CHAPTER 7

Identification of novel population-specific cell subsets in Chinese ulcerative colitis patients using Single-cell RNA sequencing

Published in *Cellular and Molecular Gastroenterology and Hepatology*, 2021.

doi: 10.1016/j.jcmgh.2021.01.020
Identification of novel population-specific cell subsets in Chinese ulcerative colitis patients using Single-cell RNA sequencing

Short Title: scRNA-seq in Chinese UC patients

Guang Li1#, Bowen Zhang2#, Jianshu Hao1, Xiaojing Chu2,3, Miriam Wiestler4, Markus Cornberg2,4, Cheng-Jian Xu2,4,5, Xinjuan Liu1 and Yang Li2,5

1Department of Gastroenterology, Beijing Chaoyang Hospital, Capital Medical University, Chaoyang District, Beijing, 100024, China
2Centre for Individualised Infection Medicine (CiIM) and TWINCORE, joint ventures between the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany
3Department of Genetics, Groningen, University of Groningen, University Medical Centre Groningen, the Netherlands
4Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany.
5Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands

#Authors share co-first authorship.
*Correspondence
Address correspondence to: Xinjuan Liu, PhD, Department of Gastroenterology, Beijing Chaoyang Hospital, Capital Medical University, Chaoyang District, Beijing, 100020, China
Email: liuxinjuan@mail.ccmu.edu.cn

Abstract

Background & Aims
Genome-wide association studies (GWAS) and transcriptome analyses have been performed to better understand the pathogenesis of ulcerative colitis (UC). However, current studies mainly focus on European ancestry, highlighting a great need to identify the key genes, pathways and cell types in colonic mucosal cells of adult UC patients from other ancestries. Here we aimed to identify key genes and cell types in colonic mucosal of UC.

Methods
We performed Single-cell RNA sequencing (scRNA-seq) analysis of 12 colon biopsies of UC patients and healthy controls from Chinese Han ancestry.

Results
Two novel plasma subsets were identified. Five epithelial/stromal and three immune cell subsets show significant difference in abundance between inflamed and non-inflamed samples. In general, UC-risk genes show consistent expression alteration in both Immune cells of inflamed and non-inflamed tissues. As one of the exceptions, IgA deflection, marking the signal of immune dysfunction, is specific to the inflamed area. Moreover, Th17 derived activation was observed in both epithelial cell lineage and immune cell lineage of UC patients as compared to controls, suggesting a systemic change of immune activities driven by Th17. The UC-risk genes show enrichment in progenitors, glial cells and immune cells, and drug-target genes are differentially expressed in antigen presenting cells.

Conclusions
Our work identifies novel population-specific plasma cell molecular signatures of UC. The transcriptional signature of UC is shared in immune cells from both inflamed and non-inflamed tissues, whereas the transcriptional response to disease is a local effect only in inflamed epithelial/stromal cells.

Keywords
Ulcerative colitis, Single-cell RNA sequencing, Genome-wide association studies
Introduction
Ulcerative colitis (UC) is a chronic inflammatory disease, which is characterized by relapsing and remitting mucosal inflammation in the rectum and extending to proximal segments of the colon. The pathogenesis of UC is complex and multifactorial, including genetic predisposition, epithelial barrier defects, and deregulated immune responses. In general, UC is thought to arise from an inappropriate activation of the intestinal mucosal immune system in response to commensal bacteria in a genetically susceptible host. The breakdown of the epithelial barrier and mucosal immune barrier homeostasis at the cellular level has been revealed to play an important role in the onset of UC.

UC is common in industrialized locations including North America and Western Europe. Recently, there is also an increasing incidence of UC in Asia due to urbanization. The recent epidemiological report showed that India and China had the highest inflammatory bowel disease (IBD) incidence in Asia. In China, the incidence of UC was reported to be positively associated with gross domestic product. Genome-wide association studies (GWAS) and transcriptome analyses have been widely used to dissect the molecular mechanisms of UC, by identifying risk alleles as well as transcriptional alterations aiming at defining the functional consequences of associated alleles for both coding and non-coding genetic variation. To better understand the pathogenesis of UC, recent studies have been focused on the transcriptome at a cellular resolution and states within the tissue of IBD lesions. However, most of the current studies were performed in individuals of European ancestry, with only one study focusing on pediatric-onset colitis in Chinese, highlighting a great demand for studying on the key genes, pathways and cell types in colonic mucosal cells of adult UC patients from other ancestries.

Here, we performed a single-cell RNA sequencing (scRNA-seq) analysis of 12 colon biopsies from 5 UC patients including 4 inflamed (UC), 4 non-inflamed (self-control, SC) biopsies and 4 healthy biopsies (HC) from healthy individuals from Chinese Han ancestry (Fig 1A, Table S1A), to uncover the pathogenesis of UC on a cellular level.

Figure 1. Single-cell expression atlas and cell typing in biopsies of Chinese UC patients and control samples. A, Experimental design. Fresh biopsy specimens were disassociated and single-cell suspensions were obtained from four samples from inflamed sigmoid colon (UC) with the non-inflamed ascending colon biopsy specimens (SC) and four healthy samples (HC). The UMAP (uniform manifold approximation and projection) plot identifies 10 epithelial cell clusters and 11 immune cells from 12 colon biopsies. C, Violin plots showing the expression distribution of selected marker genes across cell clusters. D, UMAP shows the lineage markers, PTPRC for immune cells, EPCAM for epithelial cells, COL1A1 for Fibroblasts, VWF for Endothelial cells. E, Box plots showing percentage of cell clusters of total specimens (per biopsy) in inflamed samples relative to non-inflamed samples and healthy samples (significant changes of Dirichlet-multinomial regression adjusted with the Benjamini-Hochberg method were marked out, * p<0.05, ** p<0.01).
Results

Two novel plasma cells were identified from all colon biopsies.

After quality control, 43,218 cells with an average of 1053 genes per cell remained and clustered into 21 subpopulations (Fig.1B), where cells from each of the UC, SC and HC biopsies were mapped to a comparable space in UMAP after batch correction (Fig.S1). We calculated marker genes for each cluster (Table.S1B), then annotated the clusters with both data-derived methods (Fig.1C, Fig.S2-S4) and literature-derived methods (Fig.1D, Fig.S5-S6), and end up with 10 epithelial/stromal cell types (e.g. enterocyte, enterocyte progenitors, goblet, goblet progenitors, LGR5+ stem cell, CLP/Paneth-like cells, fibroblasts, TRMP5+ tuft cell, glial cells and endothelial cells) and 11 immune cells (e.g. Naive T cells, Memory T cells, CD8+ T cell/NK cell, CD8+ T cell, naive B cell, memory B cell, monocytes/DC cells, mast cells and 3 clusters of plasma cells). Among them, 16 out of 21 subsets were replicated in the scRNA-seq study from American/European descent. Interestingly, we observed two novel immune subsets (plasma-2 and plasma-3 cells) in our data.

The proportions of five epithelial/stromal and three immune cell subsets significantly differed between UC and SC (or HC). glial cells, fibroblasts, goblet progenitors and enterocyte with its progenitors were significant decreased in inflamed UC tissues. CD8+ T cell as well as two types of plasma cells were increased in UC (Fig. 1E-1F). Tuft cells are the chemosensory cells in the gut and are enriched for taste-sensing molecules. We observed a cluster of mature tuft cells (named as TRMP5+ tuft), which highly expressed immune-related genes (AZGP1, PTPN18 and BMX) and neuronal signaling genes (AVIL, HTR3E and ITPR2). Of note, TRMP5+ tuft cell also showed high expression level on HPGD, ALOX5 and PTGST1, which function in the metabolism of arachidonic acid and prostaglandin. Moreover, TRMP5+ tuft cells also expressed IL-17RB, which may mediate the cross-talk with ILC2 and acted as a marker of tuft cell-like human colorectal cancer stem cells, as previously identified in the Chinese pediatric UC. Notably, we showed a significantly lower abundance of TRMP5+ tuft cells in UC compared to that in SC (paired student’s t-test, P=0.048) (Fig. 1E and Table S2A).

On the other hand, two out of three plasma subsets were observed with a higher proportion in UC, compared to that in either HC or SC (Fig. 1F and Table. S2A). Among the mentioned two plasma subsets, plasma-2 cells specifically expressed MZB-1 gene (Fig. S7A, Table. S1B-C), which was significantly up-regulated in UC compared to HC (Padj = 1.18×10^{-65}) or SC (Padj = 3.45×10^{-37}). Plasma-3 cells specifically expressed IGLL5 gene, which was also significantly up-regulated in UC compared to HC (Padj = 1.16×10^{-18}) or SC (Padj = 1.83×10^{-18}). Consistent with previous findings, the up-regulation of MZB-1 and IGLL5 in Crohn’s Disease (CD) has been reported in mesenteric adipose tissue. Furthermore, genes co-expressed with MZB-1 and IGLL5 in those cell-types were enriched in endoplasmic reticulum, which play important roles in cytosol transport, suggesting its involvement in the inflammatory cascades of plasma cells (Fig. S7B).

In order to directly compare our plasma signatures to those of American/European ancestry, we intersected identified DE genes between UC and SC/HC in plasmas of our study with those found in Smillie’s study (Table S3A-B). Most up-regulated genes (270 out of 502) found in our study, such as XBP-1 and MHC class II genes, have been found consistently up-regulated in Smillie’s study (Table. S3C). On the other hand, two plasma maker genes: IGLL5 and MZB-1 showed up-regulation in UC patients in our study, but were found to be down-regulated (IGLL5) and not changed (MZB-1) respectively in Smillie’s study. This observation indicates that these two genes expression alteration could be specific to the plasma of Chinese UC patients (Table. S3).

Antigen presentation pathway and IL-17 signaling pathway were activated in the inflamed tissue of ulcerative colitis.

In epithelial/stromal lineages, up-regulated genes between inflamed tissues and HC were enriched in the antigen processing and presentation pathway and MHC class II complex activity. These pathways were also showed up in UC compared with SC (Fig. 2A, 2B).

Figure 2. Gene signatures found in the inflamed samples (UC) comparing to non-inflamed samples (SC) and healthy samples (HC) in epithelial/stromal cells. A-B, Enriched KEGG and GO pathways of DE genes between UC vs HC (A) and UC vs SC (B). C, Heatmaps showing the expression changes (UC vs HC) of detected genes in IL-17 signaling pathways.
and Allograft rejection pathway in each Epithelial/stromal cell cluster. D, Violin plots showing the expression distribution of LCN2 and MHC class II genes in each Epithelial/stromal cell cluster of UC, SC and HC.

The differentially expressed genes (DEG) between UC and HC in LGR5+ stem cells, enterocyte progenitors and goblet progenitors over-represented the functions of rejective immunity, but mature cells (enterocyte and goblet) lose the functional difference of rejective reaction. Of note, the majority of genes involved in MHC class II molecules, such as HLA-DRA, HLA-DRB1 and HLA-DRC, were up-regulated in UC. but the MHC class I molecular HLA-B was down-regulated.

Moreover, IL-17 signaling pathway was significantly enriched in the up-regulated genes of goblet progenitor, endothelial cell and fibroblasts in the comparison of UC versus HC, suggesting that intestinal epithelial barrier was involved in IL-17 cytokine responses in UC. One of the major genes of the IL-17 signaling pathway in these cell types was LCN2 (Fig. 2C). LCN2, a bacteriostatic molecule, has tissue destructive effects and is pro-inflammatory with chemotactic molecule binding properties26. Earlier studies, using both DNA microarrays and RNA-seq data, showed that the LCN2 is among the top 10 up-regulated genes in UC, and correlated with disease severity25-27. In addition, IL-17A showed a synergistic effect with IL-22 and TNF-α in inducing colonic epithelial expression of LCN228.

Next, we investigated whether the DE genes showing consistent pattern between Chinese UC patients and patients of American/European ancestry. We intersected DE genes found between UC and SC/HC in epithelial cells of our study with those found in Smillie’s study14. Most up-regulated genes (299 out of 517) found in our study also showed up-regulation in the UC patients of Smillie’s study (Table. S3C), including LCN2, JUN and MHC class genes, etc. The up-regulated genes function in our study were enriched in pathways like translation initiation and structural molecule activity (Fig.S8).

Th17 cell was the major immune characteristic of UC, but IgA was the key local immune component of inflamed tissue

Similar to the activation of IL-17 signaling pathway in the epithelial lineages, the Th17 cell differentiation, as well as the T cell receptor signaling and TNF signaling pathway were activated in CD4+ cell lineages in UC compared to HC. In general, immune-mediated inflammation, the presented antigen leads to the differentiation of CD4+ helper T cells, and stimulates Th17 cells to produce proinflammatory cytokines, such as TNF and IL-1729. On the other hand, we found that the Th17 cell differentiation and its cytokines related pathways were activated in NKT cells, which is in line with the previous finding that CD8+IL-17+ cells were increased in American UC patients2. Additionally, we found that NFκBIA and JUN were predominantly presented in plasma cells, mast cells, monocytes and dendritic cells in HC, but were only expressed in B cells and T cells in both UC and SC, whereas additional MHC class II genes (e.g. HLA-DRA) were found in CD8+ T and NKT cells (Fig. 3C).

The intestinal IgA production was found to be enriched in the up-regulated genes of goblet and enterocytes cells, their progenitor cells, and glial cells in comparison between UC and SC (Fig. 3B). Most of the IgA molecules presented at mucosal sites, where they are produced by locally residing plasma cells30. As a key local immune component, IgA protects the integrity of intestinal mucosal barrier by coating the bacteria31. In line with these findings, abnormal coating capacity of IgA was reported in UC in Chinese young patients32.

195 UC-risk genes were enriched in the epithelia progenitors and immune cells with druggable targets in antigen presenting cells

Among 511 UC-risk genes published in the GWAS catalog [https://www.ebi.ac.uk/gwas/, last accessed on 22-Feb-2020], 141 were found in 195 differentially expressed genes between UC and controls, of which 90% showed up-regulation in UC. Compared to the number of height-associated loci overlapped with our DE genes, UC-risk loci showed enrichment in epithelial cells (goblet, goblet progenitors, LGR5+ stem cell and enterocyte), endothelial cell, glial cells, fibroblasts and various immune cells (plasma-1, Naive and memory B cell, CD4+ T cell and NKT cell) (Table S2B).

Among these UC-risk DE genes, eight genes were previously reported in the UC GWAS studies in the Asian population: CFB, HLA-DQA1, HLA-DOB1, HLA-DRA, HLA-DRB1, IRF8, PTPRC and SLC26A3. Particularly, HLA-C, HLA-DRA and HLA-DRB1 were up-regulated in the epithelial cells of UC compared with both SC and HC (Fig. 4A-B and Fig.S9), suggesting an increased activity of antigen presentation during
Inflammation in the inflamed tissue, which is consistent with the enrichment pathways of DEG reported above.

**Figure 4.** Comparison of signed log-P-values of differentially expressed (DE) genes from UC risk loci in epithelial/stromal cell clusters (A) and immune cells (B).

In T cell lineages, FYN, PTPRC, and CDC42SE2 were increased in UC, and they are known to be involved in the T cell receptor signaling via regulating the receptor-like tyrosine phosphataseβ. Similarly, immunosuppressive costimulatory molecules, CTLA4 with its receptor ICOS, were increased in CD4+ Naive T cell and CD4+ memory T cell, respectively. Of note, STAT3 is a recognized pro-inflammatory transcripts. It is known that the differentiation of Th17 cells typically requires the cytokine signaling via the STAT3 transcription activator. STAT3 was up-regulated in CD4+ Naive T cell of UC and SC, which was also consistent with the change of Th17 cell and IL-17 signal in systemic inflammation.

Our analysis showed that many MHC molecules were up-regulated in the epithelial/stromal cells of UC compared to either HC or SC. In immune cells, however, many dysregulated risk genes were identified between UC and HC but not between UC and SC. For example, LTB, DUSP1, and IRF8 of memory B cell and T cells were increased in UC compared to HC but not to SC. Therefore, we performed a systematic comparison of significance of DE UC-risk genes in immune and epithelial/stromal cells. It reveals contrasting patterns: the identified DE risk genes between UC and HC are more likely to be differentially expressed between UC and SC in epithelial/stromal cells than in immune cells ($P_{\text{adj}} = 1.35 \times 10^{-3}$, Fig. 4A-B and Fig. S9).

Interestingly, we also noticed that 25% of the DE genes are druggable targets. Of note, the enrichment of DE genes in the known UC-drug targets was found for Basiliximab ($P = 0.049$) and Adalimumab ($P = 0.042$) in glial cells, and Abatacept in Monocytes and dendritic cells ($P = 0.008$).

**Discussion**

Here we demonstrate novel molecular signatures of adult Chinese UC patients at a single-cell level. Most of the cell types we identified can be replicated in the previous studies in the American UC patients. The enteric nervous system (ENS) plays a pivotal role in rectifying and orchestrating the inflammatory responses in gut tract. Enteric glial cells, have been recognized as APCs, express substance P and produce TNF-α, IL-1β, and IL-6, which could induce the activation of mast cells, macrophages, and T cells, and promote lymphocyte proliferation. We identified the decreased glial cells in inflamed tissues compared with both SC and HC, similar to American UC patients, but not detected in Chinese pediatric UC patients.

Tuft cells have also been identified from UC patients. Theoretically, Tuft cells are present in the stratified squamous epithelium of the anal canal and esophagus but may increase following replacement with a metaplastic, intestine-like columnar epithelium. A tuft cell signature based on bulk profiles of TRPM5+ tuft cells contained both neuronal and inflammatory gene programs; this could reflect either co-expression in the same cells or distinct subsets. TRPM5 plays a crucial role for chemosensation in promoting tuft cell expansion in response to infection. Tuft cells also interact closely with immune cells, playing a crucial role in the cellular regulatory network coordinating responses to luminal parasites.

In the epithelial lineages, most progenitor cells were decreased in UC. We also identified LGR5+ stem cells, the major intestinal stem cell, which play a key role in regeneration of intestinal injury, in our data. In general, the intestinal epithelium is maintained by long-lived intestinal stem cells (ISCs), residing the crypt base, the specifically expressed marker is LGR5+. Above the ISC zone, there are short-lived progenitors that normally give rise to lineage-specific differentiated cell types but can dedifferentiate into ISCs in the case of injury. The short-lived enterocyte precursors just serve as a large reservoir of potential stem cells during crypt regeneration. Thus, the reducing enterocyte progenitors and matured cells suggested the deficiency of mucosal regeneration in inflamed UC.

Pericytes and enteroendocrine cells have been identified in the previous study by Smillie et al., but were largely missed in our data. A small number of pericytes and enteroendocrine cells were found but mixed with fibroblasts and LGR5+ stem cells, respectively. The lack of resolution in detecting pericytes and enteroendocrine cells is likely due to the smaller sample size of our study compared to that of Smillie’s study. Epithelial barrier and immune barrier defects are strongly implicated in the pathogenesis of ulcerative colitis with significant dysbiosis, although there were no UC-related specific bacteria identified. The DEGs were enriched in the function of antigen presentation and MHC class II complex activity in epithelial lineages of UC, and DEGs of the most progenitors were involved in the function of reactive immunity. Most differential genes code the antimicrobial proteins, such as LCN, a bacteriostatic molecule, involved in the antimicrobial immune response. CXCL1 and CXCL2 are involved as the antimicrobial humoral immune response mediated by antimicrobial peptide, acting as the recruiters of immune cells. Both of them are the risk genes and were verified up-regulated in UC.

Type 17 immunity involved in the inflammation in the pathogenesis of ulcerative colitis has been implicated, and Type 17 immunity were derived by anti-commensal response or antigen presentation in UC. We found in the DEGs in epithelial/stromal lineages that Th17 signals were increased not only in the inflamed tissue but also in the uninfamed control tissue, with the bacteriostatic molecule LCN2.
expression. In addition, Th17 cell differentiation and T cell receptor signaling were enriched in the CD4+ T cell lineage, which were consistent with the functional change in epithelial lineages. Moreover, the UC risk gene STAB3 was also increasingly expressed in CD4+ Naive T cell of UC and itself control.

Many studies have reported that the increased levels of Th17 cells and IL-17 not only in the intestinal mucosa but also in the Peripheral Blood Mononuclear Cells (PBMCs) and serum of active UC in Chinese population. It is also known that the level of Th17 cell was increased in PBMC, and the expression of IL-17A mRNA were increased in the PBMC, MLN and LPC of DSS colitis mice. Above all, type 17 immunity is one of the important systemic changes in UC.

In epithelial lineages, up-regulated genes between inflamed tissues and HC were replicated in the comparison between UC and SC, but in immune cell lineages, most DEGs identified between inflamed intestines and HC cannot be found as DEGs in the comparison between inflamed intestines and SC. These observations indicated the local intestinal damage as well as a systemic inflammation were involved in the pathogenesis of UC. Particularly, in line with that the type IL-17 immunity was increased in the PBMC, MLN and LPC of DSS colitis mice. Above all, type 17 immunity is one of the important systemic changes in UC.

In summary, most cell types could be found in the colon of both Chinese UC patents and healthy controls, while we identified two novel plasma subsets in our data. The transcriptional signature of UC is shared in immune cells from both inflamed and non-inflamed tissues, except the deflection of IgA, which is a located immunity dysfunction of inflamed area. However, in epithelial/stromal cells, the transcriptional response to disease is a local effect seen only in the inflamed biopsy. The activation of Th17 driven by anti-commensal inflammatory response is the core of the pathogenesis of UC, nearly including all the changes of epithelial cell lineage and immune cell lineage. In addition, the drug target genes were differentially expressed in antigen presenting cells. Our study serves as an important reference of the molecular mechanism behind the genetic risks of UC from transcriptional aspects, and identifies cell type specific drug targets.

**Methods**

**Patient Selection and Sample Collection**

The biopsy samples were collected from four male and one female Han Chinese patients at the Department of Gastroenterology, Beijing Chaoyang Hospital of Capital Medical University (Beijing, China). All patients have been diagnosed as left-sided moderate ulcerative colitis for at least 3 years, with relapsing course, without receiving any treatment in recent 3 months. None of them had undergone surgical resections. UC was diagnosed by the conventional clinical, radiological and endoscopic features, and eventually confirmed by histological examination of colonic biopsies. Four Han Chinese healthy controls were enrolled at the Health Examination Center of Beijing Chaoyang Hospital of Capital Medical University. Among them, three healthy controls were age- and sex-matched with three UC patients. Written informed consent was obtained and ethical approval was granted by the ethics committee of the Beijing Chaoyang Hospital, Capital Medical University. Characteristics of all patients are represented in Table S1A. For each of the five patients, one pinch biopsy specimen was collected from the inflamed sigmoid colon (the most common site of inflammation in UC) as the UC group, and one pinch biopsy specimen from the normal ascending colon of patients served as self-control (SC group), as well as four pinch biopsy specimens from sigmoid of four healthy volunteers served as healthy control (HC group). Biopsy specimens were collected into RPMI 1640 medium on ice and processed immediately.

**Tissue Processing**

Tissues were washed twice with PBS. The biopsy specimens were cut into 1mm³ pieces using sterile scalpel blades, and put into a petri dish. 2 mg/mL collagenase II and 10 U/mL DNAase were added and rotated at 37°C for a period of time.

After standing for 2-3 minutes, decant the supernatant and remove the large lumps with the filter membrane. After centrifuging the cells, the supernatant portion was poured out and discarded. The cells were suspended again with erythrocyte lysis buffer, cultured at room temperature for 2-3 minutes, and then centrifuged at 120×g for 3 minutes at 4°C. Samples were lastly resuspended with PBS.

**Single-Cell RNA Sequencing**

Cell capture and cDNA synthesis was performed using single-cell 3' Library and Gel Bead Kit V2 (10x Genomics, 120237) and Chromium Single Cell A Chip Kit (10x Genomics, 120236). The cell suspension (300-600 living cells per microliter determined by Count Star) was loaded onto the Chromium single cell controller (10x Genomics, 120236). The cell suspension (300-600 living cells per microliter determined by Count Star) was loaded onto the Chromium single cell controller (10x Genomics) to generate single-cell gel beads in the emulsion according to the manufacturer’s protocol. In short, single cells were suspended in PBS containing 0.04% BSA. About 7000 cells were added to each channel, and the target cell recovery rate was estimated to be 3000 cells. Captured cells were lysed and the released RNAs were barcoded through reverse transcription in individual Gel Beads in Emulsion (GEMs).

Using a S1000TM Touch Thermal Cycler (Bio Rad) to reverse transcribe, the GEMs were programed at 53°C for 45 min, followed 85°C for 5 min, and hold at 4°C. The cDNA was obtained and amplified, and the quality was assessed using the Agilent 4200 (performed by CapitalBio, Beijing).

Single-cell RNA-Seq libraries were prepared according to the manufacture’s introduction and the Single Cell 3’ Library Gel Bead Kit V2 was used. Sequencing was performed on the Illumina Novaseq 6000 sequencer with a sequencing depth of at least 69,000 reads per cell and 150 bp (PE150) paired-end reads (performed by CapitalBio, Beijing).

In addition, one UC sample (patient 2) and one self-control (patient 1) sample were excluded for further analyses because a high percentage of mitochondrial genes (>=0.25) and the lower number of recovered cells (688) were estimated in them.

**Reads processing and quality control**

CellRanger v3.0.2 (10x Genomics) was used to process single-cell RNA sequencing reads. To generate a digital gene expression (DGE) matrix for each sample, we...
mapped their reads to hg19 human transcriptome, and recorded the number of UMIs for each gene in each cell. For each DGE matrices, the estimated number of cells and number of genes per cell were examined based on cell barcode and UMI (Unique Molecular Identifier) barcodes.

Seurat v3.1 was applied to analyze DGE matrix from each sample61, 62. To filter out low-quality cells and doublets, empirically filtering criteria were applied to each cell: number of estimated genes should be higher than 100 and lower than 6000, and the ratio of reads mapping to the mitochondria should be lower than 25%. Only genes detected in at least 5 cells were maintained for subsequent analyses.

Clustering and identification of cell clusters

Seurat v3.1 integration workflow with SCTransform normalization method was used to cluster cells from different samples into distinct cell subsets62. We followed this workflow with the following steps: Firstly, we SCTransformed each sample and merged them into UC, SC and HC datasets. Next, we selected 2,000 variable features among three datasets and identified anchors from these features to integrate the datasets. These two steps corrected batch effects and prevented cells clustering by patients or disease phenotypes rather than by cell types or cell subsets. Principal component analysis (PCA) has then been performed on the integrated datasets, followed by Shared Nearest Neighbor (SNN) Graph construction using PC1 to 20 and k=20 nearest neighbours to identify unsupervised cell clusters. Finally, Uniform Manifold Approximation and Projection (UMAP) was used to visualize the cell clusters.

In order to keep the biological differences for downstream analyses, the above-mentioned batch correction was only used in the cell clustering and PCA related steps. For the other analyses, we used standard LogNormalization methods. The original gene counts for each cell were normalized by total UMI counts and multiplied by 10,000 (TP10K), and then log transformed by log (TP10K+1).

In order to annotate cell identity to each cluster, we used a double-checking strategy for the inference63 by comparing data-derived marker genes with public databases, and by directly visualizing the expression pattern of literature-derived marker genes. Firstly, we used the automatic cell annotation tool, SingleR64, with two reference dataset, Human Primary Cell Atlas Data and Blueprint Encode data, to generate primary annotation. Next, data-derived marker genes were detected by applying differential expression (DE) tests between cells in one cluster and all other cells in the dataset. Up-regulated genes from the cluster of interest were ranked by the Wilcoxon rank-sum test and compared with their reported cell types in human large intestines in CellMarker databases510/10/21 5:23:00 PM. Then, marker genes and cell surface markers reported in other human intestine or immune cell analyses were regarded as literature-derived markers, and we visualized their expression levels in each of the identified cell clusters to manually check the cell identities. Finally, we used the harmony algorithm65 to integrate our data with well annotated-cells from Smillie et al66, to further validate the identities.

To test significant changes of cell proportion, the Dirichlet-multinomial regression which takes compositional dependencies into account, was used. For comparison and validation, the student’s t test, the Wilcoxon signed-rank test and paired t test (when applicable) were also applied.

Differential expression genes and enrichment analysis

For each annotated cell type, differential expression genes were estimated between UC and SC, UC and HC, SC and HC, using FindMarkers function in Seurat v3.1 with the default Wilcoxon rank-sum test, and with MAST67 for validation. For each comparison, DE tests were performed only on genes that were detected in more than 10% of cells in any groups. P-value adjustment was performed using Bonferroni correction based on the total number of genes in the tested dataset.

R package clusterProfiler v3.10.168 was used for KEGG and GO enrichment analysis for over-represented pathways and GO terms on the DE genes found in UC vs HC and UC vs SC in each cell identity.

DE genes in GWAS loci

Ulcerative colitis GWAS data(2020-02-22-EFO_0000729) were downloaded from NHGRI-EBI GWAS catalog69, which contains 281 associated loci with 605 reported risk genes. Risk genes that were found to be significantly differentially expressed in any of our comparison were extracted and classified as Asian and Non-Asian risk genes based on whether they are initially estimated or have been replicated in Asian samples. In total, we identified 195 DE genes that are reported to be associated with ulcerative colitis, and only 8 of them have been reported in Asian GWAS studies, which are CFB, HLA-DOA1, HLA-DQB1, HLA-DRA, HLA-DRB1, IRF8, PTPRC and SLC26A3. The Fisher’s exact test was applied to test the over-representation of risk gene in differentially expressed genes in any cell types, using all genes expressed in >10% cells in estimated cell types as reference.

Drug target analysis

Druggable gene list was obtained from Finan C’s work36. The UC drugs information was collected from literatures70,71. Drug-targets were obtained from Drugbank72. Fisher’s exact test was applied for enrichment analysis.

Acknowledgments

The authors thank the patients-participants at the Beijing Chaoyang hospital for contributing intestinal biopsy samples.

Disclosures:

The authors declare that they have no conflicts of interest.

Author contributions:

Yang Li and Xinjuan Liu designed the study. Guang Li, Xinjuan Liu and Jianyu Hao recruited patients and performed Endoscopy and the laboratory experiments. Bowen Zhang, Xiaojing Chu, Guang Li and Xinjuan Liu performed the data analyses with the supervision of Yang Li. Xinjuan Liu, Guang Li, Bowen Zhang, Cheng-Jian Xu and Yang Li interpreted data and wrote the manuscript draft. Miriam Wiestler and Markus
Cornberg provided supports with data interpretation. All authors assisted in the writing and reviewing of the manuscript and approved the final manuscript.

**Abbreviations used in this paper:**
UC: ulcerative colitis, IBD: inflammatory bowel disease, CD: Crohn’s disease, scRNA-seq: single-cell RNA sequencing, GWAS: Genome-wide association studies, DEG: the differentially expressed genes, SC: self-control, HC: healthy biopsies, PCA: Principal component analysis, UMAP: Uniform Manifold Approximation and Projection

**Grant Support:**
Xinjuan Liu was supported by the National Natural Science Foundation of China (82070559), National Natural Science Foundation of Beijing (7192072), and the Project of Digestive Medical Coordinated Development Center of Beijing Municipal Administration of Hospitals (XXT11). Yang Li was supported by an ERC starting Grant (948207) and a Radboud University Medical Centre Hypatia Grant [2018].

**Reference**
1. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. Lancet 2017;389:1756-1770.
2. Liu TC, Stappenbeck TS. Genetics and Pathogenesis of Inflammatory Bowel Disease. Annu Rev Pathol 2016;11:127-48.
3. Meir M, Burkard N, Ungewiss H, et al. Neurotrophic factor GDNF regulates intestinal barrier function in inflammatory bowel disease. J Clin Invest 2019;129:2824-2840.
4. Chang J, Leong RW, Wasinger VC, et al. Impaired Intestinal Permeability Contributes to Ongoing Bowel Symptoms in Patients With Inflammatory Bowel Disease and Mucosal Healing. Gastroenterology 2017;153:723-731 e1.
5. Neurath MF. Targeting immune cell circuits and trafficking in inflammatory bowel disease. Nat Immunol 2019;20:970-979.
6. Brown EM, Kenny DJ, Xavier RJ. Gut Microbiota Regulation of T Cells During Inflammation and Autoimmunity. Annu Rev Immunol 2019;37:599-624.
7. Lassen KG, Xavier RJ. Genetic control of autophagy underlies pathogenesis of inflammatory bowel disease. Mucosal Immunol 2017;10:589-597.
8. Hibi T, Kobayashi T, Hisamatsu T. Pathogenesis of inflammatory bowel disease. Nihon Rinsho 2017;75:364-369.
9. Feuerstein JD, Moss AC, Farraye FA. Ulcerative Colitis. Mayo Clin Proc 2019;94:1357-1373.
10. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2018;390:2769-2778.
11. Ng SC, Kaplan GG, Tang W, et al. Population Density and Risk of Inflammatory Bowel Disease: A Prospective Population-Based Study in 13 Countries or Regions in Asia-Pacific. Am J Gastroenterol 2019;114:107-115.
12. Planell N, Lozano JJ, Mora-Buch R, et al. Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. Gut 2013;62:967-76.
13. McGovern DP, Kugathasan S, Cho JH. Genetics of Inflammatory Bowel Diseases. Gastroenterology 2015;149:1163-1176 e2.
14. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intr- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell 2019;178:714-730 e22.
15. Uniken Venema WT, Voskuil MD, Vila AV, et al. Single-Cell RNA Sequencing of Blood and Ileal T Cells From Patients With Crohn’s Disease Reveals Tissue-Specific Characteristics and Drug Targets. Gastroenterology 2019;156:812-815 e22.
16. Parikh K, Antanaviciute A, Fawkner-Corbett D, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature 2019;567:49-55.
17. Huang B, Chen Z, Geng L, et al. Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways. Cell 2019;179:1160-1176 e24.
18. Gerbe F, Jay P. Intestinal tuft cells: epithelial sentinels linking luminal cues to the immune system. Mucosal Immunol 2016;9:1353-1359.
19. Mori E, Oboki K. MITF is necessary for generation of prostaglandin D2 in mouse mast cells. J Biol Chem 2004;279:48923-9.
20. Haeggstrom JZ. Leukotriene biosynthetic enzymes as therapeutic targets. J Clin Invest 2018;128:2680-2690.

21. O’Leary CE, Schneider C, Locksley RM. Tuft Cells-Systemically Dispersed Sensory Epithelia Integrating Immune and Neural Circuitry. Annu Rev Immunol 2019;37:47-72.

22. Howitt MR, Lavoie S, Michaud M, et al. Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut. Science 2016;351:1329-33.

23. Goto N, Fukuda A, Yamaga Y, et al. Lineage tracing and targeting of IL17RB(+) tuft cell-like human colorectal cancer stem cells. Proc Natl Acad Sci U S A 2019;116:12996-13005.

24. Yesil A, Gonen C, Senates E, et al. Relationship between neutrophil gelatinase-associated lipocalin (NGAL) levels and inflammatory bowel disease type and activity. Dig Dis Sci 2013;58:2587-93.

25. Ostvik AE, Granlund AV, Torp SH, et al. Expression of Toll-like receptor-3 is enhanced in active inflammatory bowel disease and mediates the excessive release of lipocalin. Clin Exp Immunol 2013;173:502-11.

26. Dooley TP, Curto EV, Reddy SP, et al. Regulation of gene expression in inflammatory bowel disease and correlation with IBD drugs: screening by DNA microarrays. Inflamm Bowel Dis 2004;10:1-14.

27. Budzynska A, Gawron-Kiszka M, Nowakowska-Dulawa E, et al. Serum neutrophil gelatinase-associated lipocalin (NGAL) correlates with clinical and endoscopic activity in ulcerative colitis but fails to predict activity in Crohn’s disease. J Physiol Pharmacol 2017;68:859-865.

28. Stallhofer J, Friedrich M, Konrad-Zerna A, et al. Lipocalin-2 Is a Disease Activity Marker in Inflammatory Bowel Disease Regulated by IL-17A, IL-22, and TNF-alpha and Modulated by IL23R Genotype Status. Inflamm Bowel Dis 2015;21:2327-40.

29. Bunte K, Beikler T. Th17 Cells and the IL-23/IL-17 Axis in the Pathogenesis of Periodontitis and Immune-Mediated Inflammatory Diseases. Int J Mol Sci 2019;20:20.

30. Macpherson AJ, Yilmaz B, Limenitakis JP, et al. IgA Function in Relation to the Intestinal Microbiota. Annu Rev Immunol 2018;36:359-381.

31. Donaldson GP, Ladinsky MS, Yu KB, et al. Gut microbiota utilize immunoglobulin A for mucosal colonization. Science 2018;360:795-800.

32. Lin R, Chen H, Shu W, et al. Clinical significance of soluble immunoglobulins A and G and their coated bacteria in feces of patients with inflammatory bowel disease. J Transl Med 2018;16:359.

33. Zikherman J, Jenne C, Watson S, et al. CD45-Csk phosphatase-kinase titration uncouples basal and inducible T cell receptor signaling during thymic development. Immunity 2010;32:342-54.

34. Arkteg CB, Goll R, Gundersen MD, et al. Mucosal gene transcription of ulcerative colitis in endoscopic remission. Scand J Gastroenterol 2020;1:1-9.

35. Lu D, Liu L, Ji X, et al. The phosphatase DUSP2 controls the activity of the transcription activator STAT3 and regulates TH17 differentiation. Nat Immunol 2015;16:1283-73.

36. Finan C, Gaulton A, Kruger FA, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med 2017;9.

37. Li H, Fan C, Lu H, et al. Protective role of berberine on ulcerative colitis through modulating enteric glial cells-intestinal epithelial cells-immune cells interactions. Acta Pharm Sin B 2020;10:447-461.

38. Yu YB, Li YQ. Enteric glial cells and their role in the intestinal epithelial barrier. World J Gastroenterol 2014;20:11273-80.

39. O’Connor TM, O’Connell J, O’Brien Di, et al. The role of substance P in inflammatory disease. J Cell Physiol 2004;201:167-80.

40. Middelhoff M, Westphalen CB, Hayakawa Y, et al. Dclk1-expressing tuft cells: critical modulators of the intestinal niche? Am J Physiol Gastrointest Liver Physiol 2017;313:G285-G299.

41. Haber AL, Biton M, Rogel N, et al. A single-cell survey of the small intestinal epithelium. Nature 2017;551:333-339.

42. Zhang Y, Zoltan M, Riquelme E, et al. Immune Cell Production of Interleukin 17 Induces Stem Cell Features of Pancreatic Intraepithelial Neoplasia Cells. Gastroenterology 2018;155:210-223 e3.

43. Qiu W, Remotti HE, Tang SM, et al. Pancreatic DCLK1(+) cells originate distinctly from PDX1(+) progenitors and contribute to the initiation of intraductal papillary mucinous neoplasm in mice. Cancer Lett 2018;423:71-79.

44. Yi J, Bergstrom K, Fu J, et al. Dclk1 in tuft cells promotes inflammation-driven epithelial restitution and mitigates chronic colitis. Cell Death Differ 2019;26:1656-1669.

45. Bezencon C, Furholz A, Raymond F, et al. Murine intestinal cells expressing Trpm5 are mostly brush cells and express markers of neuronal and inflammatory cells. J Comp Neurol 2008;509:514-25.

46. Cheng CW, Biton M, Haber AL, et al. Ketone Body Signaling Mediates Intestinal Stem Cell Homeostasis and Adaptation to Diet. Cell 2019;178:1115-1131 e15.

47. Barker N. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. Nat Rev Mol Cell Biol 2014;15:19-33.

48. Hayakawa Y, Tsutui M, Asfaha S, et al. BLHHA15-Positive Secretory Precursor Cells Can Give Rise to Tumors in Intestine and Colon in Mice. Gastroenterology 2019;156:1066-1081 e16.

49. Tetteh PW, Basak O, Farin HF, et al. Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. Cell Stem Cell 2016;18:203-13.

50. Guo XY, Liu XJ, Hao JY. Gut microbiota in ulcerative colitis: insights on pathogenesis and treatment. J Dig Dis 2020;21:147-159.

51. Boshagh MA, Foroutan P, Moloudi MR, et al. ELR positive CXCL chemokines are highly expressed in an animal model of ulcerative colitis. J Inflamm Res 2019;12:167-174.

52. Rosen MJ, Kerns R, Vallance JE, et al. Mucosal Expression of Type 2 and Type 17 Immune Response Genes Distinguishes Ulcerative Colitis From Colon-Only Crohn’s Disease in Treatment-Naive Pediatric Patients. Gastroenterology 2017;152:1345-1357 e7.

53. Rampal R, Wari N, Singh AK, et al. Retinoic Acid Is Elevated in the Mucosa of Patients With Active Ulcerative Colitis and Displays a Proinflammatory Role by Augmenting IL-17 and IFNgamma Production. Inflamm Bowel Dis 2020.
54. Castro-Dopico T, Dennison TW, Ferdinand JR, et al. Anti-commensal IgG Drives Intestinal Inflammation and Type 17 Immunity in Ulcerative Colitis. Immunity 2019;50:1099-1114 e10.

55. Zheng K, Jia J, Yan S, et al. Paeoniflorin ameliorates ulcerative colitis by modulating the dendritic cell-mediated TH17/Treg balance. Inflammopharmacology 2020.

56. Fujino S, Andoh A, Bamba S, et al. Increased expression of interleukin 17 in inflammatory bowel disease. Gut 2003;52:65-70.

57. Long Y, Zhao X, Xia C, et al. Interleukin-17A secretion and CCR6 co-expression in Treg subsets are related to the imbalance of Treg/Th17 cells in active UC patients. Scand J Immunol 2019.

58. Ma YH, Zhang J, Chen X, et al. Increased CD4+CD45RA-FoxP3low cells alter the balance between Treg and Th17 cells in colitis mice. World J Gastroenterol 2016;22:9356-9367.

59. Ueno A, Jeffery L, Kobayashi T, et al. Th17 plasticity and its relevance to inflammatory bowel disease. J Autoimmun 2018;87:38-49.

60. Magro F, Gionchetti P, Eliakim R, et al. Third European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-Intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-anal Pouch Disorders. J Crohns Collitis 2017;11:649-670.

61. Butler A, Hoffman P, Smibert P, et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36:411-420.

62. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol 2019;20:296.

63. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 2019;15:e8746.

64. Dvir Aran, Agnieszka P Looney, Leqian Liu, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019;20:163-172.

65. Zhang X, Lan Y, Xu J, et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res 2019;47:D721-D728.

66. Ilya Korsunsky, Nghia Millard, Jean Fan, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16:1289-1296.

67. Greg Finak, Andrew McDavid, Masanaro Yajima, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol 2015;16:278.

68. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-7.

69. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47:D1005-D1082.

70. Vaughn BP, Moss AC. Novel treatment options for ulcerative colitis. Clin Investig (Lond) 2013;3:1057-1069.

71. Pastorelli L, Pizarro TT, Cominelli F, et al. Emerging drugs for the treatment of ulcerative colitis. Expert Opin Emerg Drugs 2009;14:505-21.

72. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018;46:D1074-D1082.
Supplementary figures:

Figure S1. The UMAP plots for batch correction and donor effects. A, UMAP before and after SCTransform batch correction. B, UMAP of each donor shows minimal donor effects after batch correction.

Figure S2. UMAP shows the identified labels from SingleR annotation based on HPCA(A) and Encode Blueprint (B) reference database.
Figure S3. UMAP shows the overlap of this study and Smillie, et al. separated by epithelial and stromal cells (A) and immune cells (B)

*In case some texts are difficult to read, full resolution figures can be found in supplementary files in doi: 10.1016/j.jcmgh.2021.01.020

Figure S4. Heatmap shows the top five differentially expressed (DE) genes in each cluster. DE genes were obtained by comparing expression level in cells of one cluster against to that in the rest of cells.
Figure S5. Dot-heatmap shows the expression of epithelial and stromal lineage markers.

Figure S6. Dot-heatmap shows the expression of Lymphoid, B-cells, and myeloid lineage markers.
Figure S7. Gene signatures found in three plasma subsets. A, Dot heatmap showing the expression levels of marker genes detected in the three plasma cell clusters, and their expression differences in the inflamed samples (UC) comparing to non-inflamed samples (SC) and healthy samples (HC). B, Co-expression network showing the difference of co-expressed genes to Plasma-3 specific marker gene (IGLL5) between inflamed UC and HC or SC.

Figure S8. GO and KEGG enrichment terms of the DE genes in plasma or epithelial cells found specifically in this study but not in Smillie, et al. s.Dw.HC: specific down-regulated when comparing UC to healthy control; s.Dw.SC: specific down-regulated when comparing UC to self-control; s.Up.HC: specific up-regulated when comparing UC to healthy control; s.Up.SC: specific up-regulated when comparing UC to self-control.
Figure S9. Comparison of effect sizes of differentially expressed (DE) genes from UC risk loci in epithelial cell clusters (A) and immune cells (B).