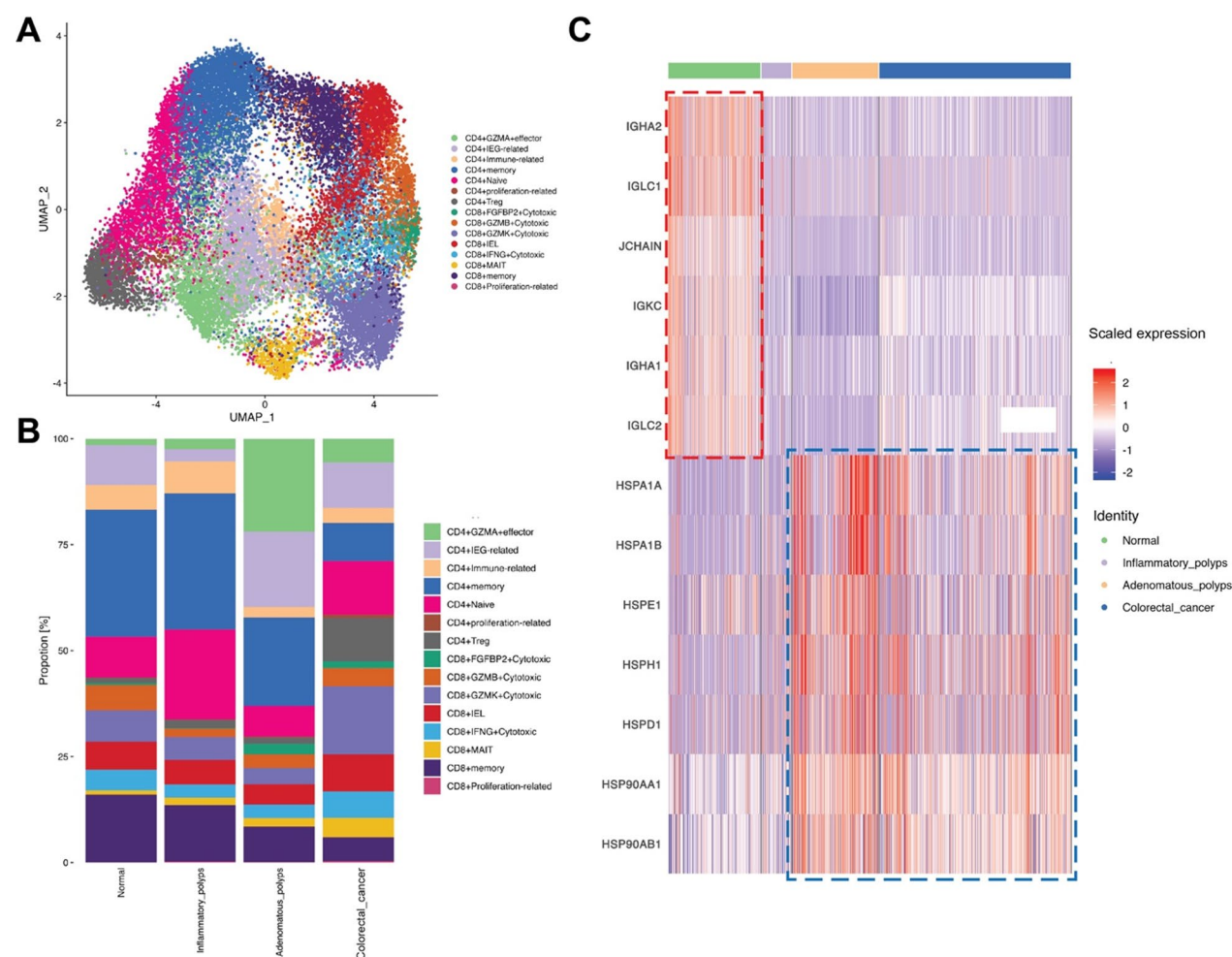


Fig. 2 Characterization of T cells. **A** UMAP plot illustrating the annotation of fifteen sub-cell types within T cells from the ten samples, color-coded accordingly. **B** Histogram presenting the proportions of the fifteen sub-cell types of T cells across the four types of human colorectal specimens. **C** Heatmap displaying the expression patterns of selected genes encoding immunoglobulins (highlighted in the red dashed square) and heat shock proteins (highlighted in the blue dashed square) among the four types of colorectal specimens

GSM4294474, GSM4294475 & GSM4294476; Normal: GSM4294486, GSM4294487 & GSM4294488). Applying identical data processing to both the external and our original data, we compared clustering using either 4000 HVGs or the resulting 128 LR-derived HVGs. Both approaches demonstrated comparable classification accuracy (Fig. 1D and F), confirming LR genes' robustness across datasets. This cross-cohort validation supports the potential universal utility of LR-based classification for colorectal tissues, with possible extension to other tissue types.

Integrating single-cell and bulk transcriptomic data from ten colorectal samples, we identified stage-specific molecular signatures across CRC progression (Fig. S6A, Table S4). Receptor-ligand analysis revealed *IGHA1* (normal), *TFF1* (inflammatory polyps), *NR1D1* (adenomatous polyps), and *SFRP2* (CRC) as predominant

markers (Fig. S6A), suggesting their diagnostic potential. Immune cell profiling (Table S4) uncovered *NR1D1* and *SFRP2* being top ranked in adenomatous and CRC samples, respectively (Fig. S6). These findings nominate *IGHA1*, *TFF1*, *NR1D1* and *SFRP2* as candidate biomarkers for pathological stratification and therapeutic targeting in CRC.

Our findings demonstrate that ligand–receptor (LR) gene expression patterns reliably identify colorectal cell types, revealing non-random organization. This cellular diversity likely reflects specialized LR-mediated communication. While prior work successfully distinguished fibroblasts from mural cells using a minimal 90-gene panel [35], broader validation across diverse tissues remains necessary. Subsequent analyses employed the 4000 HVGs, stratified by cell abundance.

High cellular heterogeneity and molecular demarcation of T cells

T cells, playing a vital role in CRC progression and clinical outcomes [36, 37], emerged as the predominant population in our single-cell analysis (Table S1, Fig. S2A), displaying remarkable heterogeneity through unsupervised clustering (Fig. 2A). Utilizing established marker genes for T cells [38–41], we identified eleven T cell clusters (Fig. S7A & D) comprising CD4+ and CD8+ subtypes (Fig. S7B), further classified into effector, memory, naïve, regulatory, and cytotoxic subsets (Fig. 2A, Figs. S7E, S8–9, Tables S7–9). Proportional analysis revealed clinically relevant patterns (Fig. 2B and Fig. S7C): (1) CD8+ MAIT cell expansion correlated with pathological progression [42] and poorer patient outcome [43], while CD8+ memory cell depletion aligned with their favorable prognosis in CRC infiltration [44, 45]; (2) inflammatory

polyps exhibited CD4+ Naïve cell predominance and lowest percentage of CD8+ cells; (3) Adenomatous polyps exhibited CD4+ GZMA+ effector and CD4+ IEGs (immediate early genes) cell enrichment [46]; (4) CRC tissues demonstrated increased CD4+ Treg cells (Fig. 3) and CD8+ GZMK+ cytotoxic cells, the latter supporting prior CRC progression associations [47]. These findings underscore the dynamic T cell landscape throughout CRC pathogenesis.

Upon comparing the single-cell expression profiles of T cells across the four types of colorectal tissues, we detected elevated expression of certain immunoglobulin (Igs) genes, including *IGHA1*, *IGHA2*, *IGKC*, *IGLC1*, *IGLC2* and *JCHAIN* in T cells from normal colorectal tissues (highlighted in the red dashed square in Fig. 2C). While Igs were historically considered exclusive to plasma-lineage cells, and the above-identified Ig

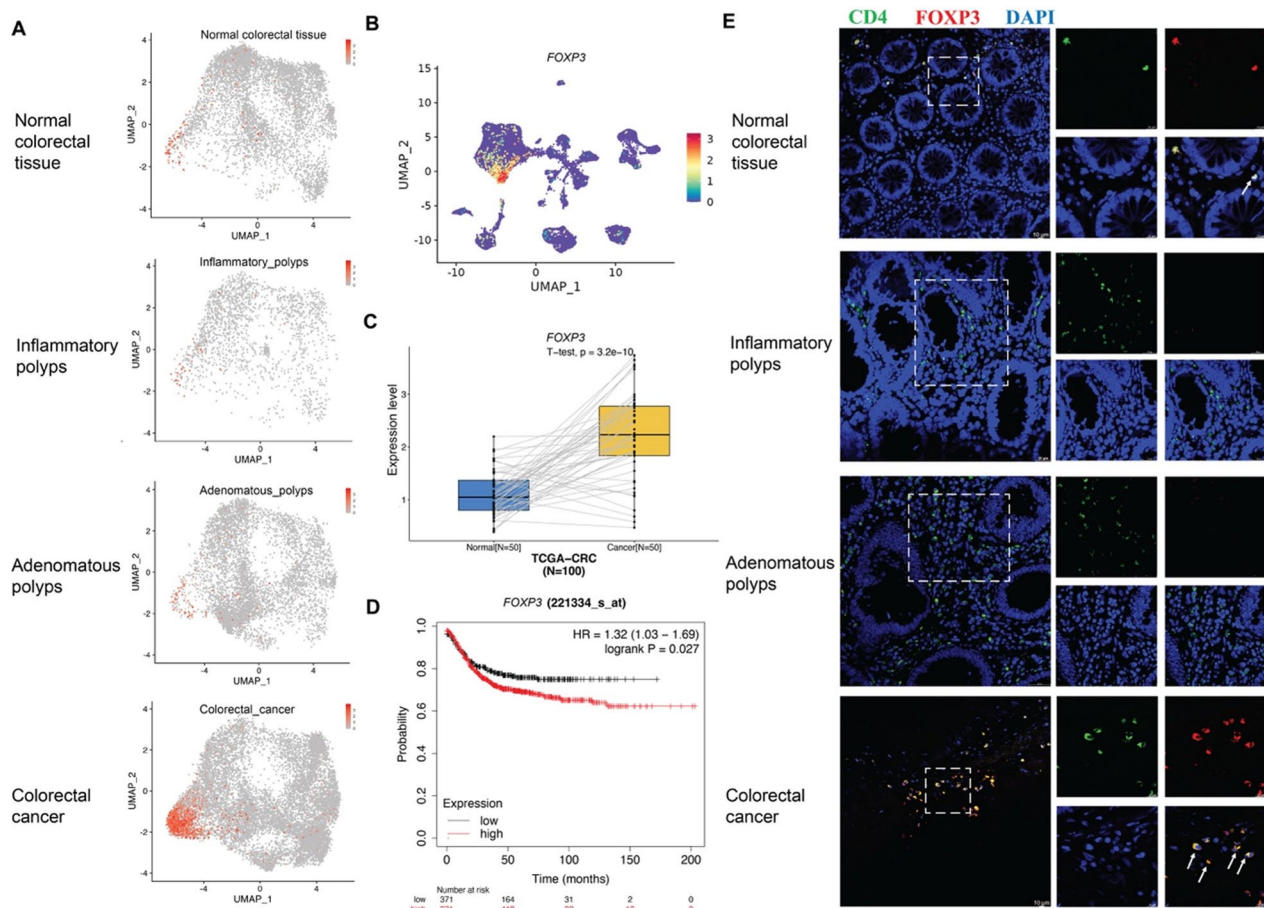


Fig. 3 CD4+ Treg cells are enriched in colorectal cancer tissues. **A** Feature plots displaying the expression of *FOXP3* in T cells across the four types of human colorectal samples. **B** A feature plot illustrating the expression of *FOXP3* in all cell types, with red indicating high expression levels. **C** A box plot presenting the expression level of *FOXP3* in paired CRC and normal tissues sourced from the TCGA database. **D** The Kaplan Meier plot demonstrating that elevated expression of *FOXP3* is indicative of poor prognosis in CRC patients. **E** Immunofluorescence analysis of Treg cells. Green: CD4; red: FOXP3. Nuclei (blue) were counter-stained with DAPI. Tissues within the white dashed squares were imaged at higher magnification as displayed on the right side. White arrows indicate Treg cells co-expressing CD4 and FOXP3. A scale bar is present in each image. Each staining was performed on tissue slides from three distinct specimens

genes typically exhibit highest expression in plasma cells based on our colon scRNA-seq database (<https://www.colon-singlecell.cn>), rigorous doublet removal (Doublet-Finder v2.0.3) and distinct clustering of T vs. plasma cells (Fig. 1C, Figs. S1C & S2B) in this study demonstrated minimal classification artifacts and assured the data validity. Previous evidence revealed Ig expression in non-plasma cells. For instance, T cells were shown to express Ig VH genes, albeit with qualitative distinctions from B-cell Igs [48], which exemplifies the existence of immunoglobulin transcripts in T cells. Furthermore, single-cell analyses reveal upregulated Ig genes (e.g., *IGKC*, *IGLC2*, *IGHA1* and *IGHG3*) in tumor-infiltrating T cells in lung adenocarcinoma and metastases [49, 50]. Functionally, tumor-derived Igs promote immune evasion [51] and are detectable in the tumor microenvironment [52]. The presence of these genes in T cells in normal colorectal tissues suggests a unique expression pattern worth further investigation. We also acknowledge the need for orthogonal validation (e.g., protein-level detection, TCR-Ig co-expression assays) and mechanistic studies to determine whether this phenotype contributes to immune surveillance or tissue homeostasis.

Adenomatous and CRC tissues showed marked upregulation of heat shock protein genes (*HSPA1A*, *HSPA1B*, *HSPD1*, *HSPE1*, *HSPH1*, *HSP90AA1*, and *HSP90AB1*; highlighted in the blue dashed square in Fig. 2C), particularly in T cells, suggesting tumor microenvironment involvement. TCGA validation (Figs. S8A–B), immunofluorescence (Fig. S8C), and IHC (Fig. S8D) confirmed CRC-specific HSP elevation, while immunoglobulin genes (*IGHA1*, *IGHA2*, *IGHC2* and *IGKC*) were downregulated. Notably, nonresponsive tumors exhibited elevated HSP expression in T cells post-immunotherapy [53], revealing a stress response signature. These stage-specific expression patterns (HSPs vs Igs) highlight their potential as complementary biomarkers when integrated with T-cell profiling for improved CRC diagnostics and prognostics [54, 55].

Our analysis revealed increased CD4+ GZMA+ effector cells in adenomatous and CRC tissues (Figs. S9B, C), consistent with GZMA's established role in CRC-associated inflammation [56]. Notably, we observed significant CD4+ FOXP3+ Treg infiltration in CRC samples (Fig. 3A, Figs. S7B, and S9C), with *FOXP3* expression specifically enriched in Tregs (Fig. 3B). High expression levels of *FOXP3* in CRC samples were confirmed by bulk RNA-seq data from the TCGA-CRC cohort (Fig. 3C). Clinical correlation analysis demonstrated that elevated *FOXP3* expression associates with poorer CRC prognosis using the Kaplan Meier plotter tool with defaulted parameters (Fig. 3D), while immunofluorescence confirmed Treg accumulation in CRC

tissues (Fig. 3E). These findings collectively implicate both GZMA+ effector T cells and FOXP3+ Tregs in CRC progression.

Complementing the analyses in Fig. 3, we further systematically profiled *FOXP3*-regulated target genes (such as *CTLA4* and *IL10*) to delineate its transcriptional regulatory role in shaping the immunosuppressive CRC microenvironment. As demonstrated in Fig. S10, *CTLA4* and *IL10* exhibit Treg cell-enriched expression patterns that align spatially and quantitatively with *FOXP3* expression in colorectal samples—a finding consistent with established mechanisms of Treg-mediated immunosuppression via checkpoint ligands (e.g., *CTLA4*) [57] and anti-inflammatory cytokines (e.g., *IL10*) [58].

Pseudotime analysis of CD4+ T cells (Fig. 4A) revealed a differentiation trajectory from naïve/memory to effector states (Fig. 4B), consistent with established T cell biology that naïve T cells differentiate into effector cells through proliferation. Module analysis identified dynamically expressed ligand–receptor genes during CRC progression (Fig. 4C & Table S5), suggesting stage-specific functions. SCENIC analysis [26] nominated *ARID5B* as a pivotal TF in CRC-associated CD4+ T cells (Fig. 4D). This finding is supported by its elevated expression in T cells from CRC patients (Fig. 4E, top panel), heightened expression in TCGA-CRC tumors (Fig. 4E, middle panel), and an association with poorer survival (Fig. 4E, bottom panel). These findings position *ARID5B* as a key regulator of CD4+ T cell dynamics in CRC pathogenesis.

Pseudotime analysis of three CD8+ T cell subtypes (Fig. 5A) revealed a differentiation trajectory from cytotoxic to memory to proliferation-related states (Fig. 5B), with the latter exclusively detected in CRC samples (Fig. S9G). Dynamic expression patterns of ligand–receptor genes (Fig. 5C, Table S6) suggested stage-specific functions during CRC progression. SCENIC analysis identified *IRF9* and *SOX4* as key transcriptional regulators in CD8+ T cells (Fig. 5D). This discovery was supported by their elevated expression in CRC patient T cells (Fig. 5E, top panel), increased levels in TCGA-CRC tumors (Fig. 5E, middle panel), and an association with poorer survival (Fig. 5E, bottom panel), implicating these TFs in CD8+ T cell-mediated CRC pathogenesis.

Finally, comparative analysis of CD4+ and CD8+ T cell ligand–receptor genes (Figs. 4C & 5C, Tables S5–6) identified five shared genes (*ANXA1*, *CCL4*, *CCR7*, *KLRB1*, and *LTB*) with distinct expression modules between these two cell subtypes. This differential regulation suggests cell-type-specific functional roles for these genes in T cell-mediated immune responses during CRC progression.

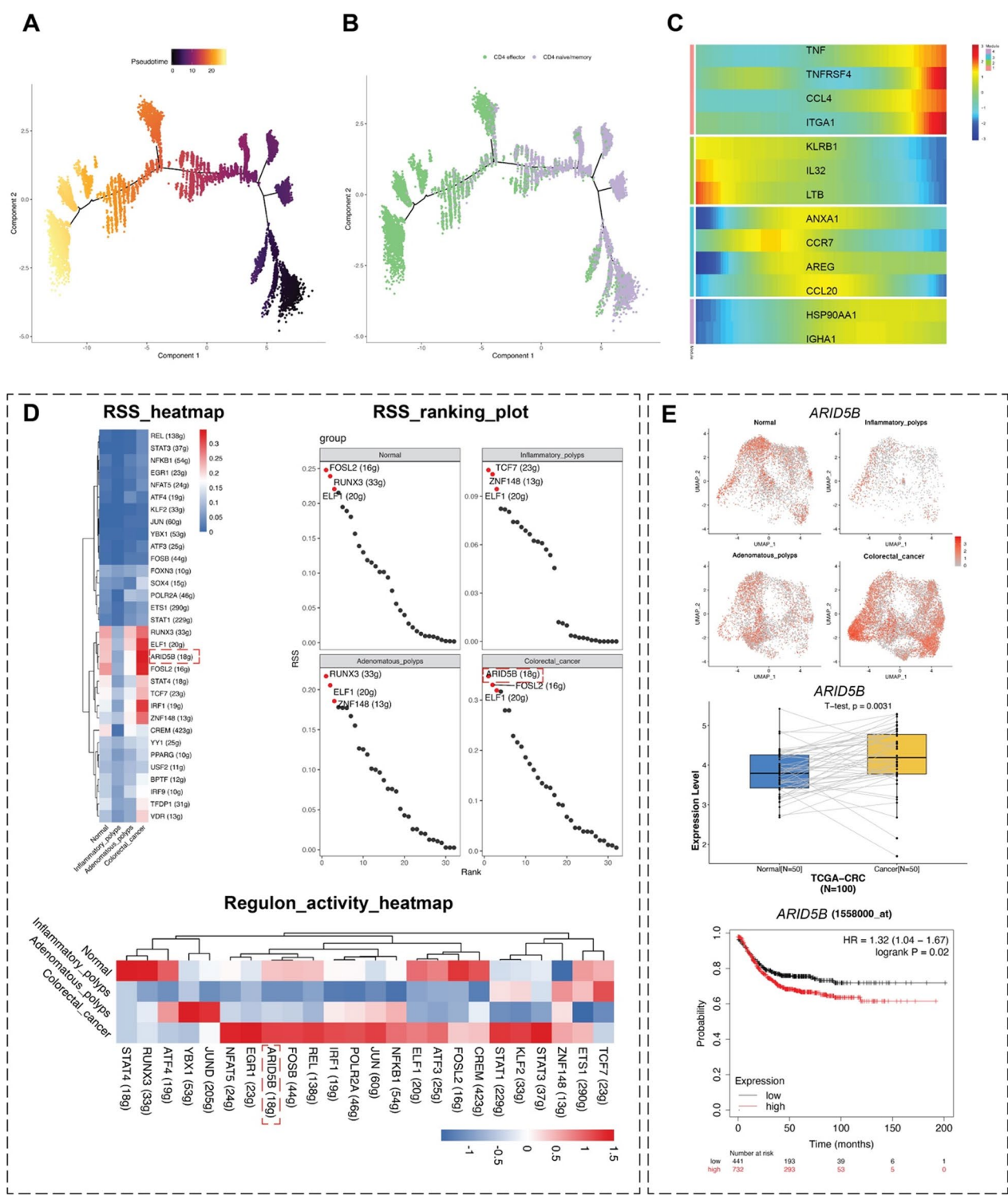
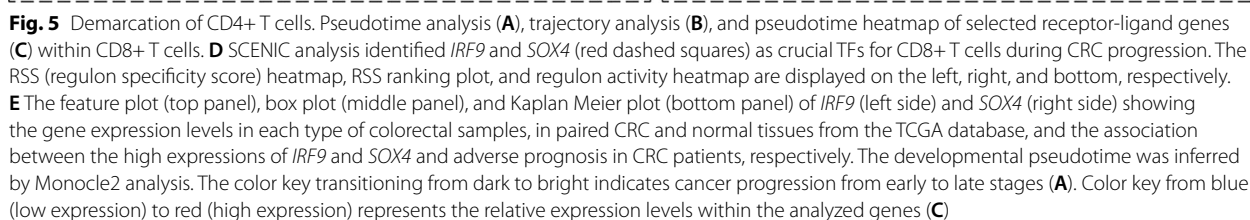


Fig. 4 Demarcation of CD4+ T cells. Pseudotime analysis (**A**), trajectory analysis (**B**), and pseudotime heatmap of selected receptors-ligand genes (**C**) for CD4+ T cells. **D** SCENIC analysis identified *ARID5B* (red dashed squares) as a key TF for CD4+ T cells during CRC progression. The RSS (regulon specificity score) heatmap, RSS ranking plot, and regulon activity heatmap are presented on the left, right, and bottom, respectively. **E** The feature plot displaying the gene expression of *ARID5B* in the four types of colorectal samples (top panel), box plot illustrating the gene expression of *ARID5B* in matched CRC and normal tissues sourced from the TCGA database (middle panel), and the Kaplan Meier plot indicating that elevated *ARID5B* expression correlates with an unfavorable prognosis in CRC patients (bottom panel). The developmental pseudotime was inferred by Monocle2 analysis. The color key transitioning from dark to bright indicates cancer progression from early to late stages (**A**). Color key from blue (low expression) to red (high expression) represents the relative expression levels within the analyzed genes (**C**)



Plasma and B cells, the second and third highest proportions according to Table S1 and Fig. 1C and Fig. S2A, were jointly analyzed due to their functional and evolutionary

relationships. Unsupervised clustering identified 15 populations (Fig. 6A), classified into plasma cells (IgA/IgG/IgM subtypes) and B cells (naïve/active) using specific marker genes (Fig. 6C, D and Fig. 7A). Distribution analysis revealed naïve B cell enrichment in inflammatory

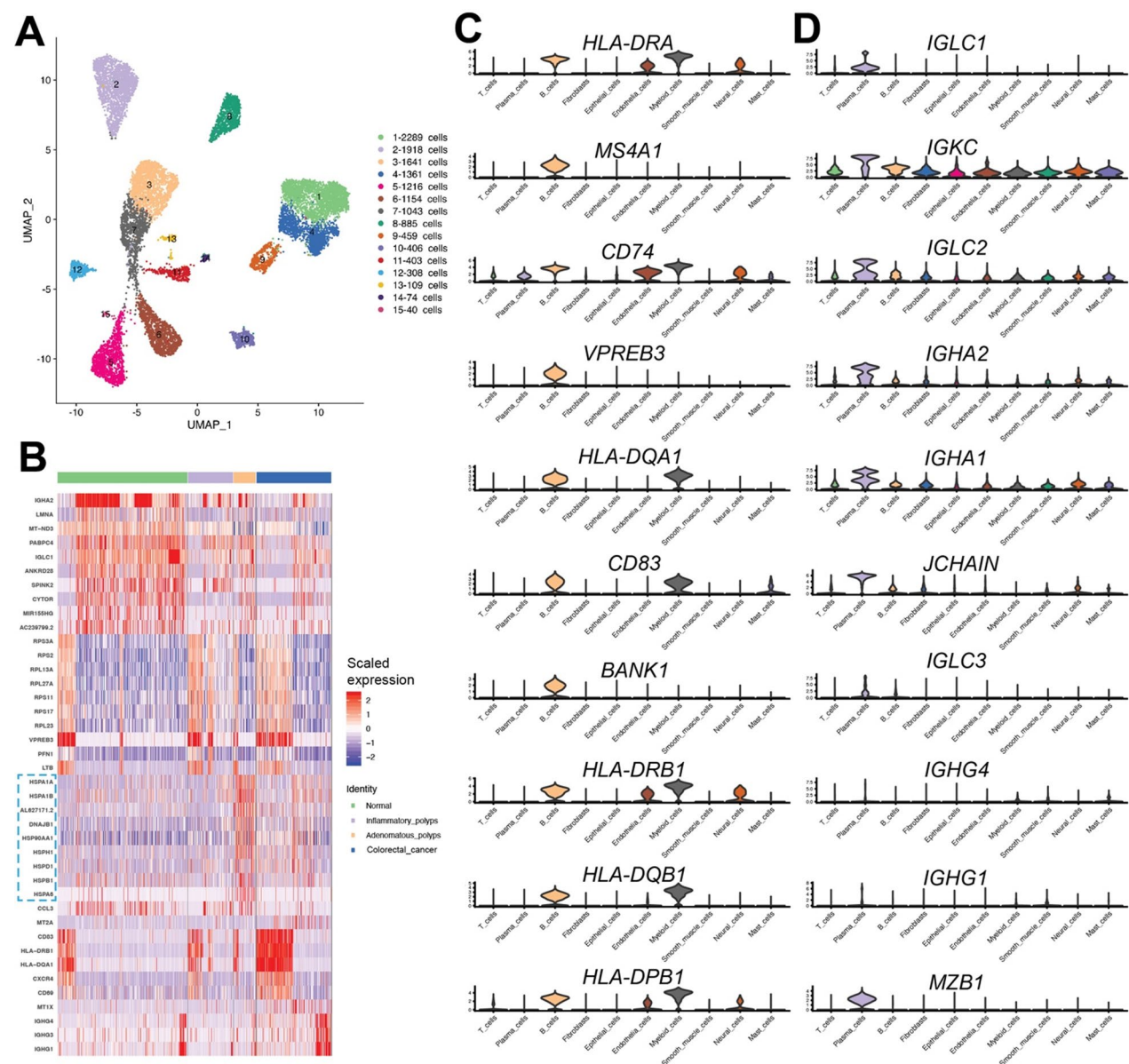


Fig. 6 Clustering and determination of B cells and plasma cells. **A** The UMAP plot visualizing the fifteen clusters of B cells and plasma cells, with corresponding color codes. **B** A heatmap illustrating the expression patterns of the top 10 marker genes specific to B cells and plasma cells across the four types of colorectal specimens. The blue dashed square highlighted the seven heat shock protein genes among the top 10 highly expressed genes within the colorectal samples featuring adenomatous polyps. **C, D** Violin plots showing the expression levels of the top 10 marker genes for B cells (**C**) and plasma cells (**D**)

polyps, while pathological progression correlated with IgA plasma cell reduction and increased active B cells and IgG plasma cells (Fig. 7C, D). While marker genes distinguished the five plasma and B cell subtypes (Fig. 7B, Table S10), none differentiated these cells across colorectal specimen types (Fig. 6B). However, adenomatous polyps showed modest upregulation of heat shock genes (*HSPA1A*, *HSPA1B*, *HSPA6*, *HSPB1*, *HSPD1*, *HSPH1*, and *HSP90AA1*; highlighted within the blue dashed square

in Fig. 6B), suggesting a stress response association with early neoplasia. As shown in Figs. 6A & 7A, we identified two IgG plasma cell subtypes (clusters 8 and 10) distinguished by elevated *IGKC* (cluster 8) versus *IGLC2/3* (cluster 10) expression. Although no single genes were identified to distinguish the six clouds of IgA plasma cells (clusters 2, 3 and 7, 5, 6, 11, and 12) from each other (Fig. 7F), they formed three groups based on the expression levels of

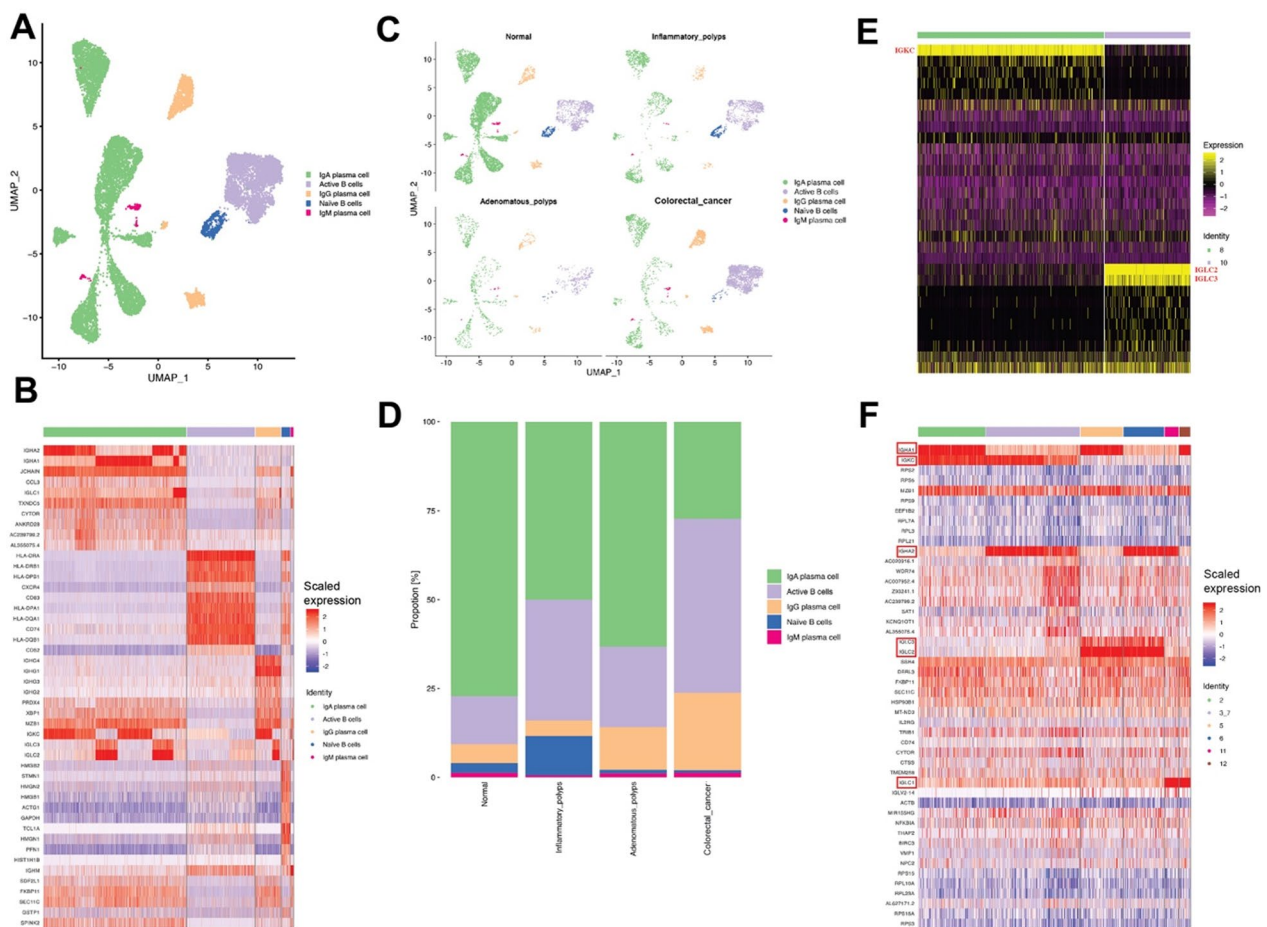


Fig. 7 Demarcation of B cells and plasma cells. **A** The UMAP plot illustrating the annotation of two sub-types of B cells and three sub-types of plasma cells, with corresponding color codes. **B** A heatmap displaying the expression patterns of the top 10 marker genes across the five populations of B cells and plasma cells. A UMAP plot (**C**) and a histogram (**D**) displaying the proportions of the five sub-cell types of B cells and plasma cells in the four types of colorectal tissue samples. **E** A heatmap suggesting the DEGs between the two clusters of IgG plasma cells (clusters 8 and 10 in Fig. 6A and cells color-coded in beige in A). The genes highlighted in red are highly differentially expressed between these two clusters. **F** A heatmap implying the DEG among the six clusters of IgA plasma cells (clusters 2, 3 and 7, 5, 6, 11, and 12 in Fig. 6A, and cells color-coded in green in A). The expression levels of the genes highlighted in red dashed squares can help differentiate these clusters of IgA plasma cells

IGKC and *IGLC2/IGLC3*: *IGKC*^{high}*IGLC2/IGLC3*^{low} (clusters 2, 3 and 7), *IGKC*^{low}*IGLC2/IGLC3*^{high} (clusters 5 and 6) and *IGKC*^{low}*IGLC2/IGLC3*^{low} (clusters 11 and 12). These immunoglobulin heterogeneity patterns in plasma cells may influence CRC prognosis and immunotherapy response [59–63].

Furthermore, considering the combined expressions of additional immunoglobulin genes, we differentiated the six clouds of IgA plasma cells (Fig. 4J) as follows: cluster 2, *IGHA1*^{high}*IGKC*^{high}; cluster 3 and 7, *IGHA2*^{high}*IGKC*^{high}; cluster 5, *IGHA1*^{high}*IGLC2*^{high}*IGLC3*^{high}; cluster 6, *IGHA2*^{high}*IGLC2*^{high}*IGLC3*^{high}; cluster 11, *IGHA2*^{high}*IGLC1*^{high}; and cluster 12, *IGHA1*^{high}*IGLC1*^{high}. We then performed Gene Set Enrichment Analysis (GSEA) for each of these

three groups versus the other two. The data (Fig. S11) shows common suppression of immunoglobulin production and immune response. The inhibition of IgA plasma cell function is not only a driving factor for CRC progression, but also a key link in immunotherapy resistance. Restoring IgA function through multidimensional intervention may provide a new direction to improve prognosis and enhance the efficacy of immune checkpoint inhibitor.

Abundant tumor-like epithelial cells and high expression of HNF4G in CRC samples

The intestinal epithelium, recognized as the origin of CRC [64] comprised 3814 epithelial cells (Fig. S12A).

They were identified based on characteristic marker genes for epithelial cells (Fig. S12C) and further classified into nine clusters (Fig. S12A) and four subtypes (Fig. 8A, B and Fig. S12B): BEST4+OTOP2+pH-sensing cells (abundant in normal tissues; dysregulated in IBD/CRC) [65], distal absorptive cells (prevalent in inflammatory

polyps), enteroendocrine cells, and tumor-like cells (CRC predominant).

Marker genes distinguished four epithelial subtypes (Fig. S12D & Table S11): *SLC12A2* for distal absorptive epithelial cells; *EGFR*, *CEACAM6*, *PRAP1*, *TRIM31*, *MUC13*, *SPINK1*, and *REG4* for tumor-like epithelial

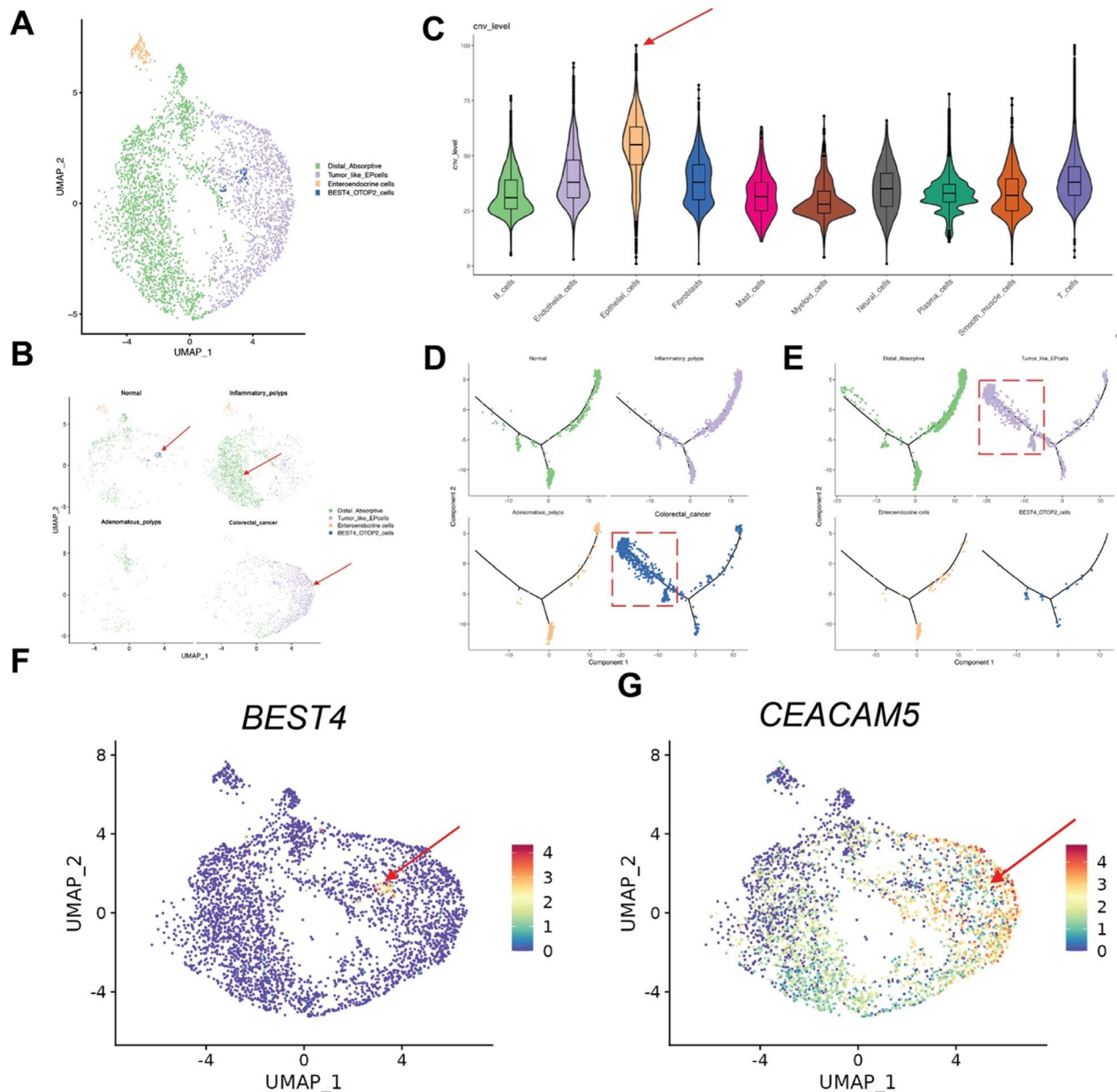


Fig. 8 Tumor-like epithelial cells identified in epithelial cells from CRC tissues. **A** The UMAP plot displaying the annotation of four sub-types of epithelial cells with corresponding color codes. **B** A UMAP plot illustrating the distribution of the four sub-cell types of epithelial cells across the four types of colorectal samples, with red arrows indicating specific cell types within the corresponding sample types. **C** InferCNV analyses shown in the violin plot indicate the genomic variation and malignancy across all cell types, with a red arrow highlighting the epithelial cells. Epithelial cells of the four sub-types (**D**) and from each sample type (**E**) are depicted in developmental trajectory analyses. Tumor-like epithelial cells from CRC tissues are highlighted in the red dashed square. Feature plots of *BEST4*+ (**F**) and *CEACAM5*+ (**G**) expression in epithelial cells are highlighted by the red arrows, respectively

cells; *CHGA*, *CRYBA2*, *CHGB*, and *CACNA1A* for enteroendocrine cells; and *GUCA2A*, *GUCA2B*, *BEST4*, and *OTOP2* for *BEST4*+*OTOP2*+epithelial cells. As shown in Figs. S12D–E (highlighted in the dark blue squares) and Table S12, CRC samples uniquely expressed tumor-like markers (*EGFR*, *CEACAM6*, *PRAP1*, *TRIM31*, *ANXA1*, *MUC13*, *SPINK1*, and *REG4*), confirming their predominance in malignant epithelium (Fig. 8B). Notably, ten ribosome proteins-encoding genes exhibited high expression in inflammatory polyps (highlighted in the red square in Fig. S12E), suggesting inflammatory regulation [66].

Tumor microenvironments comprise malignant and stromal cells with dynamic interactions [67]. This study aimed to identify malignant cells among ten cell types by analyzing copy number variations (CNVs) inferred from single-cell gene expression profiles [68]. CNV analysis identified epithelial cells as the most genomically aberrant population (Fig. 8C), with tumor-like epithelial cells predominantly deriving from CRC samples (Figs. 8D, E). CRC tissues exhibited a loss of *BEST4*/*OTOP2* cells and a gain of *CEACAM5*+tumor-like epithelial cells (Fig. 8B, F–G), validated by *CLDN4* co-staining (Fig. 9A, B). These results corroborate known CRC signatures of *BEST4* suppression [69] and *CEACAM5* overexpression [70].

SCENIC analysis identified *HOXB13* as a key TF in inflammatory polyps (Fig. 10A, B), consistent with its involvement in inflammatory processes [71], tumor-suppressive role [72] and downregulation in CRC tumors [73]. Kaplan–Meier analysis (upper panel of Fig. 10C) indicated better survival with high *HOXB13* expression. Conversely, *HNF4G* emerged as a pro-tumorigenic factor in CRC progression (Fig. 10A, B), correlating with poor prognosis (bottom panel of Fig. 10C) and supporting previous oncogenic reports [74]. Using siRNA-mediated knockdown of *HNF4G* in the CRC cell line HT29 (Fig. S13A), we observed significant inhibition of cell viability (Fig. S13B, measured by CCK-8 assay) and proliferative capacity (Fig. S13C, assessed by colony formation assay). These results collectively demonstrate that *HNF4G* plays a crucial regulatory role in CRC pathogenesis and progression, supporting its potential as a therapeutic target.

Ligand gene *INHBA* and transcription factor *TWIST1* are highly expressed in cancer-associated fibroblasts (CAFs)

We identified 4107 fibroblast cells, which were classified into six subtypes (Fig. 11A and Fig. S14D, Table S13) and further seven clusters (Figs. S14B, C): typical (PDGFRA+POSTN+), immune-related (ADAMDEC1+), CAFs (INHBA+CTHRC1+), myofibroblasts (MYH11+ACTA2+), Ig-associated (IGKC+IGHA1/2+), and antigen-presenting CAFs (CD74+) (Fig. 11A, Fig. S14A–D; Table S13) [75].

Distribution analysis (Fig. 11B and Fig. S14E) revealed: (1) reduced immune-related fibroblasts in precancerous and CRC tissues (consistent with known *ADAMDEC1* suppression), (2) myofibroblast enrichment in inflamed tissues, aligning with their known role in inflammation [76]; (3). CAF predominance in CRC samples [20], correlating with poor prognosis [77]. Combinatorial trajectory analyses (Fig. 11C–E) confirmed CAF predominance in CRC samples, while four ligand–receptor modules (Fig. 11F) revealed stage-specific functional roles of specific ligand–receptor genes during CRC progression. Differential expression analysis identified elevated *ADIRF* in fibroblasts from normal and inflammatory polyp tissues (Fig. 11G, red square; Table S14), suggesting its potential tumor-suppressive role warrants further investigation.

Fibroblasts in normal colorectal tissues exhibited elevated immunoglobulin gene expression (*IGHA1*, *IGHA2*, *IGHC1*, *IGHC2*, *IGKC* and *IGHA1N*), as indicated in Fig. 11G, highlighted in the blue squares), potentially contributing to tissue homeostasis, with *IGHA1*+fibroblasts previously reported [78]. CRC-derived fibroblasts showed upregulated CAF markers (*SFRP2*, *SFRP4*, *CTHRC1*, *INHBA*, *THBS2*, *ASPN* and *MMP11*), as shown in Fig. 11G & Fig. S14D, Tables S13 & S14), including ligand–receptor genes such as *CTHRC1* and *INHBA*, which may be useful for subtype characterization (Fig. 11F–H). The predominance of *CTHRC1*+*INHBA*+CAFs in CRC (Figs. 11G, H) confirms their established role in promoting immune evasion and metastasis [79], with immunofluorescence verifying increased *INHBA*+*COL1A1*+fibroblasts in CRC tissues (Fig. 11I).

SCENIC analysis identified *TWIST1* as a key transcription factor in CAFs during CRC progression (Fig. 12A, B), supported by its elevated expression in CAFs (Fig. 12C, top right) and CRC tissues from TCGA cohort (Fig. 12C, bottom left). Kaplan–Meier analysis revealed its negative correlation with patient survival (Fig. 12C, bottom right), underscoring its clinical relevance in CRC.

Increase in the number of endothelial cells and myeloid cells in CRC tissues

Recent reviews [80] highlight the importance of tumor-associated endothelial cells (TECs) in CRC pathogenesis. The identification of 3005 endothelial cells (ECs) based on classic EC marker gene expression (Fig. S15A) led to their classification into nine clusters (Fig. S15B) and five sub-types (Figs. S15C & E, Table S15): aEC (anterior EC, marker genes *ACE* & *GJA4*), vEC (venous EC, marker genes *ACKR1*, *CLU*, *CPE* & *SELP*), c->aEC (capillary->anterior EC, marker genes *VWA*, *IGFBP5* & *SPARC*), epithelial-like EC (marker genes *MUC12* & *KRT8*) with immune characteristics (marker genes *IGHA1*, *IGKC*,

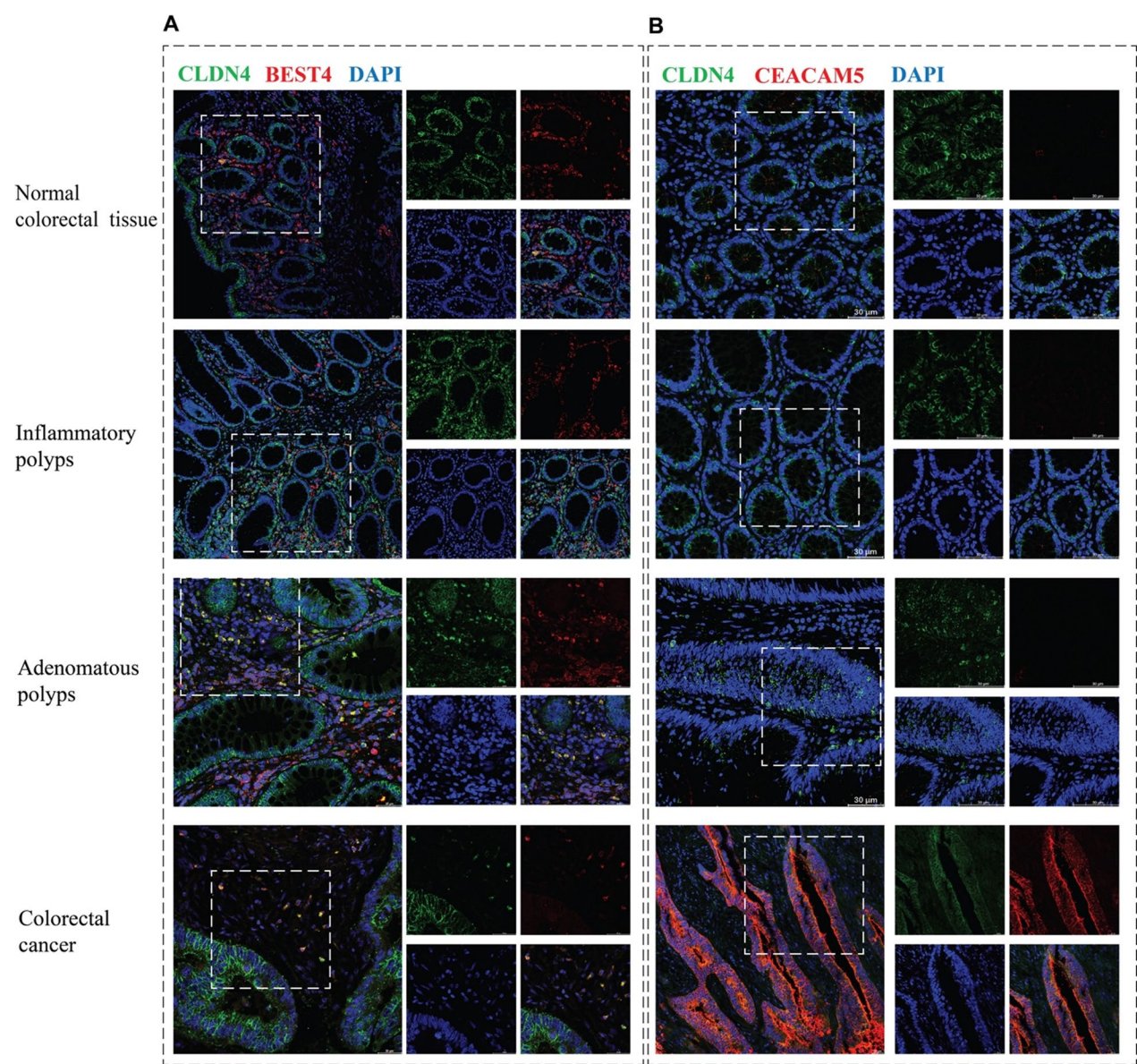


Fig. 9 Immunofluorescence staining visualizes BEST4+OTOP2+ and tumor-like epithelial cells in colorectal tissues. **A** Combined IF staining for CLDN4 (green channel, pan-marker gene for epithelial cells) and BEST4 (red channel) demonstrating the presence of BEST4+ epithelial cells in the four types of clinical colorectal samples, predominant in normal tissues. **B** Combined IF staining for CLDN4 (green channel) and CEACAM5 (red channel, pan-cancer cancer marker gene) showing the presence of CEACAM5+ epithelial cells, enriched in CRC tissue. Nuclei were counter-stained with DAPI. Tissues within the white dashed squares were imaged at higher magnification as displayed on the right side. Each staining was performed in triplicates using tissue slides sourced from three different individuals. A scale bar is present in each image

JCHAIN and *CXCL14*) and stressed EC (marker genes *RPL18A* and *RPS26*). CRC samples showed EC enrichment (Fig. S15D) with elevated *CDX1/IGFBP5/SELE* expression (Fig. S15F), potentially supporting tumor angiogenesis. DEG analysis revealed tissue-specific markers (benign: *FABP5*, *TXNIP*, *CD36*, *ADIRE*, and *ENPP2*; malignant: *AQP1*, *TGM2*, and *ADAMTS9*) with diagnostic potential (Fig. S15G; Table S16).

Myeloid cells are key regulators of tumor immunity in CRC, with implications for immunotherapy [20]. A total of 740 myeloid cells (Table S1, Fig. S16A, B, D) were annotated based on classic marker genes and classified into macrophages (*C1QA* & *C1QB*), cDC (*CD1C* & *CD1E*), tumor-associated myeloid cells (*SRRM2* and *MALAT1*) and monocytes (*S100A8*, *S100A9*, *FCN1* & *VCAN*). Predominantly found in CRC samples (Fig.

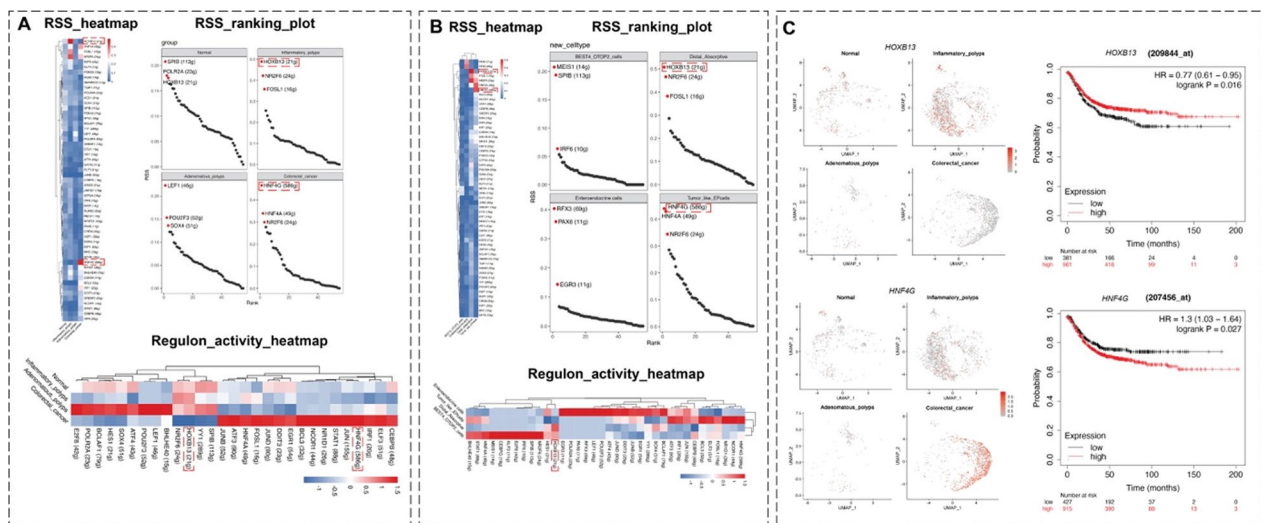


Fig. 10 Demarcation of epithelial cells. SCENIC analysis of epithelial cells based on the four types of colorectal samples (**A**) and the four cell sub-types (**B**). *HOXB13* and *HNF4G* are highlighted in the red dashed squares, with RSS (regulon specificity score) heatmap, RSS ranking plot, and regulon activity heatmap displayed on the left, right, and bottom, respectively. **C**. Feature plots (left panel) and Kaplan Meier plots (right panel) of *HOXB13* (top) and *HNF4G* (bottom) showing the gene expression levels in each type of colorectal samples and illustrating the correlation between the high expressions of *HOXB13* and *HNF4G* with the prognosis in CRC patients

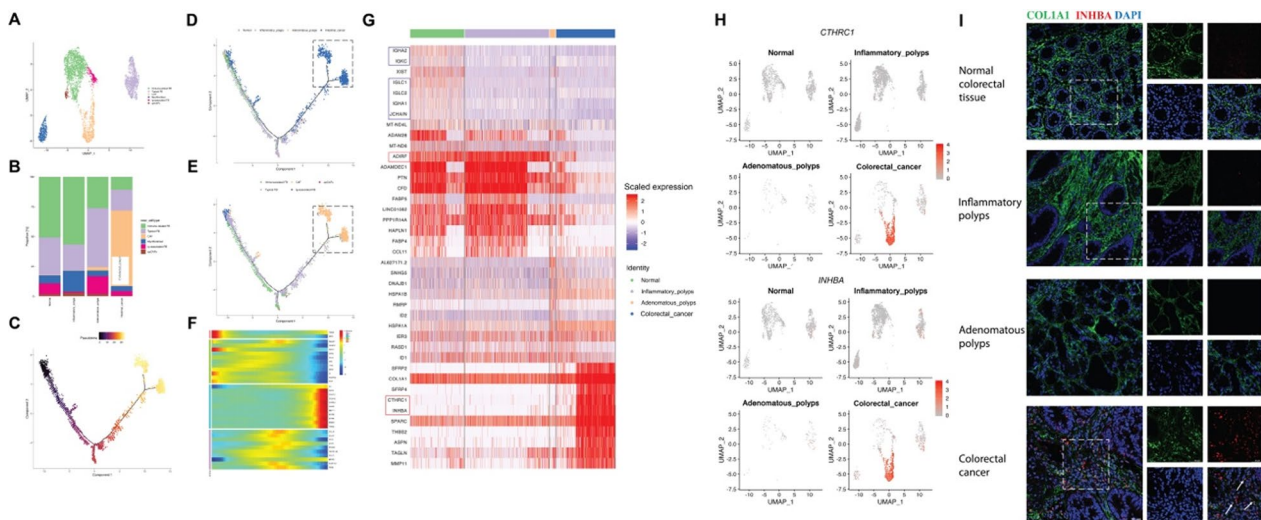


Fig. 11 Delineation of fibroblasts in colorectal tissues. **A** UMAP visualization representing the sub-type of fibroblasts with corresponding color codes. **B** A histogram presenting the proportions of the six sub-cell types of fibroblasts within the four kinds of colorectal samples. **C** Developmental pseudotime analysis of fibroblasts from all ten samples using Monocle2. The color key from dark to bright indicates pathological progression from normal colorectal tissues to CRC. **D** Fibroblasts from each type of colorectal sample displayed in the developmental trajectory analysis. **E** Developmental trajectory analysis demonstrating five selected subtypes of fibroblasts. **F** Heatmap displaying selected genes encoding ligands or receptors in fibroblasts, with a color key indicating relative expression levels from low (blue) to high (red). **G** Heatmap illustrating the DEGs of fibroblasts across the four types of colorectal samples. Immunoglobulin genes specific for normal colorectal tissues are highlighted in blue squares while *ADIRF* specific for non-cancerous colorectal samples are marked in red squares. **H** Feature plots displaying the expression of *COL1A1* (top panel) and *INHBA* (bottom panel) in fibroblasts across each kind of colorectal specimen. **I** Combined IF staining for *COL1A1* (green, pan-marker gene for fibroblasts) and *INHBA* (red) indicating the presence of *COL1A1* + *INHBA* + cancer-related fibroblasts (CAFs) in CRC tissue. Nuclei were counter-stained with DAPI. Tissues within the white dashed squares were imaged at higher magnification as shown on the right side. Each staining was performed using tissue slides from three different specimens of the same type. A scale bar is present in each image

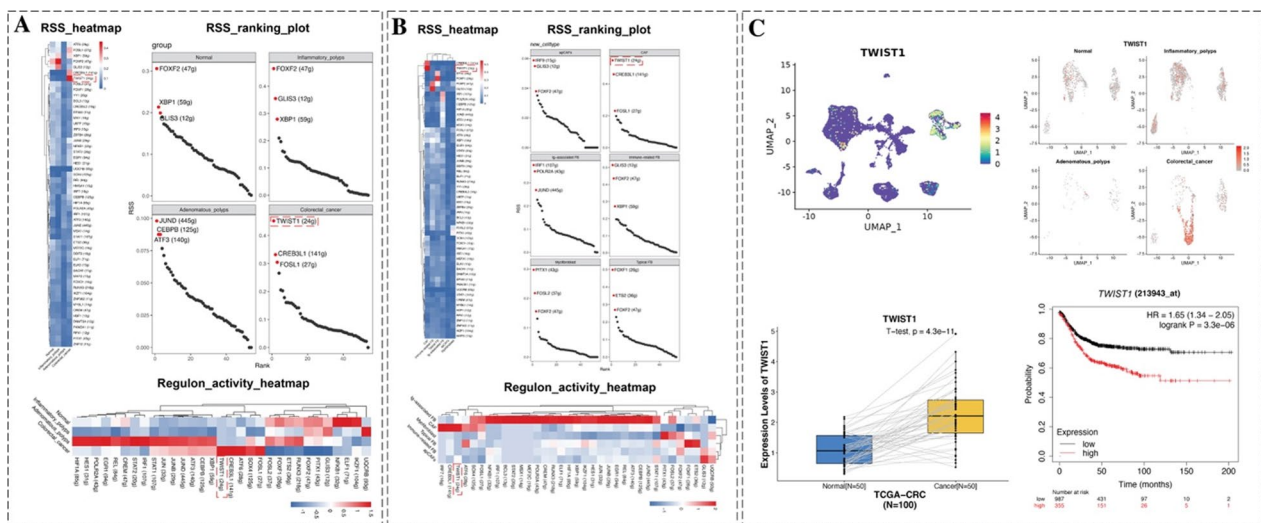


Fig. 12 Scenic analysis of fibroblasts based on the four types of colorectal samples (A) or the six sub-types of fibroblasts (B). *TWIST1* is highlighted in the red dashed squares, with RSS (regulon specificity score) heatmap, RSS ranking plot, and regulon activity heatmap displayed on the top-left, top-right, and top-bottom, respectively. C Expression of *TWIST1* and its relationship with the CRC prognosis. Top left panel: feature plot displaying the expression pattern of *TWIST1* in all cell types from the ten clinic colorectal specimens; Top right panel: *TWIST1* expression profile specifically in fibroblasts across the four types of colorectal samples; Bottom left panel: *TWIST1* expression comparison between paired CRC and normal tissues from the TCGA database; bottom right panel: Kaplan Meier plot illustrating the correlation between the high expression of *TWIST1* and poor prognosis in CRC patients

S16C), these cells exhibited distinct expression patterns (highlighted in the blue square in Fig. S16E). Tumor-associated myeloid cells were CRC-specific, consistent with their established role in immune checkpoint response and tumor microenvironment modulation [81, 82].

Cell-cell interaction analyses revealed the importance of EGFR in CRC progression

Tumors comprise complex cellular ecosystems where interactions between epithelial, stromal, and immune cells drive progression. Using CellPhoneDB [83], we mapped intercellular communication networks in CRC, identifying robust endothelial-epithelial-fibroblast cross-talk (Fig. S17A). Notably, CRC tissues exhibited specific EGFR-mediated interactions (highlighted in the red dashed square on the bottom right of Fig. S17B), with epithelial EGFR responding to multiple ligands (e.g., *AREG*, *COPA*, *GRN*, *HBEGF*, *MIF*, and *TGFB1*) from diverse cell types, reinforcing EGFR's central role in coordinating microenvironmental signaling. Subsequently, CellChat analysis [30] revealed enhanced cell-cell communication in CRC samples, with increased interaction numbers and strength among immune (T cells, B cells, plasma cells, and myeloid cells), stromal cells (endothelial cells and fibroblasts), and epithelial cells (Fig. S18A, C). Differential pair comparison showed greater communication complexity in malignant versus benign tissues (Fig.

S18B, D), highlighting evolving network dynamics during CRC progression.

Pathway analysis revealed significant EGF signaling enrichment in CRC (indicated by the red arrow in Fig. S18D), consistent with EGFR's epithelial importance (Fig. S17B). A recent integrative multi-omics profiling [84] also demonstrated elevated EGFR expression across various cancer types. Tumor-like epithelial cells showed elevated EGFR expression (Fig. 13A), correlating with poor prognosis (Fig. 13B) and confirmed by immunofluorescence (Fig. 13C). These findings underscore EGFR's critical role in CRC progression [85] and the need for improved anti-EGFR therapies [86]. Furthermore, we investigated EGFR-mediated crosstalk using Erlotinib-treated HT29 CRC cells exposed to conditioned medium from IL-1 β -stimulated HUVECs (to induce EGF secretion). Colony formation assays (Fig. 13D) revealed that Erlotinib significantly suppressed HT29 proliferation compared to controls (HUVEC-CM + IL-1 β and DMSO), demonstrating that endothelial-derived EGFR ligands promote epithelial growth. This provides functional evidence for EGFR-dependent epithelial-endothelial communication in CRC progression.

Discussion

Colorectal carcinogenesis follows a progressive “normal-polyp-adenoma-carcinoma” sequence. To elucidate underlying mechanisms, we performed scRNA-seq on

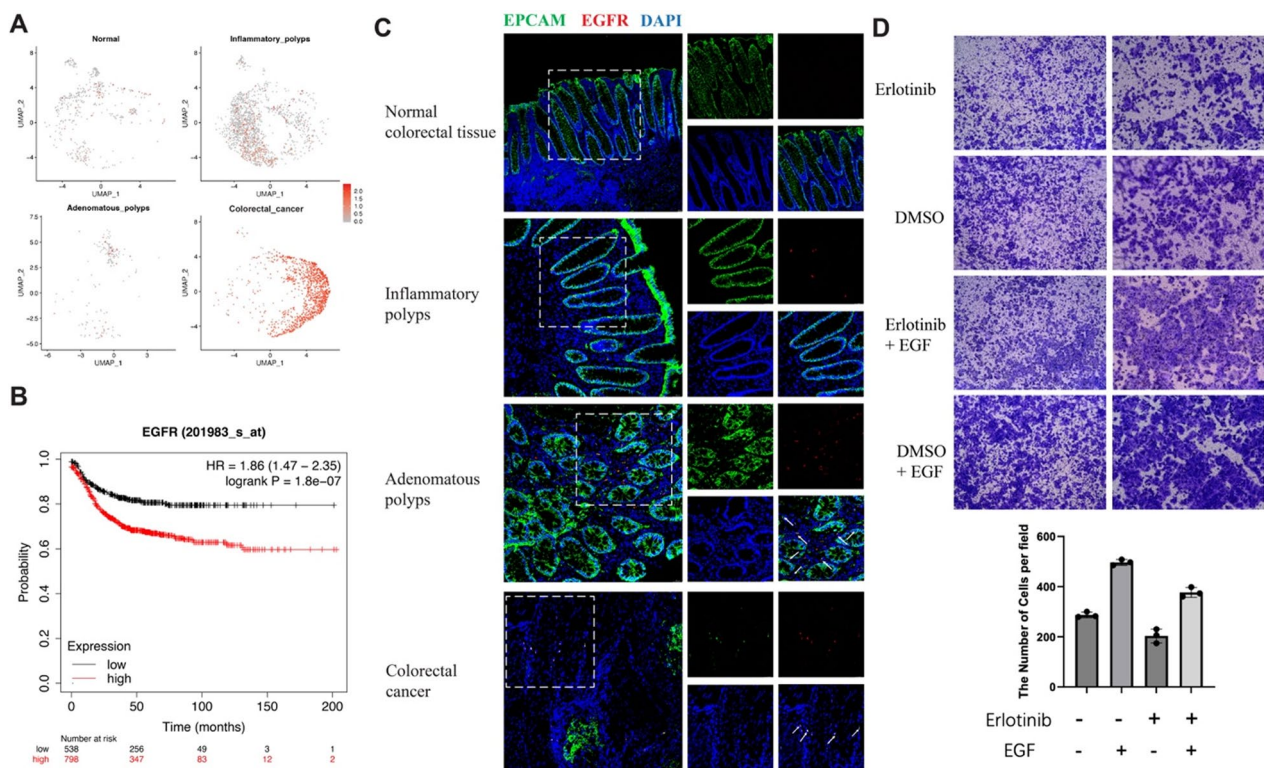


Fig. 13 EGFR is highly expressed in various cell types in CRC tissues. **A** Feature plots illustrating the expression levels of EGFR in epithelial cells across the four types of colorectal samples. **B** Kaplan Meier plots demonstrating that high *EGFR* expression is associated with poor prognosis in CRC patients, based on data from a presentative dataset. **C** Combined IF staining for EPCAM (green) and EGFR (red), in the four types of colorectal specimens (from the top to the bottom). Nuclei were counter-stained with DAPI. Tissues within the white dashed squares were imaged at higher magnification as displayed on the right side. White arrows indicate double positive cells. Each staining was performed using tissue slides from three different individuals of the same type of colorectal specimens. A scale bar is present in each image

ten clinical samples representing all stages, analyzing 51,819 single cells. This revealed molecular signatures of 10 major cell types and 39 subtypes, detailing transcriptional regulation and intercellular signaling during CRC progression. All data are publicly accessible via <http://www.colon-singlecell.cn> for interactive exploration of gene- and cell-type-specific expression patterns across disease stages.

Ligand–receptor interactions (LRIs) critically influence CRC progression and treatment [87, 88]. Our study uniquely demonstrates LRIs' utility in cell type identification, outperforming other gene sets, which was validated by an external CRC scRNA-seq data. To the best of our knowledge, no similar study has been conducted, except for one instance that utilized a 90-gene short-list to differentiate fibroblasts from mural cells [35]. Validation across more single-cell datasets is warranted to confirm these findings' generalizability. As proof of the validity and reliability of our findings, prior studies have

underscored the critical role of ligand–receptor interactions (LRIs) in tumor biology, including: PD-L1 enrichment in stromal cells [89], immunoinflammatory CRC subtypes [90], Notch-mediated stemness [91] and and fibroblast-macrophage crosstalk via COL1A1/LAMA1-ITGB1 [77]. Our findings specifically implicate epithelial *EGFR* overexpression and EGF signaling as key drivers of CRC progression, expanding the molecular understanding of LRIs in carcinogenesis.

To characterize transitional states in colorectal carcinogenesis, we integrated single-cell and bulk transcriptomic data, identifying stage-specific receptor-ligand patterns: *IGHA1* (normal), *TFF1* (inflammatory polyps), *NR1D1* (adenomatous polyps), and *SFRP2* (CRC). Immune cell analysis revealed shared *NR1D1/SFRP2* expression in advanced stages, nominating them as potential diagnostic biomarkers and therapeutic targets for CRC.

Transcription factors (TFs) critically regulate cell-type diversity in CRC [92, 93], with constitutive expression

supporting tumorigenic processes [94]. Our SCENIC analyses identified key TFs: *ARID5B* in CD4⁺ T cells, *IRF9/SOX4* in CD8⁺ T cells, *HNF4G* in tumor-like epithelial cells, and *TWIST1* in CAFs. TCGA and survival analyses validated their clinical relevance, nominating them as potential therapeutic targets or biomarkers, though further functional studies are needed to elucidate their precise roles.

CRC progression involves dynamic cellular shifts, including increased Tregs, tumor-like epithelial cells, CAFs, TAMCs, endothelial cells, and myeloid cells, alongside decreased plasma and naïve B cells. These changes highlight the evolving tumor microenvironment's complexity during disease advancement.

Our analysis of 51,819 cells revealed elevated immunoglobulin gene expression in normal colorectal tissues, e.g. *IGHA1*, *IGHA2*, *IGKC*, and *IGLC2* in T cells, *IGHA1*, *IGHA2*, *IGLC1*, *IGLC2*, *IGKC*, and *JCHAIN* in fibroblasts. On the other hand, adenomatous polyps showed increased heat shock protein (HSP) genes in T cells (*HSP90AA1*, *HSP90AB1*, *HSPA1B*, *HSPD1*, *HSPE1* and *HSPH1*) and plasma cells (*HSP90AA1*, *HSPA1A*, *HSPA1B*, *HSPA6*, *HSPB1* and *HSPD1*). This HSP enrichment mirrors the reported stress response state (T_{STR}) [53], implicating HSP modulation in CRC-associated immune responses and tumorigenesis.

Integrating heat shock proteins and immunoglobulin genes with T-cell signatures could enhance CRC diagnostic/prognostic models by capturing immune and stress response pathways, improving disease progression and treatment outcome predictions. We identified two IgG plasma cell subtypes (*IGKC*^{high} vs. *IGLC1/IGLC2*^{high}) and three IgA plasma cell groups based on *IGLC1/IGLC2*, *IGKC*, and *IGLC1-3* expression patterns, highlighting their potential functional roles in CRC immunity and microenvironment regulation.

Nevertheless, this study has limitations requiring future consideration: firstly, CRC heterogeneity (anatomical, histological, molecular) was not fully captured, as samples included varied subtypes (tubular/signet-ring adenocarcinomas) and locations (colon/rectum) without accounting for CMS classification or mutational profiles [95]. Spatial sampling disparities (e.g., normal vs. tumor site distances) may introduce expression biases. Future work should control for these variables to refine CRC molecular profiling.

Secondly, previous studies have identified various epithelial cell types in colorectal tissues, such as absorptive enterocytes/colonocytes, enteroendocrine cells, goblet cells, Paneth cells, stem cells, tuft cells and microfold cells (M-cells) [64, 65], but this study only distinguished

four subtypes of epithelial cells. This limited classification might be influenced by the sample collection locations and variations in single-cell suspension preparation methods.

An inherent limitation of scRNA-seq technique is the challenges in cell type annotation and establishing criteria for the inclusion or exclusion of individual cells within defined cell types. Establishing boundaries of sub-cell types can be intricate. In this study, we utilized the Seurat pipeline to generate unbiased cell clusters, considering these clusters as direct demarcations of cell types. However, clustering results are impacted by parameters and variables in the algorithm, as well as by the inherent heterogeneity present in the input data.

Conclusions

Our comprehensive analysis of scRNA-seq data from ten human colorectal specimens has provided valuable insights into the molecular landscape across the four stages of colorectal tissues. Leveraging this dataset, we have established a dynamic database highlighting molecular signatures associated with colorectal pathology. Findings were supported by validation with TCGA and Kaplan–Meier datasets, along with immunofluorescence staining using additional clinical samples. This research underscored the importance of ligand–receptor genes and transcription factors in colorectal carcinogenesis, emphasizing the necessity for detailed mechanistic investigations in this field.

Abbreviations

CAFs	Cancer-associated fibroblasts
cDC	Conventional dendritic cell
CNV	Copy number variation
CRC	Colorectal cancer
CSI	Connection specificity index
DEGs	Differentially expressed genes
DMEM	Dulbecco's modified eagle medium
ECs	Endothelial cells
FFPE	Formalin-fixed paraffin-embedded
H&E	Hematoxylin & Eosin
HVGs	Highly variable genes
IHC	Immunohistochemistry
LR	Ligand/receptor
LRIs	Ligand–receptor interactions
JSD	Jensen–Shannon divergence
PCA	Principal-component analysis
RSS	Regulon specificity score
SCENIC	Single-Cell rEgulatory Network Inference and Clustering
scRNA-seq	Single-cell RNA sequencing
TAMCs	Tumor-associated macrophages
TCGA	The Cancer Genome Atlas
TECs	Tumor-associated endothelial cells
TME	Tumor microenvironment
TF	Transcription factor
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecular identifier

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06785-9>.

Supplementary Material 1. Figure S1, related to Figure 1. Clustering based on the 4000 HVGs and expression of some marker genes. A. UMAP plot displaying the 23 clusters identified in this study. B. UMAP plot illustrating the distribution of the ten cell types annotated with color codes across the four types of human colorectal specimens. C. Dot plot showing the expression levels of three selected marker genes for each cell type.

Supplementary Material 2. Figure S2, related to Figure 1. Clustering of this study (GSE261388) based on 4000 HVGs or 1062 highly variable ligand–receptor genes and the expression of some marker genes. A. Histogram showing the proportions of each cell type in the four types of colorectal specimens. B. Heatmap displaying the expression of the top 10 marker genes of the ten cell types annotated based on 4000 highly variable genes. C. Dot plot presenting the expression of three selected ligand/receptor genes for each cell type. D. Heatmap exhibiting the expression of the top 10 marker genes of the ten cell types annotated based on 1062 highly variable ligand–receptor genes.

Supplementary Material 3. Figure S3, related to Figure 1. Correlation analyses of the cell type clustering based on the 4000 HVGs and the 1062 highly variable ligand and receptor genes. A. Venn diagram depicting the overlap between the 4000 highly variable genes in this study and the 2593 ligand–receptor genes extracted from the database (https://baderlab.org/CellCellInteractions?action=AttachFile&do=view&target=receptor_ligand_interactions_mitab_v1.0_April2017.txt.zip). B. Venn diagram presenting the intersection between the top 25 marker genes of the 10 cell types and the 1062 highly variable ligand–receptor genes identified in this study. UMAP plots based on the expression of the 1062 ligand–receptor genes for T cells (C), B cells & plasma cells (E), epithelial cells (G), fibroblasts (I), endothelial cells (K) and myeloid cells (M). Correlation analyses between the two methods for cell type determination based on the 4000 HVGs and the 1062 highly variable ligand and receptor genes for T cells (D), B cells & plasma cells (F), epithelial cells (H), fibroblasts (J), endothelial cells (L) and myeloid cells (N).

Supplementary Material 4. Figure S4, related to Figure 1. Venn diagrams depicting the overlaps between the 4000 HVGs in this study and the genes extracted from the provided database related to cell activation (A), transcription factors (B), cytoskeleton (C), extracellular matrix (ECM) matrisome (D), cell–cell signaling (E), cytokine stimulation (F), cell surface receptor (G), and G protein-coupled receptors (GPCRs, H).

Supplementary Material 5. Figure S5, related to Figure 1. UMAP plots based on the expression of the 87 expressed GPCRs genes identified in this study for T cells (A), B cells & plasma cells (B), epithelial cells (C), fibroblasts (D), endothelial cells (E) and myeloid cells (F).

Supplementary Material 6. Figure S6, related to Figure 1. Identification of potential ligand–receptor marker genes to differentiate the four pathological stages. The single-cell data from each sample was integrated into bulk and analyzed. A. The overall heatmap depicting the DEGs across the four pathological stages. Highlighted in red dashed squares are representative receptor–ligand genes IGHA1, TFF1, NR1D1, and SFRP2 with high expression levels in normal colorectal tissues, tissues with inflammatory polyps, adenomatous polyps, and CRC samples, respectively. B. Venn diagrams illustrating the overlaps of the top highly expressed genes among the four types of immune cells (T cells, B cells, plasma cells, and myeloid cells) for each pathological stage.

Supplementary Material 7. Figure S7, related to Figure 2. Clustering and characterization of T cells. A. UMAP plot visualizing the eleven clusters of T cells. B. UMAP plot displaying the distribution of CD4+ and CD8+ T cells, categorized and color-coded. C. UMAP plots presenting the fifteen sub-types of T cells observed in each type of colorectal sample. Heatmaps illustrating the expression of the top 10 marker genes of the eleven clusters (D) or the fifteen cell types (E) of T cells identified in this study.

Supplementary Material 8. Figure S8, related to Figure 2. The expression levels of selected immunoglobulin genes and heat shock protein genes in colorectal tissues. Box plots displaying the expression levels of six selected genes encoding heat shock proteins (A) and four selected genes encoding immunoglobulin proteins (B) in paired CRC and normal tissues sourced from the TCGA database. C. Combined IF staining for CD3 (green channel) and HSP90AA1 (red channel, left panel) or IGHA1 (red channel, right panel) showing the presence of HSP90AA1+ or IGHA1+ T cells across the four types of colorectal tissues. Tissues within the white dashed squares were imaged at higher magnification as displayed on the right side. D. Immunohistochemistry staining demonstrating the expression levels of HSP90AA1 (left panel) or IGHA1 (right panel) across the four types of colorectal tissues. C & D. Nuclei were counter-stained with DAPI. Each staining was performed in triplicates using tissue slides sourced from three different individuals. A scale bar is present in each image.

Supplementary Material 9. Figure S9, related to Figure 2. Clustering and depiction of CD4+ and CD8+ T cells. A. UMAP plot visualizing the nine clusters of CD4+ T cells. B. Histogram displaying the proportion of the seven sub-types of CD4+ T cells in each type of colorectal sample, annotated with color codes. C. UMAP plot presenting the distribution of CD4+ T cells across the four types of colorectal tissues. D. Heatmap illustrating the expression levels of the top 10 marker genes within the nine clusters of CD4+ T cells. E. UMAP plot visualizing the nine clusters of CD8+ T cells. F. Histogram showing the proportion of the nine clusters of CD8+ T cells in each type of colorectal sample, annotated with color codes. G. UMAP plot presenting the distribution of CD8+ T cells across the four types of colorectal tissues. H. Heatmap illustrating the expression profiles of the top 10 marker genes within the nine clusters of CD8+ T cells.

Supplementary Material 10. Figure S10, related to Figure 3. FOXP3's transcriptional regulatory network within the CRC immunosuppressive niche. FOXP3 target genes (CTLA4 & IL10) were characterized through integrative approaches and presented in dot plot (A) and feature plots (B).

Supplementary Material 11. Figure S11, related to Figure 7. Gene Set Enrichment Analysis (GSEA) comparing each subgroup against the others revealed a shared transcriptional signature of suppressed immunoglobulin production and impaired immune response across all subgroups.

Supplementary Material 12. Figure S12, related to Figure 8. Clustering and characterization of epithelial cells. A. UMAP plot visualizing the nine clusters of epithelial cells with color codes and cell numbers. B. Histogram depicting the proportion of the nine clusters of epithelial cells in each type of colorectal sample. C. Violin plots showing the expression level of the top 10 marker genes for epithelial cells. D. Heatmaps illustrating the expression patterns of the top 10 marker genes within the four populations of epithelial cells. E. Heatmap displaying the differentially expressed genes (DEGs) of epithelial cells across the four types of colorectal specimens.

Supplementary Material 13. Figure S13, related to Figure 10. Functional validation of HNF4G in CRC epithelial cells. siRNA-mediated knockdown of HNF4G in the CRC cell line HT29 (A) led to significant inhibition of cell viability (B, measured by CCK-8 assay) and proliferative capacity (C, assessed by colony formation assay).

Supplementary Material 14. Figure S14, related to Figure 11. Clustering and determination of fibroblasts. A. Violin plots exhibiting the expression levels of the top 10 fibroblast marker genes. B. UMAP plot visualizing the seven sub-types of fibroblasts, annotated with color codes. C. Heatmap depicting the expression profiles of the top 10 marker genes within the seven sub-types of fibroblasts. D. UMAP plot illustrating the cell distribution of the seven sub-types of fibroblasts in each type of colorectal sample. E. Heatmap showing the expression patterns of the top 10 marker genes of the six sub-types of fibroblasts. F. UMAP indicating the proportions of the six sub-types of fibroblasts in the four types of colorectal samples.

Supplementary Material 15. Figure S15. Depiction of endothelial cells in colorectal tissues. A. Violin plots showing the expression levels of the top 10 marker genes for endothelial cells. B. UMAP visualizing the eight clusters of endothelial cells. C. UMAP visualization color-coded for the five sub-types of endothelial cells. D. UMAP plots presenting the five sub-types of endothelial cells observed in each type of colorectal sample. E. Heatmap displaying the expression patterns of the top 10 marker genes for the five endothelial sub-types. F. Feature plots highlighting selected endothelial marker genes (*IGFBP5*, *SELE*, and *CDX1*) to demonstrate the high expression levels in CRC samples. G. Heatmap illustrating the expression patterns of the top 10 marker genes of combined endothelial cells for each type of colorectal sample.

Supplementary Material 16. Figure S16. Identification and analyses of myeloid cells. A. Violin plots displaying the expression levels of the top 10 marker genes for myeloid cells. B. UMAP visualization color-coded for the four sub-types of myeloid cells. C. UMAP plots presenting the four sub-types of myeloid cells observed in each type of colorectal sample. D. Heatmap illustrating the expression patterns of the top 10 marker genes for the four sub-types of myeloid cells. E. Heatmap showing the expression patterns of the top 10 marker genes of combined myeloid cells for each type of colorectal sample. Highlighted in the blue square are genes that can be potentially utilized to divide the myeloid cells from CRC specimens into two types based on expression levels.

Supplementary Material 17. Figure S17, related to Figure 13. Determination of the cell-cell interaction by CellPhoneDB analyses. A. Heatmaps presenting the numbers of interacting ligand–receptor pairs between different cell types in the four types of colorectal tissues, with red color indicating a high number of interacting pairs. B. Dot plots illustrating potential receptor–ligand interactions between different cell types across the four types of colorectal tissues. The X-axis represents the potential interacting cell types (receptor cell | ligand cell), while the Y-axis shows the interacting genes/proteins (receptor | ligand). The size of the circle signifies the significance level, and the color exhibits the expression of interacting genes (red: high; blue: low). Highlighted in a red dashed square on the bottom right are EGFR and its various interacting ligand pairs.

Supplementary Material 18. Figure S18, related to Figure 13. Characterization of the cell-cell interactions by CellChatDB analyses. A. Interaction network analyses for each type of colorectal specimens (from top to bottom: normal colorectal tissue, inflammatory polyps, adenomatous polyps, and CRC tissues). The left panels display the interaction number, while the right panels show the interaction strength. B. Differential interaction network analyses between different colorectal types. The top panels illustrate the differential number of interactions, and the bottom panels depict the differential interaction strength. The comparisons displayed are inflammatory polyps versus normal tissue, adenomatous polyps versus inflammatory polyps, and colorectal cancer versus adenomatous polyps. A & B: Network analyses are represented with color-coded dots indicating different cell types. Red lines represent high numbers or strengths, while blue lines indicate low numbers or strengths. The thickness of the lines reflects the extent of interaction, with thicker lines suggesting stronger interactions. Heatmaps display sender and receiver cell types on the Y-axis and X-axis, respectively. Red or blue colors in the heat maps signify increased or decreased interactions. Bars above and to the right of the heatmaps indicate the sum of incoming and outgoing signals, respectively. Higher bars indicate more signals. Cell types are color-coded. C. Bar plots illustrating the comparisons of cell-cell interactions. The numbers of inferred interactions are shown in the left panel and interaction strength in the right panel. The X-axis represents the four types of colorectal samples, while the Y-axis shows the corresponding numbers. D. Bar plots depicting the information flow of each signal pathway. The red arrow points to the EGF signal pathway. C & D. Each type of colorectal sample is color-coded as shown by the colors of the bars. Higher bars indicate more interactions.

Supplementary Material 19.

Supplementary Material 20.

Supplementary Material 21.

Supplementary Material 22.

Supplementary Material 23.

Supplementary Material 24.

Supplementary Material 25.

Supplementary Material 26.

Supplementary Material 27.

Supplementary Material 28.

Supplementary Material 29.

Supplementary Material 30.

Supplementary Material 31.

Supplementary Material 32.

Supplementary Material 33.

Supplementary Material 34.

Acknowledgements

We thank all the patients and their families for their generous donation of tissue samples in this study. We also thank Ms. Qingzhen Zi, Mrs. Wenyang Ding, Yongbing Ba, and Long Chen, Ms. Yao Lu from OE biotech company (Shanghai, China) for the great support of bioinformatic analyses and data-base generation.

Jianping Liu—Lead contact.

Author contributions

H.Z., Z.Y., and J.S. designed experiments. J.L. supervised the project. J.S., Y.L., Y.Z., and M.T. collected clinical samples. H.C. coordinated patient sample collections and provided a pathological diagnosis. J.S., N.L., Y.L. and Y.Z. generated human scRNA-seq data. G.Z., C.X., and Z.H. retrieved and analyzed the publicly available data. X.G. and his colleagues constructed the website. L.W. generated the artwork used in some figures and was involved in the website construction. Y.L. and P.Z. performed the immunofluorescence staining, imaging, and analysis. P.Z. uploaded raw data to the GEO database. J.S., N.L., Y.Z., J.L., and H.F. wrote and revised the manuscript with valuable input from all authors. S.L., Y.Y., Y.Z. and X. H. completed the revision. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (82360876, 82260119, 82170580, and 82260118), Academic and Technical Leader of major disciplines in Jiangxi Province (20225BCJ23021), the Natural Science Foundation of Jiangxi Province (20212BAB216016 and 20224ACB216004), the Key Laboratory Project of Digestive Diseases in Jiangxi Province (2024SSY06101), and Jiangxi Clinical Research Center for Gastroenterology (20223BCG74011).

Data availability

The accession number for the raw sequencing data in this paper is GSE261388 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE261388>). Search by gene name and subsequent visualization of analyses of the scRNA-seq dataset can be accessed at <https://www.colon-singlecell.cn>. Supplemental information can be found online at <https://doi.org/10.1186/s12967-025-06785-9>.

Material availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, J.L. (ndyfy10307@ncu.edu.cn). This study did not generate any new unique reagents.

Declarations

Ethics approval and consent to participate

Ethical approval for this study was obtained from the Ethics Committee of Yichun People's Hospital under protocol number 2021-009, with all tissue acquisitions performed following informed consent from the patients.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Gastroenterology, Yichun People's Hospital, Yichun 336000, Jiangxi, China. ²Department of Gastroenterology, The Affiliated Hospital of Yichun University, Yichun 336000, Jiangxi, China. ³Jiangxi Provincial Key Laboratory of Digestive Diseases, Department of Gastroenterology, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, China. ⁴Yichun University, Yichun 336000, Jiangxi, China. ⁵Huankui Academy, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, China. ⁶College of Animal Veterinary Medicine, Yunnan Agricultural University, Kunming 650201, Yunnan, China. ⁷College of Animal Science and Technology, Yunnan Agricultural University, Kunming 650201, Yunnan, China. ⁸Department of Medicine Huddinge, Karolinska Institutet, 141 57, Huddinge, Sweden.

Received: 21 January 2025 Accepted: 20 June 2025

Published online: 01 July 2025

References

- Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA Cancer J Clin.* 2023;73:233–54.
- Morgan E, Arnold M, Gini A, Lorenzoni V, Cabaas CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut.* 2023;72:338–44.
- Huck MB, Bohl JL. Colonic polyps: diagnosis and surveillance. *Clin Colon Rectal Surg.* 2016;29:296–305.
- Rudy DR, Zdon MJ. Update on colorectal cancer. *Am Fam Physician.* 2000;61(1759–1770):1773–1754.
- He DG, Chen XJ, Huang JN, Chen JG, Lv MY, Huang TZ, et al. Increased risk of colorectal neoplasia in inflammatory bowel disease patients with post-inflammatory polyps: a systematic review and meta-analysis. *World J Gastrointest Oncol.* 2022;14:348–61.
- Heumos L, Schaar AC, Lance C, Litnitskaya A, Drost F, Zappia L, et al. Best practices for single-cell analysis across modalities. *Nat Rev Genet.* 2023;24:550–72.
- Nofech-Mozes I, Soave D, Awadalla P, Abelson S. Pan-cancer classification of single cells in the tumour microenvironment. *Nat Commun.* 2023;14:1615.
- Van de Sande B, Lee JS, Mutasa-Gottgens E, Naughton B, Bacon W, Manning J, et al. Applications of single-cell RNA sequencing in drug discovery and development. *Nat Rev Drug Discov.* 2023;22:496–520.
- Song S, Feng L, Xi K, Sun Z, Kong D, Luo Z, et al. Single-cell profiling of the copy-number heterogeneity in colorectal cancer. *Chin Med J (Engl).* 2023;136:707–18.
- Shen Y, Ni S, Li S, Lv B. Role of stemness-related genes TIMP1, PGF, and SNAI1 in the prognosis of colorectal cancer through single-cell RNA-seq. *Cancer Med.* 2023;12:11611–23.
- Luo Y, Deng X, Liao W, Huang Y, Lu C. Prognostic value of autophagy-related genes based on single-cell RNA-sequencing in colorectal cancer. *Front Genet.* 2023;14:1109683.
- Sun R, Yang Y, Lu W, Yang Y, Li Y, Liu Z, et al. Single-cell transcriptomic analysis of normal and pathological tissues from the same patient uncovers colon cancer progression. *Cell Biosci.* 2023;13:62.
- Wen R, Zhou L, Peng Z, Fan H, Zhang T, Jia H, et al. Single-cell sequencing technology in colorectal cancer: a new technology to disclose the tumor heterogeneity and target precise treatment. *Front Immunol.* 2023;14:1175343.
- Joanito I, Wirapati P, Zhao N, Nawaz Z, Yeo G, Lee F, et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. *Nat Genet.* 2022;54:963–75.
- Shen X, Mo S, Wang Y, Lin L, Liu Y, Weng M, et al. Single-cell dissection reveals the role of DNA damage response patterns in tumor microenvironment components contributing to colorectal cancer progression and immunotherapy. *Genes Cells.* 2023;28:348–63.
- Ozato Y, Kojima Y, Kobayashi Y, Hisamatsu Y, Toshima T, Yonemura Y, et al. Spatial and single-cell transcriptomics decipher the cellular environment containing HLA-G+ cancer cells and SPP1+ macrophages in colorectal cancer. *Cell Rep.* 2023;42: 111929.
- Alzamami A. Implications of single-cell immune landscape of tumor microenvironment for the colorectal cancer diagnostics and therapy. *Med Oncol.* 2023;40:352.
- Xie S, Cai Y, Chen D, Xiang Y, Cai W, Mao J, et al. Single-cell transcriptome analysis reveals heterogeneity and convergence of the tumor microenvironment in colorectal cancer. *Front Immunol.* 2022;13: 959705.
- Tieng FYF, Lee LH, Ab Mutalib NS. Deciphering colorectal cancer immune microenvironment transcriptional landscape on single cell resolution—a role for immunotherapy. *Front Immunol.* 2022;13: 959705.
- Zhang L, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O'Brien SA, et al. Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell.* 2020;181(442–459): e429.
- Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(3573–3587): e3529.
- 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): e324.
- van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell.* 2018;174(716–729): e727.
- Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* 2019;16:1289–96.
- 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.
- Aibar S, Gonzalez-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.
- Suo S, Zhu Q, Saadatpour A, Fei L, Guo G, Yuan GC. Revealing the critical regulators of cell identity in the mouse cell atlas. *Cell Rep.* 2018;25(1436–1445): e1433.
- Puram SV, Tirosh I, Parkh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell.* 2017;171(1611–1624): e1624.
- Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. Cell PhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc.* 2020;15:1484–506.
- Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using Cell Chat. *Nat Commun.* 2021;12:1088.
- Ouyang JF, Kamaraj US, Cao EY, Rackham OJL. ShinyCell: simple and sharable visualization of single-cell gene expression data. *Bioinformatics.* 2021;37:3374–6.
- Li N, Xu X, Zhan Y, Fei X, Ouyang Y, Zheng P, et al. YAP and beta-catenin cooperate to drive *H. pylori*-induced gastric tumorigenesis. *Gut Microbes.* 2023;15:2192501.
- Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol.* 2023;42:293–304.
- Lee HO, Hong Y, Etioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet.* 2020;52:594–603.

35. Muhl L, Genove G, Leptidis S, Liu J, He L, Mocci G, et al. Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat Commun.* 2020;11:3953.
36. Zheng Z, Wiedner T, Mauerer B, Schafer L, Kesselring R, Braumuller H. T cells in colorectal cancer: unravelling the function of different T cell subsets in the tumor microenvironment. *Int J Mol Sci.* 2023;24:11673.
37. Feng M, Zhao Z, Yang M, Ji J, Zhu D. T-cell-based immunotherapy in colorectal cancer. *Cancer Lett.* 2021;498:201–9.
38. James KR, Gomes T, Elmentaite R, Kumar N, Gulliver EL, King HW, et al. Distinct microbial and immune niches of the human colon. *Nat Immunol.* 2020;21:343–53.
39. Zheng L, Qin S, Si W, Wang A, Xing B, Gao R, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science.* 2021;374:abe6474.
40. Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell.* 2017;169(1342–1356): e1316.
41. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature.* 2018;564:268–72.
42. Sundstrom P, Ahlmanner F, Akeus P, Sundquist M, Alsen S, Yrlid U, et al. Human mucosa-associated invariant T cells accumulate in colon adenocarcinomas but produce reduced amounts of IFN- γ . *J Immunol.* 2015;195:3472–81.
43. Zabijak L, Attencourt C, Guignat C, Chatelain D, Marcelo P, Marolleau JP, et al. Increased tumor infiltration by mucosal-associated invariant T cells correlates with poor survival in colorectal cancer patients. *Cancer Immunol Immunother.* 2015;64:1601–8.
44. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science.* 2006;313:1960–4.
45. Pages F, Berger A, Camus M, Sanchez-Cabo F, Costes A, Molitor R, et al. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med.* 2005;353:2654–66.
46. Lacar B, Linker SB, Jaeger BN, Krishnaswami SR, Barron JJ, Kelder MJE, et al. Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat Commun.* 2016;7:11022.
47. Tiberti S, Catozzi C, Croci O, Ballerini M, Cagnina D, Soriani C, et al. GZMK(high) CD8(+) T effector memory cells are associated with CD15(high) neutrophil abundance in non-metastatic colorectal tumors and predict poor clinical outcome. *Nat Commun.* 2022;13:6752.
48. Baer R, Forster A, Lavenir I, Rabbitts TH. Immunoglobulin VH genes are transcribed by T cells in association with a new 5' exon. *J Exp Med.* 1988;167:2011–6.
49. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun.* 2020;11:2285.
50. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallee VP, et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat Med.* 2020;26:259–69.
51. Cui M, Huang J, Zhang S, Liu Q, Liao Q, Qiu X. Immunoglobulin Expression in Cancer Cells and Its Critical Roles in Tumorigenesis. *Front Immunol.* 2021;12: 613530.
52. Wang Z, Geng Z, Shao W, Liu E, Zhang J, Tang J, et al. Cancer-derived sialylated IgG promotes tumor immune escape by binding to Siglecs on effector T cells. *Cell Mol Immunol.* 2020;17:1148–62.
53. Chu Y, Dai E, Li Y, Han G, Pei G, Ingram DR, et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. *Nat Med.* 2023;29:1550–62.
54. Liu RX, Wen C, Ye W, Li Y, Chen J, Zhang Q, et al. Altered B cell immunoglobulin signature exhibits potential diagnostic values in human colorectal cancer. *iScience.* 2023;26: 106140.
55. Tong Q, Zhou J. Construction of a 12-gene prognostic model for colorectal cancer based on heat shock protein-related genes. *Int J Hyperthermia.* 2024;41:2290913.
56. Santiago L, Castro M, Sanz-Pamplona R, Garzon M, Ramirez-Labrada A, Tapia E, et al. Extracellular granzyme A promotes colorectal cancer development by enhancing gut inflammation. *Cell Rep.* 2020;32: 107847.
57. Zhang J, Mizuuchi Y, Ohuchida K, Hisano K, Shimada Y, Katayama N, et al. Exploring the tumor microenvironment of colorectal cancer patients post renal transplantation by single-cell analysis. *Cancer Sci.* 2025;116:500–12.
58. Koi M, Carethers JM. The colorectal cancer immune microenvironment and approach to immunotherapies. *Future Oncol.* 2017;13:1633–47.
59. Onieva JL, Xiao Q, Berciano-Guerrero MA, Laborda-Illanes A, de Andrea C, Chaves P, et al. High IGKC-expressing intratumoral plasma cells predict response to immune checkpoint blockade. *Int J Mol Sci.* 2022;23:9124.
60. Fristedt R, Borg D, Hedner C, Berntsson J, Nodin B, Eberhard J, et al. Prognostic impact of tumour-associated B cells and plasma cells in oesophageal and gastric adenocarcinoma. *J Gastrointest Oncol.* 2016;7:848–59.
61. Chang YT, Tsai WC, Lin WZ, Wu CC, Yu JC, Tseng VS, et al. A novel IGLC2 gene linked with prognosis of triple-negative breast cancer. *Front Oncol.* 2021;11: 759952.
62. Wang W, Zhong Y, Zhuang Z, Xie J, Lu Y, Huang C, et al. Multiregion single-cell sequencing reveals the transcriptional landscape of the immune microenvironment of colorectal cancer. *Clin Transl Med.* 2021;11: e253.
63. Wang Q, Sun K, Liu R, Song Y, Lv Y, Bi P, et al. Single-cell transcriptome sequencing of B-cell heterogeneity and tertiary lymphoid structure predicts breast cancer prognosis and neoadjuvant therapy efficacy. *Clin Transl Med.* 2023;13: e1346.
64. Bigaeva E, Uniken Venema WTC, Weersma RK, Festen EAM. Understanding human gut diseases at single-cell resolution. *Hum Mol Genet.* 2020;29:R51–8.
65. Parikh K, Antanaviciute A, Fawcner-Corbett D, Jagielowicz M, Aulicino A, Lagerholm C, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature.* 2019;567:49–55.
66. Bianco C, Mohr I. Ribosome biogenesis restricts innate immune responses to virus infection and DNA. *Elife.* 2019;8: e49551.
67. Witz IP. The tumor microenvironment: the making of a paradigm. *Cancer Microenviron.* 2009;2(Suppl 1):9–17.
68. Mahdipour-Shirayeh A, Erdmann N, Leung-Hagesteijn C, Tiedemann RE. sciCNV: high-throughput paired profiling of transcriptomes and DNA copy number variations at single-cell resolution. *Brief Bioinform.* 2022;23:bbab413.
69. Guo S, Sun Y. OTO2, Inversely Modulated by miR-3148, Inhibits CRC Cell Migration, Proliferation and Epithelial-Mesenchymal Transition: Evidence from Bioinformatics Data Mining and Experimental Verification. *Cancer Manag Res.* 2022;14:1371–84.
70. Johnson B, Mahadevan D. Emerging role and targeting of carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) in human malignancies. *Clin Cancer Drugs.* 2015;2:100–11.
71. Noble CL, Abbas AR, Cornelius J, Lees CW, Ho GT, Toy K, et al. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut.* 2008;57:1398–405.
72. Xie B, Bai B, Xu Y, Liu Y, Lv Y, Gao X, et al. Tumor-suppressive function and mechanism of HOXB13 in right-sided colon cancer. *Signal Transduct Target Ther.* 2019;4:51.
73. Jung C, Kim RS, Zhang H, Lee SJ, Sheng H, Loehrer PJ, et al. HOXB13 is downregulated in colorectal cancer to confer TCF4-mediated transactivation. *Br J Cancer.* 2005;92:2233–9.
74. Chen L, Shi H, Wang X, Wang T, Wang Y, Wu Z, et al. Hepatocyte nuclear factor 4 gamma (HNF4G) is correlated with poor prognosis and promotes tumor cell growth by inhibiting caspase-dependent intrinsic apoptosis in colorectal cancer. *Eur J Pharmacol.* 2022;916: 174727.
75. Huang H, Wang Z, Zhang Y, Pradhan RN, Ganguly D, Chandra R, et al. Mesothelial cell-derived antigen-presenting cancer-associated fibroblasts induce expansion of regulatory T cells in pancreatic cancer. *Cancer Cell.* 2022;40(656–673): e657.
76. Bagalad BS, Mohan Kumar KP, Puneeth HK. Myofibroblasts: master of disguise. *J Oral Maxillofac Pathol.* 2017;21:462–3.
77. Qi J, Sun H, Zhang Y, Wang Z, Xun Z, Li Z, et al. Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer. *Nat Commun.* 2022;13:1742.
78. Li S, Lu R, Shu L, Chen Y, Zhao J, Dai J, et al. An integrated map of fibroblastic populations in human colon mucosa and cancer tissues. *Commun Biol.* 2022;5:1326.
79. Hu S, Qin J, Gao R, Xiao Q, Liu X, Pan Y, et al. Integrated analysis of single cell and bulk RNA sequencing identifies CTHRC1(+) INHBA(+) CAF as drivers of colorectal cancer progression. *Mol Carcinog.* 2023;62:1787–802.
80. Chen WZ, Jiang JX, Yu XY, Xia WJ, Yu PX, Wang K, et al. Endothelial cells in colorectal cancer. *World J Gastrointest Oncol.* 2019;11:946–56.

81. Cheng X, Wang H, Wang Z, Zhu B, Long H. Tumor-associated myeloid cells in cancer immunotherapy. *J Hematol Oncol*. 2023;16:71.
82. Mantovani A, Marchesi F, Jaillon S, Garlanda C, Allavena P. Tumor-associated myeloid cells: diversity and therapeutic targeting. *Cell Mol Immunol*. 2021;18:566–78.
83. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature*. 2018;563:347–53.
84. Liang WW, Lu RJ, Jayasinghe RG, Foltz SM, Porta-Pardo E, Geffen Y, et al. Integrative multi-omic cancer profiling reveals DNA methylation patterns associated with therapeutic vulnerability and cell-of-origin. *Cancer Cell*. 2023;41(1567–1585): e1567.
85. Markman B, Javier Ramos F, Capdevila J, Tabernero J. EGFR and KRAS in colorectal cancer. *Adv Clin Chem*. 2010;51:71–119.
86. Zhou J, Ji Q, Li Q. Resistance to anti-EGFR therapies in metastatic colorectal cancer: underlying mechanisms and reversal strategies. *J Exp Clin Cancer Res*. 2021;40:328.
87. Martinelli E, Ciardiello D, Martini G, Troiani T, Cardone C, Vitiello PP, et al. Implementing anti-epidermal growth factor receptor (EGFR) therapy in metastatic colorectal cancer: challenges and future perspectives. *Ann Oncol*. 2020;31:30–40.
88. Janani B, Vijayakumar M, Priya K, Kim JH, Prabakaran DS, Shahid M, et al. EGFR-based targeted therapy for colorectal cancer-promises and challenges. *Vaccines (Basel)*. 2022;10:499.
89. Ghoshdastider U, Rohatgi N, Mojtavai Naeini M, Baruah P, Revkov E, Guo YA, et al. Pan-cancer analysis of ligand-receptor cross-talk in the tumor microenvironment. *Cancer Res*. 2021;81:1802–12.
90. Lin H, Xia L, Lian J, Chen Y, Zhang Y, Zhuang Z, et al. Delineation of colorectal cancer ligand-receptor interactions and their roles in the tumor microenvironment and prognosis. *J Transl Med*. 2021;19:497.
91. Brisset M, Mehlen P, Meurette O, Hollande F. Notch receptor/ligand diversity: contribution to colorectal cancer stem cell heterogeneity. *Front Cell Dev Biol*. 2023;11:1231416.
92. Lee BK, Bhinge AA, Battenhouse A, McDaniel RM, Liu Z, Song L, et al. Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Res*. 2012;22:9–24.
93. Arvey A, Agius P, Noble WS, Leslie C. Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome Res*. 2012;22:1723–34.
94. Vishnoi K, Viswakarma N, Rana A, Rana B. Transcription factors in cancer development and therapy. *Cancers (Basel)*. 2020;12:2296.
95. Valenzuela G, Canepa J, Simonetti C, Solo Zaldivar L, Marcelain K, Gonzalez-Montero J. Consensus molecular subtypes of colorectal cancer in clinical practice: a translational approach. *World J Clin Oncol*. 2021;12:1000–8.

Publisher's Note

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