

RESEARCH

Open Access



Dynamic remodelling of epithelial plasticity in colorectal cancer from single-cell and spatially resolved perspectives

Chengyuan Xu^{1†}, Yun Pan^{1†}, Siqi Zhang^{2†}, Moubin Lin^{1*†}, Hailong Liu^{1*†} and Zicheng Yu^{2*†} 

Abstract

Background Epithelial compartments play a central role in the progression and metastasis of colorectal cancer (CRC), yet the mechanisms underlying their transcriptional reprogramming across disease stages remain incompletely understood.

Methods We constructed comprehensive single-cell and spatial transcriptomic atlases encompassing normal colorectal tissue, primary CRC, and liver metastases (LMs), profiling over 200,000 high-quality cells. Specific coexpression modules and transcriptional regulators were identified and spatially mapped. Additionally, the clinical and functional roles of SCAND1 were investigated using western blot, RT-qPCR, and immunohistochemistry in clinical samples, as well as functional assays in CRC cell lines.

Results Distinct coexpression modules were selectively enriched in normal epithelial cells and associated with vesicle transport, ion homeostasis, and barrier function, whereas LM-specific modules were linked to Wnt signalling, ribosome biogenesis, and cell cycle processes. High-CNV epithelial cells exhibited progressive activation of key transcription factors such as CEBPB, E2F, and the AP-1 family, with concurrent suppression of ARGLU1 and MTNR2L8, indicating coordinated remodelling of transcriptional and metabolic programs during metastasis. These changes were spatially validated and correlated with alterations in epithelial composition within the tumour core. Consistently, multiplex immunofluorescence (mIHC) confirmed the expression patterns of CEBPB, ARGLU1, S100A4 and FOSL1, across normal epithelium, CRC, and LM. Notably, SCAND1 was identified as a potential therapeutic target for malignant epithelial plasticity. We demonstrated for the first time that SCAND1 expression is significantly upregulated in CRC patients and is associated with superior diagnostic and prognostic value. Functional experiments revealed

[†]Chengyuan Xu, Yun Pan and Siqi Zhang contributed equally to this work and share first authorship.

[†]Moubin Lin, Hailong Liu and Zicheng Yu contributed equally to this work and share last authorship.

*Correspondence:

Moubin Lin
1500142@tongji.edu.cn

Hailong Liu
hailongliu@tongji.edu

Zicheng Yu
yuzicheng@tongji.edu.cn

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/>.

that SCAND1 promotes tumour cell proliferation, inhibits apoptosis, and enhances metastatic potential. Furthermore, co-culture of SCAND1-overexpressing tumour cells with T cells showed that SCAND1 overexpression attenuates T cell-mediated cytotoxicity, supporting an additional immune-evasive role.

Conclusions Our study elucidates the molecular logic of epithelial state transitions during CRC metastasis and underscores the promise of SCAND1 as a novel biomarker and therapeutic target in colorectal cancer.

Keywords Colorectal cancer, Single-cell transcriptomics, Spatial transcriptomics, Epithelial plasticity, Copy number variation, SCAND1

Introduction

Colorectal cancer (CRC) is one of the most common gastrointestinal malignancies worldwide, ranking third in cancer incidence and second in cancer-related mortality [1, 2]. Approximately 50% of CRC patients develop liver metastasis (LM) during disease progression, which is the main cause of CRC-related death [3, 4]. Despite recent advances in surgery, chemotherapy, radiotherapy, and targeted therapies, the prognosis of metastatic CRC remains poor. This is mainly due to the high heterogeneity of tumour cells, their capacity for immune evasion, and treatment resistance [1, 5, 6]. Large-scale gene expression studies have established molecular classification systems for CRC—most notably, the Consensus Molecular Subtypes (CMS), which stratify CRC into four subtypes (CMS1-4) [7]. However, these classifications are primarily based on bulk sequencing data and cannot precisely resolve the complex cellular heterogeneity of the tumour microenvironment.

In the pathogenesis of CRC, epithelial cells, which are the normal cell type in which oncogenic events arise and initiate tumorigenesis, are central to disease progression. In the normal intestine, epithelial cells maintain highly differentiated structures and functional homeostasis. However, during carcinogenesis, their transcriptional programs are dramatically remodelled, resulting in the progressive acquisition of stemness, proliferation, migration, and other malignant features [8, 9]. Previous studies have shown that CRC epithelial cells can increase their invasiveness and ability to adapt to distant sites by activating epithelial–mesenchymal transition (EMT), adapting to hypoxic and inflammatory stress environments, or reacquiring stem cell-like plasticity [10, 11]. These state transitions not only involve shifts in cell fate but also may drive immune evasion, chemoresistance, and the formation of a prometastatic microenvironment. However, there is still a lack of integrative research frameworks regarding the distribution patterns of these “epithelial plasticity states” across different disease stages and spatial regions, their underlying transcriptional regulatory mechanisms, and their interactions with the tumour immune microenvironment or stromal cells.

SCAND1 is a SCAN domain-only member of the conserved SCAN transcription factor family, most of which

are SCAN–zinc finger (SCAN-ZF) proteins [12–14]. The SCAN domain mediates oligomerization; SCAND1 lacks zinc fingers but hetero-oligomerizes with SCAN-ZFs such as MZF1 (ZSCAN6/ZNF42) to form repressive complexes that counteract SCAN-ZF transactivation [15–17]. Functionally, MZF1 activates the co-chaperone CDC37 to promote cancer progression, whereas SCAND1 represses CDC37 by binding MZF1, acting as a potential tumor suppressor; SCAND1–MZF1 oligomers can also reverse EMT, tumor growth, and migration by repressing EMT drivers and MAPK genes [16–18]. Clinically, MZF1 often correlates with poor outcomes, while SCAND1/MZF1 expression can associate with better prognosis in specific contexts, indicating context-dependent oncogenic versus suppressor roles [17]. Despite these advances, SCAND1’s role in colorectal cancer (CRC) remains unclear.

Recent advances in single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics have enabled detailed characterization of cellular states and their spatial organization in complex tissues [19–22]. While these technologies have been used to map the immune microenvironment and CAF subsets in CRC [23, 24], the diversity and remodelling of epithelial states during metastasis remain underexplored. In this study, we integrated single-cell and spatial transcriptomic data from primary CRC, normal colorectal tissue, and liver metastases to systematically characterize epithelial lineage composition, transcriptional state transitions, and spatial distribution. Using *hdWGCNA* and *pySCENIC*, we identified epithelial coexpression modules and key regulators, highlighting metastasis-associated programs dominated by Wnt signalling, ribosome biogenesis, and cell cycle activation. CNV inference, trajectory, and spatial analyses further revealed that epithelial cells with high CNV burdens presented increased transcriptional plasticity and heterogeneity and extensive signalling interactions with the immune microenvironment. Our work reveals the dynamic remodelling of epithelial plasticity in CRC and provides a theoretical basis for targeting key epithelial states to inhibit metastasis.

Methods

Data collection

Single-cell RNA sequencing (scRNA-seq) data were obtained from the GEO database (accession numbers: GSE132465, GSE178318, and GSE200997). The dataset includes a total of 18 normal colorectal tissue samples, 45 colorectal cancer (CRC) samples, and 6 colorectal cancer liver metastasis (LM) samples. After quality control, a total of 203,379 high-quality sequenced cells were analyzed. Spatial transcriptomics samples were obtained from <http://www.cancerdiversity.asia/scCRLM/> and GSE284061. Gene expression datasets (GSE87211, GSE71187, GSE41258, GSE44076, GSE103512 and GSE106582) were obtained from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) as of October 2023. Diagnostic curves of a single hub gene were drawn using “pROC” package to evaluate the sensitivity and specificity of a single hub gene in these datasets. Bulk RNA-seq expression profiles and associated clinical data for colorectal cancer patients were acquired from The Cancer Genome Atlas (TCGA; <https://portal.gdc.cancer.gov/>).

scRNA-seq data processing

We used Seurat v4.1.3 for preprocessing, quality control, normalization, and dimensionality reduction and clustering of single-cell data. Doublets were identified and removed using DoubletFinder v2.0.3. Quality control criteria included a minimum of 400 genes expressed per cell and a mitochondrial gene content threshold of 20%. Normalization, identification of highly variable genes, and clustering were performed using Seurat’s default parameters and workflow. Batch correction and data integration across samples were performed using Harmony v1.2.3. Cell clusters were annotated using marker genes collected from the literature and manual curation. Differentially expressed genes (DEGs) between subpopulations were identified using the FindAllMarkers function based on the Wilcoxon rank-sum test, with selection criteria of adjusted p-value < 0.05 and $|\log_2FC| > 0.5$.

Spatial transcriptomics data processing

Spatial transcriptomics data were processed using Seurat v4.3.0. Raw spatial expression matrices were normalized using SCTransform, followed by PCA for dimensionality reduction. The top 30 principal components were used for graph-based neighborhood detection and clustering. UMAP was used for spatial domain visualization and to display spatial feature distributions. For scRNA-seq data with cell type annotation, log-normalization with a scale factor of 10,000 was applied, and 3,000 highly variable genes were identified using the “vst” method. Data were scaled and reduced using PCA. UMAP embeddings were generated based on the top 30 Harmony components.

Graph-based clustering was performed using Harmony reduction and the Louvain algorithm (resolution = 0.5).

Cell type identification

We used Seurat’s FindAllMarkers function to perform differential gene expression analysis within each cluster to identify marker genes. Thresholds for defining marker genes were: adjusted p-value < 0.05, expression percentage > 0.25, and $|\log_2$ fold change > 0.25. Cell clusters were annotated using the SingleR package based on marker gene composition and reference datasets, followed by manual validation and correction using the CellMarker database.

Spatial transcriptomics spot annotation by single-cell deconvolution

To deconvolute cell type composition in spatial transcriptomics data, we used the Robust Cell Type Decomposition (RCTD) algorithm implemented in the spacexr R package. Cell type labels from annotated scRNA-seq data were used as references. The reference matrix included raw counts, cell type annotations, and total UMI counts. For each spatial sample, spot-level count matrices, spatial coordinates, and UMI counts were used to create a spatial RNA object. RCTD was run in ‘full’ mode to flexibly assign cell types to each spatial spot. The resulting cell type weight matrix was integrated into Seurat’s spatial metadata. RCTD was also rerun in ‘doublet’ mode to identify and resolve spots containing two dominant cell types. The results were integrated into the spatial object for downstream analysis and visualization.

Cell–cell communication

We used the CellChat R package (v1.1.3) to infer cell–cell communication based on ligand–receptor interactions [25]. Communication probability and number of interactions were calculated to build a communication network. Interaction strength and frequency between any two cell groups were visualized. Scatter plots were used to show major sender (signal source) and receiver (target) cell types in 2D space, helping identify the key contributors to outgoing and incoming signals within immune cell populations. A global communication model was applied to identify how multiple immune cell types and signaling pathways coordinate.

Pseudotime trajectory analysis

Pseudotime trajectories were constructed using Monocle 2 for single-cell trajectory analysis [26]. The algorithm reduces high-dimensional expression data into a low-dimensional space and arranges cells along trajectories with branching points. Dynamic expression heatmaps were generated using the `plot_pseudotime_heatmap` function.

8. Gene set enrichment analysis (GSEA)

GSEA was performed using the GSEAPy package. DEGs were ranked by log2 fold change in descending order to create a pre-ranked gene list. Hallmark gene sets were obtained from the MSigDB database (v2024.1, category: h.all) via the gseapy.Msigdb() interface. GSEA was conducted using gseapy.prerank with 1,000 permutations. Gene sets with sizes between 1 and 10,000 were included. The output included normalized enrichment scores (NES), nominal p-values, and FDR q-values. A fixed random seed ensured reproducibility.

Inference of copy number variation (CNV)

Single-cell CNV inference was performed using the inferCNV R package (v1.20.0). Gene genomic positions were curated and deduplicated. Raw count matrices and cell type annotations were extracted from the Seurat object, and cell type metadata was saved as a group annotation file. Immune cells (T cells, B cells, and myeloid cells) were used as the reference group. The analysis used a cutoff of 0.1 for UMI-based data (10x Genomics). Hierarchical clustering was performed using the “ward.D2” method. Denoising was enabled, and the HMM option was disabled to improve computational efficiency.

pySCENIC analysis

Gene regulatory network (GRN) analysis was conducted using the pySCENIC pipeline. Raw Loom-format expression data and a curated list of human transcription factors were input to GRNBoost2 to infer co-expression modules. Motif enrichment analysis was performed using a TF motif ranking database and annotation file in the ctf step with dropout masking and parallelization. Regulon activity was quantified per cell using the AUCell algorithm to generate a regulon-by-cell activity matrix.

hdWGCNA analysis

Gene co-expression networks in CRC epithelial cells were analyzed using the hdWGCNA package. A Seurat object containing epithelial cells was used to construct metacells with parameters: k=25, assay='SCT', max_shared=20, min_cells=300, reduction='umap', slot='data'. Cells of the same epithelial subtype within each sample were pooled to form metacells. Networks were built using ConstructNetwork with a soft-power threshold of 14. Functional enrichment of gene modules was performed using the EnrichR package.

Spatial transcriptomic analysis with SPATA2

Spatial transcriptomic data were analyzed using the SPATA2 R package. After log-normalization, highly variable genes were identified using the “vst” method. Spatially variable genes were detected using SPARK-X ($p < 0.01$). PCA was used for dimensionality reduction,

followed by k-means clustering to define spatial domains. Spatial trajectories were either manually drawn or defined by spatial coordinates. Genes with significant expression gradients along these trajectories were identified using spatialTrajectoryScreening ($FDR < 0.05$). Both random and non-random gradients were visualized across tissues. Top genes were selected based on model fit for downstream analysis.

Clinical samples

Forty matched samples of colorectal cancer tissues and their corresponding adjacent normal tissues were collected from patients undergoing surgical resection at Yangpu Hospital, affiliated with Tongji University, between November 2018 and November 2019. The study protocol was approved by the Ethics Committee of Yangpu Hospital (Approval No. LL-2023-LW-012), and written informed consent was obtained from all participants.

RT-qPCR

Total RNA was isolated from matched tumor and adjacent normal tissue samples using TRIzol reagent (Invitrogen), followed by reverse transcription into complementary DNA (cDNA) using a commercial kit (Takara, Dalian, China), according to the manufacturer's instructions. Quantitative real-time PCR (RT-qPCR) was subsequently carried out with gene-specific primers: SCAND1: 5'-CGCAGAGAAGCCAGAGACTT-3' and 5'-TCAGCACTGCGTCTGCACC-3'; β -actin: 5'-CCTGGC ACCCAGCACAAAT-3' and 5'-GGGCCGGACTGTCATAC-3'.

Western blot assay and immunohistochemistry (IHC)

Western blotting and immunohistochemistry (IHC) were independently performed in triplicate. For protein extraction, tissues were lysed in RIPA buffer (Solarbio, China) with protease inhibitors (1:100, Thermo Scientific). For IHC, Formalin-fixed, paraffin-embedded tissue blocks were sectioned at 4 μ m thickness. Sections were dewaxed, rehydrated, and subjected to antigen retrieval using a pressure cooker for 30 minutes. Endogenous peroxidase activity was blocked using 3% hydrogen peroxide for 20 minutes. Non-specific binding was minimized with 5% BSA for 40 minutes. Sections were incubated overnight at 4 °C and visualization was achieved using a DAB detection kit and counterstaining with hematoxylin. The primary antibodies are used as follows: SCAND1 (1:1,000; Thermofisher, PA5-49816), β -actin (1:4,000; ProteinTech, 66009-1-Ig), ZEB2 (1:1,000; ProteinTech, 14026-1-AP), Twist1 (1:1,000; ProteinTech, 25465-1-AP), N-Cadherin (1:1,000; ProteinTech, 22018-1-AP), E-Cadherin (1:1,000; ProteinTech, 20874-1-AP), α -SMA (1:1,000; ProteinTech, 14395-1-AP),

Vimentin(1:1,000; ProteinTech, 10366–1-AP), β -tubulin (1:4,000; ProteinTech, 10094–1-AP).

Multiplex immunohistochemistry staining (mIHC)

Paraffin-embedded tissue sections were deparaffinized in xylene and rehydrated using graded ethanol series. Antigen retrieval was carried out in EDTA buffer, followed by quenching of endogenous peroxidase with hydrogen peroxide. Non-specific binding was blocked using 3% bovine serum albumin (BSA). Sections were then sequentially incubated with primary antibodies against S100A4 (1:500, Proteintech, 16105–1-AP), CEBPB (1:500, Proteintech, 83791–6-RR), and ARGLU1 (1:500, Proteintech, 28155–1-AP), PanCK (1:400, ZSGB, ZM0069, 1:200), FOSL1(Proteintech, Ag25788), followed by species-matched secondary antibodies. Tyramide signal amplification was applied post-incubation, with antigen retrieval and antibody stripping repeated between cycles. Nuclei were counterstained with DAPI, and slides were mounted in antifade medium. LM samples, matched primary tumors, and corresponding adjacent normal tissues were obtained from six patients with metastatic colorectal cancer and subjected to mIHC analysis.

Cell culture and transfection

Human colorectal cancer (CRC) cell lines HCT116, LOVO, SW620, RKO along with the normal colonic epithelial cell line NCM460, were obtained from the Shanghai Institute of Biochemistry and Cell Biology. All cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS; Gibco, Carlsbad, CA, USA) and maintained at 37 °C in a humidified atmosphere containing 5% CO₂. Transfection was performed using Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA) with either a SCAND1-specific siRNA or a negative control (GeneChem, Shanghai, China). Cells were harvested for subsequent experiments 48 hours post-transfection. Each experiment was conducted in triplicate. For overexpression studies, a plasmid encoding SCAND1 was custom-synthesized by GenePharma (Shanghai, China). The sequences of the siRNAs are provided in Table S1.

Proliferation and apoptosis assays

To evaluate DNA synthesis activity, colorectal cancer (CRC) cell lines were incubated with 5-ethynyl-2'-deoxyuridine (EDU) at a final concentration of 50 μ M. After a 30-minute exposure, nuclear DNA was counterstained with Hoechst 33342, and EDU-positive cells were visualized microscopically. HCT116 cells overexpressing SCAND1 and SW620 cells with SCAND1 knockdown were dissociated into single-cell suspensions using 0.25% trypsin. Cisplatin(1 μ mol/L) was used as apoptosis inducer. Subsequently, the cells were stained with

Annexin V-APC and 7-Aminoactinomycin D (7-AAD), and apoptotic populations were quantified by flow cytometry.

Cell viability assay

Cellular viability was evaluated via the Cell Counting Kit-8 (CCK-8) assay. HCT116 cells were plated in 96-well plates at 5×10^3 cells/well and incubated for 48 hours prior to absorbance measurement. Viability was determined using the CCK-8 kit (Beyotime Biotechnology, China), with optical density at 450 nm quantified on a SpectraMax i5x microplate reader (Molecular Devices).

Migration and invasion assays

Migration and invasion assays were performed using 24-well Transwell chambers (Nest, China). Cells were seeded in serum-free DMEM (250 μ L) into the upper chamber, and 600 μ L of DMEM with 10% FBS was added to the lower chamber. For invasion assays, inserts were pre-coated with Matrigel (2 mg/mL). After 24 hours, non-migrated/invaded cells were removed, and cells on the lower membrane surface were fixed with 4% paraformaldehyde and stained with crystal violet for 10 minutes. Cells were quantified in five non-overlapping fields under a microscope (Nikon, Japan). Confluent cells were scratched using a 10 μ L pipette tip for the wound healing assay and cultured in a serum-free medium. Images were taken at 0 and 24 hours using phase-contrast microscopy to assess wound closure.

Cell co-culture assays

To investigate the functional interactions between T cells and CRC cells, indirect co-culture experiments were performed using a Transwell system (Corning, USA) with 0.4 μ m pore-size polycarbonate membrane inserts, allowing soluble factor exchange while preventing direct cell-cell contact. The system was set up in 6-well plates. Jurkat cells (Human T cells) were obtained from commercial sources (QS-BIO, #BY-0094).

For each co-culture condition, 5×10^4 HCT116–SCAND1-OE or control group were seeded into the upper chamber (insert) in 1.5 mL of RPMI 1640 medium (Yamei, CB004) supplemented with 10% fetal bovine serum (FBS, Yamei, CY106) and 1% penicillin-streptomycin (Yamei, CB010). Cells were allowed to adhere overnight at 37 °C in a humidified 5% CO₂ incubator. Concurrently, 1×10^5 Jurkat cells (at a 2:1 Jurkat cell-to-cancer cell ratio) were seeded into the lower chamber in 2.5 mL of the same medium. Co-cultures were incubated for 48 h under standard conditions. All experimental conditions were performed in biological triplicate. Cellular viability and apoptosis assays of HCT116 cells was evaluated after 48 h of indirect co-culture.



(See figure on previous page.)

Fig. 1 Integrated single-cell transcriptomic landscape of normal colorectal tissue, primary colorectal cancer, and matched liver metastasis. **(A)** UMAP projection of the integrated single-cell dataset, comprising 203,379 cells, reveals 41 clusters grouped into 8 major cell types. Each color represents a different cell cluster. **(B)** Heatmap of Spearman correlation coefficients among clusters, illustrating transcriptional similarity across cell populations. **(C)** Expression levels of canonical marker genes across major cell types. **(D)** Bar plot and UMAP visualization showing the number and distribution of each cell type. **(E)** Heatmap of the top 50 most highly expressed genes specific to each major cell type, highlighting transcriptional features. **(F)** Pie charts showing the proportions of each cell type in the CRC, Normal, and Im (liver metastasis) groups. **(G–H)** MiloR k-nearest neighbor (kNN) differential abundance analysis between groups. Each node represents a local cellular neighborhood, with color intensity indicating the log₂ Fold change in abundance between crc and normal **(G)**, or LM and normal **(H)**. White nodes denote non-significant differences (FDR > 10%). Box plots on the right summarize cell type abundance changes between groups. **(I)** UMAP plots displaying the distribution of cluster-wise differentially expressed genes (DEGs) in crc vs. normal and lm vs. normal comparisons. **(J–K)** Bar plots showing enrichment of reactome pathways among DEGs in epithelial cells (CRC vs. normal, J) and T cells (LM vs. normal, K), with pathways ranked by p-value

Statistical analysis

All statistical analyses and visualizations were performed in R v4.1.3. Pearson correlation coefficients were used to assess relationships between continuous variables. For quantitative comparisons, two-tailed unpaired Student's t-tests or one-way ANOVA with Tukey's post hoc test were used. A p-value < 0.05 was considered statistically significant.

Results

Cellular heterogeneity and transcriptomic perturbation in colorectal cancer and liver metastasis

To construct a comprehensive single-cell transcriptomic atlas of normal and malignant colorectal tissues, we systematically collected and analysed single-cell RNA sequencing (scRNA-seq) data from 18 normal colorectal tissue samples, 45 colorectal cancer (CRC) samples, and 6 liver metastasis (LM) samples [27–29]. After stringent quality control and preprocessing, a total of 203,379 high-quality cells were retained for downstream analysis (Fig. 1A, Figure S1A–B). To comprehensively delineate the cellular composition of normal colorectal tissue, we first performed batch effect correction across datasets, followed by graph-based unsupervised clustering, ultimately identifying 41 distinct cell clusters (Fig. 1A, Figure S1C–D). On the basis of Spearman correlation analysis and known cell type-specific marker genes [30–32], these clusters were further annotated into eight major cell types: T cells, B cells, epithelial cells (Epi), myeloid cells (Myc), plasma cells, NK cells, fibroblasts, and endothelial cells (Fig. 1B–C, Figure S1E–F). T cells were the most abundant population in our dataset (104,664 cells), followed by epithelial cells. Each cell type displayed typical and specific expression profiles, confirming the accuracy of the annotations (Fig. 1D–E, Table S4). We next quantified the proportions of each cell type in normal, CRC, and LM samples and used the MiloR tool to analyse changes in cell abundance between groups. Compared with normal tissue, CRC samples presented a significant increase in myeloid and epithelial cell abundance. Compared with normal tissues, B cells, fibroblasts, and plasma cells were decreased in the LM group, whereas myeloid and T cells were increased. These changes suggest substantial

remodelling of the immune microenvironment and tissue architecture during tumour progression and metastasis (Fig. 1F–H).

We further compared the number of differentially expressed genes (DEGs) numbers for each cell type across groups to assess the extent of transcriptomic perturbation (Fig. 1I). Compared with normal tissues, both CRC and LM epithelial cells presented the greatest number of DEGs. Notably, T cells presented a much greater DEG count in LM vs. normal than in CRC vs. normal, suggesting more pronounced transcriptomic reprogramming and potential functional changes in the metastatic liver microenvironment (Table S2–S3).

To explore the functional significance of these DEGs, we performed pathway enrichment analysis using the Reactome pathway database (Fig. 1J–K, Figure S2A–B). In epithelial cells, DEGs between CRC and normal tissues were enriched mainly in translation-related pathways such as translation and eukaryotic translation initiation/elongation, reflecting the high dependency of cancer cells on protein synthesis and growth. L13a-mediated translational silencing also indicated posttranscriptional regulatory abnormalities in CRC. For T cells, DEGs in LM vs. normal samples were significantly enriched in pathways related to cellular responses to stimuli and stress, implying that T cells in liver metastases are under greater metabolic stress and functional impairment. The enrichment of stress response pathways, such as HSF1 activation, further suggests that T-cell exhaustion and immune escape occur [33, 34]. Additionally, we applied the Augur algorithm, a random forest-based tool [35], to evaluate the degree of transcriptomic perturbation in each cell type. Compared with normal tissue, epithelial cells presented the most prominent perturbation in both CRC and LM samples, whereas fibroblasts presented the greatest changes between LM and CRC samples, which were associated mainly with the coagulation and glycolysis pathways (Figure S2C–E). Given that epithelial cells are the principal origin of malignancy in colorectal cancer (CRC), these observations motivated a focused analysis of the epithelial compartment in subsequent sections.

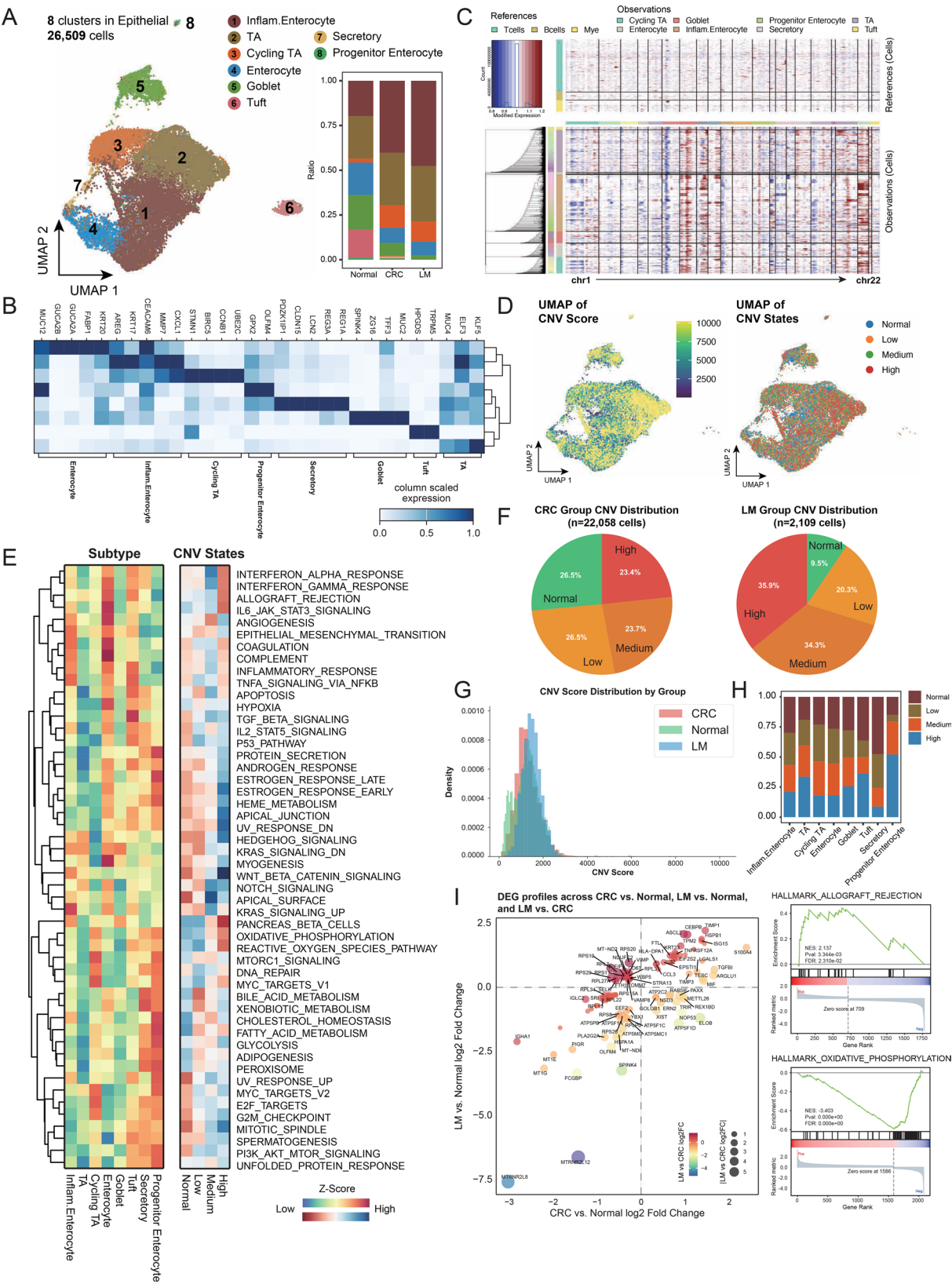


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Single-cell transcriptomic analysis of epithelial cell subtypes and CNV states in colorectal cancer and liver metastasis. **(A)** UMAP visualization of epithelial cell subtypes, including inflammatory enterocyte, transit-amplifying cell (TA), Cycling TA, Enterocyte, goblet cell, tuft cell, secretory cell, and progenitor enterocyte. The stacked bar plot shows the proportion of each subtype in normal, CRC, and LM groups. **(B)** Heatmap of canonical marker gene expression used for identification of epithelial subtypes. **(C)** Inferred copy number variation (CNV) profiles of epithelial subtypes using inferCNV, with T cells, B cells, and myeloid cells as reference populations. **(D)** UMAP plots of inferred CNV score and CNV state (Normal, Low, Medium, High) distribution across epithelial cells. **(E)** GSVA heatmaps showing pathway activity across epithelial subtypes and CNV states, based on the MSigDB Hallmark gene sets. **(F)** Pie charts displaying the distribution of CNV states in CRC and LM groups. **(G)** Density plot of CNV scores in epithelial cells across normal, CRC, and LM groups. **(H)** Stacked boxplots depicting the proportion of CNV states within each epithelial cell subtype. **(I)** Scatter plot showing shared differentially expressed genes (DEGs) among the three group comparisons: CRC High CNV vs. Normal normal CNV, LM high CNV vs. Normal normal CNV, and LM high CNV vs. CRC high CNV. The x-axis represents CRC vs. Normal, the y-axis represents LM vs. Normal, and color/size indicate LM vs. CRC. Representative GSEA plots display pathway enrichment among DEGs

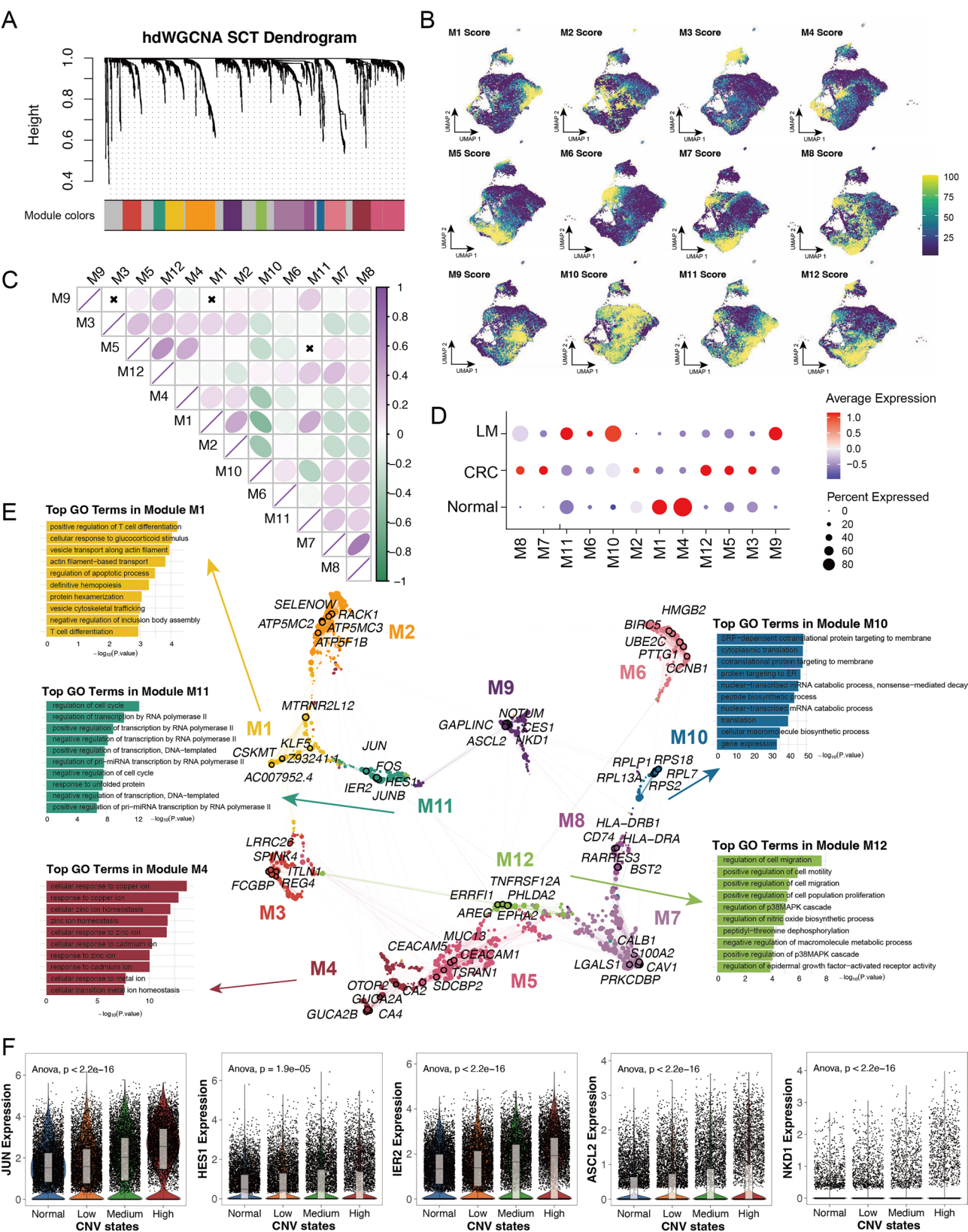
Epithelial dynamics and CNV-linked inflammatory rewiring in CRC metastasis

We performed an in-depth analysis of the epithelial lineage. A total of 26,509 epithelial cells were analysed, and unsupervised clustering identified eight distinct subpopulations. On the basis of the literature [32] and canonical marker gene expression, these subtypes were annotated as follows: inflammatory enterocyte (Inflam.Enterocyte; CXCL1, MMP7, CEACAM6, KRT17, AREG), transit-amplifying cell (TA; KLF5, ELF3, MUC4), cycling transit-amplifying cell (Cycling TA; UBE2C, CCNB1, BIRC5, STMN1), enterocyte (KRT20, FABP1, GUCA2A, GUCA2B, MUC12), goblet cell (Goblet; MUC2, TFF3, ZG16, SPINK4), tuft cell (Tuft; TRPM5, HPGDS), secretory cell (Secretory; REG1A, REG3A, LCN2, CLDN15, PDZK1IP1), and progenitor enterocyte (OLFM4). We further quantified the proportions of each subtype across disease states. In both CRC and liver metastasis (LM) samples, the fractions of inflammatory enterocytes, transit-amplifying cells, and their cycling subsets were markedly increased compared with those in normal tissue, whereas mature enterocytes, goblet cells, and tuft cells were significantly reduced (Fig. 2A–B, Figure S3, S4A–E). These changes suggest that different epithelial subtypes have distinct biological functions and malignant potential during disease progression. To explore functional features, we applied GSVA to the MSigDB hallmark gene sets and detected enrichment of epithelial-mesenchymal transition (EMT), the inflammatory response, and KRAS signalling in inflammatory enterocytes, whereas proliferative TA cells were enriched for MYC and mTORC signalling (Fig. 2E).

To further assess the malignant potential of the epithelial subtypes, we inferred copy number variation (CNV) levels using inferCNV [36]. The results revealed significant heterogeneity in CNV levels among epithelial subtypes. CNV scores were divided into five intervals (median 0, ± 1 , ± 2) and further categorized as normal, low, medium, or high, following previous approaches [37], and visualized using UMAP (Fig. 2C–D, Figure S4F–G). High-CNV epithelial cells exhibited significant activation of interferon responses (INTERFERON_ALPHA_RESPONSE,

INTERFERON_GAMMA_RESPONSE), IL6_JAK_STAT3 signalling, oxidative phosphorylation, and reactive oxygen species pathways, indicating enhanced inflammatory signalling, metabolic reprogramming, and oxidative stress in these cells (Fig. 2E). Analysis of group-specific CNV distributions revealed that the proportion of high-CNV epithelial cells was significantly greater in LM than in CRC (Fig. 2F–G). Among the subtypes, TA, cycling TA, and progenitor enterocyte had the largest fractions of high-CNV cells (Fig. 2H).

This high-resolution structure enabled us to compare the functional states of cells with similar CNV levels across different groups, providing insights into the mechanisms of tumour progression. To reveal the molecular drivers of CRC progression and liver metastasis, we performed integrated differential gene expression analysis among normal epithelial cells (normal CNV), primary CRC cells (high CNV), and liver metastasis cells (high CNV), identifying 114 intersecting DEGs (Figure S5A). Among these, CEBPB—a core transcription factor regulating inflammation and EMT [38]—was specifically upregulated in CRC and further activated in LM, providing an additional extension to previous findings. In contrast, ARGLU1, a transcriptional coactivator involved in RNA processing and chromatin regulation [39] which had been rarely reported in previous studies of cancer, was specifically downregulated in LM, reflecting further suppression of nuclear gene regulatory networks in metastatic cells. The classical EMT marker S100A4 was significantly upregulated in both CRC and LM, underscoring its central role in promoting migration and invasion. Meanwhile, the mIHC confirmed the expression pattern of those genes in the clinical samples (Figure S6). Notably, the mitochondria-derived neuroprotective gene MTRNR2L8 was markedly suppressed in LM, which may reduce the antiapoptotic capacity of metastatic cells and affect their survival and adaptation in heterogeneous microenvironments (Fig. 2I). Additionally, we conducted GSEA for group-specific DEGs. In CRC (high CNV) vs. normal (normal CNV) samples, EMT was significantly increased, whereas oxidative phosphorylation was decreased (Figure S5B–D). In LM vs. normal, the apoptosis and EMT pathways were activated, whereas oxidative



(See figure on previous page.)

Fig. 3 Gene co-expression network analysis of epithelial cells using hdWGCNA. **(A)** Dendrogram of gene modules identified by hdWGCNA, with distinct modules indicated by unique colors. **(B)** UMAP visualization of module eigengene scores (M1–M12), showing the spatial distribution of each co-expression module across epithelial cells. **(C)** Pairwise correlation matrix of module eigengenes, illustrating relationships between different gene modules. **(D)** Dot plot showing average expression and percent of cells expressing each module eigengene in Normal, CRC, and LM groups. **(E)** Network representation of representative modules, highlighting hub genes and enriched Gene Ontology (GO) terms for selected modules (M1, M2, M4, M6, M10, M11, M12). **(F)** boxplots showing the expression of representative genes across different CNV states (Normal, Low, Medium, High). Statistical significance was assessed by anova

phosphorylation and fatty acid metabolism were suppressed (Figure S5E–F). In LM vs. CRC, allograft rejection was activated, and oxidative phosphorylation was downregulated (Figure S5G–I).

Distinct epithelial coexpression modules drive metastatic adaptation

To dissect the molecular changes in epithelial cells during CRC progression and metastasis, we performed high-dimensional weighted gene coexpression network analysis (hdWGCNA) [40] on all epithelial cells, identifying 12 coexpression modules and constructing cell type-specific gene interaction networks for each (Fig. 3A–B, Figure S7A–B). Modules M1 and M4 were specifically activated in normal epithelial cells. M1 was enriched for cytoskeletal vesicle transport, immune regulation, and apoptosis, with key hub genes such as KLF5 and CSKMT. M4 was associated with copper and zinc ion homeostasis, with the core genes GUCA2B, CA4, and OTO2, suggesting a role in maintaining epithelial ion balance and barrier function [41, 42]. Notably, M1 and M4 were strongly correlated, indicating potential synergistic regulation of epithelial homeostasis (Fig. 3C–E).

The module scores map onto the UMAP distribution of epithelial subtypes, and multiple node genes within each module point to the physiological or pathological functions of these subtypes (Figure S7C). For M3—Goblet, TFF3 and SPINK4 are classic goblet cell-associated secreted molecules: TFF3 contributes to mucus-layer repair and epithelial restitution [43, 44]; SPINK4 is a serine protease inhibitor-like secreted factor that reflects goblet cell secretory activity and is implicated in intestinal barrier function and inflammation regulation [45]. In addition, REG4 is frequently overexpressed in colorectal cancer and is associated with epithelial regeneration and pro-growth signaling, suggesting a regenerative program and tumor-promoting role of the goblet/secretory lineage. For M4—Enterocyte, GUCA2A/GUCA2B indicate the ion and fluid homeostasis axis of mature absorptive epithelium [46]; OTO2 is involved in ion transport, reflecting distal colonic absorption/secretion features and potentially linking to changes in the metabolic microenvironment in tumors [47]. For M6—Cycling TA, CCNB1 and UBE2C are key drivers of the G2/M phase and mitotic progression, indicating robust proliferative activity; BIRC5 (Survivin) is an anti-apoptotic and mitotic regulator associated with stemness, invasiveness,

and poor prognosis in CRC, together delineating a highly proliferative, apoptosis-resistant pathological profile of transit-amplifying cells [48, 49].

In contrast, modules M9, M10, and M11 were specifically activated in liver metastasis (LM) epithelial cells. Functional annotation revealed that the M9 module regulated Wnt signalling and EMT, likely promoting tumour cell migration and invasion. M10 was enriched for protein synthesis and ribosome function, centred on ribosomal subunits such as RPS18 and RPL7, reflecting enhanced translational capacity in metastatic cells. M11 was linked to the cell cycle and RNA polymerase II-driven transcription, with hub genes including FOS, JUN, IER2, and JUNB, indicating increased proliferative and transcriptional reprogramming capacity in LM cells. Together, these results suggest that LM epithelial cells transition from homeostatic to malignant adaptation via activation of Wnt/EMT, protein synthesis, and cell cycle coexpression networks, conferring greater invasiveness and adaptability (Fig. 3E, Figure S7C).

Importantly, most core genes in the M9, M10, and M11 modules were closely associated with CNV levels and were upregulated in more malignant cells. For example, JUN, a classical AP-1 transcription factor, drives cell cycle progression and EMT, promoting tumour migration and drug resistance; ASCL2, a downstream Wnt target, is essential for CRC stemness and regeneration, and its upregulation may enhance self-renewal and adaptability in metastatic cells; and NKD1, a negative regulator of Wnt signalling, may reflect complex Wnt pathway regulation in tumour cells [50–52]. The high expression of these genes, which is correlated with increased CNV, highlights the molecular basis for functional diversity and adaptation in tumour progression and metastasis through coexpression network remodelling (Fig. 3F).

Transcription factor network remodelling drives CRC progression and metastatic adaptation

To systematically dissect the key regulatory mechanisms of epithelial subpopulations across CRC disease stages and metastasis, we used pySCENIC to infer transcription factor (TF) regulatory networks from single-cell transcriptomic data, revealing subtype-specific changes in TF activity (Fig. 4A, Figure S8A). Dimensionality reduction of TF activity profiles revealed clear stratification among epithelial subtypes at the regulon level (Fig. 4B). Heatmaps of the top 50 TFs (Fig. 4C) revealed that

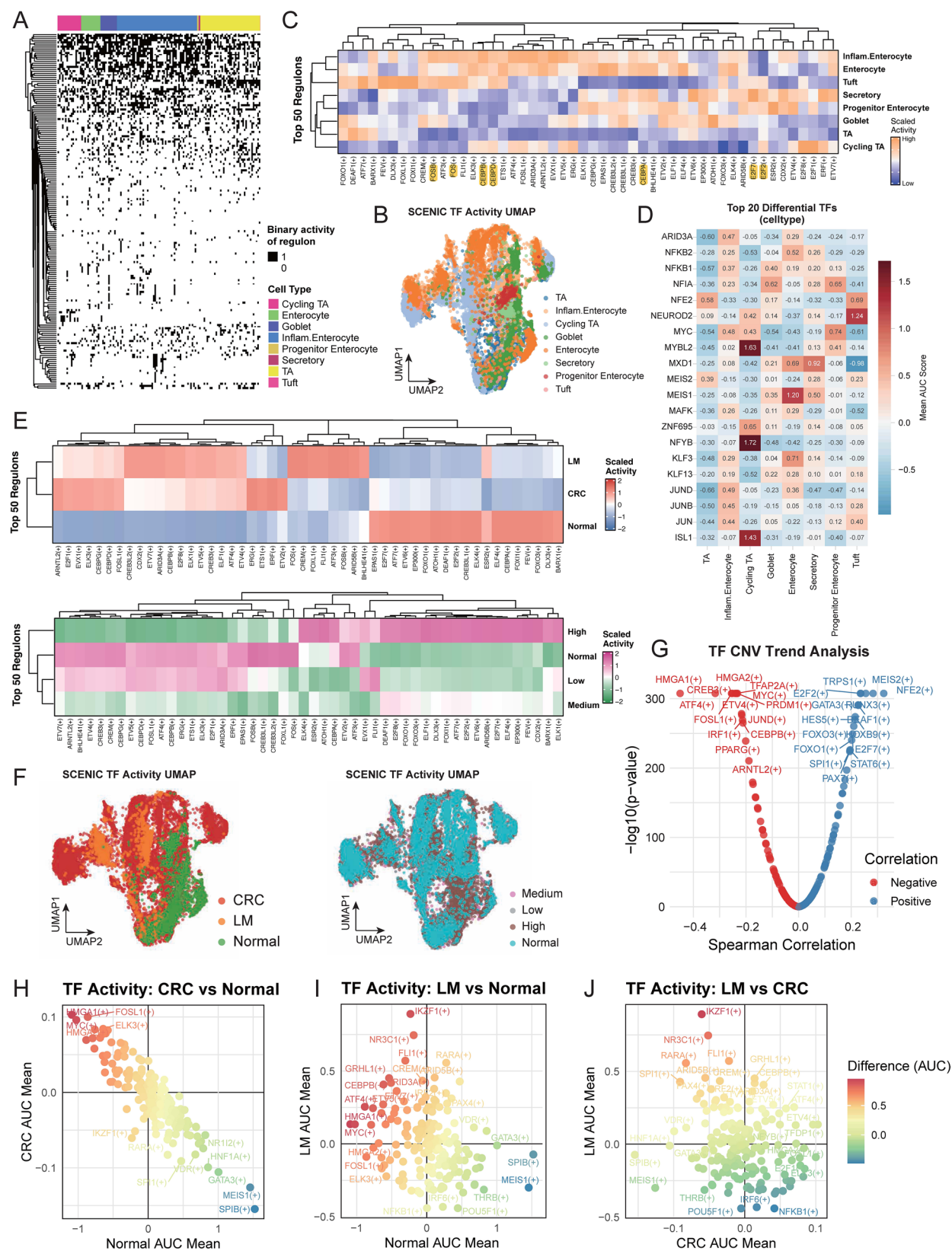


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Transcription factor regulatory network analysis of epithelial cells using pySCENIC. **(A)** Binary regulon activity matrix showing the active status of transcription factor regulons across epithelial cell subtypes. **(B)** UMAP visualization of single-cell regulon activity, colored by epithelial cell subtype. **(C)** Heatmap displaying scaled activity of the top 50 regulons across cell subtypes. **(D)** Heatmap of the top 20 differentially active transcription factors (TFs) for each epithelial cell subtype, based on mean AUC scores. **(E)** Heatmaps of regulon activity across disease groups (Normal, CRC, LM; upper panel) and CNV states (Normal, Low, Medium, High; lower panel). **(F)** UMAP plots showing regulon activity colored by disease group (left) and CNV state (right). **(G)** Scatter plot showing correlation between regulon activity and CNV score, with significantly correlated TFs highlighted. **(H–J)** Scatter plots comparing mean regulon (tf) activity (AUC) between CRC and Normal **(H)**, LM and Normal **(I)**, and LM and CRC **(J)** groups; Color scale indicates the magnitude of activity differences

CEBPB was highly activated in inflammatory and absorptive enterocytes, which is consistent with its known roles in inflammation, tumour microenvironment remodeling, and EMT—key processes in CRC metastasis and immune escape [38, 53]. The E2F family (E2F1, E2F7, and E2F8) is specifically activated in proliferative subpopulations, whereas JUN, JUNB, and JUND (AP-1 family) are enriched in inflammatory enterocytes, corresponding to EMT, migration, and drug resistance phenotypes [50, 54, 55]. MYBL2 and NFYB are activated in cycling TA cells and are associated with the cell cycle and poor prognosis (Fig. 4D) [56–58].

We further compared TF network activation across disease groups and CNV states (Fig. 4E–F, Figure S8B–D), finding distinct TF activation patterns related to both disease stage and genomic instability. Shared modules such as E2F1, CEBPB, and FOSL1 are activated in both primary and metastatic cancer, driving proliferation, inflammation, and transcriptional reprogramming [59–61]. FOSL1, in particular, is known to regulate EMT and invasiveness [62], and its expression was validated by mIHC (Figure S6). Primary CRC also showed activation of angiogenesis- and migration-related TFs such as ERG, ETS1, ERF, and ETV2, whereas metastases displayed specific activation of FOS, CREM, FOXL1, FLI1, and ATF3, the latter promoting tumour survival and drug resistance [63–65]. High-CNV epithelial cells presented broader and more specific TF activation, especially for the E2F family (e.g., E2F2 and E2F7), suggesting that genomic instability drives sustained proliferation. FOXO3 was also highly activated in high-CNV cells. While cells with normal or low CNV maintained relatively stable TF activation, medium and high CNV states were characterized by proliferative and stress-adaptive TFs.

Linear modelling revealed that the expression of TFs, including E2F2, TRPS1, MEIS2, NFE2, and the FOXO family (increasing) and HMGA1, HMGA2, CREB2, and TFAP2A (decreasing) (Fig. 4G), tended to increase with increasing CNV, suggesting that high-CNV tumour cells adapt to adverse microenvironments by driving the cell cycle and stress response programs. In the disease groups, IKZF1 and NR3C1 were specifically activated in liver metastases (LM), indicating roles in distant metastasis and microenvironmental adaptation (Fig. 4H, Figure S8E–H) [66, 67]. Overall, our results provide a comprehensive map of TF regulatory networks in CRC epithelial

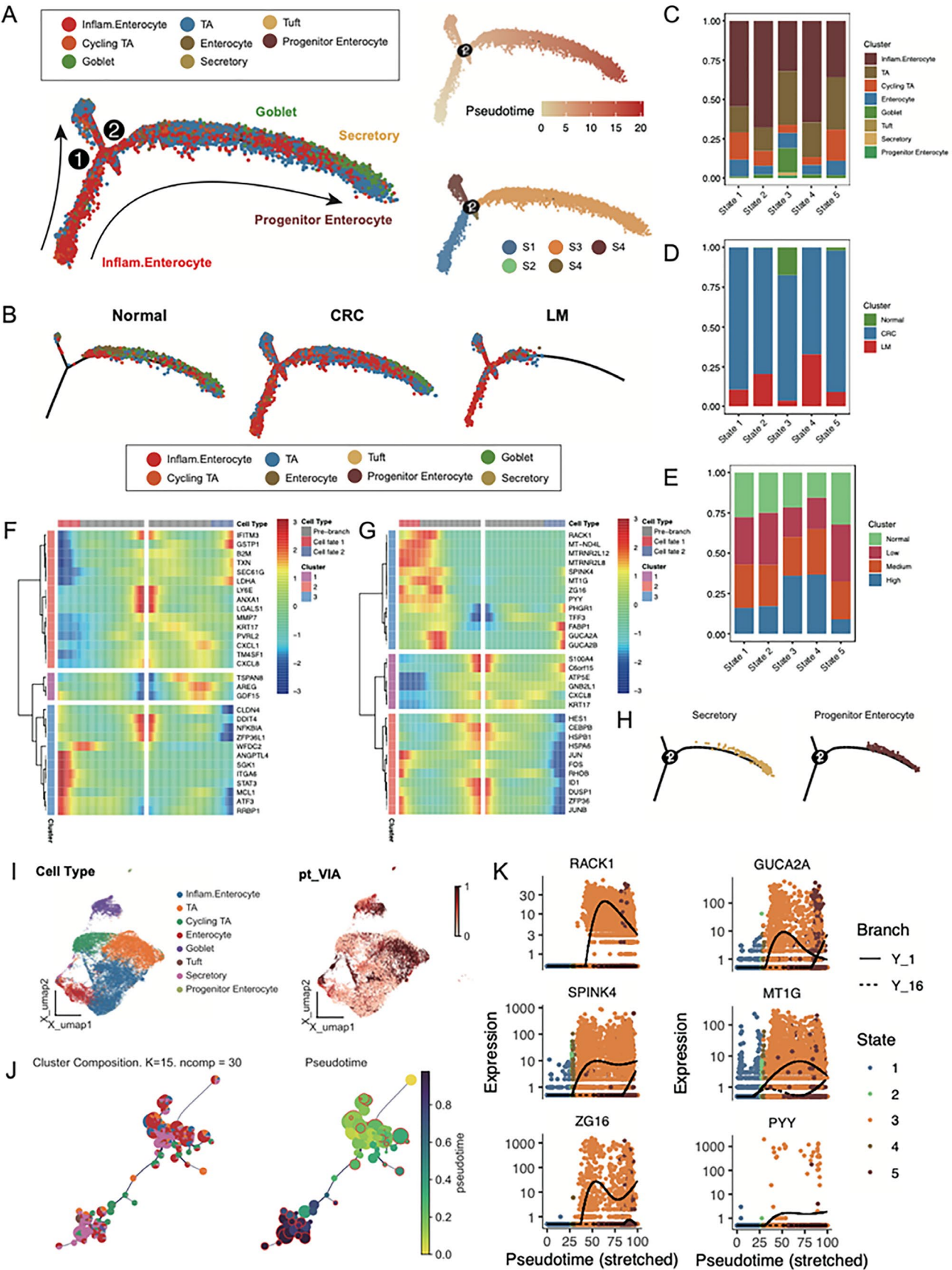
cells across disease stages, metastatic states, and genomic instability, revealing how dynamic TF activity reprogramming underlies tumour progression and adaptation.

Trajectory differentiation and state-specific molecular features of epithelial subtypes during CRC progression

To systematically characterize the dynamic changes in epithelial subtypes during disease progression, we performed pseudotime analysis of eight epithelial subtypes using Monocle. This approach reconstructed the developmental and state transition trajectories of epithelial cells and identified five major cell states (Fig. 5A–B). Analysis of the distribution of each subtype across states revealed notable differences: for example, epithelial cells from the normal group were almost exclusively found in State 3, whereas those from the liver metastasis (LM) group were predominantly in States 2 and 4. Importantly, high-CNV epithelial cells were significantly enriched in States 3 and 4, suggesting that genomic instability profoundly shapes epithelial cell states (Fig. 5C–E).

To uncover the molecular mechanisms underlying state transitions, we used BEAM analysis to examine gene expression changes at key branch points. At Node 1, Cell fate 1 (towards State 4) was associated with the upregulation of stress- and malignancy-associated genes, including the stress-response TFs ATF3 and ANGPTL4, which are involved in angiogenesis and metastasis, indicating that their activation may promote tumour cell invasiveness and adaptability (Fig. 5F–G). At Node 2, Cell fate 1 (towards State 3) was characterized by sustained upregulation of RACK1, a key signalling scaffold known to promote cancer cell migration and invasion and associated with poor prognosis in CRC, underscoring its role in tumour progression. Additionally, GUCA2A and SPINK4—markers of secretory and barrier-protective epithelial lineages—were also upregulated along this branch, reflecting the involvement of secretory and progenitor enterocyte lineages in microenvironment regulation and epithelial homeostasis (Fig. 5F–H, 5K, Table S5–S6).

To validate the robustness of these pseudotime inferences, we independently reconstructed cell trajectories using the pyVIA algorithm. The results revealed that secretory and progenitor enterocyte subpopulations were located mainly at the trajectory endpoints, whereas enterocytes were concentrated at the origin (Fig. 5I–J).



(See figure on previous page.)

Fig. 5 Trajectory and pseudotime analysis of epithelial cell differentiation using Monocle and pyVIA. **(A)** Monocle trajectory plot of epithelial cells colored by subtype (left) and cell state (right), illustrating the inferred differentiation trajectory and pseudotime progression. **(B)** Monocle trajectory plots for Normal, CRC, and LM groups, colored by cell subtype. **(C–E)** Stacked bar plots showing the distribution of **(C)** cell subtypes, **(D)** disease groups, and **(E)** CNV states across pseudotime-defined states. **(F–G)** Heatmaps of representative gene expression along pseudotime. **(H)** Monocle trajectory plots highlighting the distribution of Secretory and Progenitor enterocyte subtypes. **(I)** umap plots of epithelial cells colored by cell subtype (left) and pyVIA-derived pseudotime (right). **(J)** pyVIA trajectory graph showing cluster composition and pseudotime progression. **(K)** Expression dynamics of selected marker genes (RACK1, GUCA2A, SPINK4, MT1G, ZG16, PYY) along pseudotime and different trajectory branches

Immune subset reprogramming underlies tumour progression and metastatic adaptation

Given the marked perturbation of immune responses and the microenvironment observed in tumour samples, we systematically analysed lymphoid and myeloid cell populations. Unsupervised clustering divided lymphoid cells into eight subpopulations, which were annotated on the basis of specific marker gene expression [32] (Fig. 6A–B, Figure S9A–C). Notably, effector memory CD8⁺ T cells and Tregs were significantly expanded in both CRC and LM samples, indicating the activation of both effector and immunosuppressive mechanisms during tumour progression and metastasis (Fig. 6C–D).

In terms of transcriptomic disturbance, Tregs presented the greatest changes in CRC vs. normal, Tph cells in LM vs. normal, and Tn/Tcm cells in LM vs. CRC (Fig. 6E). GSEA revealed that in CRC vs. normal tissues, Tregs exhibited activation of the IL6_JAK_STAT3_SIGNALING pathway, which enhances Treg immunosuppressive function and promotes tumour immune evasion [68]. In LM vs. normal, Tph cells presented significant enrichment of interferon responses (IFN- γ and IFN- α signalling), suggesting roles in antiviral and inflammatory regulation within the metastatic liver environment [69]. In LM vs. CRC, the complement pathway was strongly activated in Tn/Tcm cells, indicating the involvement of the complement system in remodelling the metastatic niche (Fig. 6F–G). Together, these results highlight that Tregs are both expanded and transcriptionally remodeled in CRC and LM, suggesting a key role in tumour development, whereas Tph cells are consistently altered during liver metastasis, indicating that they are critical for metastatic adaptation (Fig. 6E, Figure S9D–F).

Further analysis revealed that LGALS1 was upregulated in Tregs from both CRC and LM, with further elevation in LM; this gene is known to promote Treg tolerance and suppress effector T-cell function [70, 71]. Additionally, MHC II molecules (e.g., HLA-DRA and HLA-DRB1) were further activated in Tregs from LM, suggesting enhanced antigen presentation capacity and a greater contribution to tumour immune escape [72, 73] (Figure S9G). In Tph cells, the upregulation of MHC-I (B2M) and MHC-II genes (HLA-DRA, HLA-DRB1), as well as DUSP4 and TNFRSF18, indicated increased immune activity (Figure S8H–I).

For myeloid cells, we identified six subpopulations (TAMs, monocytes, cDC2s, macrophages, mature DCs,

and cDC1s). Monocytes were significantly expanded in both CRC and LM, whereas cDC1s and cDC2s were reduced (Fig. 6I–J, Figure S9J). Pathway analysis revealed that macrophages and mature DCs displayed specific pathway activation in CRC and LM, participating in microenvironment remodelling via the regulation of inflammation, antigen presentation, and immunosuppression (Fig. 6K).

Inflammatory enterocytes and TAMs orchestrate key signalling hubs

On the basis of the identified epithelial and immune cell subsets in CRC, we systematically analysed cell-cell communication networks across different disease stages using CellChat. The results revealed that the number of cell-cell interactions was significantly greater in CRC and LM tissues than in normal tissues, with CRC exhibiting the highest levels (Fig. 7A). Directionality analysis revealed that in CRC, inflammatory enterocytes (Inflam. Enterocyte), tumour-associated macrophages (TAMs), and Th17 cells were the main signal receivers, whereas Inflam. Enterocyte and Th17 cells were also major signal senders. In contrast, in LM, effector memory T (Tem) cells were the main signal receivers, and TAMs were the most active signal senders (Fig. 7B), indicating a notable shift in immune regulation within liver metastases, where TAMs may promote immune adaptation by modulating Tem function [74].

Analysis of key signalling sources revealed that in CRC, Inflam. Enterocyte predominantly transmitted MIF, VEGF, CCL, and PDGF signalling axes. MIF and CCL are involved in immune cell recruitment and inflammation, whereas VEGF and PDGF directly participate in angiogenesis and tissue remodelling [75–77]. The active engagement of these pathways suggests that inflammatory epithelial cells may drive both immune evasion and tumour angiogenesis in a coordinated manner (Fig. 7C). In LM, TAMs communicate with surrounding cells mainly via SPP1 (osteopontin), GALECTIN, and BAFF signalling, indicating a role for TAMs in promoting microenvironmental adaptation and immune suppression through immune modulation, cell adhesion, and B-cell survival [78–81]. With respect to signal reception, Inflam. Enterocyte in CRC notably received IGF1BP (insulin-like growth factor-binding protein) and PAR (protease-activated receptor) signals, which may increase epithelial cell survival and regeneration [82, 83]; in LM,

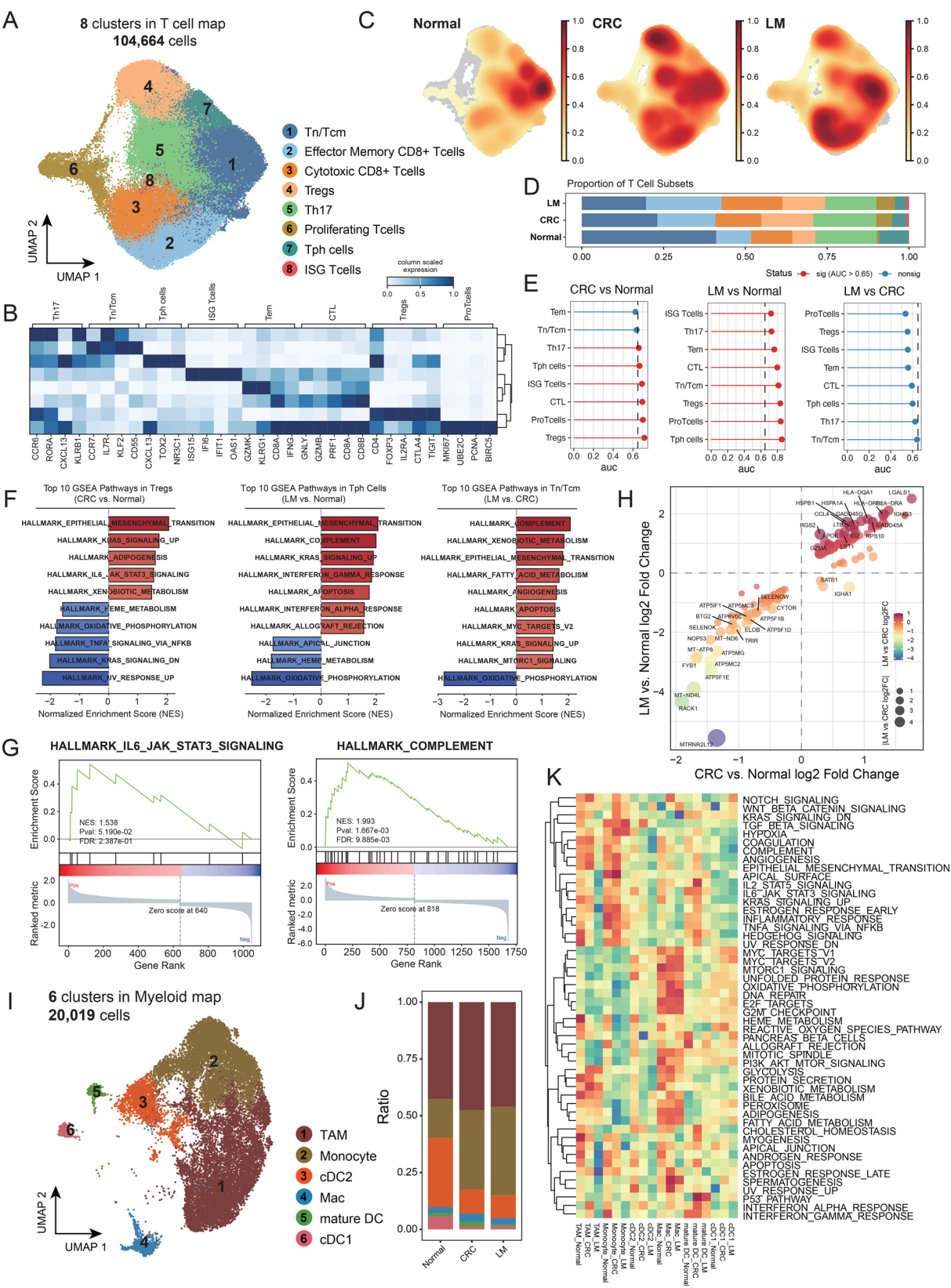


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Subclustering and functional analysis of T cell and myeloid cell populations in colorectal cancer and liver metastasis. **(A)** UMAP visualization of 104,664 T cells, revealing eight subclusters: Tn/Tcm, effector memory CD8+ T cells, cytotoxic CD8+ T cells (ctl), tregs, Th17, proliferating T cells, Tph cells, and isg T cells. **(B)** Heatmap of canonical marker gene expression for identification of T cell subtypes. **(C)** UMAP plots showing the density distribution of T cells in Normal, CRC, and LM groups. **(D)** Bar plot displaying the proportions of T cell subtypes across different disease groups. **(E)** Dot plots showing the Augur AUC scores for the most differentially enriched T cell subtypes across pairwise group comparisons (crc vs. Normal, lm vs. Normal, LM vs. CRC). **(F)** Bar plots of the top 10 GSEA hallmark pathways enriched in Tregs (CRC vs. Normal), Tph cells (lm vs. Normal), and Tn/Tcm cells (LM vs. CRC). **(G)** Representative GSEA enrichment plots for selected Hallmark pathways. **(H)** Scatter plot showing differentially expressed genes in T cells between crc vs. Normal and lm vs. Normal groups, with color indicating log2 Fold change between LM and CRC. **(I)** UMAP visualization of 20,019 myeloid cells, identifying six clusters: tam, Monocyte, cDC2, Macrophage, mature DC, and cDC1. **(J)** Stacked bar plot showing the proportion of each myeloid cell subtype across Normal, CRC, and LM groups. **(K)** Heatmap of GSVA pathway activity scores for myeloid cell subtypes, highlighting functional differences among groups

Tem cells mainly received SPP1 and GALECTIN signals, suggesting that their functional state is profoundly influenced by TAM-derived signals, potentially affecting their antitumour activity or exhaustion status (Fig. 7D).

CRC ligand-receptor pair analysis revealed that Inflammation. Enterocyte communicated with other cells via the MDK pathway axis and uniquely activated interactions such as LGALS9–P4HB/CD44, distinguishing them from TA and Th17 cells. LGALS9 is known to induce T-cell apoptosis and immune tolerance, potentially contributing to immune suppression in the TME [84]. On the receiving end, Th17 cells specifically received MIF-related signals. Previous studies have shown that MIF regulates Th17 cell differentiation and function, and our findings support a key role for MIF-driven Th17 responses in the CRC immune microenvironment [85, 86]. We observed significant enhancement of MIF and VEGF signalling in CRC. MIF is involved in inflammatory responses, immune cell recruitment, and tumour immune escape, whereas VEGF is a central driver of tumour angiogenesis [75, 77]. Our data indicate that Inflammation. Enterocyte and TA cells are the main sources of MIF signalling, primarily via the MIF-(CD74+CXCR4) and MIF-(CD74+CD44) receptor pairs. For VEGF signaling, Inflammation. Enterocytes and monocytes were the main senders, with endothelial cells as the principal receivers, predominantly via the VEGFA–FLT1 (VEGFR1) and VEGFA–KDR (VEGFR2) interactions (Fig. 7E–H).

Beyond CRC, we examined ligand–receptor pairs between intestinal epithelial subtypes and key immune communicators in Normal (B cells, Tn/Tcm) and LM (Tem, TAM, and CTL) samples (Figure S10). In Normal tissues, epithelial subtypes acted as signal senders predominantly via the MIF and LGALS9 pathways, with Cycling TA, Inflammation. Enterocyte, and Progenitor enterocyte being the main sources. MIF signals likely operate through the CD74–CXCR4 or CD74–CD44 axes to regulate recruitment and survival of lamina propria immune cells; LGALS9, via P4HB/CD44, lowers T-cell activation thresholds, thereby preventing excessive inflammation and supporting barrier function. In LM, MIF and LGALS9 remained the dominant outgoing signals from epithelial cells but were more frequent and of higher intensity, indicating a shift of epithelial programs

toward immune modulation within the hepatic metastatic niche: enhanced MIF signalling promotes inflammatory cascades and immune-cell reprogramming, while the LGALS9 axis drives T-cell suppression/exhaustion, forming a positive feedback loop with TAM enrichment. Epithelial subtypes primarily received PPIA–BSG signals, with PPIA expressed on the epithelial side responding to BSG on immune cells. This pathway likely mediates epithelial–matrix remodelling and increased motility via CD147 (BSG)-dependent activation of MAPK/ERK and MMPs [87, 88]. The same mechanism is present—and more pronounced—in LM, suggesting cooperation with the amplified MIF/LGALS9 axes to drive tissue remodelling and metastasis-related plasticity.

Tumour-associated epithelial gradients and spatial signalling networks

To further validate the spatial distribution and interactions of key epithelial subpopulations identified at the single-cell level, we analysed spatial transcriptomic data from three CRC patients in the GEO database [30, 89]. Using previously constructed single-cell annotations, we applied the RCTD algorithm to deconvolute the spatial transcriptomic data and infer the cell type composition at each spot. The results revealed that TAs (transit-amplifying cells) and Inflammation. Enterocyte (inflammatory enterocytes) were the predominant epithelial subtypes within tumour regions, highlighting their close association with tumorigenesis and microenvironmental remodelling (Fig. 8A–G, Figure S11A–B).

We then used the SPATA platform to perform spatial gradient gene expression analysis from tumour-enriched areas to distant normal regions. With increasing distance from the tumour, the expression of the chemokines CXCL2 and CXCL3 and the tight junction protein CLDN1 gradually decreased, whereas the levels of the secreted proteins SFRP2 and SFRP4 increased. CXCL2 and CXCL3 are classic inflammation-related chemokines that recruit neutrophils and other immune cells, and CLDN1 is involved in epithelial barrier alterations; their high expression in tumour zones indicates a highly inflammatory and structurally remodelled microenvironment [90, 91]. SFRP2 and SFRP4, antagonists of Wnt signalling, are upregulated in distant normal regions,

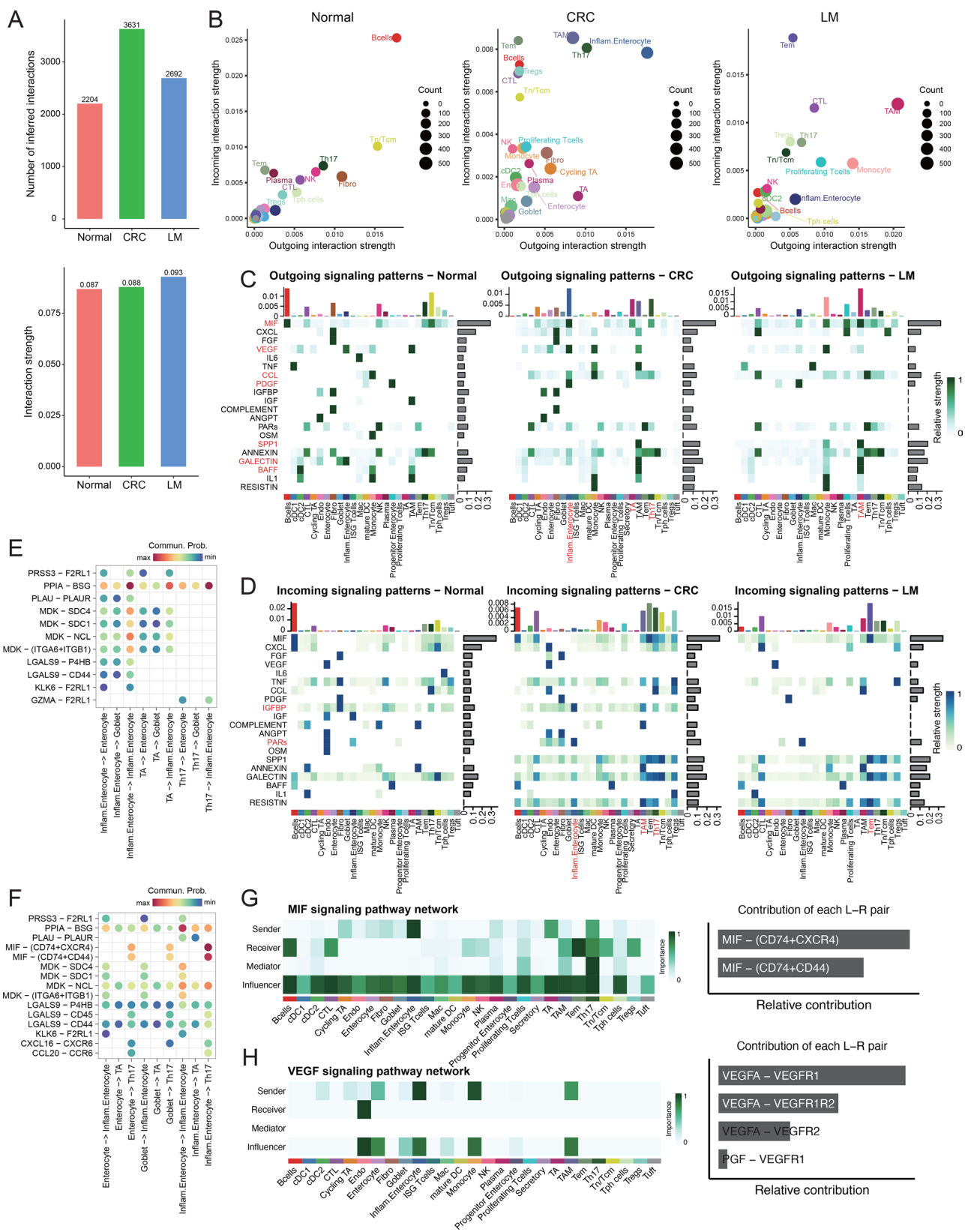


Fig. 7 (See legend on next page.)

(See figure on previous page.)

Fig. 7 Cell-cell communication analysis across normal, CRC, and liver metastasis samples using CellChat. **(A)** Bar plots showing the total number of inferred intercellular interaction numbers (top) and overall interaction strength (bottom) among cell populations in Normal, CRC, and LM groups. **(B)** Scatter plots of outgoing versus incoming interaction strength for each cell type in Normal, CRC, and LM, with point size indicating the number of inferred interactions. **(C)** Heatmaps displaying outgoing signaling patterns, highlighting the relative contribution of each signaling pathway from sender to receiver cell types in Normal, CRC, and LM. **(D)** Heatmaps showing incoming signaling patterns, illustrating the relative strength of incoming signals for each cell type across the three groups. **(E, F)** Dot plots showing the probabilities of specific ligand–receptor pairs mediating communications between epithelial and immune subtypes, with panels **(E)** and **(F)** displaying sender–receiver roles swapped. **(G, H)** Signaling pathway network analysis for MIF **(G)** and VEGF **(H)** pathways, showing sender, receiver, mediator, and influencer roles of each cell type, as well as the relative contribution of key ligand–receptor pairs. “Sender” denotes cell groups that produce and emit the pathway ligands, while “Receiver” are those expressing cognate receptors and responding to these signals. “Mediator” cells provide intermediate components that facilitate or modulate transmission between senders and receivers. “Influencer” cells express regulatory molecules that globally enhance or dampen the pathway’s communication strength across the network; VEGFR1R2” denotes a receptor complex comprising FLT1 (VEGFR1) and KDR (VEGFR2)

suggesting that normal epithelial cells may restrict aberrant proliferation through Wnt inhibition [92, 93].

To nominate genes with both malignant relevance and spatial informativeness, we intersected genes that increase monotonically from Normal to CRC to LM (Fig. 2I) with the top spatial gradient genes identified by SPATA, focusing on those expressed within epithelial compartments. For example, ASCL2 and TM4SF1 were more highly expressed near tumours and decreased with distance. ASCL2 is a key transcription factor for intestinal stem and transit-amplifying cells, indicating enhanced stemness and regenerative capacity near tumours, whereas TM4SF1 is associated with epithelial migration and invasion, supporting an active migratory state within tumour regions [94, 95] (Fig. 8H, Figure S11D–E).

Cell-cell communication analysis of the spatial transcriptomic data revealed that TA and goblet cells were central hubs in the tumour tissue network (Fig. 8I, Figure S11C). Pathway analysis revealed spatial enrichment of TGF β , WNT, CXCL, and EGF signalling, indicating that these pathways play pivotal roles in regulating intercellular interactions, tumour growth, and immune modulation (Fig. 8J–M, Figure S11F–L). Collectively, these results demonstrate that specific tumour-associated epithelial subtypes orchestrate spatial signalling networks, driving microenvironmental inflammation, cellular remodeling, and tissue homeostasis changes, thus providing new molecular insights into spatial heterogeneity and potential therapeutic targets in CRC.

SCAND1 as a novel biomarker for diagnosis and prognosis in colorectal cancer

Using SPATA, we identified genes exhibiting gradient expression patterns across the tumor parenchymal region (Fig. 8H). We then intersected this gene list with genes previously found to be upregulated in high-CNV epithelial cells from both CRC and LM groups in our scRNA-seq analysis (Fig. 2I). This analysis highlighted SCAND1, which not only displays a distinct spatial gradient but also shows progressively increased expression in high-CNV cells from CRC to LM. Moreover, SCAND1 serves as a

hub gene in the LM-specific coexpression module M9 (Figure 3E). Based on these findings, we further investigated the functional role of SCAND1. We first confirmed elevated SCAND1 expression in both paired and unpaired tumor tissues within the TCGA database (Fig. 9A). Consistent results were further validated by Western blotting ($n=24$) (Fig. 9B), RT-qPCR ($n=40$) (Fig. 9C), and immunohistochemistry ($n=6$) (Fig. 9D).

Logistic regression analysis revealed that elevated SCAND1 expression was significantly associated with advanced T stage ($p=0.045$) (Fig. 9E). TCGA and our validation cohorts confirmed higher SCAND1 levels in patients with lower T stages (Figs. 9F–9G).

Subsequently, receiver operating characteristic (ROC) curve analysis demonstrated that SCAND1 exhibited robust diagnostic performance in the TCGA colorectal cancer dataset (Fig. 9H), as well as in the separate colon and rectal cancer cohorts (Fig. 9I–J). These findings were further corroborated across multiple GSE datasets (Fig. 9K). Survival analysis confirmed that higher SCAND1 expression was associated with poorer prognosis (Fig. 9L–N), which was validated in the patient cohort from Yangpu District Central Hospital (Fig. 9O). To better assess the prognostic value of SCAND1, stratified analyses were performed across various clinical subgroups. The results revealed that patients with high SCAND1 expression had significantly poorer prognosis in the female, T3&T4, M0, age > 65, Stage II&III&IV, as well as N0&N1 subgroups (Figs. 9P–9U, Figure S12).

SCAND1 enhances malignant behaviors of CRC cells and confers resistance to T cell-mediated cytotoxicity

SCAND1 protein expression was first examined in the normal colonic epithelial cell line NCM460 and four CRC cell lines (HCT116, SW620, RKO, and LOVO). Compared with NCM460, all four cancer cell lines exhibited increased expression, with the lowest level observed in HCT116 and the highest in SW620 (Fig. 10A). In order to explore the functions of SCAND1 in CRC, it was knocked down by siRNA in SW620 and overexpressed in HCT116, and the efficiency was verified by western blotting and RT-qPCR (Fig. 10B–C). EDU results showed

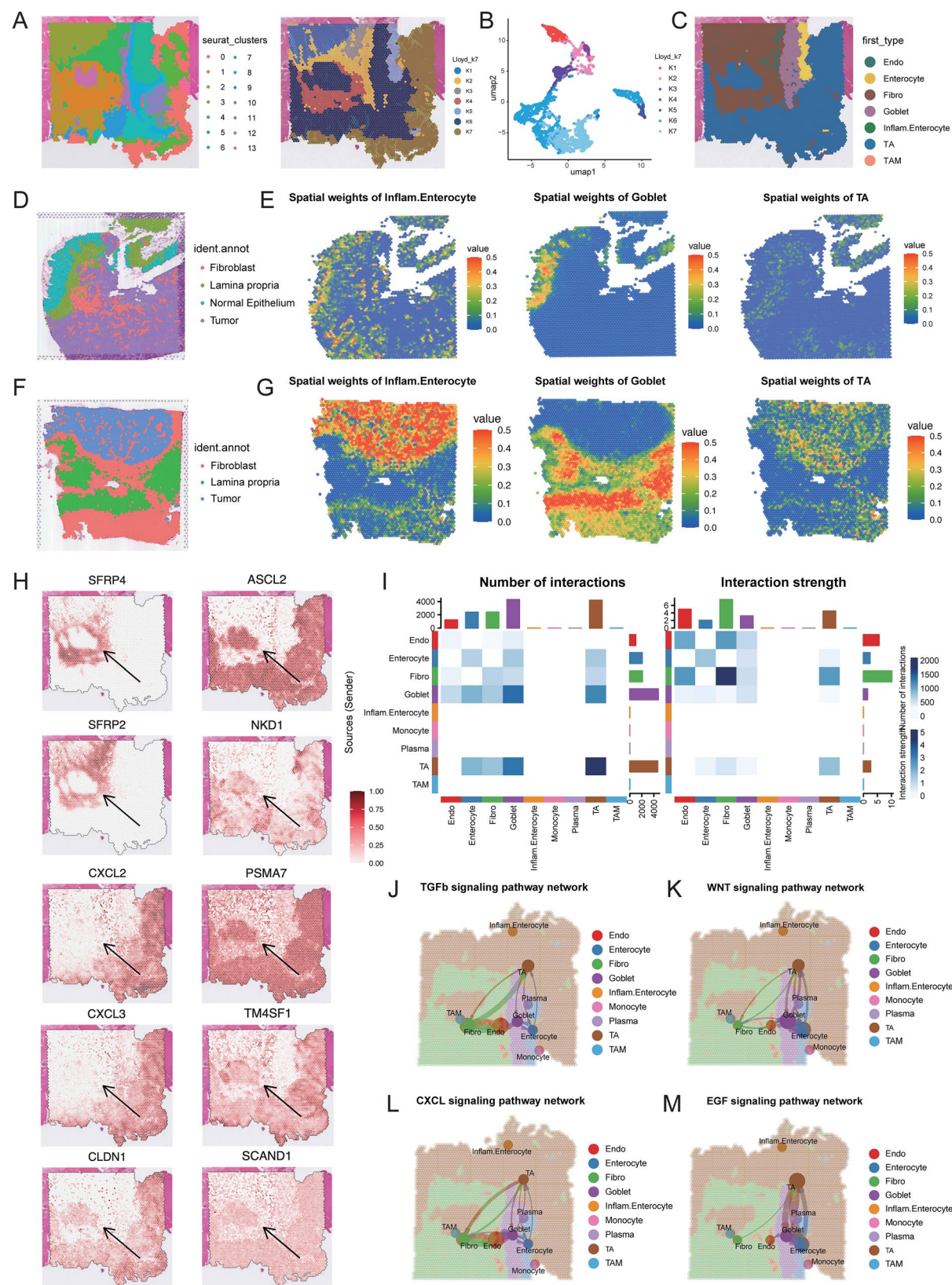


Fig. 8 (See legend on next page.)

(See figure on previous page.)

Fig. 8 Spatial transcriptomic analysis, spatial trajectory inference, and cell-cell communication in colorectal tumor tissue. **(A)** Spatial maps showing the distribution of Seurat clusters in Visium slides. **(B)** UMAP visualization of spot-level clusters identified by spatial transcriptomics. **(C)** Spatial mapping of major cell types as determined by RCTD deconvolution. **(D, F)** Annotation of spatial tissue regions, including fibroblast, lamina propria, normal epithelium, and tumor, across two spatial transcriptomic samples. **(E, G)** Spatial weights for inflammatory enterocyte, goblet, and transit amplifying (TA) cells as estimated by RCTD Spatial trajectory analysis in two samples, highlighting spatial gradients of cell-type enrichment. **(H)** Spatial expression patterns of representative genes, with arrows indicating directional signaling regions. **(I)** Heatmaps depicting the number (left) and strength (right) of inferred cell-cell interactions among major cell types, based on spatially resolved communication analysis. **(J–M)** Spatial signaling pathway networks for TGF β (**J**), WNT (**K**), CXCL (**L**), and EGF (**M**), illustrating the directionality and strength of cell-cell signaling within the tissue context

that SW620 had a decreased proliferation, while HCT116 had an increased viability (Fig. 10D). Furthermore, we detected the effect of SCAND1 on CRC apoptosis, and our study indicated that overexpression of SCAND1 significantly reduced the apoptosis rate of CRC cells, while knockdown of SCAND1 significantly increased the apoptosis rate of CRC cells (Fig. 10E). The wound healing assay showed a marked decrease in cell migration following SCAND1 knockdown (Fig. 10F) and an increase after its overexpression. Consistent with the results, transwell assays verified that SCAND1 knockdown inhibited SW620 invasion and migration, and its overexpression in HCT116 had the opposite trend (Fig. 10G). In addition, we examined the expression of key epithelial–mesenchymal transition (EMT) markers, including ZEB2, E-cadherin, N-cadherin, vimentin, α -SMA, and Twist1. We found that SCAND1 overexpression reduced E-cadherin levels while increasing the expression of ZEB2, N-cadherin, vimentin, α -SMA, and Twist1. Conversely, SCAND1 knockdown produced the opposite effects. Furthermore, to investigate the changes in immune cells within the tumor microenvironment, we performed co-culture experiments of HCT116 cells overexpressing SCAND1 with Jurkat T cells, and subsequently assessed the proliferation and apoptosis of HCT116 cells. The results demonstrated that compared with the control group, Jurkat T cells significantly suppressed CRC cell proliferation and promoted apoptosis. However, SCAND1 overexpression markedly attenuated these inhibitory and pro-apoptotic effects of T cells on CRC cells (Fig. 10I–J).

Discussion

CRC liver metastasis is a complex, multifactorial process. In this study, we leveraged large-scale integration of public datasets to revisit the biological characteristics of CRC, constructing a comprehensive, multidimensional single-cell and spatial transcriptomic atlas of colorectal cancer (CRC) and liver metastases (LM). Our dataset integrates over 200,000 high-quality cells from normal tissue, primary tumours, and metastatic lesions. Our findings reveal profound transcriptional reprogramming and cellular heterogeneity across disease stages, highlighting diverse epithelial and immune remodelling events that regulate tumour progression and metastatic adaptation.

A key observation is the dynamic remodelling of epithelial subpopulations. We found significant expansion of inflammatory enterocytes and proliferative transit-amplifying cells in CRC and LM, with concomitant loss of mature absorptive and secretory epithelial subtypes. These transitions reflect increased lineage plasticity and the acquisition of malignant traits. Notably, high-CNV epithelial cells exhibited increased activation of the EMT, oxidative phosphorylation, and interferon response pathways, suggesting a link between genomic instability and inflammatory remodelling in the metastatic microenvironment [96]. Importantly, CEBPB has emerged as a central transcriptional regulator of the inflammatory epithelial program, with progressive upregulation from CRC to liver metastasis. Multiple studies have confirmed its oncogenic roles in CRC, including promoting EMT via the TRIM2/p53 axis [38], enhancing chemoresistance through the CCL20/FOXO1/CEBPB/NF- κ B pathway [53], and driving STAT3-mediated tumour progression by upregulating SERPINA1 [97], underscoring its multifaceted effects on CRC malignancy and metastatic potential. Conversely, the nuclear coactivator and RNA processing factor ARGLU1 was downregulated in metastatic epithelial cells. Although rarely studied in CRC, ARGLU1 loss has been associated with chromatin dysregulation and therapy resistance in head and neck squamous cell carcinoma and breast cancer [98–100], suggesting its involvement in transcriptional instability during metastasis. Similarly, the mitochondrial-derived peptide MTNR2L8—known for its antiapoptotic function—was suppressed in liver metastases, potentially reflecting impaired mitochondrial stress responses and increased epithelial vulnerability in the metastatic microenvironment. Consistently, multiplex immunofluorescence (mIHC) validated the spatial and disease-stage-specific expression patterns of CEBPB, ARGLU1, S100A4, and FOSL1 across normal epithelium, CRC, and LM, providing orthogonal support for the inferred TF and effector dynamics.

Through high-dimensional gene coexpression network analysis, we mapped epithelial modules associated with either homeostatic maintenance or metastatic adaptation. Modules enriched in normal tissue featured ion transport and vesicle trafficking, whereas LM-specific modules promoted Wnt signalling, ribosome biogenesis, and cell cycle progression. These shifts were

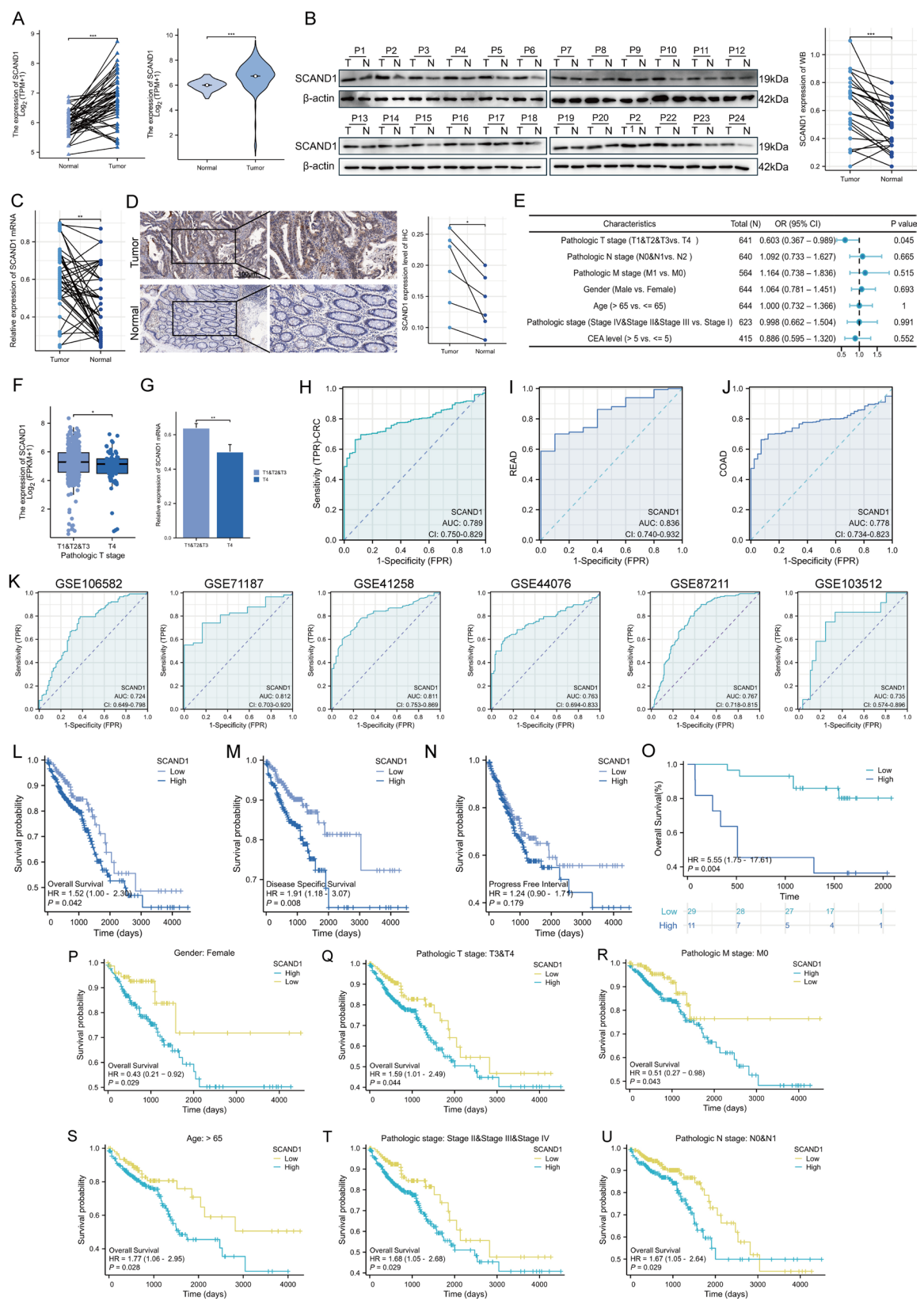


Fig. 9 (See legend on next page.)

(See figure on previous page.)

Fig. 9 Expression Pattern and Clinical Significance of SCAND1. **(A)** Expression differences of SCAND1 in paired and unpaired samples from the TCGA colorectal cancer dataset. **(B)** WB results in 24 pairs of matched samples. **(C)** The mRNA expression of SCAND1 in the enrolled forty patients. **(D)** IHC results in six pairs of matched samples. **(E)** Logistic regression analysis of SCAND1. **(F)** Box plots evaluating SCAND1 expression according to different T stages. **(G)** Validation cohort of 40 patients with varying T stages from Yangpu District Central Hospital. **(H–J)** Diagnostic ROC curves of SCAND1 in TCGA datasets of colorectal **(H)**, colon **(I)**, and rectal cancer **(J)**. **(K)** Diagnostic ROC curves of SCAND1 across multiple GSE datasets. **(L–N)** The effect of SCAND1 expression on the OS **(L)**, DSS **(M)**, and PFI **(N)** was analyzed by Kaplan–Meier method in TCGA database. **(O)** Survival rate in patients with high versus low SCAND1 in 40 patients from Yangpu District Central Hospital. **(P–U)** Survival curves of high and low SCAND1 expression groups across different gender **(P)**, T stage **(Q)**, M stage **(R)**, Age **(S)**, pathological stage **(T)**, N stage **(U)**. IHC: Immunohistochemistry, OS: overall survival; DSS: disease-specific survival, PFI: progression free interval. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ compared to the corresponding groups

accompanied by TF network remodelling, with activation of E2F, FOSL1, and AP-1 family members in high-CNV and metastatic cells, supporting a model of transcriptional plasticity underlying immune evasion and invasion. Trajectory and pseudotime analyses further revealed branched differentiation paths of epithelial cells, with distinct molecular features driving specific fates. Stress-responsive TFs (e.g., ATF3) and prometastatic factors (such as ANGPTL4 and RACK1) were enriched along metastatic trajectories, indicating early initiation of invasive programs. Spatial transcriptomics confirmed the localization of high-risk epithelial subpopulations (e.g., inflammatory enterocytes and TA cells) within tumour cores, which participate in intense signalling interactions via the MIF, VEGF, and TGF β axes, thereby reshaping the immune microenvironment.

Importantly, to validate our bioinformatic findings, we selected SCAND1—a gene identified as a key player in our integrated analysis—for further functional investigation. We confirmed elevated expression of SCAND1 in tumour tissues at both the mRNA and protein levels, and demonstrated its robust diagnostic potential across multiple independent datasets, with SCAND1 expression significantly correlated with poor prognosis. Notably, SCAND1 was differentially expressed across tumour stages and also exhibited differences in survival predictions between different subgroups, suggesting a dynamic role during CRC progression, which needs to be further validated in a larger cohort. Functional assays revealed that SCAND1 promotes colorectal cancer cell proliferation, inhibits apoptosis, and enhances migration and invasion. Mechanistically, SCAND1 upregulation was associated with increased epithelial-mesenchymal transition (EMT), further supporting its pro-metastatic function. Extending these observations to tumour-immune interactions, co-culture of SCAND1-overexpressing tumour cells with T cells demonstrated that SCAND1 overexpression attenuates T cell-mediated cytotoxicity, indicating a potential immune-evasive mechanism linked to SCAND1 activity. These results not only validate the predictions from our multi-omics analysis, but also highlight SCAND1 as a potential biomarker and therapeutic target in CRC metastasis.

A recent study profiling the “normal-inflammatory polyp-adenoma-carcinoma” continuum reported key

features of colorectal tumorigenesis, including cellular heterogeneity, the utility of ligand–receptor axes, marked Treg expansion during progression, stage-dependent elevation of heat-shock programs in immune cells, and prominent EGF/EGFR-mediated epithelial–stromal crosstalk [101]. In contrast, our work emphasizes the extended “primary–liver metastasis” dimension, with a larger dataset and complementary spatial transcriptomics. We resolve CNV-driven epithelial state stratification and TF-network rewiring (e.g., CEBPB, E2F, AP-1) linked to metastatic plasticity, define LM-specific coexpression modules (Wnt, ribosome biogenesis, cell cycle) with tumor-core spatial enrichment, and functionally validate SCAND1 as a pro-tumorigenic factor. The two studies are temporally complementary: the recent study focuses on early-to-invasive tumorigenesis, whereas our study delineates state remodeling across primary tumors and distant metastases in space, together outlining a single-cell and spatial continuum from initiation to dissemination.

In summary, our study provides a spatially resolved single-cell framework for understanding CRC progression and metastasis. We demonstrate that transcriptional, genomic, and signalling alterations collectively reshape epithelial and immune compartments, enabling malignant cells to adapt to and remodel their microenvironment. These insights may inform therapeutic strategies targeting key epithelial states, immune hubs, and intercellular signalling to inhibit CRC metastasis. Nevertheless, our study has several limitations. Our findings require further validation in larger clinical cohorts. Additionally, our work is primarily based on bioinformatic analyses and lacks functional experimental confirmation. In future studies, we plan to investigate the functional characteristics and mechanisms of epithelial subtypes in CRC progression through in vitro and in vivo experiments and assess their potential as therapeutic targets. We will also explore the relationship between these subpopulations and responses to immunotherapy, aiming to provide new avenues for improving immunotherapeutic efficacy.

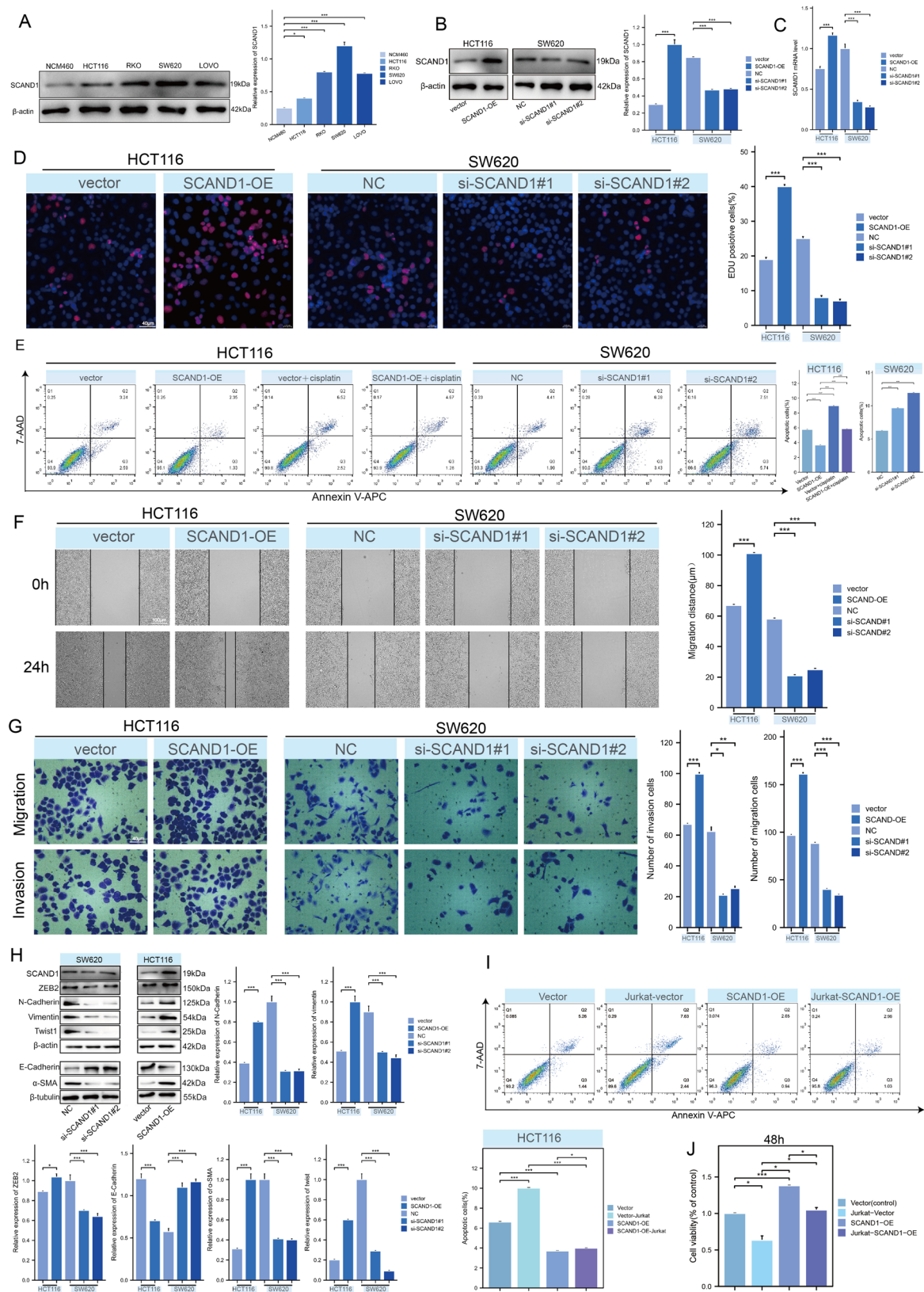


Fig. 10 (See legend on next page.)

(See figure on previous page.)

Fig. 10 SCAND1 promotes colorectal cancer cell proliferation, anti-apoptotic, migration, invasion, and EMT. **(A)** SCAND1 protein expression in normal and various colorectal cancer cell lines. **(B–C)** Protein **(B)** and mRNA **(C)** expression levels of SCAND1 were detected by western blotting and RT-qPCR. **(D)** EDU assay analysis. **(E)** Apoptosis rate was detected by flow cytometry. **(F–G)** Wound healing, migration and invasion assay analysis. **(H)** Western blotting was used to detect the expression levels of ZEB2, E-cadherin, N-cadherin, vimentin, α -SMA, and Twist1 in HCT116 and SW620 cells. **(I–J)** Apoptosis and growth assessment of HCT116 after co-culture with Jurkat cells. EMT: Epithelial–mesenchymal transition. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, compared to the corresponding groups

Abbreviation

OS	Overall survival
GSEA	Gene set enrichment analysis
DEGs	Differentially expressed genes
RT-qPCR	Quantitative real-time polymerase chain reaction
IHC	Immunohistochemistry
WB	Western blotting
ROC	Receiver operating characteristic curves
hdWGCNA	High dimensional weighted gene co-expression network analysis
CRC	Colorectal cancer
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
LMs	Liver metastases
CMS	Consensus Molecular Subtypes
EMT	Epithelial–mesenchymal transition
scRNA-seq	Single-cell RNA sequencing
Mye	Myeloid cells
Epi	Epithelial cells
CNV	Copy number variation
TAs	Transit-amplifying cells
Inflam.Enterocyte	Inflammatory enterocytes

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, as revised in 2013. The study protocol was approved by the Ethics Committee of Yangpu Hospital (Approval No. LL-2023-LW-012), and written informed consent was obtained from all participants.

Consent for publication

All authors gave their consent for publication.

Competing interests

The authors have declared that no competing interest exists.

Author details

¹Department of General Surgery, Yangpu Hospital Affiliated to Tongji University, Tongji University School of Medicine, Shanghai, China

²Department of Pharmacy, Yangpu Hospital, School of Medicine, Tongji University, Shanghai 200090, China

Received: 19 July 2025 / Accepted: 26 October 2025

Published online: 24 November 2025

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-07380-8>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7

Acknowledgements

Not applicable.

Author contributions

XCX was responsible for research design and conceptualization. XCX and PY collected the data. XCX performed the computational analyses and drafted the manuscript. PY revised the manuscript. XCX and ZSQ were consulted for molecular pathology of rectal cancer. YZC and LMB supervised and supported this research. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Funding

This work was funded by Shanghai Clinical Pharmacy Key Specialty Construction Project Support project (Shanghai Health and Pharmaceutical Administration) (201809).

Data availability

All data could be downloaded at (<https://www.cancer.gov>; <http://www.ncbi.nlm.nih.gov/geo>; <https://www.genecards.org>) and (TCGA, <https://portal.gdc.cancer.gov/>)

References

- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clinicians*. 2021;71(3):209–49.
- Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. *CA: Cancer J Clinicians*. 2023;73(1):17–48.
- Miller DK, Nogueira L, Devasia T, et al. Cancer treatment and survivorship statistics, 2022. *CA: Cancer J Clinicians*. 2022;72(5):409–36.
- Eng C, Yoshino T, Ruiz-García E, et al. Colorectal cancer. *Lancet (Lond, Engl)*. 2024;404(10449):294–310.
- Dienstmann R, Vermeulen L, Guinney J, et al. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer*. 2017;17(2):79–92.
- Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330–37.
- Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer [J]. *Nat Med*. 2015;21(11):1350–56.
- Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol*. 2017;14(10):611–29.
- Lambert AW, Weinberg RA. Linking emt programmes to normal and neoplastic epithelial stem cells. *Nat Rev Cancer*. 2021;21(5):325–38.
- Wang H, Mei Y, Luo C, et al. Single-cell analyses reveal mechanisms of cancer stem cell maintenance and epithelial–mesenchymal transition in recurrent bladder cancer. *Clin Cancer Res: Off J Am Assoc Cancer Res*. 2021;27(22):6265–78.
- Ardeljan D, Steranka PJ, Liu C, et al. Cell fitness screens reveal a conflict between LINE-1 retrotransposition and dna replication. *Nat Struct Mol Biol*. 2020;27(2):168–78.
- Tadepally HD, Burger G, Aubry M. Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. *BMC Evol Biol*. 2008;8:176.
- Edelstein LC, Collins T. The scan domain family of zinc finger transcription factors. *Gene*. 2005;359:1–17.
- Sander TL, Stringer KF, Maki JL, et al. The scan domain defines a large family of zinc finger transcription factors. *Gene*. 2003;310:29–38.
- Schumacher C, Wang H, Honer C, et al. The scan domain mediates selective oligomerization. *J Biol Chem*. 2000;275(22):17173–79.

16. Eguchi T, Prince LT, Tran MT, et al. MZF1 and SCAND1 reciprocally regulate CDC37 gene expression in prostate cancer. *Cancers (Basel)*. 2019;11(6).
17. Eguchi T, Cszimadia E, Kawai H, et al. SCAND1 reverses epithelial-to-mesenchymal transition (EMT) and suppresses prostate cancer growth and migration. *Cells*. 2022;11(24).
18. Sheta M, Yoshida K, Kanemoto H, et al. Stress-inducible scand factors suppress the stress response and are biomarkers for enhanced prognosis in cancers. *Int J Mol Sci*. 2023;24(6).
19. Mei Y, Xiao W, Hu H, et al. Single-cell analyses reveal suppressive tumor micro-environment of human colorectal cancer. *Clin Transl Med*. 2021;11(6):e422.
20. Qian J, Olbrecht S, Boeckx B, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res*. 2020;30(9):745–62.
21. Wang R, Li J, Zhou X, et al. Single-cell genomic and transcriptomic landscapes of primary and metastatic colorectal cancer tumors. *Genome Med*. 2022;14(1):93.
22. Zhang L, Li Z, Skrzypczynska MK, et al. Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell*. 2020;181(2):442–59.e29.
23. Wang Y, Qiu X, Li Q, et al. Single-cell and spatial-resolved profiling reveals cancer-associated fibroblast heterogeneity in colorectal cancer metabolic subtypes. *J Transl Med*. 2025;23(1):175.
24. Zhan Y, Sun D, Gao J, et al. Single-cell transcriptomics reveals intratumor heterogeneity and the potential roles of cancer stem cells and myCaf in colorectal cancer liver metastasis and recurrence. *Cancer Lett*. 2025;612:217452.
25. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. 2021;12(1):1088.
26. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017;14(10):979–82.
27. Lee H-O, Hong Y, Etioglu HE, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet*. 2020;52(6):594–603.
28. Che L-H, Liu J-W, Huo J-P, et al. A single-cell atlas of liver metastases of colorectal cancer reveals reprogramming of the tumor microenvironment in response to preoperative chemotherapy. *Cell Discov*. 2021;7(1):80.
29. M KA, Erdogan C, Kurt Z, et al. Refining colorectal cancer classification and clinical stratification through a single-cell atlas. *Genome Biol*. 2022;23(1):113.
30. Wu Y, Yang S, Ma J, et al. Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discov*. 2022;12(1):134–53.
31. Chu X, Li X, Zhang Y, et al. Integrative single-cell analysis of human colorectal cancer reveals patient stratification with distinct immune evasion mechanisms. *Nat Cancer*. 2024;5(9):1409–26.
32. Becker WR, Nevins SA, Chen DC, et al. Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *Nat Genet*. 2022;54(7):985–95.
33. Jacobs C, Shah S, Lu W-C, et al. HSF1 inhibits antitumor immune activity in breast cancer by suppressing CCL5 to block CD8+ T-cell recruitment. *Cancer Res*. 2024;84(2):276–90.
34. Eroglu B, Pang J, Jin X, et al. HSF1-mediated control of cellular energy metabolism and mTORC1 activation drive acute T-Cell Lymphoblastic Leukemia progression. *Mol Cancer Res: MCR*. 2020;18(3):463–76.
35. Squair JW, Skinnider MA, Gautier M, et al. Prioritization of cell types responsive to biological perturbations in single-cell data with Augur. *Nat Protoc*. 2021;16(8):3836–73.
36. Patel AP, Tirosh I, Rombetta JJ, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Sci (New York, NY)*. 2014;344(6190):1396–401.
37. Zhang S, Fang W, Zhou S, et al. Single cell transcriptomic analyses implicate an immunosuppressive tumor microenvironment in pancreatic cancer liver metastasis [J]. *Nat Commun*. 2023;14(1):5123.
38. Zhou Z, Shu Y, Bao H, et al. Stress-induced epinephrine promotes epithelial-to-mesenchymal transition and stemness of crc through the CEBPB/TRIM2/P53 axis. *J Transl Med*. 2022;20(1):262.
39. Magomedova L, Tiefenbach J, Zilberman E, et al. ARGLU1 is a transcriptional coactivator and splicing regulator important for stress hormone signaling and development. *Nucleic Acids Res*. 2019;47(6):2856–70.
40. Morabito S, Miyoshi E, Michael N, et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet*. 2021;53(8):1143–55.
41. Pattison MA, Merlino DJ, Blomain ES, et al. Guanylyl cyclase C signaling axis and colon cancer prevention. *World J Gastroenterol*. 2016;22(36):8070–77.
42. Parikh K, Antanaviciute A, Fawcner-Corbett D, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature*. 2019;567(7746):49–55.
43. Braga Emidio N, Brierley SM, Schroeder CI, et al. Structure, function, and therapeutic potential of the trefoil factor family in the gastrointestinal tract. *ACS Pharmacol Transl Sci*. 2020;3(4):583–97.
44. Bijelić N, Belovari T, Tolušić Levak M, et al. Localization of trefoil factor family peptide 3 (TFF3) in epithelial tissues originating from the three germ layers of developing mouse embryo. *Bosn J Basic Med Sci*. 2017;17(3):241–47.
45. Wang X, Yu Q, Ghareeb W M, et al. Downregulated SPINK4 is associated with poor survival in colorectal cancer. *BMC Cancer*. 2019;19(1):1258.
46. Ikpa PT, Sleddens HF, Steinbrecher KA, et al. Guanylin and uroguanylin are produced by mouse intestinal epithelial cells of columnar and secretory lineage. *Histochem Cell Biol*. 2016;146(4):445–55.
47. Lu C, Chen S, Liu S, et al. Determination of cell free BIRC5 mRNA as a serum diagnostic and prognostic biomarker for colorectal cancer. *J Surg Oncol*. 2014;109(6):574–79.
48. Lin X, Chen J-D, Wang C-Y, et al. Cooperation of MLL1 and Jun in controlling H3K4me3 on enhancers in colorectal cancer. *Genome Biol*. 2023;24(1):268.
49. Ye J, Liu S, Shang Y, et al. R-spondin1/Wnt-enhanced Ascl2 autoregulation controls the self-renewal of colorectal cancer progenitor cells. *Cell Cycle (Georget, Tex)*. 2018;17(8):1014–25.
50. Bell I, Khan H, Stutt N, et al. Nkd1 functions downstream of Axin2 to attenuate Wnt signaling. *Mol Biol Cell*. 2024;35(7):ar93.
51. Wang D, Yang L, Yu W, et al. Colorectal cancer cell-derived CCL20 recruits regulatory T cells to promote chemoresistance via FOXO1/CEBPB/NF-κB signaling. *J Immunother Of Cancer*. 2019;7(1):215.
52. Mzoughi S, Schwarz M, Wang X, et al. Oncofetal reprogramming drives phenotypic plasticity in WNT-dependent colorectal cancer. *Nat Genet*. 2025;57(2):402–12.
53. S BK, Luo Q, Murmu N, et al. Reg iv activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas. *Gastroenterology*. 2006;130(1):137–49.
54. Hui Y-J, Chen H, Peng X-C, et al. Up-regulation of ABCG2 by MYBL2 deletion drives Chlorin e6-mediated photodynamic therapy resistance in colorectal cancer. *Photodiagn Photodyn Ther*. 2023;42:103558.
55. Liu Q, Guo L, Qi H, et al. A MYBL2 complex for RRM2 transactivation and the synthetic effect of MYBL2 knockdown with WEE1 inhibition against colorectal cancer. *Cell Death Dis*. 2021;12(7):683.
56. Fang Z, Gong C, Yu S, et al. NFYB-induced high expression of E2F1 contributes to oxaliplatin resistance in colorectal cancer via the enhancement of CHK1 signaling. *Cancer Lett*. 2018;415:58–72.
57. Jing Z, Liu Q, He X, et al. NCAPD3 enhances Warburg effect through c-myc and E2F1 and promotes the occurrence and progression of colorectal cancer. *J Educ Chang Exp Clin Cancer Res: CR*. 2022;41(1):198.
58. Zhang M, Hoyle RG, Ma Z, et al. FOSL1 promotes metastasis of head and neck squamous cell carcinoma through super-enhancer-driven transcription program. *Mol Ther: J Am Soc Gene Ther*. 2021;29(8):2583–600.
59. Yang Y, Jin X, Xie Y, et al. The CEBPB+ glioblastoma subcluster specifically drives the formation of M2 tumor-associated macrophages to promote malignancy growth. *Theranostics*. 2024;14(10):4107–26.
60. Sobolev VV, Khashukoeva AZ, Evina OE, et al. Role of the transcription factor FOSL1 in organ development and tumorigenesis. *Int J Mol Sci*. 2022;23(3):1521.
61. Ku H-C, Cheng C-F. Master regulator activating transcription factor 3 (ATF3) in metabolic homeostasis and cancer. *Front Endocrinol*. 2020;11:556.
62. Wen Y, Liu Y, Li Q, et al. Spatiotemporal ATF3 expression determines VSMC fate in abdominal aortic aneurysm. *Circ Res*. 2024;134(11):1495–511.
63. Fan Z, Turiel G, Ardicoglu R, et al. Exercise-induced angiogenesis is dependent on metabolically primed ATF3/4+ endothelial cells. *Cell Metab*. 2021;33(9):1793–807.e9.
64. Feng L, Zhang H, Liu T. Multifaceted roles of IKZF1 gene, perspectives from bench to bedside. *Front Oncol*. 2024;14:1383419.
65. Yan M, Wang J, Wang H, et al. Knockdown of NR3C1 inhibits the proliferation and migration of clear cell renal cell carcinoma through activating endoplasmic reticulum stress-mitophagy. *J Transl Med*. 2023;21(1):701.
66. Johnson DE, Okeefe AR, Grandis JR. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nat Rev Clin Oncol*. 2018;15(4):234–48.

69. Wang L, Wei Y, Jin Z, et al. IFN- α /IFN- γ /IL-15 pathways identify GBP1-expressing tumors with an immune-responsive phenotype. *Clin Exp Med*. 2024;24(1):102.
70. Tan J, Fan W, Liu T, et al. TREM2+ macrophages suppress CD8+ T-cell infiltration after transarterial chemoembolisation in hepatocellular carcinoma. *J Hepatol*. 2023;79(1):126–40.
71. Cibrián D, Sánchez-Madrid F. CD69: from activation marker to metabolic gate-keeper. *Eur J Immunol*. 2017;47(6):946–53.
72. Rock KL, Reits E, Neefjes J. Present Yourself By MHC Class I and MHC Class II molecules. *Trends In Immunol*. 2016;37(11):724–37.
73. Ishina IA, Zakharova MY, Kurbatskaia IN, et al. MHC class II presentation in autoimmunity. *Cells*. 2023;12(2):314.
74. Dannenmann RS, Thielicke J, Stöckli M, et al. Tumor-associated macrophages subvert T-cell function and correlate with reduced survival in clear cell renal cell carcinoma. *Oncoimmunology*. 2013;2(3):e23562.
75. Ferrara N. VEGF and the quest for tumour angiogenesis factors. *Nat Rev Cancer*. 2002;2(10):795–803.
76. Wu X, Sun M, Yang Z, et al. The roles of CCR9/CCL25 in inflammation and inflammation-associated diseases. *Front Cell Dev Biol*. 2021;9:686548.
77. Sumaiya K, Langford D, Natarajaseenivasan K, et al. Macrophage migration inhibitory factor (MIF): a multifaceted cytokine regulated by genetic and physiological strategies. *Pharmacol Ther*. 2022;233:108024.
78. Morse C, Tabib T, Sembrat J, et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J*. 2019;54(2):1802441.
79. Zhao Y, Huang Z, Gao L, et al. Osteopontin/SPP1: a potential mediator between immune cells and vascular calcification. *Front Immunol*. 2024;15:1395596.
80. Bouffette S, Botez I, De Ceuninck F. Targeting galectin-3 in inflammatory and fibrotic diseases. *Trends Pharmacological Sci*. 2023;44(8):519–31.
81. De Jong MME, Fokkema C, Papazian N, et al. An IL-1 β -driven neutrophil-stromal cell axis fosters a BAFF-rich protumor microenvironment in individuals with multiple myeloma. *Nat Immunol*. 2024;25(5):820–33.
82. Vijayan A, Guha D, Ameer F, et al. IGFBP-5 enhances epithelial cell adhesion and protects epithelial cells from TGF β 1-induced mesenchymal invasion. *Int J Biochem Cell Biol*. 2013;45(12):2774–85.
83. Nasri I, Bonnet D, Zwarycz B, et al. PAR2-dependent activation of GSK3 β regulates the survival of colon stem/progenitor cells. *Am J Physiol Gastrointest Liver Physiol*. 2016;311(2):G221–36.
84. Lv Y, Ma X, Ma Y, et al. A new emerging target in cancer immunotherapy: Galectin-9 (LGALS9). *Genes Dis*. 2023;10(6):2366–82.
85. Yan G, Song R, Zhang J, et al. MIF promotes Th17 cell differentiation in rheumatoid arthritis through ATF6 signal pathway. *Mol Med Rep*. 2024;30(1):237.
86. Liu Z, Li Z, Yan G, et al. MIF promotes Th17 cell differentiation in Hashimoto's thyroiditis by binding HVEM and activating NF- κ B signaling pathway. *Int Immunopharmacol*. 2023;121:110494.
87. Li F, Wang J, QYY, et al. CD147 promotes breast cancer migration and invasion by inducing epithelial-mesenchymal transition via the MAPK/ERK signaling pathway. *BMC Cancer*. 2023;23(1):1214.
88. Asgari R, Vaisi-Raygani A, Aleagha MSE, et al. CD147 and MMPs as key factors in physiological and pathological processes. *Biomed Pharmacother*. 2023;157:113983.
89. Neiburger EJ. Naiveté about computers [J]. *J Retailing Am Dent Assoc* (1939), 1979;98(2):175.
90. Pope JL, Bhat AA, Sharma A, et al. Claudin-1 regulates intestinal epithelial homeostasis through the modulation of Notch-signalling. *Gut*. 2014;63(4):622–34.
91. Jiang Y, Song J, Xu Y, et al. Piezo1 regulates intestinal epithelial function by affecting the tight junction protein claudin-1 via the rock pathway. *Life Sci*. 2021;275:119254.
92. Yang E, Ding C, Zhu X, et al. Sfrp2 regulates the WNT/ β -catenin pathway to slow the development of aldosterone-producing adenoma. *Cardiovasc Diagn Ther*. 2023;13(3):523–33.
93. Lan R, Yu Y, Song J, et al. SFRP2 suppresses trophoblast cell migration by inhibiting the Wnt/ β -catenin pathway. *Mol Med Rep*. 2024;29(4):66.
94. Tang Q, Chen J, Di Z, et al. TM4SF1 promotes emt and cancer stemness via the Wnt/ β -catenin/SOX2 pathway in colorectal cancer. *J Educ Chang Exp Clin Cancer Res: CR*. 2020;39(1):232.
95. Murata K, Jadhav U, Madha S, et al. Ascl2-dependent cell dedifferentiation drives regeneration of ablated intestinal stem cells. *Cell STEM Cell*. 2020;26(3):377–90.e6.
96. Talens F, Van Vugt MATM. Inflammatory signaling in genomically unstable cancers. *Cell Cycle (Georget, Tex)*. 2019;18(16):1830–48.
97. Ma Y, Chen Y, Zhan L, et al. CEBPB-mediated upregulation of SERPINA1 promotes colorectal cancer progression by enhancing STAT3 signaling. *Cell Death Discov*. 2024;10(1):219.
98. Bachus S, Akkerman N, Fulham L, et al. ARGLU1 enhances promoter-proximal pausing of rna polymerase ii and stimulates dna damage repair. *Nucleic Acids Res*. 2024;52(10):5658–75.
99. Shi J, Yu X, Li G, et al. DTL promotes head and neck squamous cell carcinoma progression by mediating the degradation of ARGLU1 to regulate the notch signaling pathway. *Int J Biol Macromolecules*. 2024;259(Pt 2):129184.
100. Zhang D, Jiang P, Xu Q, et al. Arginine and glutamate-rich 1 (ARGLU1) interacts with mediator subunit 1 (MED1) and is required for estrogen receptor-mediated gene transcription and breast cancer cell growth. *J Biol Chem*. 2011;286(20):17746–54.
101. Sheng J, Li N, Li S, et al. A dynamic molecular landscape in colorectal cancer progression at single-cell resolution. *J Transl Med*. 2025;23(1):723.

Publisher's Note

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