Journal of Translational Medicine

RESEARCH

Open Access

Deciphering immune and cellular reprogramming during the progression from inflammatory bowel disease to colorectal cancer using multi-omics single-cell and spatial transcriptomics

Seyed Taleb Hosseini[1,2], Kimia AminianToosi[3], Farkhondeh Nemati[2], Roya BishehKolaei[2*] and Youping Deng[4*]

Abstract

Background Inflammatory bowel disease (IBD) represents a group of chronic and relapsing inflammatory disorders of the gastrointestinal tract, primarily encompassing Crohn's disease (CD) and ulcerative colitis (UC). To elucidate immune heterogeneity and molecular mechanisms underlying IBD pathogenesis, we performed an integrated high-resolution transcriptomic and experimental analysis across 1436 patients with IBD and CRC.

Methods 436 samples in transcriptomics level were analyzed, including 415 bulk RNA-seq, 18 single-cell RNA-seq (10x Genomics), and 3 spatial transcriptomic (Visium) samples from human intestinal tissue. Additionally, serum carcinoembryonic antigen (CEA) levels were evaluated in 1000 patients diagnosed with IBD (CD or UC) and colorectal cancer (CRC).

Results Single-cell analysis identified major immune and stromal cell types, among which epithelial cells, T cells, B cells, and tissue stem cells were the most abundant in both CD and UC tissues. Comparative profiling revealed an increase in epithelial and stem cell populations in diseased tissues, indicating enhanced epithelial regeneration and immune activation. Integration of bulk and single-cell RNA-seq datasets highlighted several disease-associated genes, including: CEACAM5, LGALS1, NUAK1, and PDGFRA, with CEACAM5 showing consistent upregulation across CD and UC samples. Pseudotime trajectory analysis demonstrated that CEACAM5 expression increased during the later stages of epithelial cell differentiation and suggesting its involvement in mucosal remodeling and chronic inflammation. Spatial transcriptomic mapping confirmed localized CEACAM5 overexpression in epithelial regions of colorectal cancer tissues, further supporting its role in disease progression. Serological analysis revealed that serum CEA levels were significantly higher in CRC compared to IBD and within IBD, Crohn's disease patients exhibited higher CEA levels than ulcerative colitis patients ($P < 0.05$).

*Correspondence:
Roya BishehKolaei
rbishe@gmail.com; roya.bishehkolaei@iau.ac.ir
Youping Deng
dengy@hawaii.edu

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

Conclusions Collectively, these findings indicate that CEACAM5 (CEA) serves as a key molecular marker linking epithelial activation in IBD to tumorigenic processes in colorectal cancer and providing potential diagnostic and prognostic value for distinguishing inflammatory bowel disease and colorectal cancer.

Keywords Colorectal cancer, Inflammatory bowel disease, Crohn's disease, Ulcerative colitis, Single-cell, Spatial transcriptomics, Multi-omics

Background

Crohn's disease (CD) and ulcerative colitis (UC) are the two main inflammatory gastrointestinal conditions that are included in the group of chronic and known as inflammatory bowel disease (IBD) [1]. Approximately ten million individuals worldwide suffer from IBD and the prevalence becomes higher mainly in historically high-prevalence countries such as North America and Western Europe as well as among industrialized nations in Asia, Africa and South America [2]. Although both sexes are susceptible to IBD, CD is significantly more prevalent in women and UC is significantly more common in men [2]. Particularly, between 20 and 25% of patients are identified during childhood and demonstrating the increasingly high incidence of children with IBD [3] and also IBD continues to appear in adults around the age range of 20 and 40 [3]. There are major distinctions between Crohn's disease and ulcerative colitis with regard to of their medical attitudes, histological characteristics and spatial spread [3]. Although CD may affect any part of the gastrointestinal tract, it most commonly affects the ascending colon, cecum and terminal ileum [4]. UC is limited to the colon and rectum and usually starting there and moving progressively [5]. Mainly the mucosal surface of the gastrointestinal wall has become inflammatory and the most frequently affected regions are the rectum, descending colon and sigmoid colon [5]. Progressive diarrhea (sometimes blood in UC), gut discomfort, decreased weight, exhaustion, pressure, tenesmus and extraintestinal indications including discomfort in the joints, wounds on the skin, and retinal inflammation are among the clinical signs of IBD [6]. IBD-related permanent inflammation considerably increases the risk of colorectal cancer (CRC), especially in patients with family history of the illnesses, severe colonic engagement and a widespread condition [6]. Consequently, observation colonoscopy is an essential part of ongoing treatment for people suffering from IBD [6]. Our knowledge of the pathophysiology of IBD has changed as a result of recent developments in molecular biology, especially in genomic methods [7]. By discovering certain inflammatory cell subsets, epithelial cell difficulties and stromal cell modifications that are related to inflammation, single-cell RNA seq (scRNA-seq) enables studies to analyze the intestinal mucosa's cellular composition [7]. The use of this approach is improved by spatial transcriptomics, which maintains the spatial context of gene expression inside the structure of tissues [8]. Furthermore, This method enables for the investigation of microenvironmental effects and interactions among cells across various intestinal structural regions, including the descending, transverse, and ascending colon [8]. Finally, these innovations provide strong instruments for discovering screening biomarkers, comprehending variation in patient reactions to therapy and unlocking opportunities to obtain personalized therapeutic approaches for IBD.

In this research, we conducted a high-resolution integrated investigation into immune cell populations utilizing 436 samples obtained from bulk RNA-Seq (415 samples), single-cell RNA-seq (10x Genomics, 18 samples) and spatial transcriptomic platform (3 samples, 10x visium). We first analyzed the FASTQ files of the 415 IBD samples including 65 colon healthy tissue and 25 ileum healthy tissue from healthy tissue samples and 87 colon CD tissue and 81 ileum CD tissue from Crohn's disease tissue samples and 118 colon UC tissue and 39 ileum UC tissue from ulcerative colitis tissue samples that were collected from SRA database. Consequently, we conduct an scRNA-seq assessment of 18 human IBD samples: 6 healthy (21,171 genes, 20,351 cells), 6 Crohn's disease (CD) (21,959 genes, 21,747 cells) and 6 ulcerative colitis (UC) (20,390 genes, 17,082 cells). Spatial transcriptomic analysis was performed using three publicly available colorectal cancer tissue samples from the 10x Genomics Visium platform to characterize the spatial architecture and tumor microenvironmental context. The IBD-associated cellular signatures previously defined by single-cell RNA-seq analysis were subsequently projected onto these spatial datasets to investigate how inflammation-related cell populations are spatially organized within colorectal tumor tissues. In order to complete these molecular evaluations, we conducted experimental research on a clinically recognizable cohort of 1,000 patients with pathologically diagnosed colorectal cancer and inflammatory bowel disease. This cohort included 735 IBD patients (320 Crohn's disease and 415 ulcerative colitis) and 265 patients with colorectal cancer in all clinical stages. Collectively, this integrative multi-omics and clinical framework enabled systematic dissection of immune and cellular heterogeneity across inflammatory and malignant tissues, facilitating the identification of biologically and clinically relevant molecular targets with potential diagnostic and therapeutic implications in IBD and colorectal cancer.

Methods

Collecting RNA-Seq data sets (bulk RNA-Seq, single cell RNA-Seq and spatial transcriptomic)

Figure 1 demonstrates the general framework of the data analysis and specific methods. With the goal to find studies on IBD and CRC gene expression profiles according to bulk and single cell RNA-Seq, we searched the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) [9], Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) [10] and PubMed databases (

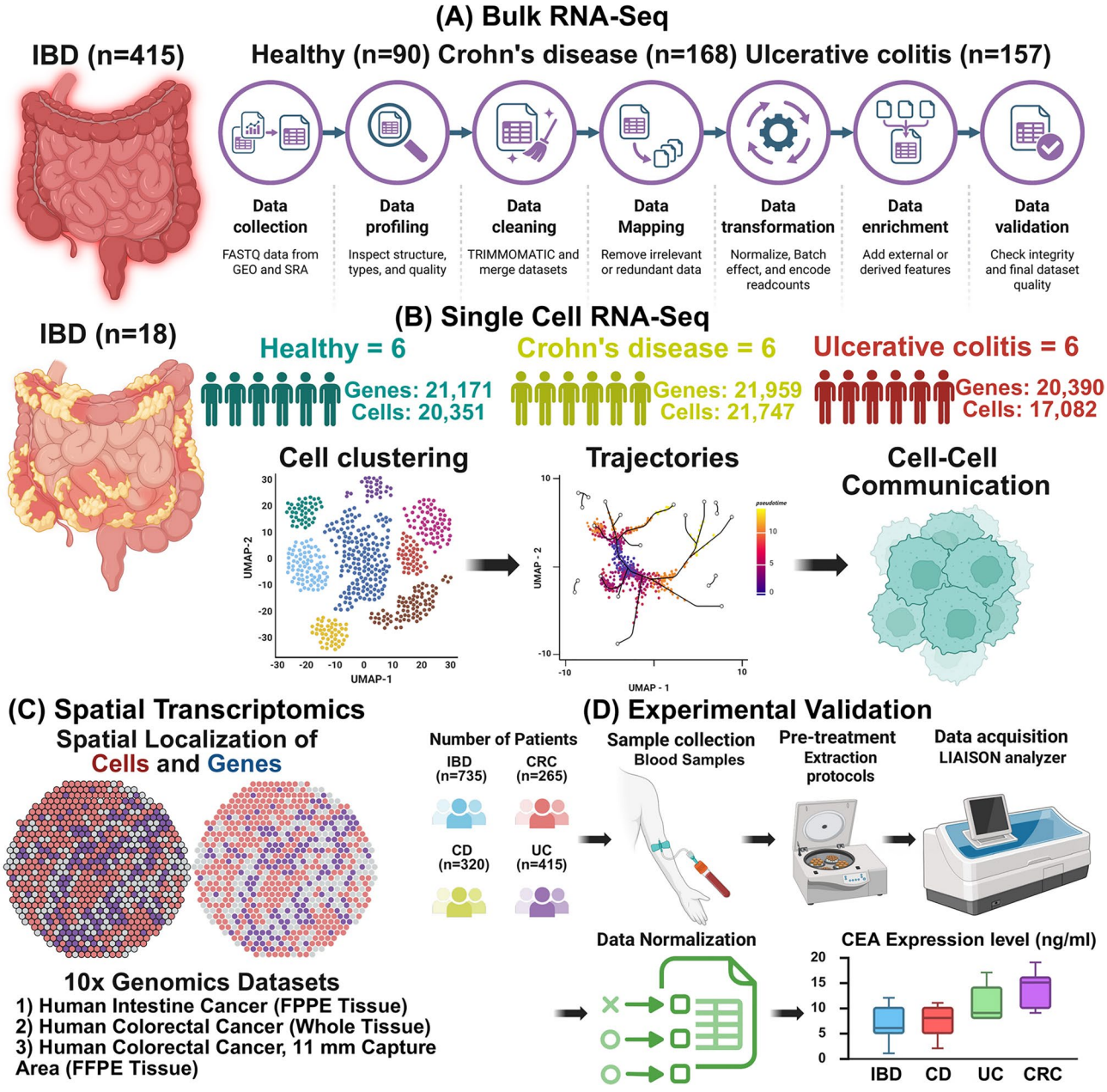

Fig. 1 Integrated workflow for bulk, single-cell, spatial transcriptomic profiling and experimental validation in inflammatory bowel disease, including Crohn's disease and ulcerative colitis and colorectal cancer patients. (**A**) Publicly available bulk RNA-seq datasets were analyzed to characterize inflammation-associated transcriptional genes across healthy intestinal tissues and inflammatory bowel disease, such as Crohn's disease and ulcerative colitis. (**B**) Single-cell RNA-seq evaluation was employed to resolve cellular and immune heterogeneity in human intestinal tissues and enabling high-resolution identification of disease-associated immune states and cell type specific transcriptional alterations. (**C**) Spatial transcriptomics of human colorectal cancer tissues was performed using Visium HD platforms to delineate tissue architecture, immune cell localization, and tumor microenvironmental heterogeneity within preserved tissue sections. (**D**) Key computational findings were evaluated through experimental and clinical analyses in a large, well-characterized patient cohort and supporting the biological relevance of the identified immune and molecular signatures across blood samples from patients with inflammatory bowel disease (IBD) and colorectal cancer (CRC)

https://pubmed.ncbi.nlm.nih.gov/). "IBD", "Crohn's disease", "Ulcerative colitis", "Bulk RNASeq", "Single cell RNASeq", "Gene expression profiling", "Spatial heterogeneity", "Visium gene expression", "Colorectal cancer" and "Spatial transcriptomics analysis" were included in the search parameters. We used the 10x Genomics database (https://www.10xgenomics.com/) as the main reference for obtaining datasets related to spatial transcriptomics. Our database search was based on established standards specific to spatial transcriptomics, which includes utilizing visium spatial gene expression technologies or comparable spatial profiling approaches that are 10x appropriate. Search for criteria contained terms including "Spatial transcriptomics", "Visium" and "Spatial gene expression" in combination with parameters for clinical metadata (such as tissue type, organisms and sequence depth) and resource format (such as raw FASTQ files, processed count matrix). The related files, which usually included raw sequencing reads, spatial barcode data and imaging data when appropriate, were downloaded in order to obtain suitable data sets. We evaluated comparability with the following analysis workflows by combining each dataset's metadata regarding our research goals in order to validate data accuracy and significance. Bulk, single cell RNA-seq and spatial transcriptomics datasets from experiments have been included in addition to data regarding the gene expression levels for every single cell type in tumor tissues from patients with colorectal cancer, healthy tissues and patients with IBD such as Crohn disease and ulcerative colitis. Single-cell and bulk RNA-seq datasets from research using models of animals, including rat and mouse investigations treating various types of cells with multiple types of drugs and antibodies and systematic review papers were not included.

Characterization of datasets: bulk, single cell and spatial transcriptomic

We obtained original expression RNA-seq datasets, including bulk RNA-seq, single cell RNA-seq and spatial transcriptomics. The collection includes 8 datasets encompassing 436 samples, with 63,520 genes and 59,180 cells from single-cell datasets and 14,878 spots from 10x Genomics spatial datasets, covering human species and spanning healthy controls, inflammatory bowel diseases (Crohn's disease and ulcerative colitis). The bulk RNA sequencing (RNA-seq) datasets from human tissue samples, consist of 415 samples from four GEO datasets. GSE165512 (SRP303290), sequenced using the Illumina HiSeq 2500 platform, comprises 168 samples: 35 healthy colon, 10 healthy ileum, 40 colon Crohn's disease (CD), 43 ileum Crohn's disease (CD), 40 colon ulcerative colitis (UC) and no ileum ulcerative colitis (UC) samples. PRJEB24645 (ERP106487), sequenced using the Illumina HiSeq 2500 platform, contains 79 samples: 16 healthy colon, 11 healthy ileum, 15 colon Crohn's disease (CD), 10 ileum Crohn's disease (CD), 16 colon ulcerative colitis (UC) and 11 ileum ulcerative colitis (UC) samples. GSE183620 (SRP336051), sequenced using the Illumina NovaSeq 6000 platform, includes 114 samples without healthy controls: 18 colon Crohn's disease (CD), 22 ileum Crohn's disease (CD), 46 colon ulcerative colitis (UC) and 28 ileum ulcerative colitis (UC) samples [11]. GSE191234 (SRP351557), sequenced using the Illumina HiSeq 2500 platform, consists of 54 samples: 14 healthy colon, 4 healthy ileum, 14 colon Crohn's disease (CD), 6 ileum Crohn's disease (CD), 16 colon ulcerative colitis (UC) and no ileum ulcerative colitis (UC) samples. These datasets collectively including 90 healthy, 168 Crohn's disease (CD) and 157 ulcerative colitis (UC) samples across colon and ileum tissues. GSE214695 [12], sequenced using the Illumina NextSeq 500 platform, includes 18 human IBD samples: 6 healthy (21,171 genes, 20,351 cells), 6 Crohn's disease (CD) (21,959 genes, 21,747 cells) and 6 ulcerative colitis (UC) (20,390 genes, 17,082 cells). These datasets offer cellular resolution for inflammatory bowel diseases in humans. 10x Genomics datasets for spatial transcriptomics analysis such as (1) Human Colorectal Cancer datasets (Whole Transcriptome) sequenced using the Illumina NovaSeq 6000 platform (https://www.10xgenomics.com/datasets/human-colorectal-cancer-whole-transcriptome-analysis-1-standard-1-2-0), 1 slide (V10A13-206) with 3,138 spots, 112,228 mean reads per cell, a median of 3,538 genes per spot and 8,906 UMI counts per spot, related to invasive adenocarcinoma; (2) Human Intestine Cancer (FPPE), sequenced using the Illumina NovaSeq 6000 platform (https://www.10xgenomics.com/datasets/human-intestine-cancer-1-standard), includes 1 slide (V10L13-021) with 2,660 spots, 77,288 reads per spot, a median of 7,438 genes per spot and 25,312 UMI counts per spot, associated with large intestine colorectal cancer; (3) Human Colorectal Cancer (11 mm Capture Area (FFPE)), sequenced using the Illumina NovaSeq 6000 platform (https://www.10xgenomics.com/datasets/human-colorectal-cancer-11-mm-capture-area-ffpe-2-standard), 1 slide (V52Y10-310) with 9,080 spots, 911,044,482 total reads at 46.2% saturation, a median of 9,560 genes per spot, and 49,282 UMI counts per spot, related to colorectal cancer. All information for datasets including bulk, single cell and spatial transcriptomics were reported in Table 1.

Sample processing and differential expression analysis in bulk-transcriptomics

Fastq files were converted from .sra format to paired files with fastq format utilizing the SRAToolkits (v3.0.0) (https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit) package [9]. The FASTQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was

Table 1 Basic characteristics of the gene expression profile data in inflammatory bowel disease and colorectal cancer

Datasets/ BioProject/ Species (GEO)	SRP Datasets	Platform	Healthy		Crohn's disease		Ulcerative colitis		Data Type
			Colon-Healthy	Ileum-Healthy	Colon-CD	Ileum-CD	Colon-UC	Ileum-UC	
GSE165512 (*N* = 168) (Human)	SRP303290	Illumina HiSeq 2500	35	10	40	43	40	0	Bulk RNASeq
PRJEB24645 (*N* = 79) (Human)	ERP106487	Illumina HiSeq 2500	16	11	15	10	16	11	Bulk RNASeq
GSE183620 (*N* = 114) (Human)	SRP336051	Illumina NovaSeq 6000	0	0	18	22	46	28	Bulk RNASeq
GSE191234 (*N* = 54) (Human)	SRP351557	Illumina HiSeq 2500	14	4	14	6	16	0	Bulk RNASeq
GSE214695 (*N* = 18) (Human)	-	Illumina NextSeq 500	6 (Total Genes: 21,171 - Total Cells: 20,351)		6 (Total Genes: 21,959 - Total Cells: 21,747)		6 (Total Genes: 20,390 - Total Cells: 17,082)		Single Cell RNASeq
10x Genomics Datasets (Spatial Transcriptomics)	**Sequencing instrument**	**Sequencing depth**	**Slide**	**Spots detected**	**Median genes per spot**	**Median UMI counts per spot**	**Disease State**		**Species**
Human Colorectal Cancer: Whole Transcriptome Analysis	Illumina NovaSeq 6000, flow cell HHYWHDSXY (lanes 1-4)	112,228 mean reads per cell	V10A13-206	3,138	3,538	8,906	Invasive Adenocarcinoma		Human
Human Intestine Cancer (FPPE)	Illumina NovaSeq 6000, flow cell HCCVHDSX2 (lanes 1-2)	77,288 reads per spot	V10L13-021	2,660	7,438	25,312	large intestine colorectal cancer		Human
Human Colorectal Cancer, 11 mm Capture Area (FFPE)	Illumina NovaSeq 6000, flow cells HJWFGDSX5 and HJW2WDSX5 (lane 1-4)	911,044,482 reads / 46.2% saturation	V52Y10-310	9,080	9,560	49,282	colorectal cancer		Human

utilized in order to evaluate the samples' read quality [13]. We employed the TRIMMOMATIC software (V-0.39) to eliminate and reduce the sequence of samples [14]. For all samples, the following settings were used to minimize the sequenced reads: LEADING:15, TRAILING:15, SLIDINGWINDOW:4:25, and MINLEN:50. The human reference genome GRCH38 was utilized for alignment of the cleaned RNA-Seq data using the HISAT2 (v2.2.1) (https://daehwankimlab.github.io/hisat2/) mapping software [15]. Read counts for gene expression in each sample were calculated using the HT-Seq software and hg38.ncbiRefSeq.gtf annotation file [16]. All of the samples collected from these four data sets has been integrated using the SVA package (v3.50) in the R programming language to generate a combined data set of 415 samples [17]. The "ComBat_seq" function was employed for obtaining the batch effect from the count data. The R package DESeq2, (v1.42.1), was used to normalize the levels of gene expression to detect differences in gene expression across all samples [18]. The default threshold was defined as $|\log2(FC)| \geq 1$ and false discovery rate (FDR) < 0.05. The most crucial enrichment pathways and biological processes of DEGs were subsequently identified by utilizing the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) studies, which was carried out utilizing the ClusterProfiler package in the R programming language [19].

Quality control and data integration in single-cell transcriptomics analysis

The Seurat package [20] was utilized for analyzing the 10x Genomice scRNA-seq data (63,520 genes and 59,180 cells) in the R programming language. Quality control (QC) was carried out on the input matrix, and

lower-quality cells were filtered according to the following criteria in order to generate a high-quality scRNA-seq expression matrix: (1) To create a seurat object, cells express over 200 several distinct genes and genes expressed in at least three distinct cells considered to be acceptable. (2) In regard to total diversity, just cells with levels of gene expression greater than 200 and fewer than 6000 for normal objects, 6000 for objects with Crohn's disease, and 5000 for objects with ulcerative colitis have been evaluated. (3) To discover the proportion of genes related with both ribosome or mitochondrial functions that were actually found in each cell, the "Percentage-FeatureSet" function was employed. The data collected from scRNA-seq has been normalized utilizing the "LogNormalize" method in the "NormalizeData" function. After evaluation of quality, the top 2000 highly variable genes were identified and displayed using the "FindVariableFeatures" and "VariableFeaturePlot" functions. The "SelectIntegrationFeatures", "FindIntegrationAnchors" and "IntegrateData" approaches were used to integrate all seurat objects (normal, Crohn's disease, and ulcerative colitis) using the CCA algorithm. Following that, data for every gene expression was subsequently adjusted and centered using the "ScaleData" function. Consequently, the cell obtained a variance of 1 and a mean expression of 0.

Nonlinear dimension reduction, clustering and marker identification

Principal component analysis (PCA) was used for evaluating 2000 genes utilizing the "RunPCA" function in the Seurat package [20]. Utilizing the initial 20 main components, a comprehensive cellular clustering evaluation was carried out. The "VizDimLoadings", "DimPlot" and "DimHeatmap" functions were used to display gene expression for principal component analysis (PCAs). Following that, cellular clustering was identified using the "JackStraw, num.replicate = 100", "ScoreJackStraw, dims = 1:20", "JackStrawPlot", "ElbowPlot", "FindNeighbors, dims = 1:20" and "FindClusters" functions in the Seurat package [20], with the criterion "resolution" set at 0.5. Additionally, the uniform manifold approximation and projection (UMAP) approach with the "RunUMAP" function was used to reduce dimension and discover clusters of cells. The "FindAllMarkers" function was employed to estimate the false discovery rates (Adj.Pvalue) and log2foldchange in order to detect the genes that displayed differential expression (DEGs) in each cluster. DEGs with logfc. threshold = 1 and min.pct = 0.25 were considered the marker genes for each cluster. Utilizing the "FeaturePlot" functions with reduction "umap" in the Seurat package [20], the gene expression patterns of marker genes across multiple clusters throughout each cell types were also determined. Finally, the SingleR [21], celldex [21] and SingleCellExperiment [22] R packages were utilized to computationally clusters and categorize different types of cells utilizing the "HumanPrimaryCellAtlasData" function. To further resolve lymphocyte heterogeneity, B cell and T cell subsets were extracted from the main single-cell dataset and re-clustered using SingleR [21] annotation referenced to the Human Primary Cell Atlas and enabling refined subtype identification within each lineage. The ClusterProfiler package [19] in the R programming language was applied to detect the GO and KEGG characteristics of the marker genes of critical cells.

Single cell pseudotime trajectories analysis

The R package Monocle2 (v2.30.0) [23] was utilized to conduct single-cell trajectory evaluation with the aim of determining the cell state changes. We imported RDS data including cell types into R programming language. The "newCellDataSet" function has been utilized to build a new object with the inputs "expressionFamily = negbinomial.size" and "lowerDetectionLimit = 0.5". The options "reduction_method = "DDRTree" and "max_components = 2" were used to decrease dimensionality by utilizing the "reduceDimension" function. The cell lineage trajectories were then determined through pseudotime and cell clustering, utilizing the standard parameters of the Monocle2 package [23]. The findings were then displayed by using the "plot_cell_trajectory" function. Furthermore, dynamic changes in pseudotime-dependent gene expression over pseudotime were displayed using the "plot_genes_in_pseudotime" function. Additionally, pseudotime trajectories were constructed separately for each major cell type to examine lineage-specific gene expression dynamics. The directionality of each trajectory was inferred from progressive transitions in expression patterns of previously identified biological markers rather than preset annotations. Accordingly, early pseudotime phases represented immature or low-activity cellular states, whereas late phases corresponded to elevated expression of differentiation and activation markers, reflecting progressive cellular maturation within each lineage.

Investigation of cell-to-cell communication

The intercellular communication networks are comprehensively calculated utilizing scRNA-seq data by employing a new package named CellChat [24]. CellChat (https://github.com/sqjin/CellChat) may anticipate the essential communication inputs and results for cells and discover exactly these cells and signaling are connected based on human communication between ligand and receptor collections and detection of patterns algorithms [24]. The potential cell-cell interaction networks were discovered for every group. CellChat was built on a foundation of an extensive database containing secreted communication,

ECM-receptor, and cell-cell contact connections. Following the guidelines, we employed the R programming language to load the RDS files including the normalized count and cell types from the Seurat package [20] into CellChat (v2.1.2) package [24]. Normal parameter preprocessing approaches including "identifyOverExpressedGenes" and "identifyOverExpressedInteractions" were employed. Critical functions such as "aggregateNet", "computeCommunProb" and "computeCommunProbPathway" were utilized for the main analysis based on comparable parameters. Communications were demonstrated utilizing the R programming language.

Integrative analysis of gene biomarkers: ROC, survival and TCGA expression profiles

To assessment the biomarker potential of selected genes, ROC curves were created utilizing the pROC [25] package and "roc" function applied to compare tumor and normal samples and calculate AUC values for each gene. The findings were visualized to highlight the most discriminative genes and their relative performance. AUC values between 0.7 and 0.8 were considered acceptable, 0.8-0.9 good, and 0.9-1 excellent, demonstrating strong biomarker potential. Subsequently, survival evaluation was performed using the survival [26] and survminer [27] packages. Kaplan-Meier curves were generated using "Surv" and "survfit" functions and stratifying patients into high and low expression groups based on the median gene expression and hazard ratios with corresponding p-values were calculated to assess the prognostic significance of each gene. Finally, gene expression levels in tumor and normal tissues were examined using TCGA-COAD and TCGA-READ data obtained through the TCGAbiolinks [28] package. Differences in expression were visualized to validate the relevance of the selected genes and providing a comprehensive overview of their diagnostic and prognostic potential in colorectal cancer.

Decoding tissue architecture using spatial transcriptomics and cell deconvolution

We used an extensive collection of packages that complemented the processing, visualization, and computational components of our research for the spatial deconvolution evaluation. These contained GraphST for graph-based spatial analysis [29], OS for management of files, Torch for deep learning assistance [30], Skmisc and Scikit-learn for additional statistical and metric calculations [31], OT for optimal transportation analysis [32], Matplotlib and its submodule Pyplot for visualization [33], Pandas [34] and NumPy [35] for data manipulation and Scanpy [36] and AnnData for processing spatial transcriptomics data. The GraphST package [29] in a Python programming language was used to integrate single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics data in order to determine the spatial organization and cellular composition of IBD (Crohn's disease and ulcerative colitis) and colorectal cancer tissues. The scRNA-seq data, originally stored in RDS format, were converted to the AnnData-compatible h5ad format using the R packages SeuratData [37] and SeuratDisk [38]. The R programming language was chosen in order to perform this conversion. Following that, the "scanpy.read_h5ad" function from the Scanpy package [36] was used to import the single-cell dataset into Python. The filtered feature-barcode matrix (filtered_feature_bc_matrix.h5) and related images with high resolution were also included in the spatial transcriptomics data that was loaded using "scanpy.read_visium" from the visium platform. Using the "var_names_make_unique" function, unique variable names were allocated to the single-cell and spatial datasets to facilitate integration and prevent naming differences. To improve the accuracy of data and ensure analytical strength, the single-cell and spatial datasets completed a thorough preparation process. By applying "sc.pp.normalize_total" to normalize the spatial dataset (called as "data_spatial") with a target sum of 10,000 counts per spot, log-transformation was subsequently carried out utilizing "sc.pp.log1p" in order to stabilize variance. In order to identify highly variable genes (HVGs), "sc.pp.highly_variable_genes" was used. The top 2,000 genes were chosen based on their dispersion, and the dataset was then subset to these genes. To prepare the spatial data for GraphST analysis, we applied "GraphST.preprocess" for initial preprocessing, constructed spatial interactions with "GraphST.construct_interaction" to capture neighborhood relationships and added contrastive labels using "GraphST.add_contrastive_label" to improve feature discrimination. The single-cell dataset, called "data_singlecell" completed a similar processing approach. "sc.pp.normalize_total(target_sum=1e4)" was employed for normalization, while "sc.pp.log1p" was used for log-transformation. "data_singlecell.X[data_singlecell.X < 0] = 0" was used to establish negative expression variables to zero in order to eliminate data objects, and "np.nan_to_num" with the parameters "nan = 0.0", "posinf = 0.0", and "neginf = 0.0" was applied in order to replace incomplete or infinity values with zeros. "sc.pp.filter_cells(min_genes = 1)" was utilized to filter cells expressing a minimum of one gene, and "sc.pp.filter_genes(min_cells = 1)" was employed to filter genes expressed in at least one cell. The dataset was filtered in accordance with the selection of HVGs utilizing "sc.pp.highly_variable_genes(n_top_genes = 2000)". The single-cell data was then aligned using the spatial approaches using the "GraphST.preprocess" method. The "GraphST.preprocess.filter_with_overlap_gene" was used to harmonize the feature spaces of "data_spatial" and "data_singlecell" in order to identify and preserve

overlapping genes and ensure dataset compatibility. To create a representation appropriate for downstream modeling, "GraphST.get_feature" was used to extract features from the spatial data for model training and deconvolution. The GraphST model was initialized with a configuration of 1,200 epochs, a random seed of 50 for reproducibility, and computation delegated to a CUDA-enabled GPU (`torch.device('cuda:1')`) if available, defaulting to CPU otherwise. The single-cell reference data was utilized to determine the proportions of all types of cells inside each spatial spot by setting the "deconvolution=True" criterion. By mapping cellular characteristics onto the spatial context, "model.train_map" combined the single-cell and spatial data to create updated "data_spatial" and "data_singlecell" objects. The single-cell data was annotated with cell types according to specified frequency that were maintained in a list of terms (referred as "frequencies"). The "data_singlecell.obs['cell_type']" column was updated with a list of cell types ("cell_type_list") that was created by repeating each cell type based on its frequency. "GraphST.utils.project_cell_to_spot" was then used to project the single-cell annotations onto the spatial spots, keeping 15% of the top-ranked features ("retain_percent = 0.15") to optimize noise reduction and guarantee reliable deconvolution. We used the "scanpy.pl.spatial" function from the Scanpy [36] package to demonstrate the spatial distribution of deconvoluted cell types. This plotting function was used to the "data_spatial" object, highlighting specific cell types of interest for normal-Crohn's disease group such as: Epithelial cell, T cell (CD4+ central memory T cells, CD8+ central memory T cells, CD4+ effector memory T cells, CD8+ naïve T cells, CD4+ T cells and gamma-delta T cells), B cell (Plasma B cells, Memory B cells, Naïve B cells and Immature B cells), Tissue stem cell, Macrophage, Monocyte, Neutrophil, NK cell, DC, Endothelial cell and Astrocyte. For normal-ulcerative colitis such as: B cell (Plasma B cells, Memory B cells, Naïve B cells and Immature B cells), Epithelial cell, T cell (CD8+ central memory T cells, CD4+ effector memory T cells, CD8+ naïve T cells and gamma-delta T cells), Tissue stem cell, NK cell, Macrophage, Neutrophils, Hepatocytes and Endothelial cell. By modeling the distribution of cells according to spatial coordinates, we were able to show the number of cells in various colorectal and IBD tissues regions using the magma and viridis color palettes.

Spatial transcriptomics profiling of gene expression patterns

For gene expression analysis, 10x Genomics spatial transcriptomics data was employed to identify the spatial gene expression of colorectal cancer. Spatial transcriptomics data (ST) was analyzed and demonstrated using the Seurat (v5.0.1) R package [20]. The "SCTransform" function was used to normalize the data, the "ScaleData" function was utilized to scale the data, and the "RunPCA", "FindNeighbors", "FindClusters", and "RunUMAP" functions were employed to reduce the dimensionality and evaluate the clustering process. The results interpretation for integrated cell types were employed to provide the scRNA-seq reference in the normal options. By applying the R package SpacexR (v.2.2.1) [39], the spatial gene expression and spatial deconvolution of cell types from reference scRNA-seq data to the spatial transcriptomic data was obtained utilizing the RCTD (Robust Cell Type Decomposition) approach. For each slice of the Visium HD dataset, matrix containing raw counts and spatial coordinates were used to build the spatialRNA object in RCTD. After that, the reference scRNA-seq and spatialRNA objects were processed in parallel using the "create.RCTD" function, and the RCTD algorithm was carried out using the "doublet_mode = doublet" option using the "run.RCTD" function. The "SpatialFeaturePlot" function was used to identify specific genes.

Spatial autocorrelation analysis of gene expression and cell type frequencies

Spatial autocorrelation analysis of gene expression and cell type frequencies was conducted in R utilizing the Seurat [20], dplyr [40] and spdep [41] packages. A preprocessed single-cell dataset stored in H5Seurat format and imported with "LoadH5Seurat" function and normalized gene expression values for marker genes was extracted from the RNA assay using "GetAssayData" function, followed by subsetting to genes available in the dataset. These expression profiles were merged with the active cell type identities from H5Seurat object to construct a unified data matrix for downstream analyses. For each tissue dataset, spatially matched cells were retained and grouped by their annotated cell types. Spatial autocorrelation of gene expression within each cell type was evaluated using Moran's I approach from the spdep [41] package. Spatial neighbor relationships were defined through a 4-nearest-neighbor graph computed using "knearneigh" function and "knn2nb" function and corresponding weight matrices was generated utilizing "nb2listw(style = "W")" function. Moran's I statistics and p-values were calculated using "moran.test" function with "zero.policy = TRUE" to accommodate isolated spatial points.

IBD and CRC patients

A total of 1000 patients diagnosed with inflammatory bowel disease (IBD) and colorectal cancer (CRC) were included in this study from the Medical Diagnostic Laboratory in Golestan and Mazandaran Provinces, Iran. All the IBD and CRC patients involved in the study were diagnosed with pathological proof and has not been

received chemotherapy or radiotherapy before the surgery. Among them, 735 patients diagnosed with IBD and 265 patients diagnosed with CRC. The IBD group consisted of 348 males and 387 females, aged between 18 and 59 years. Within this group, 320 patients were diagnosed with Crohn's disease (CD) including 143 males and 177 females and 415 patients were diagnosed with ulcerative colitis (UC) including 205 males and 210 females. The CRC group included 265 patients, aged between 45 and 89 years, comprising 142 males and 123 females. According to clinical staging, 79 patients (36 males and 43 females) were classified as Stage I, 68 patients (35 males and 33 females) as Stage II, 80 patients (33 males and 47 females) as Stage III and 38 patients (24 males and 14 females) as Stage IV. Clinical information of the patients, including Patient ID, Pathological Diagnosis, TNM or IBD Group, Age, Gender, and CEA level, is provided in Supplementary Table S1.

Serum CEA evaluation in CRC and IBD patients

Serum Carcinoembryonic Antigen (CEA) levels evaluated in inflammatory bowel disease (IBD) and colorectal cancer (CRC) patients. The study population included patients diagnosed with or suspected of having IBD and CRC based on physician referral and clinical findings. Individuals with other indications for laboratory testing were excluded. Additional exclusion criteria included a history of jaundice, cirrhosis, cholelithiasis, pancreatitis, or other malignancies. All participants were informed about the purpose and procedures of the study, and written consent was obtained prior to blood collection. Demographic data such as age, sex, and relevant clinical characteristics were recorded for correlation analysis. From each participant, 5 mL of fasting venous blood was collected under sterile conditions in plain tubes without anticoagulant. Samples were centrifuged at 1500 rpm for 5 min to separate serum, which was then aliquoted into microtubes and stored at -20 °C until analysis. To ensure consistency, all samples were processed immediately after collection following standard laboratory procedures. Serum levels of CEA were quantitatively determined using a fully automated LIAISON analyzer (Diasorin, Italy), employing the chemiluminescent immunoassay (CLIA) method, known for its high sensitivity and specificity. In some comparable studies, enzyme-linked immunosorbent assay (ELISA) kits, including the CEA Human ELISA Kit (Thermo Fisher), were used for reference. Normal reference ranges were defined as < 4 µg/mL for CEA, consistent with international clinical laboratory standards. The clinical staging of CRC patients was determined based on postoperative pathological findings following the TNM classification system of the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC). The study included patients across all stages (I-IV) and cases undergoing neoadjuvant chemotherapy or with incomplete data were excluded. All laboratory approaches including blood collection, sample handling and measurements were conducted according to standardized operating protocols. The testing was performed under supervision of a qualified clinical laboratory specialist. Data were analyzed to evaluate the relationship between tumor marker levels and patient demographic or clinical variables. This study was performed in accordance with the Helsinki Declaration and all patients participating in the study provided written informed consent. This study is approved by the research ethics committees of Islamic Azad University-Sari Branch with the following ethic code IR.IAU.SARI.REC.1404.365.

Statistical analysis

All statistical analyses were performed using Ubuntu 22.04 LTS, R (v4.3.2, R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org/) and Python programming languages. Various packages and toolkits were employed for data processing and visualization, including SRAToolkits, FASTQC, TRIMMOMATIC, HISAT2, HTSeq, SVA, DESeq2, ClusterProfiler, Seurat, Celldex, SingleCellExperiment, SingleR, Monocle2, CellChat, SeuratDisk, SeuratData, SpacexR, dplyr, spdep, TCGAbiolinks, pROC, survival, survminer, Scanpy, Matplotlib, OS, Torch, Pandas, NumPy, Skmisc, Scikit-learn, OT, and GraphST. Descriptive statistics were used to summarize the demographic and biochemical characteristics of study participants. Differences between groups were assessed using the chi-square test, unpaired Student's t-test, or ANOVA, as appropriate. The serum carcinoembryonic antigen (CEA) levels were compared across diagnostic and demographic subgroups. Cell frequencies measured by Xenium were compared between the two groups using the Mann-Whitney test, with multiple comparisons controlled using the two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli to control the false discovery rate (FDR). A p-value < 0.05 was considered statistically significant.

Results

Preprocessing of bulk-RNA seq data

FASTQC software was used to calculate the raw processed read accuracy from bulk-RNAseq analysis, and the results indicated that the level of quality was excellent. To make sure that none of the bases with a phred quality lower than 15 were identified, every sample were processed using standard data cleaning approaches in accordance with their initial data good. We conducted a comprehensive analysis of transcriptomic data derived from four publicly available datasets: SRP303290, ERP106487, SRP336051, and SRP351557, encompassing

a total of 11,583,879,924 sequences. The SRP303290 dataset initially included 6,613,782,588 sequences, which were refined to 5,712,391,130 after quality trimming. Similarly, the ERP106487 dataset started with 1,511,710,813 sequences, retaining 970,158,026 post-trimming. For SRP336051, we processed 1,945,450,027 sequences, of which 1,465,643,660 remained after trimming. Lastly, the SRP351557 dataset contained 1,512,936,496 sequences, with 1,120,611,457 sequences retained following the trimming process. In total, after applying quality control measures, 9,268,804,273 sequences were carried forward for subsequent transcriptomic analyses, ensuring robust data integrity for downstream investigations (Supplementary Figure S1). The bulk-RNASeq analysis statistics are listed in Supplementary Table S2.

Differential gene expression analysis

Following preprocessing, we performed differential gene expression analysis to identify upregulated and downregulated genes across distinct groups, comparing normal samples to those from Crohn's disease and ulcerative colitis across various colon tissue regions. In the normal vs. Crohn's disease group for colon tissue, we identified 140 upregulated and 185 downregulated genes. For the normal vs. ulcerative colitis group in colon tissue, 1,010 genes were upregulated, while 116 were downregulated. In the normal vs. Crohn's disease group for ileum tissue, 1,089 genes were upregulated, and 48 were downregulated. For the normal vs. Crohn's disease group in ascending colon tissue, we observed 366 upregulated and 401 downregulated genes, whereas the normal vs. ulcerative colitis group in ascending colon tissue showed 631 upregulated and 420 downregulated genes. In the normal vs. Crohn's disease group for sigmoid colon tissue, 767 genes were upregulated, and 662 were downregulated, while the normal vs. ulcerative colitis group in sigmoid colon tissue exhibited 678 upregulated and 663 downregulated genes. Lastly, in the normal vs. Crohn's disease group for terminal ileum tissue, we identified 719 upregulated and 1,297 downregulated genes, and the normal vs. ulcerative colitis group in terminal ileum tissue revealed 206 upregulated and 317 downregulated genes. The full list of DEGs for all groups can be found in Supplementary Table S3.

Exploring cellular diversity through single-cell transcriptome analysis

We performed a scRNA-seq analysis of 63,520 genes and 59,180 cells from 18 human IBD samples: 6 healthy (21,171 genes, 20,351 cells), 6 Crohn's disease (CD) (21,959 genes, 21,747 cells) and 6 ulcerative colitis (UC) (20,390 genes, 17,082 cells). Supplementary Table S4 indicates the number of cells and genes for 18 samples both before and after quality controls. After that, scRNA-seq data normalization and quality evaluation, lower-quality cells were removed. Lastly, for additional analysis, the top 2000 genes with the largest variance variations were selected. Next, we utilized the "RunPCA" function to reduce the PCA dimensional of the top 2000 highly variable genes as a component of a dimension reduction examination across samples. Many "important" PCs with low p-values were identified. After utilizing the JackStrawPlot method to demonstrate the top 20 principal components, the variance of each PC was measured and compared to the median dispersion. The data of highly diverse genes was fully reflected by the "important" PCs, which typically had a low p-value. whereas combined, the "ElbowPlot" function demonstrated that, although the flexible point was approximately the 10th PC, the variation decreased gradually after the 20th PC. At this point we generated a heatmap of the original PCs, highlighting their most significant genes in order to demonstrate how their DEGs differentiated from that of other PCs. Overall, we selected the top 20 PCs for additional UMAP analysis (Supplementary Figure S2).

Inflammation associated modulation of cellular landscape in Crohn's disease

Numerous cell clusters between normal and Crohn's disease were examined by combining the samples, as demonstrated by the umap plot (Fig. 2A). Furthermore, cells were clustered using the "FindCluster" function to generate 24 clusters (Fig. 2A). Following that, we identified 11 different cell types inside these clusters using SingleR, a computational annotation program based on the Human Primary Cell Atlas database. These cell types included: epithelial cell (cluster annotated: C0, C3, C4, C9 and C11), T cell (cluster annotated: C1, C5, C10, C12, C14 and C20), B cell (cluster annotated: C2, C7 and C18), tissue stem cell (cluster annotated: C6, C19 and C23), macrophage (cluster annotated: C8), monocyte (cluster annotated: C13), neutrophil (cluster annotated: C15), NK cell (cluster annotated: C16), DC (cluster annotated: C17), endothelial cell (cluster annotated: C21) and astrocyte (cluster annotated: C22) (Fig. 2A, Supplementary Table S5). Next, we detected 4 subpopulation for B cells such as: immature B cells (cluster annotated: C5), naïve B cells (cluster annotated: C4), memory B cells (cluster annotated: C3) and plasma B cells (cluster annotated: C0, C1 and C2) from 7 subclusters and 6 subpopulation for T cells such as: CD4+ central memory T cells (cluster annotated: C0, C2 and C8), CD8+ central memory T cells (cluster annotated: C1 and C5), CD4+ effector memory T cells (cluster annotated: C3 and C9), CD8+ naïve T cells (cluster annotated: C4), CD4+ T cells (cluster annotated: C6 and C7) and gamma-delta T cells (cluster annotated: C10) from 11 subclusters (Supplementary Table S5). The frequency analysis revealed significant alterations in epithelial cell, T cell, B cell and tissue stem cell populations

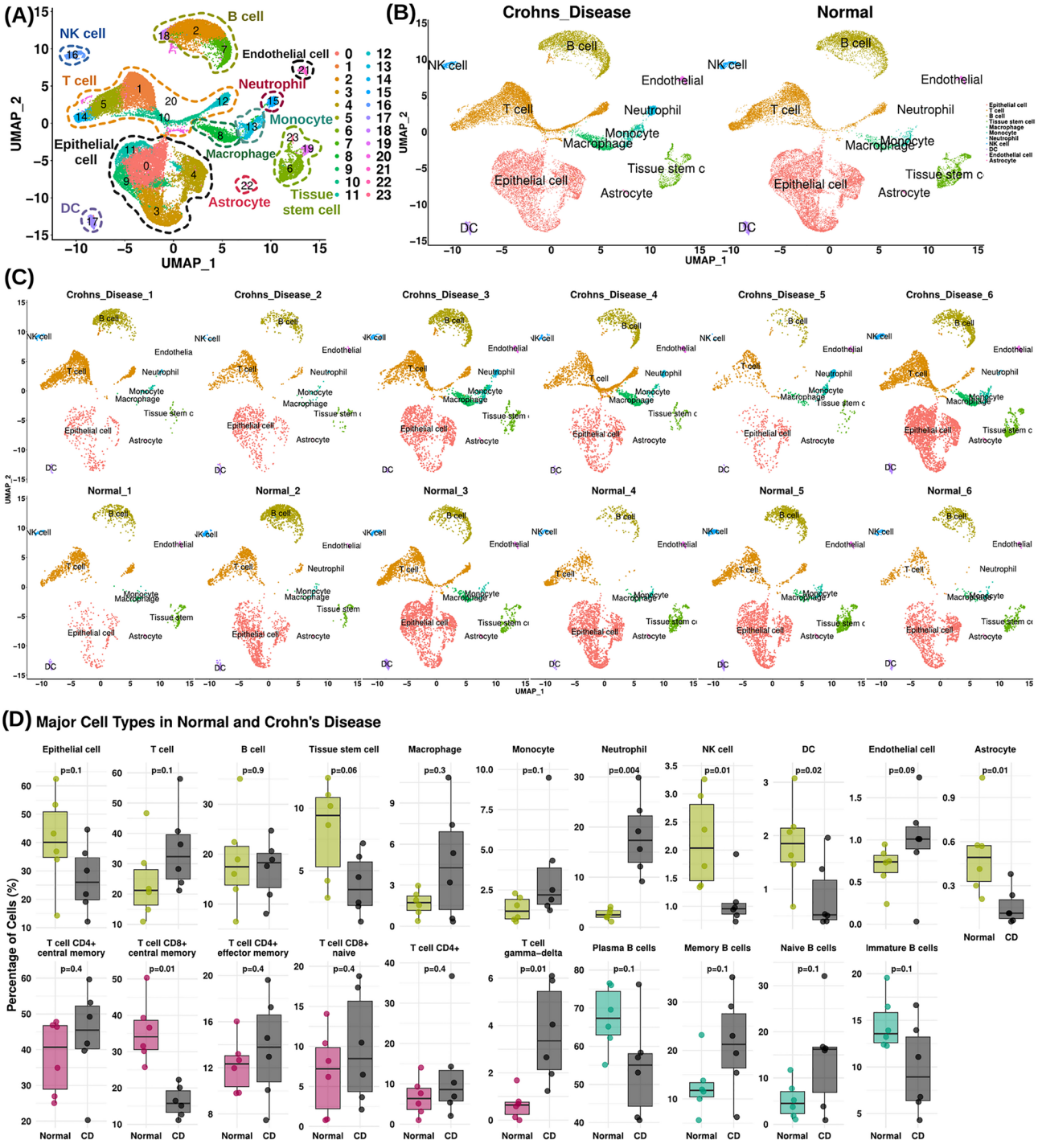

Fig. 2 High resolution of single cell transcriptomic atlas in Crohn's disease compared to normal intestinal tissue reveals distinct reshaping of the mucosal immune landscape. (**A**) UMAP visualization of 24 integrated clusters from combined normal and Crohn's disease samples, colored by computationally annotated cell types using SingleR and the Human Primary Cell Atlas reference. Major cell types identified include epithelial cells (C0, C3, C4, C9, C11), T cells (C1, C5, C10, C12, C14, C20), B cells (C2, C7, C18), tissue stem cells (C6, C19, C23), macrophages (C8), monocytes (C13), neutrophils (C15), NK cells (C16), dendritic cells (DC; C17), endothelial cells (C21), and astrocytes (C22). (**B**) UMAP plot showing the relative frequencies of major cell types in Crohn's disease compared to normal samples, highlighting significant depletion of epithelial cells and tissue stem cells, and enrichment of T cells in Crohn's disease tissue. (**C**) UMAP plot displaying the proportional composition of all 11 annotated cell types in each individual normal and Crohn's disease samples. (**D**) Comparative box plots of cell-type proportions between Crohn's disease and normal groups, indicating decreased percentages of epithelial cells, B cells, tissue stem cells, NK cells, DC, and astrocytes, and increased percentages of T cells, macrophages, monocytes, neutrophils, and endothelial cells in Crohn's disease samples and subpopulation analysis of immune cells revealed marked shifts in T and B cell differentiation states in Crohn's disease: increased proportions of CD4+ central memory, CD4+ effector memory, CD8+ naïve, CD4+, and gamma-delta T cells, with decreased CD8+ central memory T cells; and elevated memory and naïve B cells alongside reduced plasma and immature B cells in Crohn's disease compared to normal tissue

between Crohn's tissue and normal controls and demonstrating that the most common cell types in samples of normal and Crohn's disease are epithelial cells, T cells, B cells, and tissue stem cells. Specifically, epithelial cells exhibited a significant decrease in Crohn's tissue compared to normal samples. In contrast, T cells showed a significant increase in Crohn's tissue relative to normal tissue. Moreover, both B cells and tissue stem cells demonstrated a significant decrease in Crohn's tissue compared with the normal group (Fig. 2B). The proportion of 11 cell types in tissue samples from Crohn's disease and normal tissue was then analyzed, as illustrated in Fig. 2C. The percentages of each population in the normal and Crohn's disease samples were subsequently compared and it was discovered that in the immune cell population, the proportion of epithelial cell (CD vs. normal: 29.94% vs. 41.92%) ($p = 0.1$), B cell (CD vs. normal: 16.85% vs. 18.36%) ($p = 0.9$), tissue stem cell (CD vs. normal: 4.33% vs. 7.92%) ($p = 0.06$), NK cell (CD vs. normal: 1.24% vs. 2.19%) ($p = 0.01$), DC (CD vs. normal: 0.86% vs. 1.95%) ($p = 0.02$) and astrocyte (CD vs. normal: 0.20% vs. 0.56%) ($p = 0.01$) in the Crohn's disease tissue samples was deceased, while the proportion of T cell (CD vs. normal: 33.65% vs. 23.32%) ($p = 0.1$), macrophage (CD vs. normal: 5.36% vs. 1.74%) ($p = 0.3$), monocyte (CD vs. normal: 3.10% vs. 1.34%) ($p = 0.1$), neutrophil (CD vs. normal: 3.51% vs. 0.01%) ($p = 0.004$), endothelial cell (CD vs. normal: 0.90% vs. 0.64%) ($p = 0.09$) was increased (Fig. 2D, Supplementary Table S6). Several distinct changes were observed in T cell and B cell subpopulations when comparing normal and Crohn's disease samples. The percentages of each T cell subset in the normal and Crohn's disease samples were subsequently compared, revealing distinct distribution profiles across multiple T cell populations. The proportion of CD4+ central memory T cells (CD vs. normal: 43.63% vs. 36.51%) ($p = 0.4$), CD4+ effector memory T cells (CD vs. normal: 13.17% vs. 12.92%) ($p = 0.4$), CD8+ naïve T cells (CD vs. normal: 9.48% vs. 6.42%) ($p = 0.4$), CD4+ T cells (CD vs. normal: 12.95% vs. 6.06%) ($p = 0.4$), and gamma-delta T cells (CD vs. normal: 3.66% vs. 0.65%) ($p = 0.01$) was increased in Crohn's disease compared to normal controls. In contrast, the proportion of CD8+ central memory T cells (CD vs. normal: 16.28% vs. 35.28%) ($p = 0.01$) was decreased in Crohn's disease tissue samples (Fig. 2D, Supplementary Figure S3A, Supplementary Table S6). The percentages of each B cell subset in the normal and Crohn's disease samples were subsequently compared and revealing distinct distribution patterns among the four B cell subpopulations. The proportion of plasma B cells (CD vs. normal: 54.79% vs. 67.10%) ($p = 0.1$) and immature B cells (CD vs. normal: 9.49% vs. 14.64%) ($p = 0.1$) was decreased in Crohn's disease tissue samples. In contrast, the proportion of memory B cells (CD vs. normal: 21.39% vs. 12.86%) ($p = 0.1$) and naïve B cells (CD vs. normal: 14.33% vs. 5.56%) ($p = 0.1$) was increased in Crohn's disease compared to normal controls (Fig. 2D, Supplementary Figure S3A, Supplementary Table S6). We utilized the umap approach for unsupervised clustering to identify gene expression profiles and discovered numerous important genes, including: (epithelial cell: 389), (T cell: 185), (B cell: 104), (tissue stem cell: 588), (macrophage: 87), (monocyte: 198), (neutrophil: 88), (NK cell: 67), (DC: 47), (endothelial cell: 198) and (astrocyte: 132). The complete list of markers for the 11 cell types can be found in Supplementary Table S7.

Uncovering inflammatory driven cellular diversity in ulcerative colitis

The umap plot demonstrates that many cell clusters between normal and ulcerative colitis were investigated via integrating the samples (Fig. 3A). Additionally, 20 clusters were created by clustering cells using the "FindCluster" function (Fig. 3A). Utilizing these clusters, we then employed SingleR, a computational annotation program based on the Human Primary Cell Atlas database, to identify 9 distinct cell types. These cell types consisted of: B cell (Cluster annotated: C0, C2 and C11), epithelial cell (Cluster annotated: C1, C6, C7 and C18), T cell (Cluster annotated: C3, C4, C8 and C12), tissue stem cell (Cluster annotated: C5 and C19), NK cell (Cluster annotated: C9 and C14), macrophage (Cluster annotated: C10 and C17), neutrophils (Cluster annotated: C13), hepatocytes (Cluster annotated: C15) and endothelial cell (Cluster annotated: C16) (Supplementary Table S5). The frequency analysis revealed significant alterations in epithelial, T, B, and tissue stem cell populations between ulcerative colitis tissue and normal controls. Specifically, epithelial cells exhibited a significant decrease in ulcerative colitis tissue compared to normal samples. In contrast, T cells and B cells showed a significant increase in ulcerative colitis tissue relative to normal tissue. Moreover, tissue stem cells demonstrated a significant decrease in ulcerative colitis tissue compared with the normal group (Fig. 3B). Following that, we identified 4 B cell subpopulations, including: plasma B cells (cluster annotated: C0, C2, C6, C7, C8 and C9), memory B cells (cluster annotated: C1 and C4), naïve B cells (cluster annotated: C3) and immature B cells (cluster annotated: C5) from 10 subclusters and CD8+ central memory T cells (cluster annotated: C0, C2, C4 and C7), CD4+ effector memory T cells (cluster annotated: C1 and C6), CD8+ naïve T cells (cluster annotated: C3) and gamma-delta T cells (cluster annotated: C8) from 9 subclusters (Supplementary Table S5). Following that, the proportion of nine different cell types in samples of normal and ulcerative colitis tissues was examined and shown in (Fig. 3C). After comparing the proportions of each population in the normal and ulcerative colitis

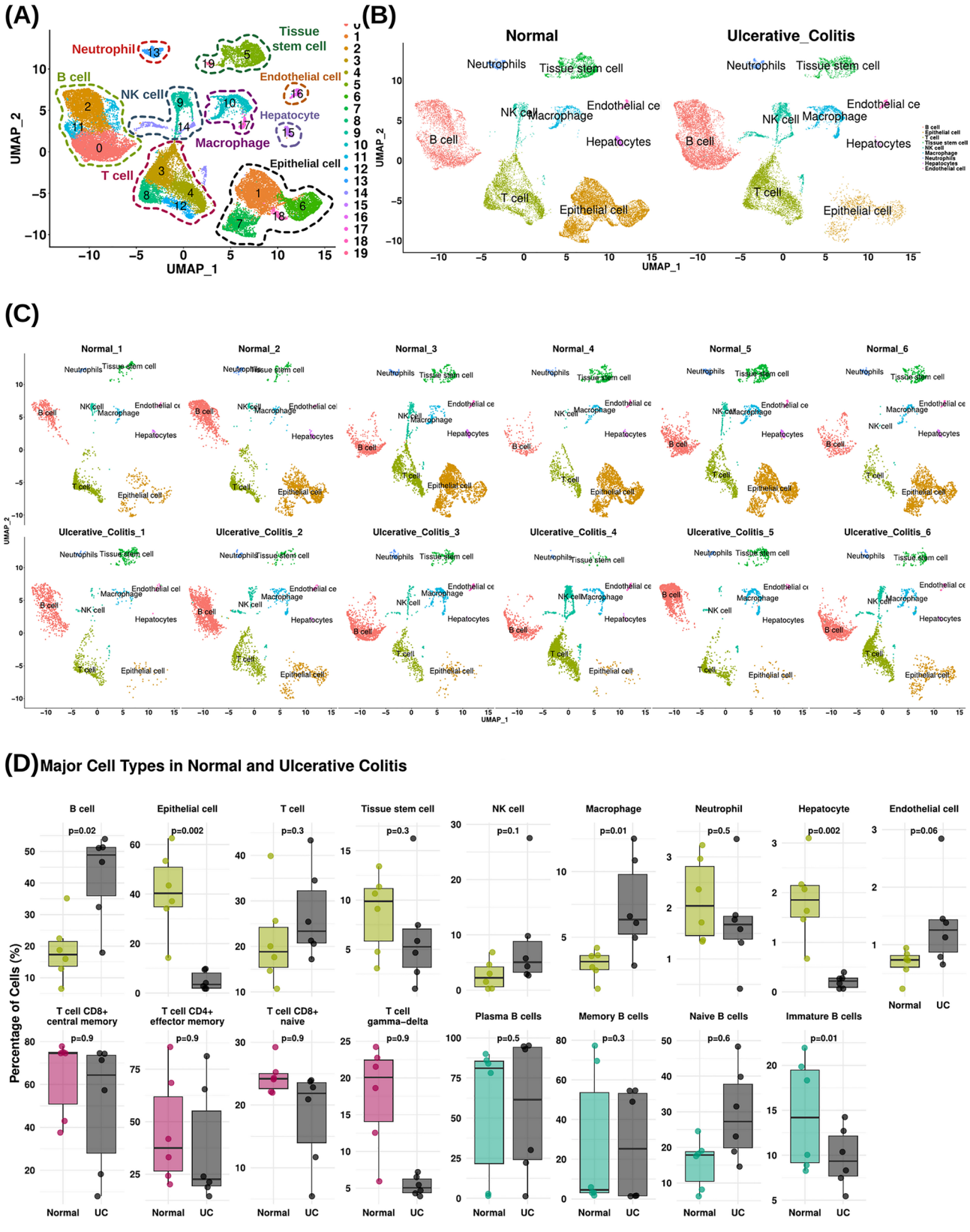

Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 High-resolution single-cell transcriptomic atlas of ulcerative colitis compared to normal intestinal tissue demonstrates distinct reshaping of the mucosal immune landscape. (**A**) UMAP visualization of 20 integrated clusters from combined normal and ulcerative colitis samples, colored by computationally annotated cell types using SingleR and the Human Primary Cell Atlas reference. Nine major cell types were identified: B cells (C0, C2, C11), epithelial cells (C1, C6, C7, C18), T cells (C3, C4, C8, C12), tissue stem cells (C5, C19), NK cells (C9, C14), macrophages (C10, C17), neutrophils (C13), hepatocytes (C15), and endothelial cells (C16). (**B**) UMAP plot showing relative frequencies of major cell types in ulcerative colitis compared to normal samples, demonstrating significant depletion of epithelial cells and tissue stem cells accompanied by marked enrichment of B cells and T cells in ulcerative colitis tissue. (**C**) UMAP plot illustrating the proportional composition of all nine annotated cell types across individual normal and ulcerative colitis samples. (**D**) Comparative box plots of cell type proportions between ulcerative colitis and normal groups, showing decreased percentages of epithelial cells, tissue stem cells, neutrophils, and hepatocytes, and increased percentages of B cells, T cells, NK cells, macrophages, and endothelial cells in ulcerative colitis samples and subpopulation analysis indicated shifts in B and T cell differentiation states in ulcerative colitis: increased naïve B cells and slightly elevated plasma B cells with reduced memory and immature B cells; stable proportions of most T-cell subsets except a decrease in gamma-delta T cells

samples, it was found that among the immune cell population, the proportion of epithelial cell (UC vs. normal: 5.15% vs. 42.07%) ($p = 0.002$), tissue stem cell (UC vs. normal: 5.33% vs. 8.47%) ($p = 0.3$), neutrophils (UC vs. normal: 1.64% vs. 2.19%) ($p = 0.5$), hepatocytes (UC vs. normal: 0.22% vs. 1.95%) ($p = 0.002$) in the ulcerative colitis tissue samples was reduced, while the proportion of B cell (UC vs. normal: 40.72% vs. 18.28%) ($p = 0.02$), T cell (UC vs. normal: 28.07% vs. 20.61%) ($p = 0.3$), NK cell (UC vs. normal: 9.95% vs. 2.70%) ($p = 0.1$), macrophage (UC vs. normal: 7.59% vs. 3.03%) ($p = 0.01$) and endothelial cell (UC vs. normal: 1.30% vs. 0.64%) ($p = 0.06$) was increased (Fig. 3D, Supplementary Table S8). The percentages of each T cell subset in the normal and ulcerative colitis samples were subsequently compared, revealing distinct distribution patterns among the four measured T cell populations. The proportion of CD8+ central memory T cells (UC vs. normal: 43.57% vs. 41.83%) ($p = 0.9$), CD4+ effector memory T cells (UC vs. normal: 32.40% vs. 30.39%) ($p = 0.9$), and CD8+ naïve T cells (UC vs. normal: 18.70% vs. 19.30%) ($p = 0.9$) showed no notable change in ulcerative colitis tissue. In contrast, the proportion of gamma-delta T cells (UC vs. normal: 5.33% vs. 8.48%) ($p = 0.9$) was decreased in ulcerative colitis compared with normal controls, with the most pronounced reduction observed in gamma-delta T cells (Fig. 3D, Supplementary Figure S3B, Supplementary Table S8). The percentages of each B cell subset in the normal and ulcerative colitis samples were subsequently compared, revealing distinct distribution patterns among the four measured B cell populations. The proportion of plasma B cells (UC vs. normal: 37.69% vs. 36.68%) ($p = 0.5$) showed a slight increase in ulcerative colitis tissue. In contrast, the proportion of memory B cells (UC vs. normal: 17.45% vs. 20.63%) ($p = 0.3$) and immature B cells (UC vs. normal: 13.32% vs. 21.75%) ($p = 0.01$) was decreased in ulcerative colitis, while naïve B cells (UC vs. normal: 29.36% vs. 23.12%) ($p = 0.6$) was increased in ulcerative colitis compared with normal controls (Fig. 3D, Supplementary Figure S3B, Supplementary Table S8). These findings indicate a shift in T cell and B cell composition in ulcerative colitis, potentially reflecting altered immune dynamics and T cell and B cell responses in the disease state. Employing unsupervised clustering depending on the umap method, 9 cell types were identified across normal and ulcerative colitis samples and many significant genes were found, such as: (B cell: 72), (epithelial cell: 417), (T cell: 150), (tissue stem cell: 347), (NK cell: 62), (macrophage: 268), (neutrophils: 65), (hepatocytes: 39) and (endothelial cell: 200). A comprehensive list of markers for each of the nine cell types are provided in Supplementary Table S9.

Identification of pathway enrichment analysis in Crohn's disease and ulcerative colitis

GO and KEGG features have been identified in the main cell marker genes for normal ulcerative colitis and normal Crohn's disease. In the normal-Crohn's disease group, marker genes of T cell, NK cell, macrophage, astrocyte, B cell and monocyte were associated with biological process such as: regulation of T cell activation, activating cell surface receptor signaling pathway and B cell mediated immunity (Supplementary Table S10, Supplementary Figure S4A). In the molecular function category, the marker genes of T cell, macrophage, astrocyte and monocyte related to MHC protein complex binding and MHC class II protein complex binding and antigen binding. The marker genes of T cell, NK cell, monocyte, tissue stem cell and neutrophils are connected to cytokine activity and immunoglobulin binding (Supplementary Table S10, Supplementary Figure S4B). In the cellular component category, the marker genes of epithelial cell, NK cell, macrophage, astrocyte, B cell, monocyte, tissue stem cell and endothelial cell related to collagen-containing extracellular matrix, external side of plasma membrane and endoplasmic reticulum lumen (Supplementary Table S10, Supplementary Figure S4C). The marker genes of epithelial cell, macrophage, monocyte and tissue stem cell is related to the Mineral absorption and the marker genes of T cell, macrophage, astrocyte and monocyte are associated to the Antigen processing and presentation and Cytokine-cytokine receptor interaction. The marker genes of DC, monocyte, tissue stem cell and neutrophils are related to the NF-kappa B signaling pathway; found in the category of KEGG pathways (Supplementary Table S10, Supplementary Figure S4D). Supplementary Table

S10 fully includes all GO BP, GO MF, GO CC and KEGG Pathway gene markers for all cell types in the normal-Crohn's disease group. In the normal-ulcerative colitis group, marker genes of B cell, T cell, NK cell, neutrophils and macrophage are related to the immune response-activating cell surface receptor signaling pathway, immune response-regulating cell surface receptor signaling pathway, B cell mediated immunity and immunoglobulin mediated immune response; found in biological process category (Supplementary Table S11, Supplementary Figure S4E). In the molecular function category, marker genes of B cell, T cell, NK cell, macrophage are related to the antigen binding, immunoglobulin receptor binding, MHC protein complex binding. The marker genes of T cell, neutrophils, macrophage and tissue stem cell is connected to the cytokine activity. Marker genes of NK cell and macrophage are associated to the MHC class II protein complex binding (Supplementary Table S11, Supplementary Figure S4F). The marker genes of B cell are related to the immunoglobulin complex, IgG immunoglobulin complex and endoplasmic reticulum lumen. The marker genes of T cell, NK cell, neutrophils, macrophage, tissue stem cell and endothelial cell are related to the external side of plasma membrane and MHC class II protein complex; found in the cellular component category (Supplementary Table S11, Supplementary Figure S4G). In the KEGG pathway category, marker genes of NK cell and T cell are associated to the Intestinal immune network for IgA production, Cytokine-cytokine receptor interaction, Natural killer cell mediated cytotoxicity and Apoptosis. The marker genes of epithelial cell are related to the Mineral absorption and Bile secretion and finally the marker genes of hepatocytes are connected to the MicroRNAs in cancer and EGFR tyrosine kinase inhibitor resistance (Supplementary Table S11, Supplementary Figure S4H). All GO BP, GO MF, GO CC and KEGG Pathway gene markers for every cell type in the normal-ulcerative colitis group are fully included in Supplementary Table S11.

Pseudotime trajectories between Crohn's disease and ulcerative colitis

In order to study the patterns of development, pseudotemporal evaluation was conducted independently for each type of cell. The overall lineage progression trajectories were then created using the Monocle2 package in the R programming language by categorizing distinct cells according to the clusters normal-Crohn's disease, pseudotime, and classification (Fig. 4A). Cells underwent three distinct states with a subsequent trajectories progress: the pre-branch (the start of branching) and two more branches (Fig. 4A). B cell and T cell were appeared primarily into the beginning of the trajectory. Neutrophils and NK cells were most common around the final point of trajectory branch 1, while a greater proportion of tissue stem cells and astrocytes were found near the end of trajectory branch 2 (Fig. 4A). Our findings indicates that the cells responsible for inflammation to spread may be B cells, T cells, neutrophils, NK cells and tissue stem cells. The T cell subcategories including: CD8+ naïve T cells, gamma-delta T cells and CD4+ T cells are mostly seen during the first phase of the development (Fig. 4B). In Crohn's disease, it was found that the CD8+ central memory T cells, CD4+ effector memory T cells, and CD8+ naïve T cells remained around the end of the trajectory branches (Fig. 4B). The B cell subcategories such as: naïve B cells and immature B cells are mainly observed in the initial stage of the progression (Fig. 4C). The memory B cells and plasma B cells cell were discovered near the final stage of trajectories branches 2 (Fig. 4C). According to our studies, plasma B cells, memory B cells, immature B cells, CD8+ central memory T cells, CD4+ effector memory T cells and CD8+ naïve T cells might be the cells causing inflammatory processes to develop in Crohn's disease. In order to construct the tree-like structure of the complete lineage development for the normal-ulcerative colitis category, each cell was classified according to its class, pseudotime and normal-ulcerative colitis cluster (Fig. 4D). The cells with following trajectory development went through three different states: the pre-branch (beginning of branching) and two additional branches (Fig. 4D). Mostly at the initial point of the trajectory, B and T cells were detected. Neutrophils, B cell and macrophage cells were most common around the final stage of trajectory branch, while a biggest abundance of epithelial cell and hepatocytes were identified around the final stage of trajectory branch (Fig. 4D). According to our study findings, B cells, T cells, neutrophils, epithelial cells and hepatocytes can represent the cells which lead inflammation to propagate. In ulcerative colitis, the T cell subtypes such as CD4+ effector memory T cells and CD8+ central memory T cells is primarily observed in the initial stage of differentiation (Fig. 4E). The gamma-delta T cells and CD8+ naïve T cells were discovered to have maintained during the final stage of the trajectory branches (Fig. 4E). The B cell subtypes including: plasma B cells and memory B cells are mainly detected in the first stage of the trajectories (Fig. 4F). The immature B cells and memory B cells were identified near the final point of trajectories branches (Fig. 4F). Based on our research, immature B cells, memory B cells, plasma B cells, CD4+ effector memory T cells, CD8+ central memory T cells, and CD8+ naïve T cells may be the cells responsible for the development of inflammatory processes in ulcerative colitis.

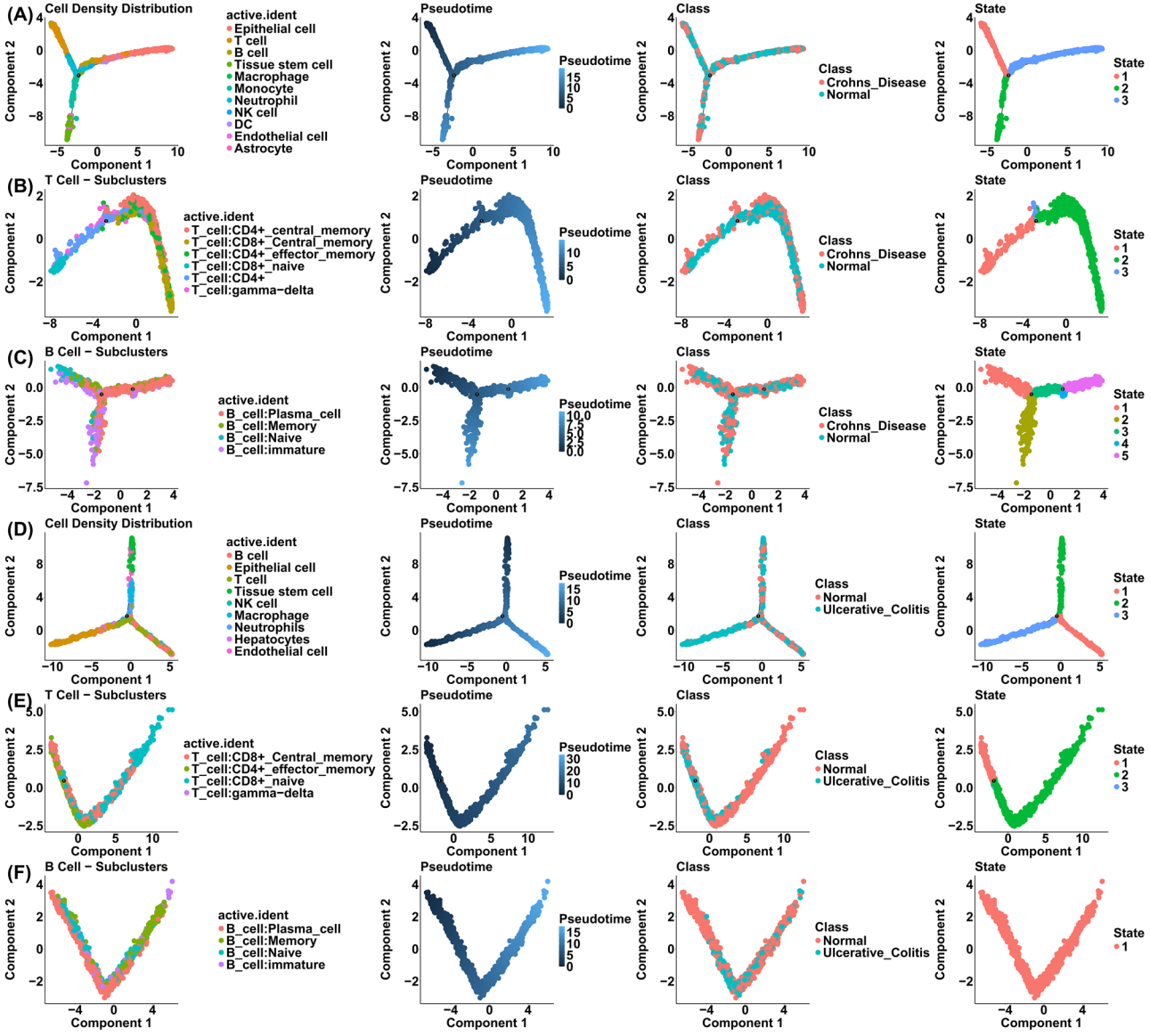

Fig. 4 Pseudotime trajectory analysis reveals lineage progression and inflammatory cell dynamics in Crohn's disease and ulcerative colitis. (**A**) Monocle2-derived pseudotime trajectory of integrated Crohn's disease and normal samples, colored by major cell type. Cells progress through a pre-branch state followed by bifurcation into two distinct branches. B cells and T cells dominate the root (early pseudotime), neutrophils and NK cells enrich branch 1 terminal fate, whereas tissue stem cells and astrocytes predominate branch 2 terminal fate. (**B**-**C**) Distribution of B and T cell subpopulations along Crohn's disease pseudotime. Naïve and immature B cells together with CD8+ naïve, gamma-delta, and CD4+T cells localize to early pseudotime, whereas memory and plasma B cells, CD8+ central memory, and CD4+ effector memory T cells occupy late pseudotime, suggesting their potential roles in sustaining chronic inflammation in Crohn's disease. (**D**) Pseudotime trajectory of integrated ulcerative colitis (UC) and normal samples, colored by major cell type. B and T cells initiate the trajectory at the root; neutrophils, additional B cells, and macrophages accumulate at the end of one branch, while epithelial cells and hepatocytes dominate the terminal fate of the second branch. (**E**-**F**) Distribution of B and T cell subpopulations along UC pseudotime. Plasma and memory B cells, CD4+ effector memory, and CD8+ central memory T cells appear early, whereas immature B cells, naïve CD8+, and gamma-delta T cells localize to late pseudotime, highlighting distinct differentiation programs driving inflammation in ulcerative colitis compared to Crohn's disease

Immune cell network interactions highlight key inflammatory pathways in Crohn's disease and ulcerative colitis

Cellchat assessment showed that the tissue stem cells, astrocytes and endothelial cells in the normal-Crohn's disease group had highest signal output capabilities (Fig. 5A). Furthermore, the strongest incoming and outgoing signals were detected by tissue stem cells, monocytes and endothelial cells (Fig. 5B). The analysis of cell-cell communication across eight major signaling pathways such as: COLLAGEN, LAMININ, APP, CypA, CD99, GALECTIN, VISFATIN and MIF reveals specific dominant interactions that define the intercellular landscape within the tissue microenvironment (Fig. 5C). In

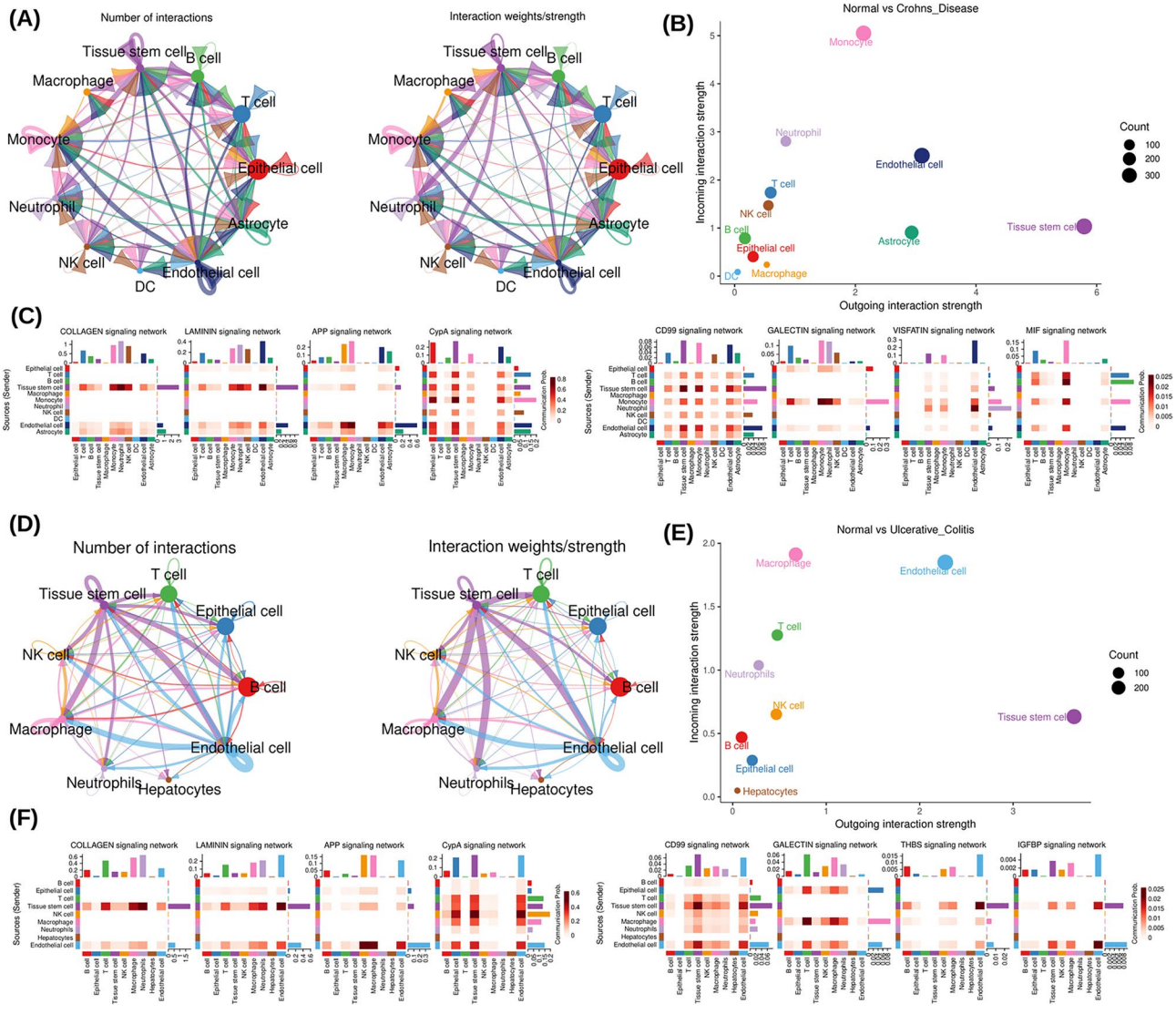

Fig. 5 Cell-cell communication networks inferred by Cell Chat in Crohn's disease and ulcerative colitis compared to normal intestinal tissue. (**A**) Overall cell-cell communication network in the normal-Crohn's disease group showing interaction number (left) and interaction strength (right). (**B**) Incoming and outgoing signaling strength per cell type in the normal-Crohn's disease group, with tissue stem cells, astrocytes, monocytes, and endothelial cells emerging as the most active communication hubs. (**C**) Major signaling pathways in the normal-Crohn's disease group. Tissue stem cells serve as the dominant signal senders in the COLLAGEN, LAMININ, and CD99 pathways, primarily directing strong outgoing signals toward neutrophils, monocytes, NK cells, T cells, endothelial cells, and astrocytes. Monocytes act as central senders in the GALECTIN and VISFATIN pathways, targeting a broad range of immune populations including neutrophils, NK cells, T cells, and B cells. Endothelial cells predominantly drive the APP pathway toward monocytes and macrophages. The MIF pathway exhibits complex, multidirectional interactions that integrate B cells, T cells, monocytes, and endothelial cells. (**D**) Overall cell-cell communication network in the normal-ulcerative colitis group showing interaction number (left) and interaction strength (right). (**E**) Incoming and outgoing signaling strength per cell type in the normal-ulcerative colitis group, with tissue stem cells and endothelial cells displaying the highest overall signaling activity. (**F**) Major signaling pathways in the normal-ulcerative colitis group. Tissue stem cells again function as the primary signal senders in the COLLAGEN, LAMININ, and CD99 pathways, directing strong signals to neutrophils, macrophages, T cells, and endothelial cells. Macrophages dominate the GALECTIN pathway toward neutrophils, NK cells, and T cells. NK cells and endothelial cells are the main contributors to the CypA pathway, targeting tissue stem cells and epithelial cells. Endothelial cells serve as the principal receivers in the THBS and IGFBP pathways, receiving strong incoming signals from macrophages, B cells, and tissue stem cells

the COLLAGEN signaling network, tissue stem cells acted as major signal senders, establishing strong interactions with neutrophils, monocytes, NK cells, and T cells (Fig. 5C). This suggests that tissue stem cells might use collagen-mediated signaling pathways to coordinate the recruitment and modification of immune cells. Within the LAMININ signaling network, tissue stem cells similarly played a central role, exhibiting strong outgoing signals toward endothelial cells, neutrophils, and monocytes (Fig. 5C). Through laminin-associated pathways, these

connections suggest that tissue stem cells may play a role in vascular remodeling and innate immune responses. The APP signaling network was characterized by prominent signaling from endothelial cells toward monocytes, macrophages, and even autocrine signaling among endothelial cells themselves (Fig. 5C). Using APP-mediated interaction, this connection suggests that endothelial cells play a crucial role in establishing inflammation and regenerative processes. In the CypA signaling network, monocytes emerged as primary signal initiators, communicating strongly with epithelial cells, tissue stem cells, and endothelial cells (Fig. 5C). Because of this extensive signaling, monocytes may use CypA-driven pathways to control the tissue environment's immunological and structural elements. Tissue stem cells exhibited robust outgoing interaction toward monocytes, endothelial cells, NK cells, and astrocytes within the CD99 signaling network, underscoring their diverse function in regulating the vascular and immunological components of the environment (Fig. 5C). Furthermore, T cells acted as active indicators, exhibiting strong interactions with monocytes, endothelial cells, and tissue stem cells (Fig. 5C). This suggests that T cells dynamically regulate the stromal and immunological compartments through CD99 signaling. The GALECTIN signaling network was dominated by monocytes, which engaged in potent communication with neutrophils, NK cells, T cells, and B cells (Fig. 5C). This comprehensive pattern of interactions suggests that monocytes play a crucial part in regulating both innate and adaptive immune responses via galectin-mediated pathways. In the VISFATIN signaling network, both monocytes and neutrophils acted as major communicators (Fig. 5C). Monocytes primarily targeted endothelial cells and tissue stem cells, while neutrophils exhibited strong signaling toward endothelial cells, tissue stem cells, and monocytes (Fig. 5C). These findings suggest that VISFATIN signaling may serve as a bridge linking innate immune activation with stromal and vascular remodeling. The MIF signaling network revealed complex and multi-directional communication patterns. B cells showed strong interactions with monocytes and T cells, whereas T cells primarily targeted monocytes (Fig. 5C). Moreover, endothelial cells were involved in significant signaling toward monocytes, T cells, and B cells, underscoring the integration of vascular and immune responses via MIF-driven pathways (Fig. 5C). The MIF and CD99 signaling networks in the B cell subpopulation demonstrated robust connections between immature B cells and plasma B cells (Supplementary Figure S5A). Within the T cell subpopulation, the MHC-I signaling network was dominated by strong communication from CD8+ central memory T cells to multiple T cell subsets, including CD4+ central memory T cells, CD4+ effector memory T cells, CD8+ naïve T cells, and gamma-delta T cells (Supplementary Figure S5B). This pattern indicates that CD8+ central memory T cells may exert widespread immunomodulatory influence across diverse T cell populations via antigen presentation pathways. In the MIF signaling network, CD8+ naïve T cells demonstrated significant outgoing communication toward gamma-delta T cells and CD4+ central memory T cells, suggesting a role for CD8+ naïve T cells in modulating adaptive and innate-like T cell responses through inflammatory signaling (Supplementary Figure S5B). Additionally, the CD45 signaling network revealed robust interactions initiated by CD8+ naïve T cells, targeting CD8+ central memory T cells, CD4+ central memory T cells and CD4+ effector memory T cells (Supplementary Figure S5B). In the normal-ulcerative colitis group, cellchat evaluation indicated that tissue stem cell and endothelial cell shown the strongest signals output capability (Fig. 5D). Furthermore, the strongest incoming and outgoing signals were collected by tissue stem cells and endothelial cells (Fig. 5E). COLLAGEN, LAMININ, APP, CypA, CD99, GALECTIN, THBS, and IGFBP are among the eight key signaling pathways that are analyzed in cell-cell communication. These interactions shape the intercellular landscape within the tissue microenvironment (Fig. 5F). In the COLLAGEN signaling network, tissue stem cells acted as major signal initiators, exhibiting strong communication toward neutrophils, macrophages, and T cells (Fig. 5F). Within the LAMININ signaling network, tissue stem cells continued to serve as central communicators, sending strong signals to endothelial cells, neutrophils, macrophages, and T cells, further emphasizing their involvement in vascular support and immune regulation via laminin-associated pathways (Fig. 5F). Communication from endothelial cells to macrophages and NK cells was the predominant characteristic of the APP signaling network, suggesting that endothelial cells play a part in regulating innate immune responses and preserving tissue integrity via APP-driven processes (Fig. 5F). NK cells established important signalers in the CypA signaling network, exhibiting substantial interactions with endothelial, tissue stem, and epithelial cells. Furthermore, endothelial cells demonstrated strong outgoing interactions with tissue stem cells and epithelial cells, underscoring the wide-ranging controlling effect of CypA signaling on both immunological and structural components (Fig. 5F). The CD99 signaling network revealed tissue stem cells as dominant communicators, targeting endothelial cells, T cells, macrophages, and neutrophils (Fig. 5F). Effective interactions between macrophages and T cells, neutrophils, and NK cells were observed in the GALECTIN signaling network, indicating that macrophages participate a key role in coordinating both innate and adaptive immune responses via galectin-driven processes (Fig. 5F). Conversely, the THBS

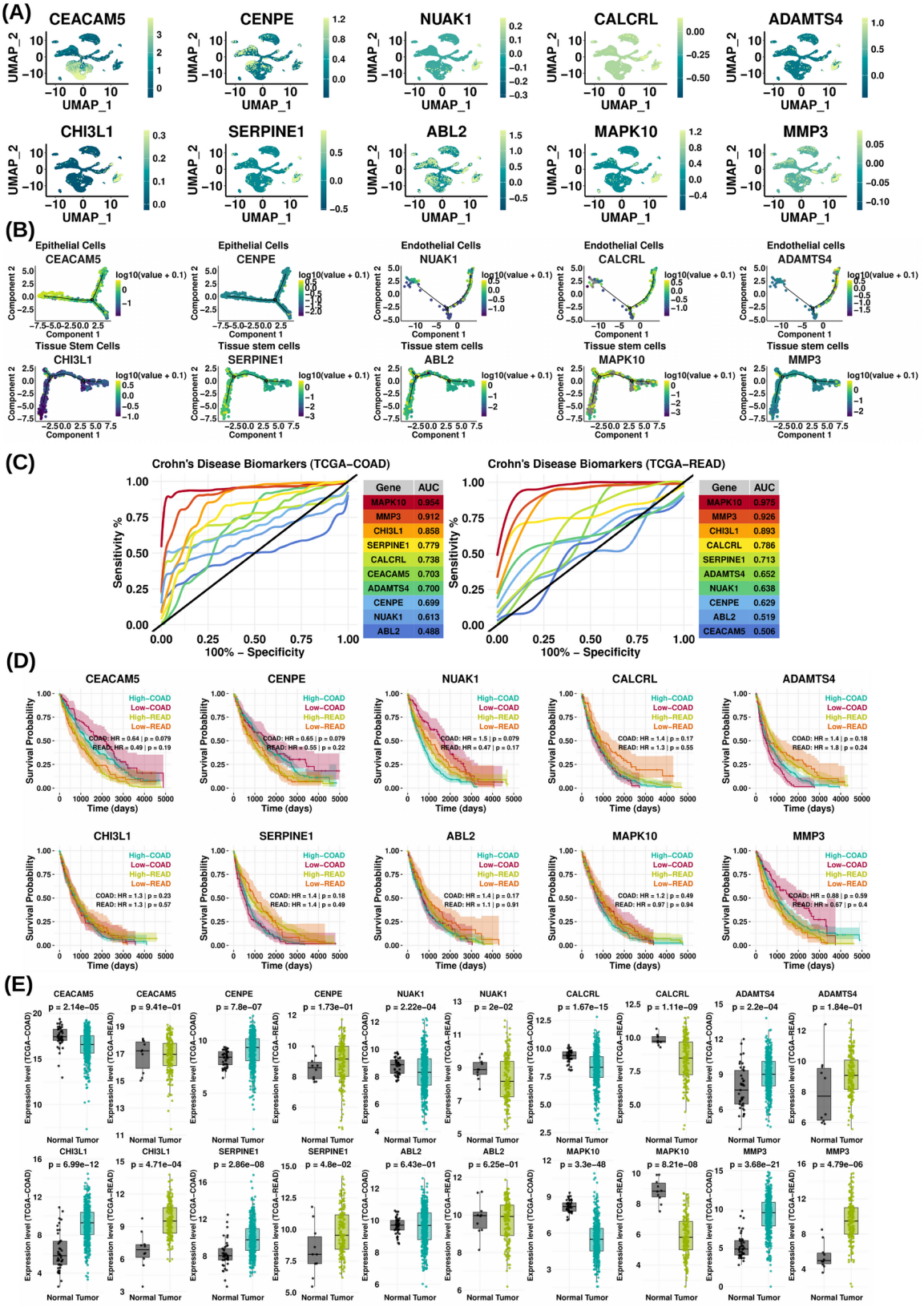

Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Cell type specific expression and translational relevance of Crohn's disease-associated marker genes. (**A**) UMAP projections showing expression patterns of key marker genes in the integrated Crohn's disease dataset. CEACAM5 and CENPE are predominantly expressed in epithelial cells; NUAK1, CHI3L1, SERPINE1, ABL2, MAPK10, and MMP3 are enriched in tissue stem cells; CALCRL and ADAMTS4 are mainly detected in endothelial cells. (**B**) Pseudotime-dependent expression dynamics of marker genes along differentiation trajectories in Crohn's disease. CEACAM5 and NUAK1 increase toward later stages in epithelial and endothelial lineages; CENPE increased in pseudotime trajectories in epithelial cells; CHI3L1, SERPINE1, ABL2, and MMP3 gradually decrease in tissue stem cells; CALCRL and ADAMTS4 progressively rise in endothelial cells, whereas MAPK10 shows a transient mid-trajectory increase in tissue stem cells. (**C**) Diagnostic performance of the identified marker genes in colorectal cancer using TCGA-COAD and TCGA-READ datasets. ROC analysis reveals strong discriminatory power for ileum-associated genes (CHI3L1, SERPINE1, CALCRL, MMP3) and sigmoid colon-derived MAPK10, with moderate performance for CENPE, CEACAM5, NUAK1, and ADAMTS4 ($p < 0.05$). (**D**) Kaplan-Meier overall survival analysis (TCGA-COAD and TCGA-READ) showing no significant prognostic association for any of the Crohn's disease-related marker genes (NUAK1, CENPE, CEACAM5, CHI3L1, CALCRL, SERPINE1, ABL2, MAPK10, ADAMTS4, MMP3). (**E**) Differential expression of the marker genes in tumor compared to normal samples from TCGA-COAD and TCGA-READ. Most genes (CENPE, NUAK1, CALCRL, ADAMTS4, CHI3L1, SERPINE1, ABL2, MMP3) are significantly upregulated in tumor tissue; MAPK10 shows upregulation in colon adenocarcinoma but downregulation in rectal adenocarcinoma, whereas CEACAM5 expression remains largely unchanged in colon adenocarcinoma

signaling network displayed a connection pattern, tissue stem cells received strong incoming signals from endothelial cells, B cells, and macrophages (Fig. 5F). According to this, endothelial cells primarily participated as signal receivers in the IGFBP signaling network, with significant contributions from macrophages and tissue stem cells. This suggests that IGFBP signaling may be essential for regulating the activity of endothelial cells and vascular homeostasis (Fig. 5F). In the B cell subpopulation, the MIF and CD99 signaling networks revealed strong outgoing interactions from memory B cells to naïve B cells, indicating a possible role of memory B cells in regulating the maturation or activation of naïve B cells through inflammatory and adhesion-related pathways (Supplementary Figure S6A). Additionally, within the CypA signaling network, plasma B cells exhibited significant communication directed toward naïve B cells, suggesting a potential mechanism by which differentiated plasma B cells modulate early B cell responses (Supplementary Figure S6A). In the T cell compartment, the MHC-I signaling pathway demonstrated strong communication from gamma-delta T cells toward both CD8+ central memory T cells and CD4+ effector memory T cells, indicating a role for gamma-delta T cells in shaping memory T cell populations through antigen presentation mechanisms (Supplementary Figure S6B). Similarly, in the MIF signaling network, gamma-delta T cells maintained prominent outgoing interactions with CD4+ effector memory T cells, further emphasizing their influence in regulating effector T cell functions during immune responses (Supplementary Figure S6B). The CypA signaling network highlighted CD8+ central memory T cells as major signalers toward gamma-delta T cells, indicating interactions between innate-like gamma-delta T cells and conventional memory T cells (Supplementary Figure S6B). Meanwhile, the CLEC signaling pathway showed strong communication initiated by CD4+ effector memory T cells toward gamma-delta T cells, underscoring the complex cross-talk between adaptive and innate T cell subsets mediated through CLEC-related pathways (Supplementary Figure S6B). Additionally, in the CD99 signaling network, gamma-delta T cells again conduct as main communicators, targeting CD4+ effector memory T cells, highlighting their role in regulating T cell migration and immune activation via adhesion molecules (Supplementary Figure S6B).

Tissue specific gene intersections in Crohn's disease and ulcerative colitis transcriptomics

The intersection analysis of differentially expressed genes between bulk RNA-seq and single-cell RNA-seq across various intestinal tissues in Crohn's disease (CD) and ulcerative colitis (UC) identified. CEACAM5 was identified as a common gene across all tissues and comparisons. In the colon, NUAK1 and CENPE were specific to the normal vs. Crohn's disease group, while LGALS1 and PDGFRA were identified for the normal vs. ulcerative colitis group. In the ileum, CHI3L1, CALCRL, and SERPINE1 were specific to the normal vs. Crohn's disease group. For the ascending colon, ABL2 was identified in the normal vs. Crohn's disease group, and MMP12 in the normal vs. ulcerative colitis group. In the sigmoid colon, MAPK10 was specific to the normal vs. Crohn's disease group, and PDGFRA to the normal vs. ulcerative colitis group. In the terminal ileum, ADAMTS4 and MMP3 were identified for the normal vs. Crohn's disease group, and NAMPT for the normal vs. ulcerative colitis group (Supplementary Figure S7). The total number of biomarkers identified in bulk RNA seq and single cell RNA-seq analyses is provided in the Supplementary Table S12.

Single cell expression and pseudotime trajectories uncover distinct molecular signatures in Crohn's disease

In the Crohn's disease (CD) group, UMAP visualization revealed distinct cell type specific expression patterns. CEACAM5 and NUAK1 were mainly expressed in epithelial and endothelial cells, respectively. CENPE expression was concentrated within epithelial cell clusters, while CHI3L1, SERPINE1, ABL2, MAPK10, and MMP3 were predominantly expressed in tissue stem cells. CALCRL and ADAMTS4 showed higher expression in endothelial cells, consistent with their endothelial

association (Fig. 6A). In Crohn's disease group, the pseudotime trajectories showed an increase in CEACAM5 and NUAK1 expression in epithelial cells and endothelial cells and suggesting a possible activity during the latter phases of cellular development. The flexible distribution of CENPE expression in epithelial cells, which peaked in the middle stage of the pseudotime trajectories and then decreased suggests that it plays a role in the middle phases of epithelial cell proliferation. CHI3L1 expression in tissue stem cells decreased along the pseudotime trajectory, indicating a potential downregulation during stem cell differentiation. During the pseudotime trajectory, CALCRL expression in endothelial cells developed continuously, indicating that it was more active in later phases of endothelial cell development. According to the pseudotime progression, SERPINE1 expression in tissue stem cells amounted at a transition point and then started to reduction, suggesting a temporary involvement underlying stem cell transformation. ABL2 expression in tissue stem cells decreased along the pseudotime trajectory, suggesting a reduction in activity as stem cells progress. MAPK10 expression in tissue stem cells showed a slight increase in the middle of the pseudotime trajectory before stabilizing, indicating a potential role in mid-stage stem cell differentiation. ADAMTS4 expression in endothelial cells increased along the pseudotime trajectory, suggesting its involvement in final point of endothelial cell development. MMP3 expression in tissue stem cells decreased along the pseudotime trajectory, indicating a potential downregulation during stem cell development (Fig. 6B). ROC curve analysis was performed to evaluate the diagnostic potential of Crohn's disease-associated biomarkers in colon and rectal adenocarcinoma tissues using TCGA-COAD and TCGA-READ datasets. The results revealed varying discriminatory power among different gene candidates depending on their intestinal localization. In colon-derived tissues, NUAK1 (colon tissue) exhibited a moderate diagnostic ability (AUC = 0.6129, 95% CI = 0.5557-0.6701, $p = 0.0164$), while CENPE (colon tissue) and CEACAM5 showed higher discrimination values (AUC = 0.6994 and 0.7031, respectively; $p < 0.0001$). Genes associated with ileal tissues, including CHI3L1, CALCRL, and SERPINE1, demonstrated robust diagnostic accuracy in COAD, with AUC values of 0.8575, 0.7385, and 0.7791, respectively ($p < 0.0001$). Similarly, in rectal cancer (READ), CHI3L1 (AUC = 0.8928, $p < 0.0001$) and CALCRL (AUC = 0.7856, $p = 0.0024$) retained strong diagnostic value. Among markers originating from the ascending colon, ABL2 showed poor discrimination (COAD AUC = 0.5117, $p = 0.8043$). Conversely, MAPK10, linked to sigmoid colon tissue, exhibited exceptional accuracy in distinguishing tumor from normal samples, with AUC values of 0.9538 (COAD) and 0.9754 (READ), both highly significant ($p < 0.0001$). In addition, ADAMTS4 and MMP3, derived from terminal ileum, showed considerable diagnostic performance (COAD AUC = 0.6995 and 0.9120, respectively; $p < 0.0001$). MMP3 also demonstrated strong accuracy in READ (AUC = 0.9257, $p < 0.0001$). Overall, the analysis suggested that ileum-associated markers (CHI3L1, CALCRL, SERPINE1, and MMP3) and MAPK10 from sigmoid colon possess high sensitivity and specificity for differentiating tumor and normal samples in both colon and rectal cancers, emphasizing their potential diagnostic relevance in Crohn's disease-related molecular patterns (Fig. 6C). According to the survival analysis based on the TCGA-COAD and TCGA-READ datasets, high expression of NUAK1, CENPE, and CEACAM5 was not considerably related to reduced survival rates of patients with colon and rectal adenocarcinomas. Similarly, elevated expression levels of CHI3L1, CALCRL, and SERPINE1, which are associated with ileum tissues, showed no significant correlation with patients' overall survival. Expression of ABL2 from ascending colon tissue also did not reveal a meaningful relationship with survival outcomes. Furthermore, high expression of MAPK10, ADAMTS4, and MMP3 did not show any considerable association with overall survival in either colon or rectal cancer cases (Fig. 6D). Overall, these findings suggest that none of the Crohn's disease-related biomarkers exhibited a statistically significant relationship with patient prognosis in colon or rectal adenocarcinomas. We analyzed the expression of Crohn's disease-related genes in tumor samples from TCGA-COAD and TCGA-READ datasets. Overall, most of these genes, including CENPE, NUAK1, CALCRL, ADAMTS4, CHI3L1, SERPINE1, ABL2, and MMP3, showed higher expression in tumor tissues compared to normal controls. Interestingly, MAPK10 displayed different patterns in different tissues, being upregulated in colon adenocarcinoma but downregulated in rectal adenocarcinoma, while CEACAM5 showed little change in colon adenocarcinoma (Fig. 6E). These results highlight tissue specific variations in the expression of Crohn's disease-associated genes in tumors.

Single cell expression and pseudotime trajectories highlight distinct molecular signatures in ulcerative colitis

In the ulcerative colitis (UC) group, CEACAM5 and LGALS1 were enriched in both epithelial cells and tissue stem cells. PTPRB expression was mainly localized in endothelial cells, whereas MMP12 and NAMPT were highly expressed in macrophages. PDGFRA expression was confined to tissue stem cell clusters. Together, these UMAP patterns highlight distinct gene expression profiles across epithelial, endothelial, stem, and immune cell populations in Crohn's disease and ulcerative colitis (Fig. 7A). In the ulcerative colitis group,

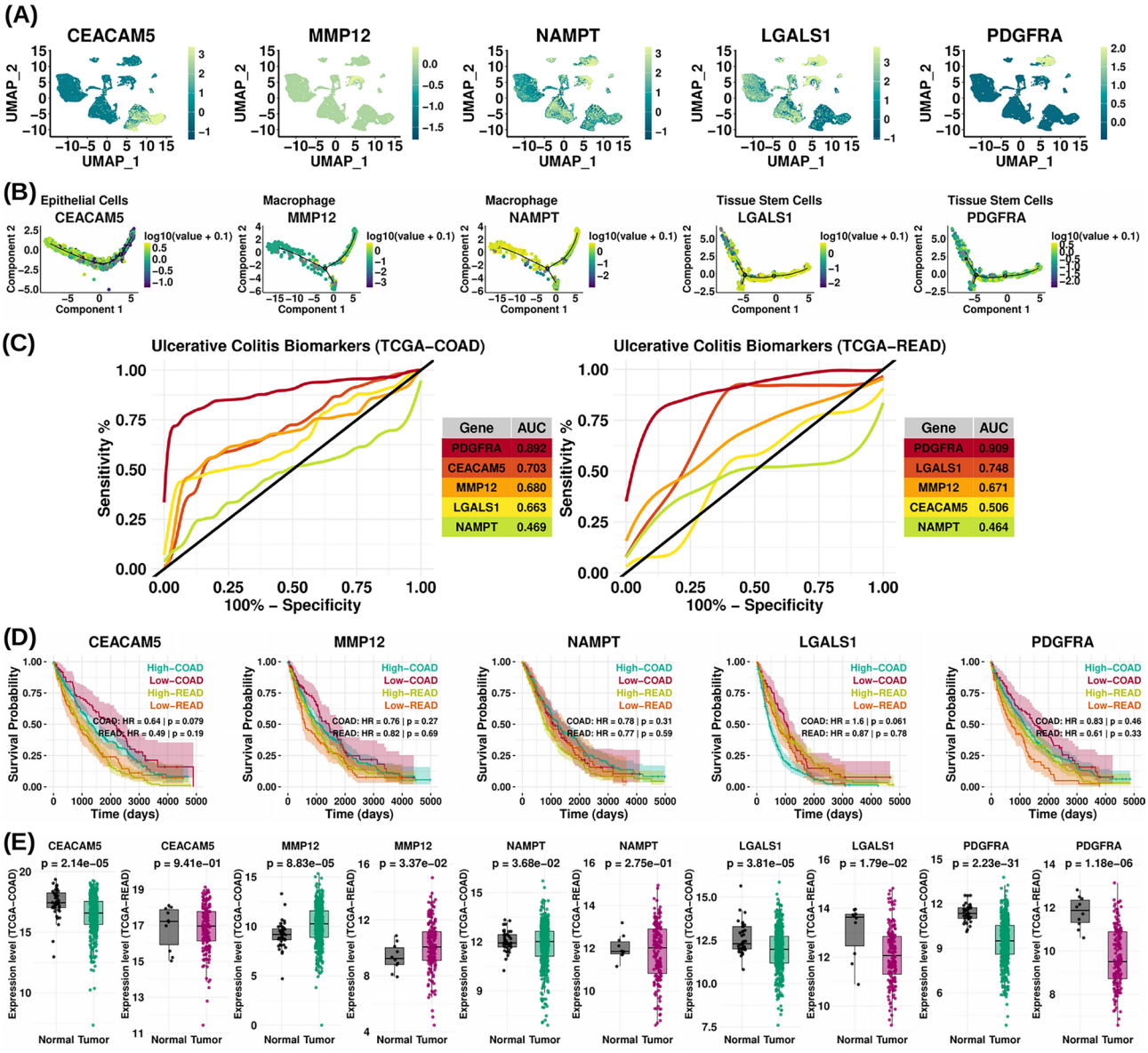

Fig. 7 Cell type specific expression and translational relevance of ulcerative colitis-associated marker genes. (**A**) UMAP projections showing expression patterns of key marker genes in the integrated ulcerative colitis dataset. CEACAM5 and LGALS1 are enriched in both epithelial cells and tissue stem cells; PTPRB is predominantly expressed in endothelial cells; MMP12 and NAMPT are highly expressed in macrophages; PDGFRA is largely confined to tissue stem cell clusters. (**B**) Pseudotime-dependent expression dynamics along differentiation trajectories in ulcerative colitis. CEACAM5 and LGALS1 progressively increase toward late pseudotime in epithelial and tissue stem cell lineages; PDGFRA shows strong upregulation at terminal stages of tissue stem cell differentiation; MMP12 and NAMPT rise steadily in macrophages, peaking at the end of the trajectory, indicating roles in late-stage macrophage maturation. (**C**) Diagnostic performance of the identified marker genes in colorectal cancer (TCGA-COAD and TCGA-READ). ROC analysis demonstrates excellent discriminatory power for PDGFRA, moderate to good performance for LGALS1 and MMP12, while NAMPT shows no significant diagnostic value. (**D**) Kaplan-Meier overall survival analysis revealing no statistically significant prognostic association for any of the ulcerative colitis related marker genes (LGALS1, PDGFRA, MMP12, NAMPT, CEACAM5). (**E**) Differential expression of the marker genes in tumor compared to normal samples from TCGA-COAD and TCGA-READ. MMP12 and NAMPT are significantly upregulated in tumor tissue; PDGFRA is upregulated in colon adenocarcinoma only; LGALS1 and CEACAM5 show higher expression selectively in rectal adenocarcinoma, highlighting tissue-specific expression patterns in colorectal cancer

throughout the pseudotime trajectory, CEACAM5 and LGALS1 expression in epithelial cells and tissue stem cells increased and amounted near the end, indicating a potential involvement in later phases of stem cell development. During the pseudotime trajectory, MMP12 expression in macrophages elevated indicating that it was active in later phases of macrophage development. In macrophages, NAMPT expression appeared according to the pseudotime trajectory and amounted at the final stage, suggesting a possible function in the latter phases of macrophage maturation. PDGFRA expression in tissue stem cells increased along the pseudotime

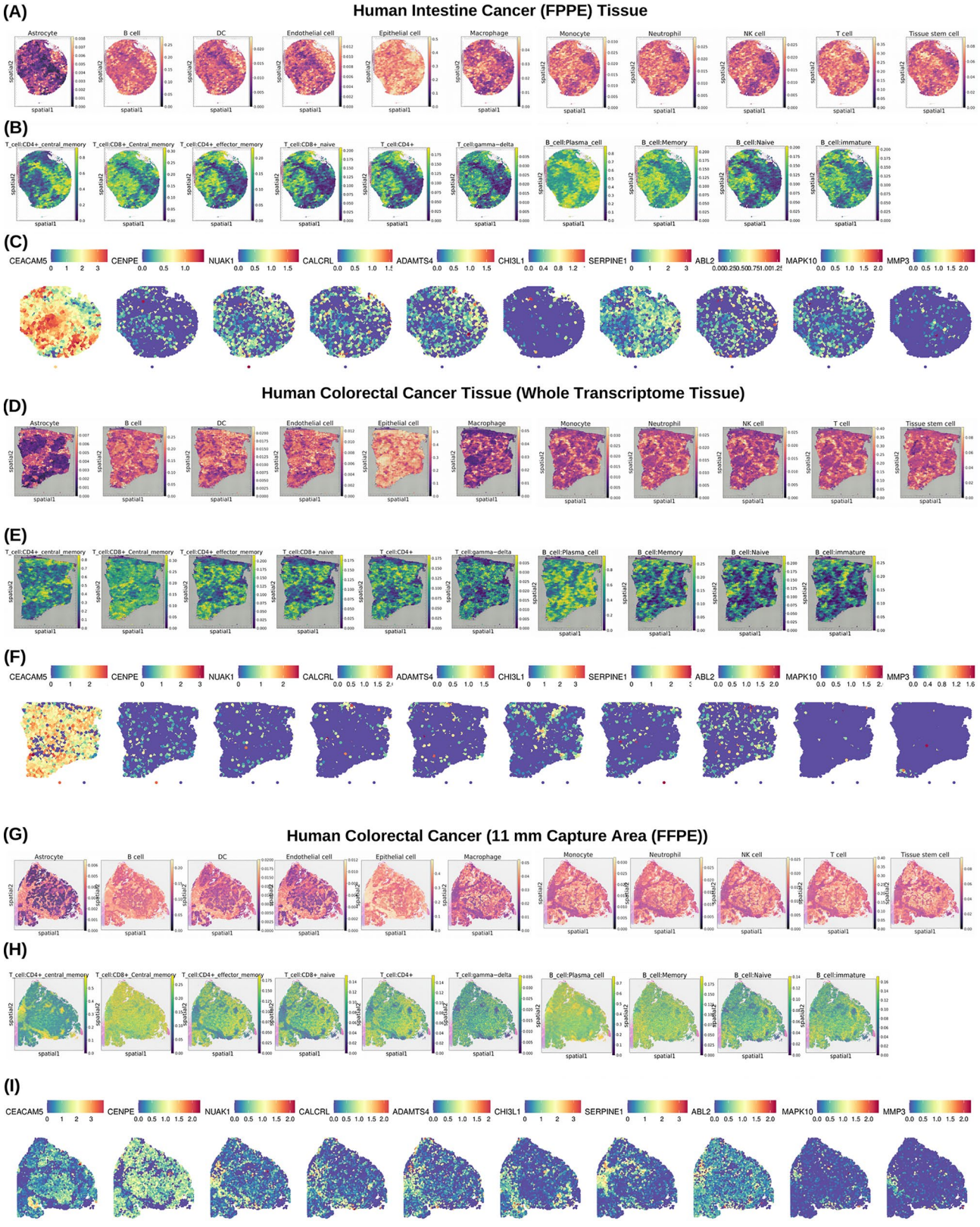

Fig. 8 (See legend on next page.)

(See figure on previous page.)

Fig. 8 Spatial transcriptomic analysis of human colorectal and intestinal cancer tissues (FFPE) showing compartmentalized distribution of immune and stromal cell populations together with expression patterns of Crohn's disease associated marker genes. In human intestinal cancer (FFPE) tissue, (**A**) epithelial cells, monocytes, T cells, neutrophils, and B cells are the most abundant populations and predominantly concentrated in the central region, while DCs, endothelial cells, macrophages, NK cells, and tissue stem cells accumulate in the central and upper right areas; astrocytes remain rare and sparsely localized to the upper right zone. (**B**) Among T cell subsets, CD8+ central memory and CD4+ effector memory T cells dominate the central and upper left regions, whereas gamma-delta T cells are present at low levels and restricted to the upper left area and B cell subpopulations show strong enrichment of plasma B cells in the upper right region and memory B cells in the upper left region. (**C**) Key marker genes CEACAM5, NUAK1, CALCRL, SERPINE1, and ADAMTS4 exhibit broad and high expression; ABL2 and MAPK10 display moderate expression in central and upper left zones; CENPE, CHI3L1, and MMP3 are expressed at low levels or regionally restricted. In human colorectal cancer whole transcriptome FFPE tissue, (**D**) B cells, dendritic cells, endothelial cells, and epithelial cells are highly abundant and tightly clustered in the central region, followed by tissue stem cells, T cells, neutrophils, monocytes, and NK cells; macrophages and astrocytes are present only at low abundance. (**E**) CD8+ central memory T cells are strongly enriched centrally, with other T cell subsets showing moderate distribution in central and upper right areas and plasma B cells are highly abundant in the central region, accompanied by moderate levels of memory, naïve, and immature B cells. (**F**) Among marker genes, CEACAM5 shows high central expression; NUAK1, CENPE, CALCRL, SERPINE1, and ADAMTS4 are generally low and central; CHI3L1 and ABL2 retain moderate central levels, whereas MAPK10 and MMP3 are very low. In human colorectal cancer 11 mm capture area FFPE tissue, (**G**) epithelial cells, B cells, tissue stem cells, T cells, NK cells, neutrophils, and monocytes are highly enriched in the central and upper left regions, while DCs and endothelial cells show moderate upper left localization. (**H**) CD8+ central memory T cells are widely abundant across the tissue, with CD4+ effector memory and CD8+ naïve T cells concentrated in central and upper right zones and plasma and memory B cells are highly abundant throughout the tissue, whereas naïve and immature B cells are moderate and restricted to the upper right area. (**I**) Crohn's disease related markers CEACAM5 and CENPE are uniformly high; NUAK1, SERPINE1, ABL2, and ADAMTS4 show moderate expression mainly in left and central areas; CHI3L1 and CALCRL are low and peripheral; MAPK10 and MMP3 are very low and predominantly left sided

trajectory and suggesting its involvement in the later stages of stem cell differentiation (Fig. 7B). Similarly, ROC curve analysis of ulcerative colitis-related biomarkers using TCGA-COAD and TCGA-READ data revealed that several genes have promising diagnostic utility. For colon-derived markers, LGALS1 showed moderate discrimination power in COAD (AUC = 0.6632, 95% CI = 0.5942-0.7321, $p = 0.0005$) and slightly higher performance in READ (AUC = 0.7479, $p = 0.0085$). PDGFRA exhibited excellent diagnostic accuracy in both datasets (COAD AUC = 0.8922; READ AUC = 0.9087; both $p < 0.0001$), suggesting its potential as a reliable biomarker in UC-associated colorectal neoplasia. Among ascending colon-associated genes, MMP12 demonstrated fair accuracy (COAD AUC = 0.6797, $p = 0.0001$), while its performance in READ was moderate (AUC = 0.6713, $p = 0.0692$). However, NAMPT, derived from terminal ileum tissue, did not exhibit significant diagnostic capability (COAD AUC = 0.5307, $p = 0.5146$; READ AUC = 0.5359, $p = 0.7030$) (Fig. 7C). Collectively, these findings indicate that PDGFRA and LGALS1 are the most promising UC-associated biomarkers for distinguishing colorectal cancer samples from normal tissues, particularly within the TCGA-COAD and TCGA-READ datasets. Survival investigation demonstrated that high expression of LGALS1 was not considerably related to reduced survival rates of individuals with colon and rectal adenocarcinomas, although a slight decreasing trend in survival was observed in colon cancer cases. Similarly, expression levels of PDGFRA, MMP12, and NAMPT showed no significant relationship with overall survival in either TCGA-COAD or TCGA-READ datasets (Fig. 7D). Altogether, these results indicate that ulcerative colitis-associated biomarkers were not significantly correlated with survival outcomes in colon and rectal cancers. We analyzed the expression of ulcerative colitis-related genes in tumor samples from TCGA-COAD and TCGA-READ datasets. MMP12 and NAMPT showed higher expression in tumor tissues compared to normal controls. PDGFRA was upregulated in colon adenocarcinoma but showed no change in rectal adenocarcinoma, while LGALS1 displayed little change in colon adenocarcinoma but higher expression in rectal adenocarcinoma. CEACAM5 showed minimal change in colon adenocarcinoma but higher expression in rectal adenocarcinoma (Fig. 7E). These results indicate tissue specific differences in the expression of these ulcerative colitis-associated genes in tumors.

Integrative spatial transcriptomics and deconvolution to uncover cellular heterogeneity

We performed spatial transcriptomics analysis using 10x Genomics datasets on three tissue samples from patients with colorectal cancer such as: Human Intestine Cancer (FPPE) tissue, Human Colorectal Cancer tissue (Whole Transcriptome Tissue), Human Colorectal Cancer (11 mm Capture Area (FFPE)) to evaluate the composition of immune cell types, B cell and T cell subtypes in Crohn's disease and ulcerative colitis.

Spatial profiling of Crohn's disease associated immune cells within colorectal cancer regions

In the Human Intestine Cancer (FPPE) tissue, epithelial cells, monocytes, T cells, neutrophils and B cells demonstrated high abundance and predominantly localized within the central regions of the intestinal cancer tissue. In the tumor microenvironment, this significant elevation in the tissue center indicates being conducted immunological and epithelial activation (Fig. 8A). Significantly abundant cells, such as DCs, endothelial cells, macrophages, NK cells and tissue stem cells, were mostly found in the central and upper right regions, suggesting

that these spots could function as concentrates for immune activation. In contrast, astrocytes displayed a low abundance and were sparsely localized to the upper right regions of intestinal cancer tissue (Fig. 8A). Regarding T cell subtypes, CD8+ central memory T cells and CD4+ effector memory T cells were highly abundant and especially concentrated in the central and upper left parts of the intestinal cancer tissue. CD4+ central memory T cells, CD8+ naïve T cells and CD4+ T cells were moderately abundant and mainly distributed in the upper right and upper left regions. Gamma-delta T cells were present at low abundance and primarily localized to the upper left regions (Fig. 8B). Further analysis of B cell subtypes demonstrated that plasma B cells were highly enriched in the upper right regions, whereas memory B cells predominantly localized to the upper left regions. Immature B cells showed moderate abundance within the upper left areas, while naïve B cells were detected at low levels in the same regions (Fig. 8B). Human Intestine Cancer (FPPE) tissue demonstrated a marked upregulation of several genes. CEACAM5, NUAK1, CALCRL, SERPINE1 and ADAMTS4 showed high expression levels across broad regions of the tissue (Fig. 8C). ABL2 and MAPK10 showed moderate expression in central and upper left areas, while CENPE displayed reduced expression and limited to central regions of tissue (Fig. 8C). CHI3L1 and MMP3 indicated very low expression in the central regions of tissue (Fig. 8C). In the Human Colorectal Cancer tissue (Whole Transcriptome Tissue), B cells, dendritic cells (DCs), endothelial cells and epithelial cells demonstrated high abundance and all primarily concentrated within the central regions of colorectal cancer tissue (Fig. 8D). Cells with significant proportions included: tissue stem cells, T cells, neutrophils, monocytes and NK cells, similarly localized within the central regions. Conversely, astrocytes and macrophages were present at low abundance, also within the central of colorectal cancer tissue (Fig. 8D). Regarding T cell subtypes, CD8+ central memory T cells were highly enriched in the central regions of colorectal cancer tissue (Fig. 8E). CD4+ central memory T cells, CD4+ effector memory T cells, CD8+ naïve T cells and CD4+ T cells demonstrated moderate abundance and concentrated predominantly in the central and partially upper right regions of colorectal cancer tissue (Fig. 8E). Gamma-delta T cells were detected at low abundance and localized within the central regions of colorectal cancer tissue (Fig. 8E). Analysis of B cell subtypes revealed a high abundance of plasma B cells in central of colorectal cancer tissue, while memory B cells exhibited moderate abundance across both the central and upper left regions. Naïve B cells and immature B cells were significantly proportion and primarily distributed within the central regions (Fig. 8E). In the Human Colorectal Cancer tissue (Whole Transcriptome Tissue), overall gene expression levels were comparatively lower. CEACAM5 gene demonstrated high expression and enriched to the central regions (Fig. 8F). Most genes including NUAK1, CENPE, CALCRL, SERPINE1, and ADAMTS4 demonstrated low expression and localized to central regions (Fig. 8F). CHI3L1 and ABL2 maintained significant expression in central of the tissue (Fig. 8F). MAPK10 and MMP3 showed very low expression and enriched in the central of the tissue (Fig. 8F). In the Human Colorectal Cancer (11 mm Capture Area (FFPE)), epithelial cells, B cells, tissue stem cells, T cells, NK cells, neutrophils and monocytes were present at high abundance and predominantly enriched in the central and upper left regions of the tissue (Fig. 8G). A moderate abundance was observed for dendritic cells (DCs) and endothelial cells (both enriched in the upper left region), as well as macrophages and astrocytes located within the central and upper left region, respectively (Fig. 8G). Regarding T cell subtypes, the analysis demonstrated high levels of CD8+ central memory T cells across the tissue (Fig. 8H). Additionally, CD4+ T cells, CD4+ effector memory T cells and CD8+ naïve T cells showed enriched presence in the central and upper right regions (Fig. 8H). Gamma-delta T cells and CD4+ central memory T cells were detected at moderate abundance and primarily localized in the upper left regions (Fig. 8H). Within the B cell subpopulations, both plasma B cells and memory B cells exhibited high abundance throughout the tissue (Fig. 8H). Immature B cells and naïve B cells were found at moderate levels, with a specific spatial restriction to the upper right region (Fig. 8H). For spatial genes expression related to the Crohn's disease group in the Human Colorectal Cancer (11 mm Capture Area (FFPE)), CEACAM5 and CENPE showed high expression uniformly across the tissue. NUAK1, SERPINE1, ABL2 and ADAMTS4 demonstrated moderate expression and localization primarily in the left and central regions of the tissue (Fig. 8I). CHI3L1 and CALCRL exhibited low expression and particularly confined to the left and right at the bottom of the tissue (Fig. 8I). Notably, MAPK10 and MMP3 showed very low expression and enriched in left regions of the tissue (Fig. 8I).

Spatial profiling of ulcerative colitis associated immune cells within colorectal cancer regions

In the Human Intestine Cancer (FPPE) tissue, B cells and epithelial cells exhibited high abundance and predominantly distributed across the central and upper left regions of the tissue. Additionally, tissue stem cells showed significant enrichment within the upper right regions (Fig. 9A). Moderate abundance was observed for T cells, macrophages, neutrophils, hepatocytes and endothelial cells, primarily localized in the upper right regions of the tissue (Fig. 9A). NK cells were detected at

low abundance and defined to the upper right regions (Fig. 9A). For T cell subtypes, CD8+ central memory T cells and CD8+ naïve T cells showed high abundance and localized respectively to the upper left and upper right areas of the intestinal cancer tissue (Fig. 9B). CD4+ effector memory T cells and gamma-delta T cells exhibited moderate abundance (Fig. 9B). Additionally, CD4+ effector memory T cells were mainly found in the upper right region, while gamma-delta T cells were present in the upper left (Fig. 9B). Analysis of B cell subtypes revealed that plasma B cells were highly abundant and particularly enriched in the central and upper right regions of intestinal cancer tissue, while memory B cells, naïve B cells and immature B cells exhibited significant abundance (Fig. 9B). Memory B cells and naïve B cells were predominantly localized in the upper right regions, whereas immature B cells were more common in the upper left regions (Fig. 9B). In the Human Intestine Cancer (FPPE) tissue, CEACAM5 and LGALS1 maintained a very high expression level across the entire tissue (Fig. 9C). NAMPT also showed strong expression throughout the tissue (Fig. 9C). PDGFRA were expressed at moderate levels and enriched in the whole tissue (Fig. 9C). MMP12 was expressed at very low levels and limited to the central region of the tissue (Fig. 9C). In the Human Colorectal Cancer tissue (Whole Transcriptome Tissue), epithelial cells and B cells demonstrated high abundance and predominantly localized in the upper left and upper right regions of the colorectal cancer tissue (Fig. 9D). Moderate abundance was identified for T cells, tissue stem cells, macrophages, endothelial cells, hepatocytes and neutrophils and mainly distributed within the central regions of the colorectal cancer tissue (Fig. 9D). NK cells were detected at low abundance and confined to the central regions of colorectal cancer tissue (Fig. 9D). For T cell subtypes, CD8+ central memory T cells were highly abundant along the left and right sides of the colorectal cancer tissue (Fig. 9E). CD4+ effector memory T cells and CD8+ naïve T cells showed significant abundance within the central regions of the colorectal cancer tissue and gamma-delta T cells were detected at low abundance across the central and upper regions (Fig. 9E). Additional analysis of B cell subtypes showed that plasma B cells were highly enriched and distributed across the central and upper regions of the colorectal cancer tissue (Fig. 9E). Memory B cells also demonstrated high abundance and localized predominantly in the upper left and upper right regions of the colorectal cancer tissue (Fig. 9E). Naïve B cells exhibited moderate abundance within the upper left and right regions of the tissue, whereas immature B cells were detected at low abundance and mainly in the upper right region (Fig. 9E). In the Human Colorectal Cancer tissue (Whole Transcriptome Tissue), a general decrease in gene expression was detected. CEACAM5 and LGALS1 remained highly expressed and restricted to central regions (Fig. 9F). NAMPT displayed moderate expression throughout the tissue, while PDGFRA, PTPRB and MMP12 showed low expression levels in the whole tissue (Fig. 9F). In the Human Colorectal Cancer (11 mm Capture Area (FFPE)), high-abundance cell types included B cells and epithelial cells, localized primarily in the central and upper right regions of the tissue (Fig. 9G). Additionally, T cells, tissue stem cells, NK cells and macrophages were enriched in the central and upper left regions (Fig. 9G). Neutrophils, hepatocytes and endothelial cells were found at moderate abundance and distributed mainly throughout the central and upper left regions (Fig. 9G). Within the T cell subpopulations, the most prominent subtypes such as: CD8+ central memory T cells, CD4+ effector memory T cells and CD8+ naïve T cells were uniformly highly abundant across the tissue (Fig. 9H). The gamma-delta T cells were the only T cell subtype observed at low abundance and although were dispersed throughout the tissue (Fig. 9H). Among the B cell subtypes, plasma B cells and memory B cells showed high abundance across the entire tissue (Fig. 9H). Naïve B cells were observed at moderate levels in the upper left region, while immature B cells were detected at low abundance throughout the tissue (Fig. 9H). For spatial gene expression associated to the ulcerative colitis group in the Human Colorectal Cancer (11 mm Capture Area (FFPE)), CEACAM5 and LGALS1 revealed very high expression specifically in the central and upper left regions of the tissue (Fig. 9I). MMP12 exhibited high expression within the central region of the tissue and NAMPT and PDGFRA were significantly expressed with spatial enrichment in the upper right and upper left regions, respectively and finally, PTPRB showed uniformly very low expression in the whole tissue (Fig. 9I).

Spatial autocorrelation mapping of gene expression and cell type distributions in human intestinal tissues

To investigate local gene expression aligns with immune cell distribution for Crohn's disease, we examined spatial co-occurrence patterns across colorectal tissues. In the Human Intestine Cancer (FPPE) dataset, CEACAM5 showed spatial association with CD4+ effector memory T cells regions ($p = 0.03$), while CHI3L1 exhibited extremely strong colocalization with tissue stem cells ($p = 3.93E-16$). ABL2 aligned with both T cells ($p = 0.001$) and DCs ($p = 1.97E-06$), and MAPK10 demonstrated a pronounced spatial relationship with CD4+ central memory T cells populations ($p = 4.93E-09$). Comparable associations were observed in the Human Colorectal Cancer-Whole Transcriptome dataset, where CEACAM5 localized to CD4+ effector memory T cells areas ($p = 0.01$), and CHI3L1 remained strongly enriched in tissue stem cell niches ($p = 3.93E-16$). ABL2 was associated with T

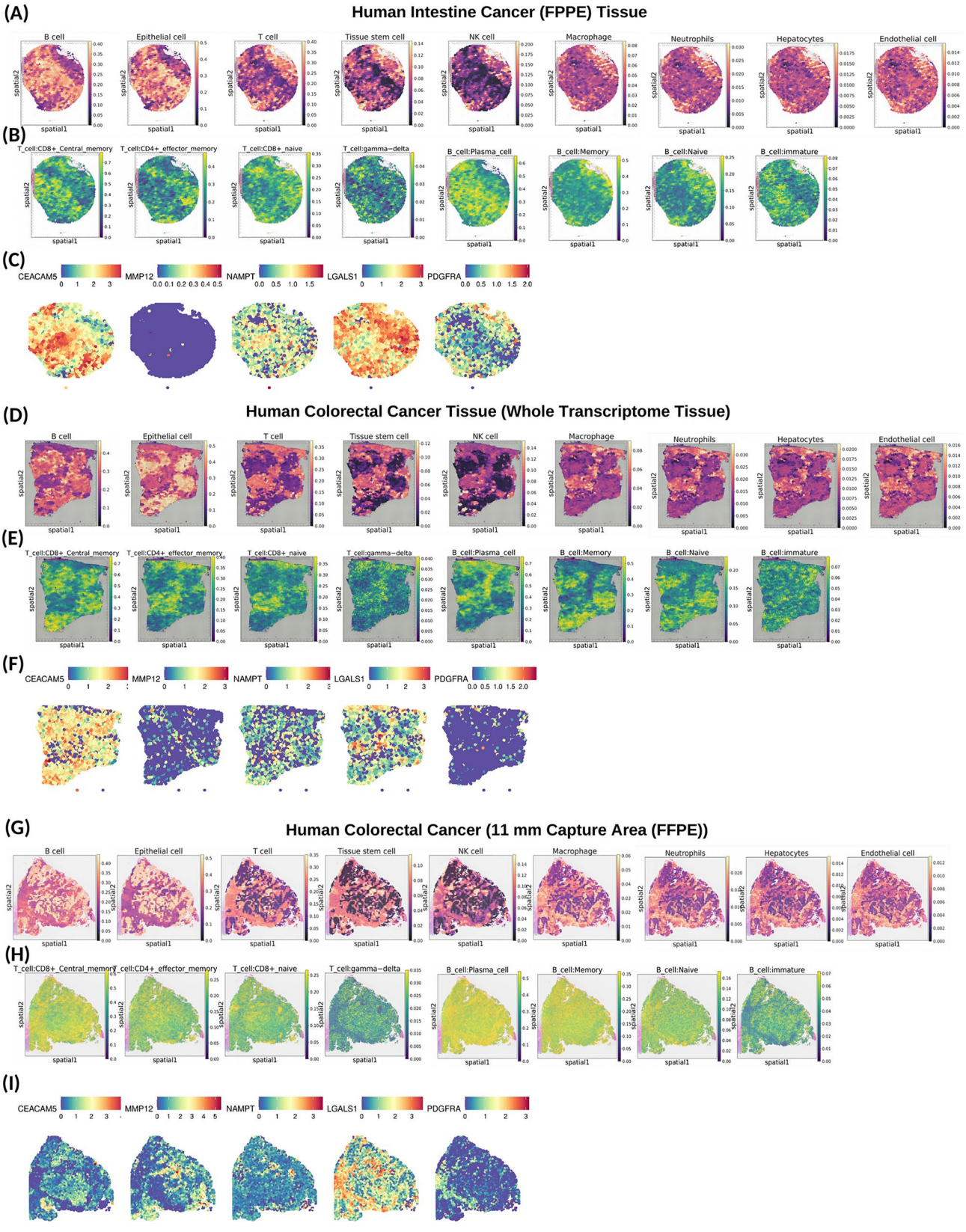

Fig. 9 (See legend on next page.)

(See figure on previous page.)

Fig. 9 Spatial transcriptomic analysis of human intestinal and colorectal cancer tissues (FFPE) reveals regional enrichment of immune and stromal populations together with expression patterns of ulcerative colitis associated marker genes. In human intestinal cancer (FFPE) tissue, (**A**) B cells and epithelial cells are the most abundant populations and predominantly distributed across the central and upper left regions, while tissue stem cells are strongly enriched in the upper right region; T cells, macrophages, neutrophils, hepatocytes, and endothelial cells show moderate abundance mainly in the upper right zone, and NK cells are present at low levels in the upper right. (**B**) Among T cell subsets, CD8+ central memory and CD8+ naïve T cells are highly abundant in upper left and upper right areas, respectively, while CD4+ effector memory and gamma-delta T cells are moderately represented and plasma B cells are highly enriched in central and upper right regions; memory and naïve B cells are prominent in the upper right, and immature B cells are more common in the upper left. (**C**) Marker genes CEACAM5 and LGALS1 exhibit very high expression across the entire tissue; NAMPT is strongly expressed throughout the tissue; PDGFRA shows moderate uniform expression; MMP12 is very low and restricted to the central region. In human colorectal cancer whole transcriptome FFPE tissue, (**D**) epithelial cells and B cells are highly abundant in upper left and upper right regions; T cells, tissue stem cells, macrophages, endothelial cells, hepatocytes, and neutrophils are moderately abundant in central areas, whereas NK cells are low and confined centrally. (**E**) CD8+ central memory T cells are highly abundant along left and right sides; CD4+ effector memory and CD8+ naïve T cells are prominent centrally; gamma-delta T cells are low and plasma and memory B cells are highly enriched centrally and in upper regions; naïve B cells are moderate in upper left and right zones; immature B cells are low in upper right. (**F**) Overall gene expression is reduced compared to intestinal cancer; CEACAM5 and LGALS1 remain highly expressed; NAMPT is moderate expression throughout the tissue; PDGFRA, PTPRB, and MMP12 are low expression. In human colorectal cancer 11 mm capture area FFPE tissue, (**G**) B cells and epithelial cells are highly abundant in central and upper right regions; T cells, tissue stem cells, NK cells, and macrophages are enriched in central and upper left; neutrophils, hepatocytes, and endothelial cells are moderate in central and upper left zones. (**H**) CD8+ central memory, CD4+ effector memory, and CD8+ naïve T cells are uniformly highly abundant across the tissue; gamma-delta T cells are the only low abundance T cell subset and plasma and memory B cells are highly abundant throughout the tissue; naïve B cells are moderate localization in upper left; immature B cells are low. (**I**) Ulcerative colitis associated markers CEACAM5 and LGALS1 show very high expression in central and upper left regions; MMP12 is high centrally; NAMPT is significantly expressed in upper right; PDGFRA is prominent in upper left; PTPRB remains uniformly very low

cells ($p = 0.0006$) and DC ($p = 1.97E-06$), while MAPK10 showed highly significant spatial clustering within CD4+ central memory T cells regions ($p = 3.76E-31$). In the Human Colorectal Cancer-11 mm Capture Area (FFPE) dataset, CENPE expression colocalized with tissue stem cells ($p = 0.002$) and naïve B cells ($p = 0.02$). ADAMTS4 aligned with tissue stem cells ($p = 0.001$) and memory B cells ($p = 0.005$). SERPINE1 ($p = 0.03$) and ABL2 ($p = 0.03$) both associated with monocytes, and ABL2 was additionally related to CD4+ effector memory T cells regions ($p = 0.04$). MAPK10 demonstrated extremely strong non-random spatial clustering with epithelial cells ($p = 1.37E-56$), as well as associations with tissue stem cells ($p = 0.04$) and plasma B cells ($p = 0.004$). MMP3 showed a pronounced spatial relationship with CD4+ effector memory T cells ($p = 1.17E-11$) (Supplementary Table S13). To assess spatial interactions between gene expression and immune cell distribution in ulcerative colitis, we evaluated multiple tissue datasets. In the Human Intestine Cancer (FPPE) dataset, NAMPT showed strong spatial association with macrophages ($p = 0.0001$) and with plasma B cells ($p = 3.05E-07$). LGALS1 demonstrated extensive colocalization across several compartments, including hepatocytes ($p = 0.001$), CD4+ effector memory T cells ($p = 1.80E-14$), CD8+ central memory T cells ($p = 8.48E-05$), memory B cells ($p = 2.87E-12$), naïve B cells ($p = 0.0003$), and plasma B cells ($p = 1.86E-20$). In the Human Colorectal Cancer-Whole Transcriptome dataset, NAMPT localized to macrophages ($p = 0.0002$), memory B cells ($p = 0.02$), and plasma B cells ($p = 1.27E-08$). LGALS1 was associated with B cells ($p = 0.04$), T cells ($p = 0.03$), hepatocytes ($p = 0.001$), CD4+ effector memory T cells ($p = 3.85E-14$), CD8+ central memory T cells ($p = 5.03E-05$), memory B cells ($p = 5.30E-13$), naïve B cells ($p = 0.0002$), and plasma B cells ($p = 1.49E-20$). In the Human Colorectal Cancer-11 mm Capture Area (FFPE) dataset, CEACAM5 was enriched in epithelial cells ($p = 0.05$) and CD8+ naïve T cells regions ($p = 0.03$). MMP12 showed spatial clustering with CD4+ effector memory T cells ($p = 0.0009$). NAMPT displayed spatial distribution across NK cells ($p = 0.002$), tissue stem cells ($p = 3.23E-13$), macrophages ($p = 8.77E-06$), gamma-delta T cells ($p = 0.01$), CD4+ effector memory T cells ($p = 6.25E-06$), memory B cells ($p = 0.0008$), naïve B cells ($p = 0.008$), and plasma B cells ($p = 8.21E-06$). LGALS1 showed highly significant associations with T cells ($p = 1.99E-19$), neutrophils ($p = 0.008$), CD8+ naïve T cells ($p = 0.03$), CD4+ effector memory T cells ($p = 1.24E-22$), CD8+ central memory T cells ($p = 2.01E-08$), memory B cells ($p = 9.77E-07$), naïve B cells ($p = 0.0002$), and plasma B cells ($p = 2.66E-10$) (Supplementary Table S13). Overall, spatial correlation evaluation demonstrated that genes in Crohn's disease including (CEACAM5, CHI3L1, ABL2, MAPK10, CENPE, ADAMTS4, SERPINE1, MMP3) and ulcerative colitis such as (NAMPT, LGALS1, CEACAM5, MMP12) consistently colocalize with major immune cell populations such as T cell subsets, B cell subsets, macrophages, monocytes, dendritic cells, neutrophils, NK cells, epithelial cells, and tissue stem cells. These patterns demonstrate an integrated interaction between immune infiltration and gene expression in several regions. Collectively, our findings highlight that the inflammatory development of Crohn's disease and ulcerative colitis can be significantly affected by the local microenvironmental context. The integrated summary of gene-level evidence across these analyses is provided in Supplementary Table S14.

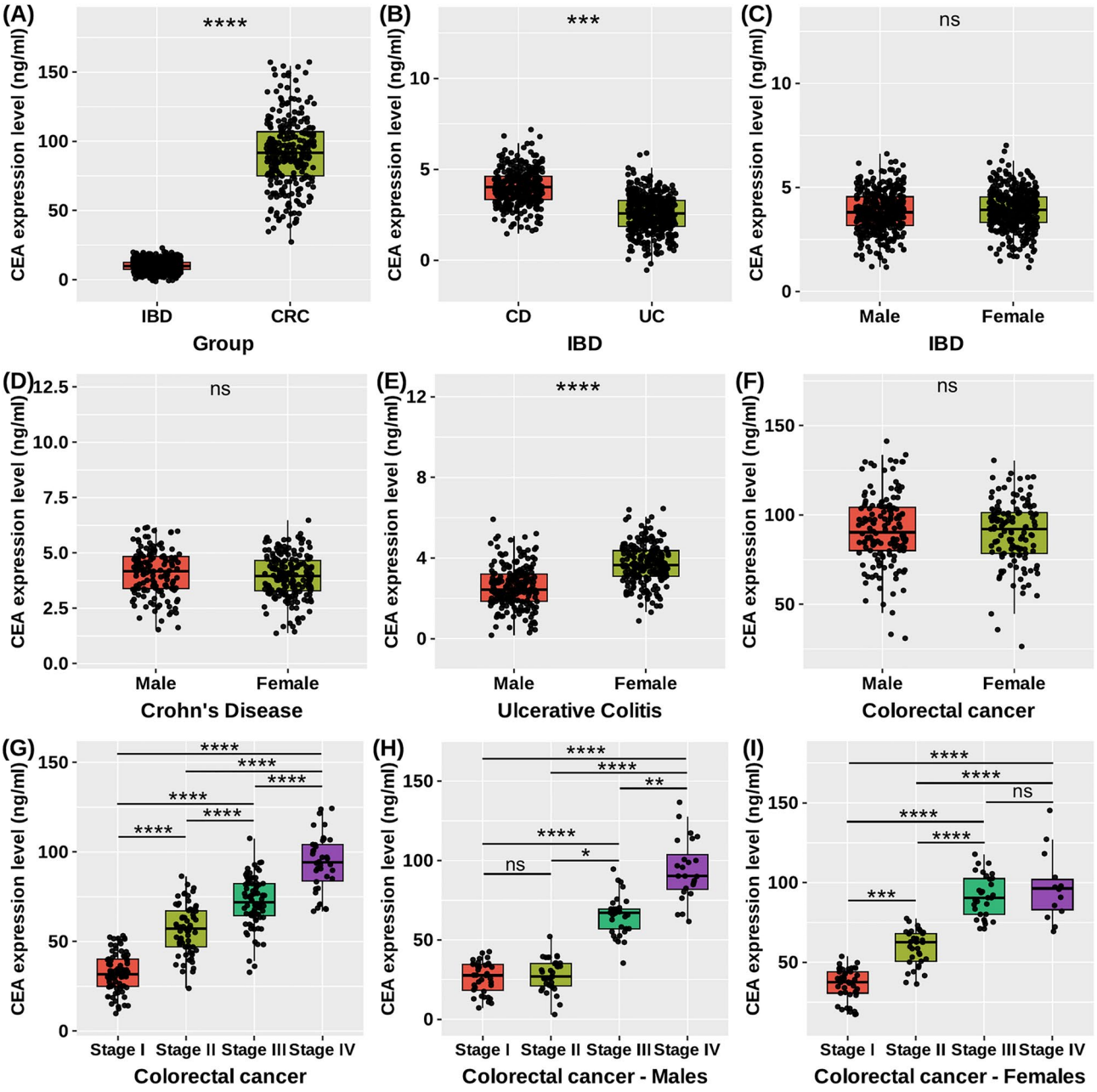

Fig. 10 Serum carcinoembryonic antigen (CEA) levels in colorectal cancer (CRC) and inflammatory bowel disease (IBD) including Crohn's disease (CD) and ulcerative colitis (UC). (**A**) Serum CEA is significantly higher in CRC patients than in IBD patients ($P < 0.05$). (**B**) Within IBD, CEA levels are significantly higher in Crohn's disease than in ulcerative colitis ($P < 0.05$). (**C**) No significant gender difference in serum CEA across all IBD patients ($P > 0.05$). (**D**) No significant gender difference in serum CEA in Crohn's disease patients ($P > 0.05$). (**E**) In ulcerative colitis, serum CEA is significantly higher in female than in male patients ($P < 0.05$). (**F**) No significant gender difference in serum CEA in CRC patients ($P > 0.05$). (**G**) Serum CEA increases with CRC stage, with stage IV significantly higher than stages I-III ($P < 0.05$). (**H**) In male CRC patients, stage IV CEA levels are significantly higher than stages I-III, with no difference between stages I and II ($P > 0.05$). (**I**) In female CRC patients, stage IV CEA levels are significantly higher than stages I and II ($P < 0.05$), but not significantly different from stage III ($P > 0.05$).

Evaluation of serum CEA expression in colorectal cancer and inflammatory bowel disease patients

The serum CEA expression level was significantly higher in colorectal cancer (CRC) patients compared to those with inflammatory bowel disease ($P < 0.05$) (Fig. 10A). Moreover, the serum CEA level was significantly higher in patients with Crohn's disease compared to those with ulcerative colitis ($P < 0.05$) (Fig. 10B). Regarding gender differences in IBD, no significant change in serum CEA level was observed between male and female patients ($P > 0.05$) (Fig. 10C). Similarly, in Crohn's disease, there was no significant gender related difference in serum

CEA expression ($P > 0.05$) (Fig. 10D). However, in ulcerative colitis, the serum CEA level was significantly higher in female patients compared to males ($P < 0.05$) (Fig. 10E). In contrast, in CRC patients, no significant difference in serum CEA expression was found between males and females ($P > 0.05$) (Fig. 10F). In the analysis of CRC stages, the CEA expression level was significantly elevated in stage IV compared with stages III, II, and I ($P < 0.05$) (Fig. 10G). When CRC stages were analyzed separately by gender, male patients in stage IV showed significantly higher CEA levels compared with stages III, II, and I, while no significant difference was detected between stages I and II ($P > 0.05$) (Fig. 10H). Among female CRC patients, the CEA expression was significantly higher in stage IV compared with stages I and II ($P < 0.05$) (Fig. 10I), whereas no significant difference was observed between stage IV and stage III ($P > 0.05$) (Fig. 10I).

Discussion

Crohn's disease (CD) and ulcerative colitis (UC) are both part of IBD, a chronic gastrointestinal disorder that is characterized by ongoing tissue remodeling, abnormal mucosal immunology and epithelial barrier dysfunction [42]. Patients that suffer extended phases of colonic mucosal inflammation and regeneration are at increased risk for colorectal cancer (CRC) due to stromal activation, abnormal epithelial proliferation, and genetic susceptibility [42]. The molecular interaction between chronic inflammation and neoplasia remains poorly recognized mainly because of the complex cellular heterogeneity and spatiotemporal interactions among immune, stromal, and epithelial compartments [43]. To discover pathological pathways and establish targeted therapeutic approaches to address inflammation-associated carcinogenesis, it is crucial to completely understand this complex cellular context [44]. Recent advances in spatial transcriptomics and single-cell RNA sequencing (scRNA-seq) have changed our understanding of IBD and CRC by revealing disease-specific cell populations, regulatory pathways, and potential biomarkers that were previously ignored by bulk studies [45]. These technologies have produced high-resolution maps of cellular diversity, signaling interactions, and spatial organization within diseased tissues [45]. The integration of multi-omics approaches has further emphasized the prognostic and therapeutic advantages of cellular biomarkers, which can range from immune cell subpopulations to extracellular matrix (ECM) remodeling pathways, providing novel opportunities for personalized treatment [46]. These approaches provide insights how inflammatory regions change into favorable to tumors microenvironments in colorectal cancer and contribute to determining the balance between pathogenic inflammation and protective immune responses in the context of IBD [46]. The single-cell transcriptome investigation showed that Crohn's disease significantly alters the mucosal cellular landscape, with several immune subsets increasing and epithelial and tissue stem cell regions significantly decreasing. The loss of the epithelial barrier, an essential component of CD pathophysiology, can be associated with a reduction in epithelial cells [47]. In contrast, the increase in T cells, monocytes, neutrophils, and macrophages reflects enhanced inflammatory infiltration and sustained immune activation [47]. Chronic immunological activation is further supported by altered distributions within T cell and B cell subpopulations, including enhanced naïve and memory categories [48]. While a decrease of CD8+ central memory T cells may reflect deficient adaptive control, the increase in gamma-delta T cells indicates an innate-like response to epithelium damage [49, 50]. The cellular profile of ulcerative colitis demonstrated a strong inflammatory signature with significant epithelial depletion and proliferation of B cells, T cells, NK cells, and macrophages. Ulcerative colitis has been demonstrated to exhibit widespread patterns of barrier damage, which are consistent with the considerable reduction in tissue stem cells and epithelial cells [51]. Increased humoral and cytotoxic responses within the inflammatory mucosa are highlighted by increased B cell and NK cell proportions [52]. Changes in gamma-delta T cells and other T cell subsets indicate dysregulated mucosal immunity and impaired tissue-protective responses [53]. Alterations in B cell composition, including increased naïve B cells and reduced immature subsets, may reflect local differentiation dynamics driven by chronic inflammation [54]. Pseudotime trajectory investigation provided more understanding into how biological states change across duration. The conditions showed that adaptive immune populations (B and T cells) characterized the initial pseudotime phases, followed by initial effectors (neutrophils, NK cells), and epithelial, stem, and astrocyte-like populations around the final stage. This progressive phase, which is comparable to the chronic inflammatory cycle characteristic of IBD, probably represents the transition from acute immune activation to tissue transformation and repair [55, 56]. Plasma B cells and memory B cells developed at final pseudotime branches in CD, demonstrating continued humoral activation, but epithelial and hepatocyte-like cells appeared at the end of pseudotime trajectories in UC, emphasizing metabolic reprogramming and regeneration of epithelial tissue [56]. These trajectories confirm the concept that cancer cell transformation is made possible by a combination of regenerative proliferation and chronic immune activation. CellChat findings demonstrated that cytokine and extracellular matrix (ECM)-related pathways, specifically collagen, laminin, GALECTIN, and MIF signaling, are in

responsible of active intercellular communication. These connections between immune system, epithelial, endothelial, and stem cells demonstrated the essential function of stem and endothelial cells as communication centers in ulcerative colitis and Crohn's disease. Strong tissue remodeling can be detected by the prevalence of ECM pathways [57], whereas immune-metabolic interaction including monocytes and neutrophils is demonstrated by GALECTIN and MIF networks [58, 59]. Collectively, these signaling pathways could support chronic inflammation and facilitate angiogenesis and immune evasion, demonstrating that focusing on MIF, GALECTIN, or ECM regulators could be useful in destabilizing interactions that contribute to disease. These modifications indicate to a highly inflammatory and dysregulated mucosal microenvironment which could be an origin of intestinal inflammation. Spatial transcriptomic mapping provided a contextual understanding of how these cell types are organized within tissue architecture. T and B lymphocytes, neutrophils, monocytes, and epithelial cells were found mainly in the tumor central region and areas of active inflammation, while tissue stem cell and endothelial cells were located in distinct surrounding regions that indicated sites of stromal remodeling, angiogenesis, and regeneration [60]. Subpopulation mapping showed that memory B cells and CD8+ central memory T cells localized to adjacent regions and suggesting organized adaptive responses [61], while plasma B cells spread in distinct regions and possibly corresponding to tertiary lymphoid structures (TLS) [62]. These spatial patterns could function as regional centers for tissue healing and immunological activation, but when imbalanced, they can also create chronic inflammation, influence the extracellular matrix, and accelerate the spread of tumors. Integrating bulk and single-cell transcriptome data demonstrated that a number of genes associated with ECM remodeling, inflammation, and proliferation were continuously elevated. These genes included NUAK1, CENPE, LGALS1, PDGFRA, CHI3L1, CALCRL, SERPINE1, ABL2, MMP12, MAPK10, ADAMTS4, MMP3, NAMPT and CEACAM5. Furthermore, inflammatory IBD tissues also had higher levels of the CRC marker CEACAM5 and indicating that chronic inflammation could initiate tumor-promoting pathways before to the onset of dysplasia. This is confirmed by higher blood CEA levels in CRC and some CD patients, but they also emphasize the importance for caution when evaluating tumor markers while inflammation is occurring.

NUAK1, a serine/threonine kinase involved in cellular metabolism and the regulation of oxidative stress, has been increasingly linked to gastrointestinal inflammation and tumorigenesis [63]. According to previous studies by Xin Ni [64] and Jennifer Port [65], NUAK1 is elevated in several types of cancers, including colorectal cancer. Its high expression is associated with aggressive tumor characteristics and a lower overall survival rate, probably as an outcome of altered antioxidant defense pathways. Yajie Wang and et al., suggested that NUAK1 plays a role in IBD by regulating cell adhesion and calcium signaling activities [66]. According to our single-cell RNA-seq data, NUAK1 was mainly expressed in clusters of fibroblasts and epithelial cells and associated with Crohn's disease patients' intestinal mucosa. This suggests that NUAK1 plays an important role in epithelial-stromal interactions during inflammatory processes. Pseudotime trajectory evaluation demonstrated a progressive increase in NUAK1 expression along the differentiation path of epithelial progenitors toward inflammatory states and suggesting activation in response to oxidative stress during chronic inflammation [67]. Spatial transcriptomic analysis confirmed strong localization of NUAK1 in inflamed colonic crypt regions, consistent with its contribution to mucosal remodeling and epithelial stress response during Crohn's disease [68]. CENPE, a centromere-associated motor protein essential for chromosome segregation, has been linked to cancer progression and genomic instability [69]. Yuan Yuan [70] and Yuan-Xiang Shi [71] revealed that CENPE overexpression in colon cancer associated with poor prognosis, immune evasion and increased tumor mutational burden and suggesting a role in both cell cycle regulation and immune modification. Our findings indicated that CENPE was upregulated in epithelial cell clusters in single-cell datasets from Crohn's disease patients. This indicates abnormal cell proliferation and epithelial regeneration under chronic inflammatory stress [72]. Pseudotime trajectory evaluation showed CENPE activation during late inflammatory stages and associated with an increase in epithelial proliferation [73]. CENPE expression was further localized by spatial transcriptomic data to hyperproliferative epithelial regions adjacent to ulcerated mucosa, confirming its role in excessive regenerative activities and possible vulnerability to dysplasia in Crohn's disease [74]. CHI3L1 (chitinase-3-like 1) has been consistently associated with inflammatory and neoplastic processes in the gut [75, 76]. Kamba [77] and Chen [78] demonstrated that CHI3L1 as highly expressed in IBD mucosa with dysplasia and facilitating the proliferation of epithelial cells by activating NF-κB. Deutschmann [79] and Levantovsky [80] revealed that CHI3L1 is related to Crohn's disease progression and particularly in fibroblast and myeloid populations associated with fistula formation. Our single-cell assessment confirmed elevated CHI3L1 expression in fibroblasts and macrophage clusters within the ileal mucosa of Crohn's disease patients and suggesting its participation in chronic inflammation and tissue remodeling [80]. Pseudotime trajectory indicated that CHI3L1 upregulation at intermediate to late

inflammatory states, consistent with fibroblast activation and immune recruitment [81]. Spatial transcriptomics demonstrated that CHI3L1 was expressed to inflamed submucosal regions and around immune follicles and reflecting its involvement in both epithelial repair and fibrotic remodeling in Crohn's disease [82, 83]. The spatial co-localization of CHI3L1 expressing inflammatory niches with epithelial stem cell regions suggests that stem cells may be exposed to persistent inflammatory and proliferative signals within the local microenvironment [84, 85]. Given the high responsiveness of intestinal stem cells to microenvironmental signals, such conditions may promote sustained epithelial proliferation and altered differentiation dynamics, potentially creating a microenvironment predisposed to dysplasia development [74]. CALCRL, a gene encoding the calcitonin receptor-like receptor involved in vascular and neuroimmune signaling, has demonstrated context-dependent roles in intestinal homeostasis [86]. Jakob Wiese and colleagues demonstrated that minimal transcriptional changes in CALCRL expression in inflamed IBD colon tissues [87]. Additionally, Ran Wang and colleagues highlighted its beneficial lymphatic role and indicating that intestinal inflammation after mucosal damage is exacerbated through a lack in Calcrl gene [88]. CALCRL expression was mainly identified in epithelial and endothelial clusters of Crohn's disease tissue in our single-cell RNA-seq dataset, which is consistent with its demonstrated neuro-epithelial regulation function [89]. Pseudotime trajectory analysis indicated stable expression during early epithelial differentiation and reflecting a potential homeostatic rather than reactive function [89]. According to our spatial transcriptomics findings, low to moderate CALCRL expression was found in the central and peripheral mucosal regions of colorectal cancer tissue and demonstrating its role in maintaining epithelial lymphatic interaction in chronic inflammatory circumstances [90]. SERPINE1, encoding plasminogen activator inhibitor-1 (PAI-1), is a key regulator of extracellular matrix remodeling and fibrosis [91]. De Santis and et al., discovered that gastrointestinal fibrosis in Crohn's disease is driven by IL-33-induced SERPINE1 overexpression [92]. Its increased expression has been significantly related with adverse clinical outcomes, advanced tumor stage, and immune system infiltration in both active IBD and colorectal cancer, according to several investigations such as: B Jójárt and et al. [93], Tang and et al. [94], Wang and et al. [95], Yuan and et al. [96]. Notably, according to Tang and colleagues [94], SERPINE1 may be used as a biomarker for Crohn's disease activity and treatment response. Our single-cell RNA-seq data demonstrated that SERPINE1 was primarily expressed in fibroblast and endothelial populations within inflamed ileal tissue, consistent with its pro-fibrotic and pro-inflammatory role [97, 98]. Analysis of pseudotime trajectories revealed that SERPINE1 was over time upregulated during the fibroblast activation trajectory and suggesting a role in chronic tissue remodeling [98]. The moderate expression of SERPINE1 in the central regions of colorectal cancer was further validated by spatial transcriptomic results, which also associated with histologic areas of inflammation and submucosal fibrosis [99]. Inflammation-associated neoplasia has been associated with ABL2, a non-receptor tyrosine kinase related to endothelial function and cytoskeletal dynamics [100]. Chakrabarty and et al., demonstrated that ABL2 mutations in patients with ulcerative colitis who were at high risk of developing colorectal cancer, indicating that the protein may be involved in inflammation-driven carcinogenesis [101]. ABL2 expression in endothelial and fibroblast populations from inflammatory intestinal areas was found by Levitte [102] and Ke [103], suggesting a possible function in protecting blood vessel function during inflammation [104]. ABL2 expression was mainly detected in endothelial and fibroblast clusters of Crohn's ascending colon tissue, according to our single-cell and bulk RNA-seq findings. Pseudotime trajectory analysis demonstrated dynamic ABL2 expression during fibroblast activation and related to endothelial remodeling [97]. Spatial transcriptomic revealed that ABL2 was expressed in the central mucosal regions of Crohn's tissue and associated to vascularized and fibrotic regions [105]. MAPK10 (JNK3), a stress-activated kinase, regulates apoptosis and inflammatory signaling [106]. Quaglio and colleagues demonstrated that colonic tissues with inflammation had elevated MAPK10 expression, indicating an adaptive reaction to stress initiated through inflammation [107]. According to our single-cell RNA-seq data, Crohn's sigmoid colon samples showed modest but particular expression of MAPK10 within epithelial clusters. Pseudotime trajectory evaluation demonstrated that late-stage induction and suggesting MAPK10 activation under prolonged inflammation [106]. Spatial transcriptomic mapping revealed that very low expression of MAPK10 in left mucosal regions of colorectal cancer tissue and demonstrating its specific involvement in inflammation [106]. The metalloproteinase and a disintegrin related to the thrombospondin motifs (ADAMTS) family play a role in the pathophysiology of numerous chronic inflammations as well as the formation of the extracellular matrix [108]. Buran and et al. [109], and Lin and et al. [110], demonstrated that ADAMTS4 was upregulated in IBD, including Crohn's and ulcerative colitis and increased cytoplasmic protein localization in inflamed tissues. Our single-cell RNA-seq analysis revealed that ADAMTS4 was expressed primarily in fibroblast clusters from Crohn's terminal ileum and supporting its profibrotic activity [111]. According to pseudotime trajectory findings, there was persistent upregulation during

fibroblast differentiation, which is consistent with chronic ECM degradation [112]. The core regions of colorectal cancer tissue demonstrated moderate ADAMTS4 expression, which is consistent with fibrotic remodeling areas and supports its role in matrix disintegration during chronic inflammation [111], as demonstrated by spatial transcriptomics. In inflammatory bowel disease, damage to the gut and infiltration of immune cells have been associated with MMP3, another important matrix metalloproteinase [113, 114]. Jimbo and et al. [115], , Gordon and et al. [114], , and Pan and et al. [116], discovered that immune cell infiltration and disease progression were associated with increased MMP3 expression in active UC and CD, especially in epithelial and plasma cells. Li and et al. [117], and de Bruyn and et al. [118], demonstrated that MMP3 contributes to colitis-related carcinogenesis. In Crohn's disease ileal tissue, our single-cell RNA-seq data showed that MMP3 expression was detected in clusters of macrophages and epithelial cells, contributing it to mechanisms of mucosal damage and healing [119]. Pseudotime trajectory assessment indicated highest expression of MMP3 in the final phases of inflammation. The crucial role of MMP3 in chronic mucosal remodeling was supported by spatial transcriptomic data, which showed low but distinct MMP3 expression and localized mainly in the inflamed regions of colorectal cancer [120].

LGALS1, which codes for galectin-1, is essential for mucosal defense, apoptosis, and immunological modulation [121]. In experimental models, Fernandez-Perez and et al. [122], , Papa-Gobbi and et al. [123], , and Morosi and et al. [124], demonstrated that Gal-1 contributes to colitis via modulating T-cell activity, especially in the Th17/Th1 proportion. Furthermore, Cagnoni and et al. [121], discovered that immunosuppressive CD8+ regulatory T cells proliferation is facilitated by increased Gal-1 expression in colorectal cancer, which contributes to a poor prognosis. The immunological function of LGALS1 in the inflammatory microenvironment was shown by our single-cell study, which revealed that it mainly expressed in stromal fibroblasts and macrophages in ulcerative colitis tissue [121]. The pseudotime trajectory indicated that LGALS1 originated at the beginning of stromal lineages before immune infiltration, which is consistent with its recognized anti-inflammatory signaling pathway. High expression of LGALS1 has been demonstrated by spatial transcriptomics in regions of immune cell aggregation and subepithelial stromal tissues, showing its dual function in tissue healing and suppression of immune cells underlying ulcerative colitis development [119]. The receptor tyrosine kinase known as PDGFRA (Platelet-Derived Growth Factor Receptor Alpha) is mostly expressed in mesenchymal cells, such as fibroblasts and particular stromal cell types in the gastrointestinal tract [125]. Pellon-Cardenas and et al. [126], reported that PDGFRA+ fibroblasts secrete paracrine ligands that support epithelial healing under oncogenic stress, while Jia Li and colleagues [127] identified these fibroblasts in both endoscopic biopsies and surgical ulcerative colitis specimens. According to our single-cell findings, fibroblast clusters in UC sigmoid colon samples had higher levels of PDGFRA expression, which confirms their contribution in matrix conversion and epithelial maintenance [51]. Pseudotime trajectory examination showed persistent PDGFRA activity across the fibroblast differentiation trajectory and suggesting continual fibroblast activation as the disease progressed [126]. High expression of PDGFRA has been demonstrated by spatial transcriptomics and determining their anatomical function in ulcerative colitis-related tissue fibrosis and mucosal healing [99]. In inflammation, MMP12 (matrix metalloproteinase-12), which codes for macrophage elastase, facilitates tissue remodeling and extracellular matrix disintegration [128]. According to Nighot and et al. [128], , Arosa and et al. [129], , and Di Sabatino and et al. [130], , MMP12 is significantly increased in ulcerative colitis and Crohn's disease, and it is correlated with the advanced stage of the disease, macrophage proliferation, and epithelial damage, while Mäkitalo and colleagues [131], confirmed its expression in colonic epithelium and infiltrating neutrophils of ulcerative colitis samples. Our single-cell RNA-seq results indicated that MMP12 is mainly expressed by macrophage populations within ulcerative colitis tissues and highlighting its immune-mediated pathogenic role [131]. Pseudotime trajectory evaluation demonstrated that high expression of MMP12 in activated macrophages at late inflammatory stages and associated with the phase of chronic tissue damage [132]. High MMP12 expression was shown by spatial transcriptomics and enriched in the core areas of colorectal cancer tissue and highlighting its role in mucosal damage and active inflammation [132]. In intestinal inflammation, NAMPT, an essential enzyme for NAD^+ production, connects immunological and metabolic reactions [133]. NAMPT overexpression has been consistently reported in IBD by Hong et al. [134], Colombo et al. [135], and Neubauer et al. [136], which have associated it to immunological infiltration, macrophage activation, and a poor therapeutic response to anti-TNF medications. Lei and colleagues [137] demonstrated that bacterial NAMPT regulation increases the release of inflammatory cytokines through the MAPK pathway, and Starr and et al. [138], showed that IBD patients' inflamed mucosa had higher NAMPT levels. Our single-cell RNA-seq results indicated that NAMPT was significantly expressed in UC ileal tissue macrophage clusters, highlighting a role for it in metabolic immune interaction [139]. Pseudotime trajectory assessment demonstrated that progressive

NAMPT induction during macrophage activation and consistent with SPP1⁺ tumor-associated macrophages in chronic inflammation [140]. Spatial transcriptomic evaluation confirmed moderate NAMPT expression across colorectal cancer tissues, particularly in immune-rich regions and highlighting its contribution to persistent inflammation in ulcerative colitis [141]. CEACAM5 (Carcinoembryonic Antigen, CEA) is a cell adhesion glycoprotein that plays a dual function in the stability of epithelial cells and proliferation of tumors, widely known as a diagnostic and prognostic biomarker in colorectal malignancies [142]. Several research such as Liu and et al. [143], , Holmer and et al. [144], demonstrated that CEACAM5 regulated by inflammatory cytokines like IL-6 and is significantly expressed in a variety of epithelial tumors, particularly gastric and colorectal cancers. Additionally, inhibition of TGF-β/SMAD3 signaling has been related to its expression and demonstrating a biological role in immunological modulation and epithelial differentiation [145, 146]. According to Saiz-Gonzalo and colleagues [142], CEACAM5 has been shown to exhibit inconsistent patterns in inflammatory bowel disease (IBD), including downregulated in Crohn's disease (CD) and upregulated in adult ulcerative colitis (UC). Our experimental findings demonstrated that serum CEA levels were significantly higher in colorectal cancer (CRC) patients compared to IBD patients and among IBD patients, CEA expression was significantly higher in Crohn's disease compared to ulcerative colitis. The higher CEACAM5 levels observed in Crohn's disease compared with ulcerative colitis may be related to the distinct pathological characteristics of these two IBD subtypes. Crohn's disease is characterized by transmural inflammation and repeated cycles of epithelial injury and regeneration, which may enhance epithelial proliferation and differentiation processes associated with CEACAM5 expression [147, 148]. In addition, stronger activation of pro-inflammatory cytokine pathways, including IL-6 mediated signaling, may further contribute to the increased expression and systemic release of CEACAM5 in Crohn's disease [149, 150]. Furthermore, CEA levels increased progressively across CRC stages, particularly in stage IV and establishing that it is associated with the growth of tumors. Gender-based comparisons revealed no significant differences in serum CEA levels between males and females in CD or CRC. Additionally, UC female patients exhibited higher serum CEA levels compared to males. According to Gao and colleagues [151], increased CEA levels are associated with tumor proliferation, lymphatic and vascular invasion, and the spread of metastases in colorectal cancer. At the transcriptomic level, CEACAM5 expression was detected as a differential biomarker in both Crohn's disease and ulcerative colitis. In the single-cell RNA-seq dataset, CEACAM5 was predominantly expressed in epithelial cell clusters. Trajectory analysis revealed that CEACAM5 expression progressively increased along the pseudotime trajectory and indicating a function in late-stage epithelial differentiation and potential transition toward a tumorigenic state [152]. Spatial transcriptomic evaluation demonstrated that high CEACAM5 expression across extensive regions of the intestinal and colorectal cancer tissues, particularly enriched in the central areas of the colorectal capture area tissue and intestinal cancer regions, while maintaining moderate expression in whole transcriptome colorectal tissue. Strong epithelial localization from spatial transcriptomic findings and upregulation of CEACAM5 in Crohn's disease and advanced stages of colorectal cancer demonstrate its diagnostic importance and possibilities as a biomarker for determining disease progression and the progression from inflammatory to neoplastic stages in the colon and ileum [153]. Collectively, our multi-omics findings indicated that CEACAM5 represents a crucial biological connection between colorectal carcinogenesis and chronic intestinal inflammation.

Several limitations of this study should be acknowledged. One limitation is the limited number of publicly available spatial transcriptomics datasets for colorectal cancer. Although the selected datasets provide high quality spatial gene expression profiles and cellular heterogeneity, they may not fully represent the heterogeneity across all CRC clinical stages. Another limitation is the absence of longitudinal biopsy samples that follow patients over time from IBD to colorectal cancer. Consequently, the inferred disease progression is based on integrative analyses of bulk, single cell, and spatial transcriptomic datasets rather than real time temporal observation. While this approach provides biologically consistent evidence of cellular and molecular transitions, it represents an indirect inference of progression rather than direct longitudinal validation. An additional limitation is the lack of direct mechanistic validation of CEACAM5 function at the tissue and cellular levels. Due to substantial financial and infrastructural constraints, experimental approaches such as real time PCR, Western blotting, immunohistochemistry, gene knockdown or overexpression assays, and in vivo modeling could not be performed. Future studies should prioritize functional validation using in vitro cell based models, targeted manipulation of CEACAM5 expression, and tissue level analyses to clarify its causal role in colorectal cancer initiation and progression. Finally, although our single cell RNA sequencing dataset includes several innate immune cell populations (myeloid and epithelial compartments) that are known to express pattern recognition receptors (PRRs), we did not perform a systematic analysis specifically focused on PRR families such as Toll like receptors (TLRs), NOD like receptors (NLRs), or other innate

immune sensing pathways. The primary objective of this study was to characterize the overall immune cellular landscape of IBD and to further resolve disease associated heterogeneity within B and T cell subpopulations. Future studies could build upon the current dataset to investigate the expression patterns and regulatory networks of PRRs across specific innate immune cell populations, which may provide additional insights into the mechanisms of innate immune dysregulation and host-microbiota interactions in inflammatory bowel disease.

Conclusions

Our integrative multi-omics analysis demonstrates that chronic intestinal inflammation in inflammatory bowel disease induces profound immune, epithelial, and stromal reprogramming, thereby establishing a microenvironment favorable to colorectal carcinogenesis. Single-cell and pseudotime trajectory analyses reveal sustained immune activation, epithelial attrition, and dysregulated regeneration as key features of Crohn's disease and ulcerative colitis and providing mechanistic insight into the progression from chronic mucosal damage to dysplasia susceptible states. Spatial transcriptomic profiling further demonstrated that inflammation-associated genes involved in extracellular matrix remodeling, immune evasion, angiogenesis, and metabolic adaptation are not uniformly distributed, but instead localize to distinct tissue niches that function as hubs of disease progression. The spatial co-occurrence of immune infiltration, fibroblast activation, and epithelial stress responses reveals shared microenvironmental programs between inflammatory bowel disease and colorectal cancer that represent actionable therapeutic targets. The spatially resolved characterization of CEACAM5 and LGALS1 as candidate biomarkers underscores their potential clinical utility in risk stratification, early detection, and precision therapy selection. Collectively, our findings establish a comprehensive cellular and spatial framework that elucidates inflammation-driven colorectal tumorigenesis in the context of inflammatory bowel disease, including Crohn's disease and ulcerative colitis and supports the rational development of precision therapeutic strategies targeting both pathogenic inflammation and cancer-promoting tissue remodeling in colorectal cancer.

Abbreviations

CD	Crohn's Disease
UC	Ulcerative Colitis
IBD	Inflammatory Bowel Disease
CRC	Colorectal Cancer
GI	Gastrointestinal
scRNA-seq	Single-Cell RNA Sequencing
ST	Spatial Transcriptomics
SRA	Sequence Read Archive
GEO	Gene Expression Omnibus
FPPE	Fixed Paraffin-Embedded
FFPE	Formalin-Fixed Paraffin-Embedded
QC	Quality Control
log2 (FC)	Log2 Fold Change
FDR	False Discovery Rate
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
CCA	Canonical Correlation Analysis
PCA	Principal Component Analysis
UMAP	Uniform Manifold Approximation and Projection
DEGs	Differentially Expressed Genes
RDS	R Data File
RCTD	Robust Cell Type Decomposition
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
TCGA-COAD	The Cancer Genome Atlas-Colon Adenocarcinoma
TCGA-READ	The Cancer Genome Atlas-Rectum Adenocarcinoma
CEA	Carcinoembryonic Antigen
CLIA	Chemiluminescent Immunoassay
ELISA	Enzyme-Linked Immunosorbent Assay
AJCC	American Joint Committee on Cancer
UICC	Union for International Cancer Control
MHC I	Major Histocompatibility Complex Class I
GO BP	Gene Ontology Biological Process
GO MF	Gene Ontology Molecular Function
GO CC	Gene Ontology Cellular Component
COLLAGEN	Collagen
LAMININ	Laminin
APP	Amyloid Precursor Protein
CypA	Cyclophilin A
CD99	Cluster of Differentiation 99
GALECTIN	Galectin Family Proteins
VISFATIN	Visfatin (Nicotinamide Phosphoribosyl transferase)
MIF	Macrophage Migration Inhibitory Factor
THBS	Thrombospondin
IGFBP	Insulin-Like Growth Factor Binding Protein
CLEC	C-Type Lectin Domain-Containing Protein
95% CI	95% Confidence Interval
ECM	Extracellular Matrix
TME	Tumor Microenvironment
NUAK1	NUAK Family Kinase 1
CENPE	Centromere Protein E
LGALS1	Galectin-1
PDGFRA	Platelet-Derived Growth Factor Receptor Alpha
CHI3L1	Chitinase-3-Like Protein 1
CALCRL	Calcitonin Receptor-Like Receptor
SERPINE1	Serpin Family E Member 1
ABL2	Abelson Tyrosine-Protein Kinase 2
MMP12	Matrix Metallopeptidase 12
MAPK10	Mitogen-Activated Protein Kinase 10
ADAMTS4	ADAM Metallopeptidase with Thrombospondin Type 1 Motif 4
MMP3	Matrix Metallopeptidase 3
NAMPT	Nicotinamide Phosphoribosyl transferase
CEACAM5	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 5

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-026-08158-2.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We gratefully acknowledge the support and collaboration of the following institutions: Ospedale San Raffaele, Department of Gastroenterology, Milan, Italy; IDIBAPS, Department of Gastroenterology, Barcelona, Spain; Department of Paediatrics, University of Cambridge, Cambridge, UK; The MRC Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK; SciLifeLab, Gene Technology, Stockholm, Sweden; and the Micky & Madeleine Arison Family Foundation Crohn's & Colitis Discovery Lab,

University of Miami, Miami, FL, USA. We also thank the providers of the Visium HD Spatial Gene Expression datasets, the 10x Genomics leadership team, and the members of the 10x Genomics microscopy, flow, sequencing, and histopathology core facilities for their technical support and for making essential resources available to this study.

Author contributions
Conceptualization, S.T.H.; methodology, S.T.H.; software, S.T.H. and K.A.T.; validation, S.T.H.; formal analysis, S.T.H., K.A.T., F.N., R.B. and Y.D.; investigation, S.T.H.; resources, S.T.H.; data curation, S.T.H.; writing—original draft preparation, S.T.H.; writing—review and editing, S.T.H., F.N., R.B. and Y.D.; visualization, S.T.H.; supervision, R.B. and Y.D.; project administration, S.T.H. All authors have read and agreed to the published version of the manuscript.

Funding
This research was conducted without external funding.

Data availability
The datasets supporting the conclusions of this article are publicly available in the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) and Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra/). Bulk RNA-Seq, Single Cell RNA-Seq and Spatial transcriptomics data can be accessed at (GSE165512, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165512, ***N*** = 168, Human; SRA: SRP303290, Illumina HiSeq 2500), (PRJEB24645, https://www.ncbi.nlm.nih.gov/Traces/study/?acc=ERP106487&o=acc_s%3Aa, ***N*** = 79, Human; SRA: ERP106487, Illumina HiSeq 2500), (GSE183620 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183620, ***N*** = 114, Human; SRA: SRP336051, Illumina NovaSeq 6000), (GSE191234, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191234, ***N*** = 54, Human; SRA: SRP351557, Illumina HiSeq 2500), (GSE214695, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE214695, ***N*** = 18, Human; Illumina NextSeq 500). In addition, all 10x Genomics datasets are publicly available via 10x Genomics database (https://www.10xgenomics.com/datasets). Spatial transcriptomic datasets from 10x Genomics are available, including (Human Colorectal Cancer: Whole Transcriptome Analysis, https://www.10xgenomics.com/datasets/human-colorectal-cancer-whole-transcriptome-analysis-1-standard-1-2-0, Illumina NovaSeq 6000, flow cell HHYWHDSXY, lanes 1-4; 112,228 mean reads per cell), (Human Intestine Cancer, (FPPE) https://www.10xgenomics.com/datasets/human-intestine-cancer-1-standard, Illumina NovaSeq 6000, flow cell HCCVHDSX2, lanes 1-2; 77,288 reads per spot) and (Human Colorectal Cancer, 11 mm Capture Area (FFPE), https://www.10xgenomics.com/datasets/human-colorectal-cancer-11-mm-capture-area-ffpe-2-standard, Illumina NovaSeq 6000, flow cells HJWFGDSX5 and HJW2WDSX5, lanes 1-4; 911,044,482 total reads / 46.2% saturation). All other data supporting the findings of this study are available within the article, the Supplementary Information, or from the corresponding author upon reasonable request.

Code availability
All codes used for pre-processing of the bulk-RNA-Seq data in Linux environment, single-cell RNA-Seq and spatial transcriptomics data are available at https://github.com/seyedtalebhosseini/IBD-to-CRC-Multi-Omics-Analysis-Bulk-RNA-Seq-Single-Cell-Spatial-Transcriptomics-. Other data analysis approaches, including algorithm and packages are described in methods section.

Declarations

Ethics approval and consent to participate
This research was conducted in full accordance with the ethical standards of the Declaration of Helsinki and was approved by the Research Ethics Committee of Islamic Azad University-Sari Branch (Approval Code: IR.IAU.SARI.REC.1404.365). Written informed consent was obtained from all participants prior to enrollment, and consent for the publication of anonymized data and results was also secured. All identifiable information was removed to maintain confidentiality and protect participant privacy.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Immunogenetics Research Center, Mazandaran University of Medical Sciences, Sari, Iran
[2]Department of Biology, QaS.C., Islamic Azad University, Qaemshahr, Iran
[3]Clinical Research Development Unit, Faculty of Medicine, Ghaem Hospital, Mashhad University of Medical Sciences, Mashhad, Iran
[4]Department of Quantitative Health Sciences, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI 96813, USA

Received: 4 February 2026 / Accepted: 12 April 2026
Published online: 18 April 2026

References

1. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. Lancet. 2007;369(9573):1627–40.
2. Ananthakrishnan AN. Epidemiology and risk factors for IBD. Nat reviews Gastroenterol Hepatol. 2015;12(4):205–17.
3. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2017;390(10114):2769–78.
4. Dolinger M, Torres J, Vermeire S. Crohn's disease. Lancet. 2024;403(10432):1177–91.
5. Feuerstein JD, Cheifetz AS, editors. Ulcerative colitis: epidemiology, diagnosis, and management. Mayo Clinic Proceedings; 2014: Elsevier.
6. Fanizza J, Bencardino S, Allocca M, Furfaro F, Zilli A, Parigi TL, et al. Inflammatory bowel disease and colorectal cancer. Cancers. 2024;16(17):2943.
7. Thomas MF, Slowikowski K, Manakongtreecheep K, Sen P, Samanta N, Tantivit J, et al. Single-cell transcriptomic analyses reveal distinct immune cell contributions to epithelial barrier dysfunction in checkpoint inhibitor colitis. Nat Med. 2024;30(5):1349–62.
8. Baars MJ, Floor E, Sinha N, Ter Linde JJ, van Dam S, Amini M, et al. Multiplex spatial omics reveals changes in immune-epithelial crosstalk during inflammation and dysplasia development in chronic IBD patients. Iscience. 2024;27(8).
9. Leinonen R, Sugawara H, Shumway M, Collaboration INSD. The sequence read archive. Nucleic Acids Res. 2010;39(suppl1):D19–21.
10. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2012;41(D1):D991–5.
11. Jacobsen GE, Fernández I, Quintero MA, Santander AM, Pignac-Kobinger J, Damas OM, et al. Lamina propria phagocyte profiling reveals targetable signaling pathways in refractory inflammatory bowel disease. Gastro Hep Adv. 2022;1(3):380–92.
12. Garrido-Trigo A, Corraliza AM, Veny M, Dotti I, Melón-Ardanaz E, Rill A, et al. Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease. Nat Commun. 2023;14(1):4506.
13. Andrews S. FastQC: a quality control tool for high throughput sequence data. No Title; 2010.
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
15. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11(9):1650–67.
16. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.
17. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882–3.
18. Anders S, Huber W. Differential expression analysis for sequence count data. Nat Precedings. 2010:1.
19. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
20. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–87. e29.
21. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163–72.

22. Amezquita RA, Lun AT, Becht E, Carey VJ, Carpp LN, Geistlinger L, et al. Orchestrating single-cell analysis with Bioconductor. Nat Methods. 2020;17(2):137–45.
23. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14(10):979–82.
24. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12(1):1088.
25. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+to analyze and compare ROC curves. BMC Bioinformatics. 2011;12(1):77.
26. Therneau T. A package for survival analysis in R. R package version 3.2-7. 2020.
27. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using'ggplot2'. CRAN: Contributed Packages; 2016.
28. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44(8):e71–e.
29. Long Y, Ang KS, Li M, Chong KLK, Sethi R, Zhong C, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. Nat Commun. 2023;14(1):1155.
30. Paszke A, Pytorch. An imperative style, high-performance deep learning library. arXiv preprint arXiv:191201703. 2019.
31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
32. Flamary R, Courty N, Gramfort A, Alaya MZ, Boisbunon A, Chambon S, et al. Pot: Python optimal transport. J Mach Learn Res. 2021;22(78):1–8.
33. Hunter JD, Matplotlib. A 2D graphics environment. Comput Sci Eng. 2007;9(03):90–5.
34. McKinney W. Data structures for statistical computing in Python. SciPy. 2010;445(1):51–6.
35. Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62.
36. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19:1–5.
37. Satija R, Hoffman P, Hao Y, Hartman A, Molla G, Butler A, Stuart T. SeuratObject: data structures for single cell data. 2023.
38. Hoffman P. SeuratDisk: Interfaces for HDF5-based single cell file formats. Github. 2022.
39. Cable DM, Murray E, Zou LS, Goeva A, Macosko EZ, Chen F, Irizarry RA. Robust decomposition of cell type mixtures in spatial transcriptomics. Nat Biotechnol. 2022;40(4):517–26.
40. Wickham H. dplyr: A grammar of data manipulation. R package version 04. 2015;3:p156.
41. Pebesma E, Bivand R. Spatial data science: With applications in R. Chapman and Hall/CRC; 2023.
42. Zhang Y-Z, Li Y-Y. Inflammatory bowel disease: pathogenesis. World J gastroenterology: WJG. 2014;20(1):91.
43. Greten FR, Grivennikov SI. Inflammation and cancer: triggers, mechanisms, and consequences. Immunity. 2019;51(1):27–41.
44. Todoric J, Antonucci L, Karin M. Targeting inflammation in cancer prevention and therapy. Cancer Prev Res. 2016;9(12):895–905.
45. Gudiño V, Bartolomé-Casado R, Salas A. Single-cell omics in inflammatory bowel disease: recent insights and future clinical applications. Gut. 2025.
46. Ivanisevic T, Sewduth RN. Multi-omics integration for the design of novel therapies and the identification of novel biomarkers. Proteomes. 2023;11(4):34.
47. Martin JC, Chang C, Boschetti G, Ungaro R, Giri M, Grout JA, et al. Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. Cell. 2019;178(6):1493–508. e20.
48. McHeyzer-Williams M, Okitsu S, Wang N, McHeyzer-Williams L. Molecular programming of B cell memory. Nat Rev Immunol. 2012;12(1):24–34.
49. Nielsen MM, Witherden DA, Havran WL. γδ T cells in homeostasis and host defence of epithelial barrier tissues. Nat Rev Immunol. 2017;17(12):733–45.
50. Sheridan BS, Romagnoli PA, Pham Q-M, Fu H-H, Alonzo F, Schubert W-D, et al. γδ T cells exhibit multifunctional and protective memory in intestinal tissues. Immunity. 2013;39(1):184–95.
51. Dotti I, Mora-Buch R, Ferrer-Picón E, Planell N, Jung P, Masamunt MC, et al. Alterations in the epithelial stem cell compartment could contribute to permanent changes in the mucosa of patients with ulcerative colitis. Gut. 2017;66(12):2069–79.
52. Tandel N, Negi S, Dalai SK, Tyagi RK. Role of natural killer and B cell interaction in inducing pathogen specific immune responses. Int Rev Immunol. 2023;42(4):304–22.
53. Jacobse J, Li J, Rings EH, Samsom JN, Goettel JA. Intestinal regulatory T cells as specialized tissue-restricted immune cells in intestinal immune homeostasis and disease. Front Immunol. 2021;12:716499.
54. Courey-Ghaouzi A-D, Kleberg L, Sundling C. Alternative B cell differentiation during infection and inflammation. Front Immunol. 2022;13:908034.
55. Van Praet L, Jacques P, Van den Bosch F, Elewaut D. The transition of acute to chronic bowel inflammation in spondyloarthritis. Nat Rev Rheumatol. 2012;8(5):288–95.
56. Saez A, Herrero-Fernandez B, Gomez-Bris R, Sánchez-Martinez H, Gonzalez-Granado JM. Pathophysiology of inflammatory bowel disease: innate immune system. Int J Mol Sci. 2023;24(2):1526.
57. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Biol. 2014;15(12):786–801.
58. Zmora N, Bashiardes S, Levy M, Elinav E. The role of the immune system in metabolic health and disease. Cell Metabol. 2017;25(3):506–21.
59. Im S, Kim H, Jeong M, Yang H, Hong JY. Integrative understanding of immune-metabolic interaction. BMB Rep. 2022;55(6):259.
60. Koutroubakis IE, Tsiolakidou G, Karmiris K, Kouroumalis EA. Role of angiogenesis in inflammatory bowel disease. Inflamm Bowel Dis. 2006;12(6):515–23.
61. Duckworth BC, Qin RZ, Groom JR. Spatial determinates of effector and memory CD8+T cell fates. Immunol Rev. 2022;306(1):76–92.
62. Le Rochais M, Hémon P, Ben-Guigui D, Garaud S, Le Dantec C, Pers J-O, et al. Deciphering the maturation of tertiary lymphoid structures in cancer and inflammatory diseases of the digestive tract using imaging mass cytometry. Front Immunol. 2023;14:1147480.
63. Cao L, Lin G, Fan D, Weng K, Chen Y, Wang J, et al. NUAK1 activates STAT5/GLI1/SOX2 signaling to enhance cancer cell expansion and drives chemoresistance in gastric cancer. Cell Rep. 2024;43(7).
64. Ni X, Pan F, Lang YK, Zhang W. Prognostic significance of NUAK1 and its association with immune infiltration in stomach adenocarcinoma. Discover Oncol. 2024;15(1):800.
65. Port J, Muthalagu N, Raja M, Ceteci F, Monteverde T, Kruspig B, et al. Colorectal tumors require NUAK1 for protection from oxidative stress. Cancer Discov. 2018;8(5):632–47.
66. Wang Y-j, Xiao L, Pan Y-s. Potential common molecular mechanisms between inflammatory bowel disease and erectile dysfunction in men: genetic screening and in vivo validation. 2024.
67. Lugrin J, Rosenblatt-Velin N, Parapanov R, Liaudet L. The role of oxidative stress during inflammatory processes. Biol Chem. 2014;395(2):203–30.
68. Kaser A, Blumberg RS, editors. Endoplasmic reticulum stress in the intestinal epithelium and inflammatory bowel disease. Seminars in immunology. Elsevier; 2009.
69. El-Arabey AA, Salama SA, Abd-Allah AR. CENP-E as a target for cancer therapy: Where are we now? Life Sci. 2018;208:192–200.
70. Yuan Y, Deng X, Wang S, Han S. Comprehensive pan-cancer analysis identifies centromere-associated protein E as a novel prognostic and immunological biomarker in human tumors. Biochim et Biophys Acta (BBA)-General Subj. 2023;1867(6):130346.
71. Shi Y-X, Dai P-H, Jiang Y-F, Wang Y-Q, Liu W. A pan-cancer landscape of centromere proteins in tumorigenesis and anticancer drug sensitivity. Translational Oncol. 2023;31:101658.
72. Ma X, Dai Z, Sun K, Zhang Y, Chen J, Yang Y, et al. Intestinal epithelial cell endoplasmic reticulum stress and inflammatory bowel disease pathogenesis: an update review. Front Immunol. 2017;8:1271.
73. Landén NX, Li D, Ståhle M. Transition from inflammation to proliferation: a critical step during wound healing. Cell Mol Life Sci. 2016;73(20):3861–85.
74. Ullman T, Odze R, Farraye FA. Diagnosis and management of dysplasia in patients with ulcerative colitis and Crohn's disease of the colon. Inflamm Bowel Dis. 2009;15(4):630–8.
75. Mizoguchi E. Chitinase 3–like-1 exacerbates intestinal inflammation by enhancing bacterial adhesion and invasion in colonic epithelial cells. Gastroenterology. 2006;130(2):398–411.
76. Deutschmann C, Roggenbuck D, Schierack P. The loss of tolerance to CHI3L1–a putative role in inflammatory bowel disease? Clin Immunol. 2019;199:12–7.
77. Kamba A, Lee I-A, Mizoguchi E. Potential association between TLR4 and chitinase 3-like 1 (CHI3L1/YKL-40) signaling on colonic epithelial cells in inflammatory bowel disease and colitis-associated cancer. Curr Mol Med. 2013;13(7):1110–21.

78. Chen C-C, Pekow J, Llado V, Kanneganti M, Lau CW, Mizoguchi A, et al. Chitinase 3-like-1 expression in colonic epithelial cells as a potentially novel marker for colitis-associated neoplasia. Am J Pathol. 2011;179(3):1494–503.
79. Deutschmann C, Sowa M, Murugaiyan J, Roesler U, Röber N, Conrad K, et al. Identification of chitinase-3-like protein 1 as a novel neutrophil antigenic target in Crohn's disease. J Crohn's Colitis. 2019;13(7):894–904.
80. Levantovsky RM, Tastad C, Zhang J, Gettler K, Sabic K, Werner R, et al. Multimodal single-cell analyses reveal mechanisms of perianal fistula in diverse patients with Crohn's disease. Med. 2024;5(8):886–908. e11.
81. Davidson S, Coles M, Thomas T, Kollias G, Ludewig B, Turley S, et al. Fibroblasts as immune regulators in infection, inflammation and cancer. Nat Rev Immunol. 2021;21(11):704–17.
82. Li C, Kuemmerle JF. Mechanisms that mediate the development of fibrosis in patients with Crohn's disease. Inflamm Bowel Dis. 2014;20(7):1250–8.
83. Zidar N, Langner C, Jerala M, Boštjančič E, Drobne D, Tomažič A. Pathology of fibrosis in Crohn's disease—contribution to understanding its pathogenesis. Front Med. 2020;7:167.
84. Fan Y, Meng Y, Hu X, Liu J, Qin X. Uncovering novel mechanisms of chitinase-3-like protein 1 in driving inflammation-associated cancers. Cancer Cell Int. 2024;24(1):268.
85. Low D, Subramaniam R, Lin L, Aomatsu T, Mizoguchi A, Ng A, et al. Chitinase 3-like 1 induces survival and proliferation of intestinal epithelial cells during chronic inflammation and colitis-associated cancer by regulating S100A9. Oncotarget. 2015;6(34):36535.
86. Uddin J, Yano H, Parkhurst CN, Zhang W, Ahmed A, Ribeiro de Godoy V, et al. CGRP-related neuropeptide adrenomedullin 2 promotes tissue-protective ILC2 responses and limits intestinal inflammation. Nat Immunol. 2025:1–11.
87. Wiese JJ, Manna S, Kühl AA, Fascì A, Elezkurtaj S, Sonnenberg E, et al. Myenteric plexus immune cell infiltrations and neurotransmitter expression in Crohn's disease and ulcerative colitis. J Crohn's Colitis. 2024;18(1):121–33.
88. Wang R, Moniruzzaman M, Wong KY, Wiid P, Harding A, Giri R, et al. Gut microbiota shape the inflammatory response in mice with an epithelial defect. Gut Microbes. 2021;13(1):1887720.
89. Ayala G. Neuroepithelial interactions in cancer. Annu Rev Pathol. 2023;18(1):493–514.
90. Pawelec G, Goldeck D, Derhovanessian E. Inflammation, ageing and chronic disease. Curr Opin Immunol. 2014;29:23–8.
91. Shaikh SB, Balaya RDA, Dagamajalu S, Bhandary YP, Unwalla H, Prasad TSK, Rahman I. A signaling pathway map of plasminogen activator inhibitor-1 (PAI-1/SERPINE-1): a review of an innovative frontier in molecular aging and cellular senescence. Cell Communication Signal. 2024;22(1):544.
92. De Santis S, Vidmar K, Williams J, Mattioli B, Chieppa M, Del Pinto R, et al. Upregulation of serpine1/Pai-1 induced by Il-33 contributes to intestinal fibrosis in a Crohn's disease-like ileitis model. Inflamm Bowel Dis. 2024;30:S62.
93. Jójárt B, Molnár T, Szabó V, Varga Á, Resál T, Szántó K, et al. P459 Expression of SerpinE1, a potential new disease activity marker, reflects therapeutic response in inflammatory bowel disease. J Crohn's Colitis. 2021;15(Supplement1):S453–S.
94. Tang D, Huang Y, Che Y, Yang C, Pu B, Liu S, Li H. Identification of platelet-related subtypes and diagnostic markers in pediatric Crohn's disease based on WGCNA and machine learning. Front Immunol. 2024;15:1323418.
95. Wang Y, Wang J, Gao J, Ding M, Li H. The expression of SERPINE1 in colon cancer and its regulatory network and prognostic value. BMC Gastroenterol. 2023;23(1):33.
96. Yuan K, Hu D, Mo X, Zeng R, Wu B, Zhang Z, et al. Novel diagnostic biomarkers of oxidative stress, immune-infiltration characteristics and experimental validation of SERPINE1 in colon cancer. Discover Oncol. 2023;14(1):206.
97. Slany A, Bileck A, Kreutz D, Mayer RL, Muqaku B, Gerner C. Contribution of human fibroblasts and endothelial cells to the hallmarks of inflammation as determined by proteome profiling. Mol Cell Proteom. 2016;15(6):1982–97.
98. Flavell S, Hou T, Lax S, Filer A, Salmon M, Buckley C. Fibroblasts as novel therapeutic targets in chronic inflammation. Br J Pharmacol. 2008;153(S1):S241–6.
99. Gordon IO, Agrawal N, Willis E, Goldblum JR, Lopez R, Allende D, et al. Fibrosis in ulcerative colitis is directly linked to severity and chronicity of mucosal inflammation. Aliment Pharmacol Ther. 2018;47(7):922–39.
100. Amado-Azevedo J, van Stalborch A-MD, Valent ET, Nawaz K, van Bezu J, Eringa EC, et al. Depletion of Arg/Abl2 improves endothelial cell adhesion and prevents vascular leak during inflammation. Angiogenesis. 2021;24(3):677–93.
101. Chakrabarty S, Varghese VK, Sahu P, Jayaram P, Shivakumar BM, Pai CG, Satyamoorthy K. Targeted sequencing-based analyses of candidate gene variants in ulcerative colitis-associated colorectal neoplasia. Br J Cancer. 2017;117(1):136–43.
102. Levitte S, Yarani R, Ganguly A, Martin L, Gubatan J, Nadel HR, et al. Case series of precision delivery of methylprednisolone in pediatric inflammatory bowel disease: feasibility, clinical outcomes, and identification of a vasculitic transcriptional program. J Clin Med. 2023;12(6):2386.
103. Ke B-J, Abdurahiman S, Biscu F, Zanella G, Dragoni G, Santhosh S, et al. Intercellular interaction between FAP+ fibroblasts and CD150 + inflammatory monocytes mediates fibrostenosis in Crohn's disease. J Clin Investig. 2024;134(16).
104. Britzen-Laurent N, Weidinger C, Stürzl M. Contribution of blood vessel activation, remodeling and barrier function to inflammatory bowel diseases. Int J Mol Sci. 2023;24(6):5517.
105. Ippolito C, Colucci R, Segnani C, Errede M, Girolamo F, Virgintino D, et al. Fibrotic and vascular remodelling of colonic wall in patients with active ulcerative colitis. J Crohn's Colitis. 2016;10(10):1194–204.
106. Kyriakis JM, Avruch J. Mammalian MAPK signal transduction pathways activated by stress and inflammation: a 10-year update. Physiol Rev. 2012;92(2):689–737.
107. Quaglio AEV, Castilho ACS, Di Stasi LC. Experimental evidence of MAP kinase gene expression on the response of intestinal anti-inflammatory drugs. Life Sci. 2015;136:60–6.
108. Tang BL. ADAMTS: a novel family of extracellular matrix proteases. Int J Biochem Cell Biol. 2001;33(1):33–44.
109. Buran T, Batır MB, Çam FS, Kasap E, Çöllü F, Çelebi HBG, Şahin M. Molecular analyses of ADAMTS-1,-4,-5, and IL-17 a cytokine relationship in patients with ulcerative colitis. BMC Gastroenterol. 2023;23(1):345.
110. Lin P, Hua J, Teng Z, Lin C, Liu S, He R, et al. Screening of hub inflammatory bowel disease biomarkers and identification of immune-related functions based on basement membrane genes. Eur J Med Res. 2023;28(1):247.
111. Rieder F, Mukherjee PK, Massey WJ, Wang Y, Fiocchi C. Fibrosis in IBD: from pathogenesis to therapeutic targets. Gut. 2024;73(5):854–66.
112. Derkacz A, Olczyk P, Olczyk K, Komosinska-Vassev K. The role of extracellular matrix components in inflammatory bowel diseases. J Clin Med. 2021;10(5):1122.
113. He L, Kang Q, Chan KI, Zhang Y, Zhong Z, Tan W. The immunomodulatory role of matrix metalloproteinases in colitis-associated cancer. Front Immunol. 2023;13:1093990.
114. Gordon JN, Pickard KM, Di Sabatino A, Prothero JD, Pender SL, Goggin PM, MacDonald TT. Matrix metalloproteinase-3 production by gut IgG plasma cells in chronic inflammatory bowel disease. Inflamm Bowel Dis. 2008;14(2):195–203.
115. Jimbo K, Ohtsuka Y, Kojima Y, Hosoi K, Ohbayashi N, Ikuse T, et al. Increased expression of CXCR3 axis components and matrix metalloproteinase in pediatric inflammatory bowel disease patients. Pediatr Int. 2014;56(6):873–83.
116. Pan Z, Lin H, Fu Y, Zeng F, Gu F, Niu G, et al. Identification of gene signatures associated with ulcerative colitis and the association with immune infiltrates in colon cancer. Front Immunol. 2023;14:1086898.
117. Li R, Chi H, Liao X, Cen S, Zou Y. The Glabridin from Huangqin decoction prevents the development of ulcerative colitis into colitis-associated colorectal Cancer by modulating Mmp1/Mmp3 activity. Int Immunopharmacol. 2024;135:112262.
118. de Bruyn M, Breynaert C, Arijs I, De Hertogh G, Geboes K, Thijs G, et al. Inhibition of gelatinase B/MMP-9 does not attenuate colitis in murine models of inflammatory bowel disease. Nat Commun. 2017;8(1):15384.
119. Villablanca EJ, Selin K, Hedin CR. Mechanisms of mucosal healing: treating inflammatory bowel disease without immunosuppression? Nat Reviews Gastroenterol Hepatol. 2022;19(8):493–507.
120. Neurath MF, Travis SP. Mucosal healing in inflammatory bowel diseases: a systematic review. Gut. 2012;61(11):1619–35.
121. Cagnoni AJ, Giribaldi ML, Blidner AG, Cutine AM, Gatto SG, Morales RM, et al. Galectin-1 fosters an immunosuppressive microenvironment in colorectal cancer by reprogramming CD8 + regulatory T cells. Proc Natl Acad Sci. 2021;118(21):e2102950118.
122. Fernandez-Perez R, Lopez-Santalla M, Sánchez-Domínguez R, Alberquilla O, Gutiérrez-Cañas I, Juarranz Y, et al. Enhanced susceptibility of galectin-1 deficient mice to experimental colitis. Front Immunol. 2021;12:687443.
123. Papa-Gobbi R, Muglia CI, Rocca A, Curciarello R, Sambuelli AM, Yantorno M, et al. Spatiotemporal regulation of galectin-1-induced T-cell death in lamina propria from Crohn's disease and ulcerative colitis patients. Apoptosis. 2021;26(5):323–37.

124. Morosi LG, Cutine AM, Cagnoni AJ, Manselle-Cocco MN, Croci DO, Merlo JP, et al. Control of intestinal inflammation by glycosylation-dependent lectin-driven immunoregulatory circuits. Sci Adv. 2021;7(25):eabf8630.
125. Heldin CH, Lennartsson J, Westermark B. Involvement of platelet-derived growth factor ligands and receptors in tumorigenesis. J Intern Med. 2018;283(1):16–44.
126. Pellon-Cardenas O, Rout P, Hassan S, Fokas E, Ping H, Patel I, et al. Dynamic reprogramming of stromal Pdgfra-expressing cells during WNT-mediated transformation of the intestinal epithelium. bioRxiv. 2025.
127. Li J, Simmons AJ, Hawkins CV, Chiron S, Ramirez-Solano MA, Tasneem N, et al. Identification and multimodal characterization of a specialized epithelial cell type associated with Crohn's disease. Nat Commun. 2024;15(1):7204.
128. Nighot M, Ganapathy AS, Saha K, Suchanec E, Castillo EF, Gregory A, et al. Matrix metalloproteinase MMP-12 promotes macrophage transmigration across intestinal epithelial tight junctions and increases severity of experimental colitis. J Crohn's Colitis. 2021;15(10):1751–65.
129. Arosa L, Camba-Gómez M, Lorenzo-Martín LF, Clavaín L, López M, Conde-Aranda J. RNA Expression of MMP12 Is Strongly Associated with Inflammatory Bowel Disease and Is Regulated by Metabolic Pathways in RAW 264.7 Macrophages. Int J Mol Sci. 2024;25(6):3167.
130. Di Sabatino A, Saarialho-Kere U, Buckley MG, Gordon JN, Biancheri P, Rovedatti L, et al. Stromelysin-1 and macrophage metalloelastase expression in the intestinal mucosa of Crohn's disease patients treated with infliximab. Eur J Gastroenterol Hepatol. 2009;21(9):1049–55.
131. Mäkitalo L, Kolho K-L, Karikoski R, Anthoni H, Saarialho-Kere U. Expression profiles of matrix metalloproteinases and their inhibitors in colonic inflammation related to pediatric inflammatory bowel disease. Scand J Gastroenterol. 2010;45(7–8):862–71.
132. Oishi Y, Manabe I. Macrophages in inflammation, repair and regeneration. Int Immunol. 2018;30(11):511–28.
133. Garten A, Schuster S, Penke M, Gorski T, De Giorgis T, Kiess W. Physiological and pathophysiological roles of NAMPT and NAD metabolism. Nat Reviews Endocrinol. 2015;11(9):535–46.
134. Hong SM, Lee A-Y, Hwang S-M, Ha Y-J, Kim M-J, Min S, et al. NAMPT mitigates colitis severity by supporting redox-sensitive activation of phagocytosis in inflammatory macrophages. Redox Biol. 2022;50:102237.
135. Colombo G, Clemente N, Zito A, Bracci C, Colombo FS, Sangaletti S, et al. Neutralization of extracellular NAMPT (nicotinamide phosphoribosyltransferase) ameliorates experimental murine colitis. J Mol Med. 2020;98(4):595–612.
136. Colombo G, Caviglia GP, Ravera A, Tribocco E, Frara S, Rosso C, et al. NAMPT and NAPRT serum levels predict response to anti-TNF therapy in inflammatory bowel disease. Front Med. 2023;10:1116862.
137. Lei J, Lv L, Zhong L, Xu F, Su W, Chen Y, et al. The Gut Microbiota Affects Anti-TNF Responsiveness by Activating the NAD+ Salvage Pathway in Ulcerative Colitis. Adv Sci. 2025;12(8):2413128.
138. Starr AE, Deeke SA, Ning Z, Chiang C-K, Zhang X, Mottawea W, et al. Proteomic analysis of ascending colon biopsies from a paediatric inflammatory bowel disease inception cohort identifies protein biomarkers that differentiate Crohn's disease from UC. Gut. 2017;66(9):1573–83.
139. Tian L, Gao H, Yao T, Chen Y, Gao L, Han J, et al. Interactions between NAD+ metabolism and immune cell infiltration in ulcerative colitis: subtype identification and development of novel diagnostic models. Front Immunol. 2025;16:1479421.
140. Xu L, Chen Y, Liu L, Hu X, He C, Zhou Y, et al. Tumor-associated macrophage subtypes on cancer immunity along with prognostic analysis and SPP1-mediated interactions between tumor cells and macrophages. PLoS Genet. 2024;20(4):e1011235.
141. Rogler G. Chronic ulcerative colitis and colorectal cancer. Cancer Lett. 2014;345(2):235–41.
142. Saiz-Gonzalo G, Hanrahan N, Rossini V, Singh R, Ahern M, Kelleher M, et al. Regulation of CEACAM family members by IBD-associated triggers in intestinal epithelial cells, their correlation to inflammation and relevance to IBD pathogenesis. Front Immunol. 2021;12:655960.
143. Liu J-N, Wang H-B, Zhou C-C, Hu S-Y. CEACAM5 has different expression patterns in gastric non-neoplastic and neoplastic lesions and cytoplasmic staining is a marker for evaluation of tumor progression in gastric adenocarcinoma. Pathology-Research Pract. 2014;210(10):686–93.
144. Holmer R, Wätzig GH, Tiwari S, Rose-John S, Kalthoff H. Interleukin-6 trans-signaling increases the expression of carcinoembryonic antigen-related cell adhesion molecules 5 and 6 in colorectal cancer cells. BMC Cancer. 2015;15(1):975.
145. Li Y, Cao H, Jiao Z, Pakala SB, Sirigiri DNR, Li W, et al. Carcinoembryonic antigen interacts with TGF-β receptor and inhibits TGF-β signaling in colorectal cancers. Cancer Res. 2010;70(20):8159–68.
146. Han SU, Kwak TH, Her KH, Cho Y-H, Choi C-w, Lee HJ, et al. CEACAM5 and CEACAM6 are major target genes for Smad3-mediated TGF-β signaling. Oncogene. 2008;27(5):675–83.
147. Geyl S, Guillo L, Laurent V, d'Amico F, Danese S, Peyrin-Biroulet L. Transmural healing as a therapeutic goal in Crohn's disease: a systematic review. lancet Gastroenterol Hepatol. 2021;6(8):659–67.
148. Fernandes SR, Rodrigues RV, Bernardo S, Cortez-Pinto J, Rosa I, da Silva JP, et al. Transmural healing is associated with improved long-term outcomes of patients with Crohn's disease. Inflamm Bowel Dis. 2017;23(8):1403–9.
149. Li Y, de Haar C, Chen M, Deuring J, Gerrits MM, Smits R, et al. Disease-related expression of the IL6/STAT3/SOCS3 signalling pathway in ulcerative colitis and ulcerative colitis-related carcinogenesis. Gut. 2010;59(2):227–35.
150. Gui X, Iacucci M, Ghosh S. Dysregulation of IL6/IL6R-STAT3-SOCS3 signaling pathway in IBD-associated colorectal dysplastic lesions as compared to sporadic colorectal adenomas in non-IBD patients. Pathology-Research Pract. 2020;216(11):153211.
151. Gao Y, Wang J, Zhou Y, Sheng S, Qian SY, Huo X. Evaluation of serum CEA, CA19-9, CA72-4, CA125 and ferritin as diagnostic markers and factors of clinical parameters for colorectal cancer. Sci Rep. 2018;8(1):2732.
152. Guarino M, Rubino B, Ballabio G. The role of epithelial-mesenchymal transition in cancer pathology. Pathology. 2007;39(3):305–18.
153. Kelleher M, Singh R, O'Driscoll CM, Melgar S. Carcinoembryonic antigen (CEACAM) family members and inflammatory bowel disease. Cytokine Growth Factor Rev. 2019;47:21–31.

Publisher's note