Multiomic analysis reveals cellular and epigenetic plasticity in intestinal pouches of ulcerative colitis patients.

Yu Zhao*[2], Ran Zhou*[1], Bingqing Xie[1], Cambrian Y Liu[1], Martin Kalski[1], Candace M Cham[1], Jason Koval[1], Christopher R Weber[3], David T Rubin[1,3], Mitch Sogin[4], Sean Crosson[5], Jun Huang[2], Aretha Fiebig[5], Sushila Dalal[1], Eugene B Chang[1], Anindita Basu[1], Sebastian Pott[1]

[1] University of Chicago, Department of Medicine, Chicago, IL
[2] University of Chicago, Pritzker School of Molecular Engineering, Chicago, IL
[3] University of Chicago, Department of Pathology, Chicago, IL
[4] Marine Biological Laboratory, Woods Hole, MA
[5] Michigan State University, East Lansing, MI
* Contributed equally

Correspondence: obasu@uchicago.edu, spott@uchicago.edu

Background & Aims: Total proctocolectomy with ileal pouch anal anastomosis (IPAA) is the standard of care for patients with severe treatment resistant ulcerative colitis (UC). Despite improvements in patient outcomes, about 50% of patients will develop inflammation of the pouch within 1-2 years following surgery. Establishment of UC pouches is associated with profound histological changes of the mucosa. A detailed characterization of these changes on a cellular and molecular level is crucial for an improved understanding of pouch physiology and diseases management.

Methods: We generated cell-type-resolved transcriptional and epigenetic atlases of UC pouches using scRNA-seq and scATAC-seq data from paired biopsy samples from the ileal pouch and ileal segment above the pouch (pre-pouch) of UC-IPAA patients (n=6, female=2) without symptoms. We also collected data from paired biopsies of the terminal ileum (TI) and ascending colon (AC) from healthy controls (n=6, female=3).

Results: We identified novel populations of colon-like absorptive and secretory epithelial cells, constituting a significant proportion of the epithelial cell fraction in the pouch but not in matched pre-pouch samples. Pouch-specific enterocytes expressed colon-specific genes, including CEACAM5, CA2. However, in contrast to normal colonic epithelium, these cells also expressed a range of inflammatory and secretory genes, similar to previously detected gene expression signatures in IBD patients. Comparison to longitudinal bulk RNA-seq data from UC pouches demonstrated that colon-like epithelial cells are present early after pouch functionalization and independently of subsequent pouchitis. Finally, single cell chromatin accessibility revealed activation colonic transcriptional regulators, including CDX1, NFIA, and EHF.

Conclusion: UC pouches are characterized by partial colonic metaplasia of the epithelium. These data constitute a resource of transcriptomic and epigenetic signatures of cell populations in the pouch and provide an anchor for understanding the underlying molecular mechanisms of pouchitis.

Introduction

Patients with medically refractory Ulcerative colitis (UC) often undergo ileal pouch anal anastomosis (IPAA), which removes the colon in its entirety and thus the organ affected by UC[1]. Establishment of an ileal pouch (pouch) from the terminal segment of the ileum restores

continence and improves quality of life[1]. Unfortunately, up to 50% of patients will develop inflammation of the pouch (pouchitis) within 1-2 years following surgery[2,3]. In most cases, acute pouchitis can be managed with antibiotic treatments and biologics; however, about 60% of these patients subsequently develop chronic pouchitis[4]. The etiology for pouchitis remains unclear, hampering efforts to develop specific treatment strategies[5].

While these patients previously presented with UC, the pouch tissue has been derived from ileum which is not involved in UC. Establishment of the pouch in UC patients is frequently associated with histological changes of the mucosa involving both the epithelium and immune cells[6,7], including some degree of inflammation (neutrophil infiltration) and villus atrophy[6,8]. Notably, histological changes alone do not appear to be predictive of pouchitis[8]. Transcriptional characterization of newly established ileal pouches for UC patients using bulk RNA-seq revealed a colon-like expression profile in pouch compared to pre-pouch samples[9,10]. However, bulk RNA-seq is unable to provide insights into the cellular composition of the pouch epithelium or cell-type-specific gene expression differences between pouch and pre-pouch. The only single cell RNA-seq study of pouch tissue focused on immune cells (CD45+)[11].

Therefore, it remains unclear whether the ileum-derived mucosa of the pouch undergoes changes in cell type composition and how gene expression changes in specific cell types after pouch functionalization. Addressing these questions is an important first step to understanding the cellular processes that shape this unique tissue and potentially identifying molecular characteristics that predispose patients to pouchitis. We used single cell RNA-seq (scRNA-seq) and single cell ATAC-seq (scATAC-seq) to generate a cellular atlas of pouch samples and samples from the ileal segment immediately above the pouch (pre-pouch) obtained from UC patients without clinical pouchitis, and healthy control samples from the ascending colon (AC) and the terminal ileum (TI). Using scRNA-seq, we observed substantial compositional changes in the epithelial lineage and identified colon-like absorptive and secretory epithelial populations that co-existed with corresponding ileal-like epithelial cell types in pouches, but not pre-pouch. Using scATAC-seq, we identified cell-type-resolved regulatory elements and found that cells with colonocyte-like transcriptomes are associated with key colonic regulators while differentiation of ileal-like cells in the pouch was associated with regulators of ileal cell fate. Together, our data suggested the differentiation of a distinct colon-like epithelial cells within pouches leading to partial colonic metaplasia. Unlike regular colonocytes, however, these cells also express increased levels of inflammation-associated genes.

Results

scRNA-seq profiling reveals cellular composition of ileal pouch tissue

We aimed to create a cellular atlas of ileal pouches in UC patients, who previously underwent IPAA surgery, to identify their constituent cell types and compare their cellular composition and transcription profiles to related control tissues. We performed single-cell RNA sequencing (scRNA-seq; 10x Genomics) on paired biopsy samples from the pouch and the ileal segment above the pouch (pre-pouch) from 6 UC-IPAA patients with long established pouches (mean 17 yrs post surgery, range 9-31 yrs) without clinical symptoms of pouchitis. Of note, histopathological assessment of these biopsies suggested that mild to significant enteritis was present in the pouch but not pre-pouch samples of all scored individuals (see below **Figure 4B)** in agreement with previous assessments of UC pouch mucosa[8]. For comparison, we profiled paired biopsies from the TI and AC from 6 healthy individuals undergoing routine screening (**Figure 1A, Supplementary Figure 1A, Supplementary Table 1)**. After quality-control filtering, we obtained 103,772 cells from all scRNA datasets combined (**Figure 1B**).

Using community-based clustering and marker-gene analysis we identified 5 major clusters comprising cells of the following lineages: epithelial (59,751 cells, marked by *EPCAM*), T-cells (16,389 cells, *CD3D*), B-cells (17,317 cells, *CD79*), myeloid cells (5,375 cells, *TYROP*), and stromal cells (4,940 cells, *IGFBP7*) (**Supplementary Figure 2B)**. Each individual and sample contributed to all cell lineages (**Supplementary Figure 1B**).

Each of the major lineages further separated into multiple cell-types. By combining computational annotation leveraging previously published data with manual annotation employing well-defined marker genes, we annotated 47 cell types across all lineages (**Figure 1C and D, Supplementary Figure 11-14,** see **Methods** for detailed description of classification).

We identified all major cell types previously reported for the absorptive and secretory epithelial lineages (**Figure 1C, D, Supplementary Figure 3A, 11**)[12]. This included stem cells (LGR5, ASCL2), Transit amplifying cells (TA cells) (MKI67), absorptive epithelial cells and secretory cells such as, Goblet cells (MUC2), Paneth cells (DEFA5), Microfold cells (M-cells) (CCL20). Several absorptive and secretory cell populations from the TI and AC clustered separately based on their location of origin. To distinguish these ileal and colonic cell populations we denoted the Enterocyte cell clusters as EC1 (ileal) and EC2 (colon), and similarly we identified Goblet1 and Goblet2 and Stem1 and Stem2 clusters. Furthermore, we distinguished progenitor and mature cells within the EC1 and EC2 cell clusters and denoted them as EC1-1 (SI, GSTA2) and EC2-1 (CD24, CA2) and EC1-2 (SLC15A1, APOA4) and EC2-2 (CEACAM5), respectively (**Figure 1 D, Supplementary Figure 3A)**.

All samples contained a range of adaptive and innate immune cells representing T-cell, B-cell, and myeloid lineages. (**Figure 1C, D, Supplementary Figure 7, 8, 9**). T cells were further clustered into CD4, CD8, and gamma-delta T cells. CD4 T cells include central memory (CCR7), regulatory (FOXP3), CD103- and CD103+ resident memory (CD69+) T cells. CD8 T cells include CD103- and CD103+ resident memory T cells. B cells readily separated into germinal center B cells (BCL6), naive B cells (IGHD), plasma B cells (SDC1) and memory/effector B cells (CD27). (**Figure 1C, D, Supplementary Figure 8, 13A**). Innate immune cells included macrophages (CD163), monocytes (FCN1), dendritic cells (CLEC9A, CD1C), mast cells (CPA3), and neutrophils (FCGR3B) (**Figure 1C, D, Supplementary Figure 13B**). Stromal cell types and others constituted 4.8% (4,940 out of 103,772) of cells. We distinguished several fibroblast populations and identified small clusters of the following cell types: Endothelial cells from arteries (EFNB2) and veins (ACKR1), pericytes (NOTCH3), smooth muscle (DES), lymphatic endothelial cells (LYVE1), and glial cells (S100B) (**Supplementary Figure 14**).

These data constitute the first cellular atlas of UC pouch and pre-pouch tissues, which enabled us to compare the cellular and transcriptional changes between pouch and pre-pouch and to identify similarities and differences to TI and AC samples from non-IBD controls.

Profound differences in the cellular composition of the epithelial lineages between pouch and pre-pouch

To systematically assess the cellular identities and composition of UC pouches, we used data from all 4 regions (pouch, pre-pouch, AC, TI). We noticed stark differences in cellular composition of the epithelial lineage between pouch and pre-pouch samples (**Figure 2**). Notably, we found that pouch samples differed in cellular composition compared to all other samples and featured two different enterocyte, goblet cell, and stem cell populations corresponding to EC1 and EC2, Goblet1 and Goblet2 and Stem1 and Stem2, respectively. (**Figure 2B, C, Supplementary Figure 5A**). This compositional change was observed in all 6 individual suggesting that partial colonic metaplasia might be a common feature of UC pouches. In contrast to the pouch, and in agreement with their anatomic location, the cellular composition of pre-pouch samples was similar to TI.

In addition, we observed smaller proportional shifts in additional epithelial cell types. The proportion of TA cells across lineages was higher in pouch compared to pre-pouch samples (14% vs 4.8%, p-value=0.0087) and this difference was also observed in AC compared to TI (12% vs 1.9%, p-value=0.0022). The proportion of TA cells in pre-pouch and TI samples differed slightly (4.8% vs 1.9%, p-value=0.041) (**Supplementary Figure 3**). The proportion of M-cells was elevated in pouch and pre-pouch compared to AC (**Supplementary Figure 3**). We detected Paneth cells in TI, pre-pouch, and pouch but not in AC (**Supplementary Figure 3**).

Pouch-specific changes in non-epithelial cell types

Given the profound cellular changes in the pouch epithelium, we also tested whether similar changes were observed in the immune and stromal compartments. We noted that the changes in the immune compartment were less uniform across patients compared to the epithelial lineage. For example, while we observed an increase in the proportion of IgG plasma B-cells in 3 patients[13], the proportion of these cells in the remaining patients was indistinguishable from pre-pouch and control samples. Samples from pouch and pre-pouch contained a slightly higher proportion of CD103+ CD4 Trm cells compared to AC (13%, 11% vs 4.3%, p-value=0.015, 0.0043, respectively), while the proportion of KLRG1+ CD8 Trm was higher in AC, pouch, and pre-pouch compared to TI. Of note, NK T cells were specifically increased in pouch samples compared to pre-pouch samples (1.8%, 0.4%, p-value=0.0087) as were regulatory T cells (13%, 5.1%, p-value=0.0043) (**Supplementary Figure 7**). These differences suggest compositional changes of immune cells in pouch samples possibly reflecting the tissue origin (CD103+ CD4 Trm) and potentially disease status (IgG plasma cells, NK T cells).

Among stromal cell types, we found a significantly higher proportion of the fibroblast 2 (NRG1+) population in the pouch (3.4%) compared to pre-pouch. This population is almost absent in pre-pouch and TI samples (0.2% and 0.3%, respectively) and highest in AC (13%). These differences potentially indicate a remodeling of the stromal niche.

In comparison to the observed changes in the pouch epithelium, the differences in the non-epithelial clusters are smaller and more variable across samples. Increasing the number of cells and samples would likely improve power to detect additional differences in the immune and stromal lineages. In this study, we focused our analysis on the characterization of the partial metaplasia phenotype observed in epithelial cell types in the pouch.

Differential expression in pouch enterocytes suggested partial colonic metaplasia.

We sought to compare and characterize the transcriptional programs of enterocytes in the pouch (EC1 and EC2) and to compare these differences to expression differences between matched enterocyte populations from pre-pouch and control samples. Specifically, we identified genes differentially expressed between ileal-like (EC1) and colon-like (EC2) enterocyte populations (**Figure 2D,** Methods) and found a total of 5,286 differentially expressed genes (DEG) in at least one of the comparisons (**Supplementary Figure 4A**).

First, we detected a substantial number of DEG between pouch EC2 and EC1 (1,240 genes up in EC2 and 952 genes up in EC1, FDR < 0.05, log fold change > 1) and between AC (EC2) and TI (EC1) from controls (1,883 genes up in AC, 2,141 genes up in TI). Notably, a substantial proportion of these DEGs were shared across these two comparisons (836 genes upregulated in pouch EC2 vs pouch EC1 and AC vs TI, and 770 genes upregulated between pouch EC1 vs pouch EC2 and TI vs AC). This analysis suggested that EC1 and EC2 present in the pouch recapitulate at least a portion of the expression differences observed between enterocytes from colon and ileum. We further confirmed these common expression differences between EC1 and

EC2 cell types through pairwise comparisons across all other conditions and samples (**Figure 2D, Supplementary Figure 4A**). Note that the limited numbers of Goblet 1 and 2 and Stem 1 and 2 cells in the pouch did not allow us to compare these populations directly within the pouch samples. However, using data from all samples in our dataset revealed a significant number of genes shared between these comparisons and the DEG detected between EC1 and EC2, suggesting the establishment of a region-specific gene expression pattern that is shared across enterocytes, goblets, and stem cells (**Supplementary Figure 4C, 4D, 6C, 6D**).

Second, we asked whether the sets of DEGs between populations of EC1 and EC2 cells represented specific cellular functions. DEGs associated with EC2 in the pouch were enriched for immune and antimicrobial pathways and shared enrichment for proliferation associated terms and oxidative phosphorylation with enterocytes from AC (**Figure 2E, Supplementary Table 2**). We found that DEGs associated with EC1 populations were strongly enriched in lipid and bile acid metabolism terms (**Figure 2E, F**). This comparison suggested that compared to EC1, EC2 cells appear to lose metabolic functions associated with the ileal epithelium and functionally resemble AC enterocytes.

Third, despite the transcriptional similarity between pouch and AC EC2 cells, clustering of the DEGs suggested expression differences between these cells (**Figure 2D**). We identified 1442 differentially expressed genes between these locations (770 up in EC2 pouch, 672 up in AC EC, FDR < 0.05, log fold change > 1) (**Supplementary Figure 5 D, Supplementary Table 5)**. Among the most strongly up-regulated genes in the pouch were genes with known ileal expression and function, including FABP6, APOA4 and APOB. In contrast, AQP8, HMGCS2, SELENBP1, SATB2 showed increased expression in AC EC2 compared to pouch. In addition, we noticed a number of genes previously reported as part of an expression signature in UC/IBD epithelium, including REG1A[14], DMBT1[15,16], LYZ[17], and REG4[18]. This finding indicates that while the transcriptional identity of pouch EC2 cells resembles that of EC2 enterocytes in the colon, they show differences in expression pattern possibly due to different tissues of origin, differing microbial and metabolic environments, and disease status.

Motivated by the observation that some UC-associated genes were more highly expressed in pouch EC2 compared to AC controls, we compared the set of DEGs between AC EC2 and Pouch EC2 to a previous scRNA-seq study of UC samples[19]. We found concordant overlap between a subset of pouch EC2-specific genes and UC-associated genes from Smillie et al. (81 out of 770 pouch EC2 genes). Conversely, AC EC2-specific genes were enriched for genes down-regulated in colonic epithelium of UC patients in that study (29 out of 672 AC EC genes) (**Figure 3A, Supplementary Table 3**). This comparison suggested a UC-like inflammatory signature in pouch EC2 cells, highlighting a potentially persistent inflammatory environment. It is worthwhile noting that expression of these markers appears variable between individuals (**Figure 3A**). We also identified DEGs between the pouch Goblet2 with AC goblet cells and found similar inflammation-related changes in these secretory cells (**Figure 3B, Supplementary Table 3)**.

Together, these analyses established the presence of distinct absorptive and secretory enterocyte populations with colon-like transcription profiles in UC pouches and suggested that the pouch epithelium undergoes partial metaplasia.

Histopathology confirmed presence of colon-like cells in pouch epithelium

We sought to validate the metaplastic phenotype and to characterize the spatial distribution of EC1 and EC2 cells in the pouch epithelium. We obtained tissue sections from pouch and pre-pouch samples from 5 out of 6 individuals profiled with scRNA-seq for histopathological assessment and immunohistochemistry (IHC). In addition to mild to severe enteritis (Figure 4B),

we observed differences of the epithelial architecture in pouch biopsies, including villus atrophy, (**Figure 4A, Supplementary Figure 15**), in agreement with previous reports[6].

Assessment of sections from pre-pouch biopsies showed staining for the ileal enterocyte marker FABP6 and complete absence of colonic markers CEACAM5 and CA2 (**Figure 4A, Supplementary Figure 15A, B**). In contrast, expression of FABP6 was greatly diminished in pouch samples and confined to a small subset of cells. In addition, we observed expression of CEACAM5 and CA2 in a subset of epithelial cells from pouch samples (**Figure 4A, Supplementary Figure 15A, B**). These observations validated our scRNA-seq finding by confirming the expression of colonic markers in a subset of epithelial cells while showing diminished or no expression of ileal marker genes.

Pouch metaplasia occurs early after functionalization and independently of subsequent inflammation.

We observed colon-like EC2 cells in all our pouch samples; however, our cross-sectional study was limited to six patients with well-established pouches and without overt clinical symptoms. To generalize our observation, we took advantage of a previously generated bulk RNA-seq dataset[9] which reported a colon-like expression signature in pouch samples[9]. These data allowed us to address when EC2 cells appear after pouch functionalization, and to ask whether their presence relates to the development of pouchitis. This longitudinal study included 19 patients with newly established pouches, sampled at three timepoints within the first year after UC IPAA surgery (4-mo, 8-mo, 12-mo post functionalization); 6 out of the 19 individuals were subsequently diagnosed with pouchitis[9].

We first assessed the expression of EC1- and EC2-specific markers obtained with our scRNA-seq data in bulk samples from pre-pouch and pouch regions. EC2 markers showed increased expression in bulk samples from pouches, while EC1 markers were increased in pre-pouch samples (**Figure 4C, D**). We then used the cell types identified in UC pouch individuals (i.e. from pouch and pre-pouch samples) with our scRNA-seq data as reference to deconvolve the bulk RNA-seq data from Huang et al using MuSiC[20] (**Methods**). This approach allowed us to infer the constituent cell types and to estimate their respective proportions in each bulk sample. We found that a population of EC2 cells was present in the majority of pouch samples (58%, 21/36) (**Figure 4E, F**) and in only 3 out of 35 (9%) pre-pouch samples (**Supplementary Figure 15C, D**). We detected EC2 cells (>1%) in 14 out of 19 individuals in at least one time point. The fraction of EC2 cells in pouch samples (4.0% - 51.9%) was generally inversely correlated to the fraction of EC1 cells (spearman correlation coefficient = -0.82, p-value = 7.4e-10) in the same sample.

EC2 cells were detected in pouch samples at the earliest time point (4 month) and there was no significant difference in the proportion of samples or the proportion of cells per sample across timepoints (**Figure 4E, F**). Similarly, the proportion of samples with EC2 cells was comparable between healthy individuals and those diagnosed with pouchitis (**Figure 4E, F**), suggesting that the presence of EC2 cells alone is not a causing pouchitis.

Together these data suggest the presence of separate colon-like EC2 enterocytes in most pouches providing a cellular basis for previously reported colon-like expression within pouch samples.

Integrative scATAC-seq profiling in ileal pouch and neighboring tissues

To identify key transcription factors, gene regulatory elements, and target genes associated with pouch metaplasia, we measured chromatin accessibility using scATAC-seq[21]. To minimize biological and technical variation, scATAC-seq was performed on cells from the same 24 biopsies used for scRNA-seq (**Methods**).

After quality-control filtering, we retained 88,595 cells, (sample mean = 3,691, range: 1,704-5,524). Dimensionality reduction and clustering following previously established procedures[22] identified cells from all major lineages (Epithelium, T cells, B cells, Myeloid, and Stromal/others) (**Supplementary Figure 17**). We assigned cell-type labels to each cell by computationally transferring cluster labels from scRNA-seq and observed that with few exceptions, specific cell-type labels mapped to discrete locations in the embedding suggesting good correspondence between the modalities (**Figure 5A, Methods**). To assess the accuracy of our cell-type annotations, we used gene activity scores as a proxy for gene expression[22] (**Methods**) and compared the scores for marker genes identified by scRNA-seq across clusters. This comparison further supported our assignment (**Supplementary Figure 16A**). Finally, we confirmed that cell-type proportions identified separately by each modality (scRNA-seq and scATAC-seq) were similar for each sample (**Supplementary Figure 16B**).

Focusing on cells from the epithelial lineage alone, we retained 59,987 cells across all samples, containing 12,474 cells from pouch samples (**Figure 5B, C**). Similar to scRNA-seq, we detected absorptive cells (EC1-2, EC2-2, BEST4), secretory cells (Goblet1, Goblet2, Paneth, EEC, Tuft) and progenitor cells (Stem, EC1-1, EC2-1). We excluded TA-cells and M-cells from annotation since these were more ambiguous, likely due to differences between gene scores and RNA expression for informative markers of these cell types (***see Methods***). Loci harboring marker genes for the annotated cell types showed highly specific chromatin accessibility (**Figure 5E**).

Comparison of clusters from different sampling locations confirmed the presence of both ileum-like (EC1) and colon-like enterocyte (EC2) populations in the pouch (**Figure 5C**). As observed with scRNA-seq, the proportion of EC2 cells was significantly increased in pouch compared to pre-pouch (42.47% vs 11.6%) (**Figure 5D, Supplementary Figure 17B**). These data confirmed the presence of EC2 cells in the pouch using an independent, chromatin-based assay and thus providing the basis to identify and characterize regulatory regions and transcription factors associated with colonic metaplasia in UC pouches.

Identification of regulatory regions and sequence motifs associated with colon-like cells in the pouch

To identify cis-regulatory elements (CREs) in each epithelial cell type, we called peaks on each cell-type cluster separately (**Methods**). We identified a union set of 262,177 peaks across all clusters with the number of peaks per cluster ranging from 37,127 to 150,000 (**Supplementary Figure 17D**). While most peaks were shared across multiple cell types, we observed significant variation in chromatin accessibility across cell types (**Figure 6A**). Enrichment analysis within cell-type-specific peaks identified motifs of known master transcription factors (TFs) for these cell types and lineages (**Figure 6B**). For example, motifs for HNF4A/G were strongly enriched in ileal-like EC1 cells, while motifs for CDX1/2 were enriched in colon-like EC2 cells (**Figure 6B**). This analysis established the union set of candidate CREs in the epithelial lineage and their use across cell types and helped to infer sequence motifs associated with cell-type-specific TFs.

We then used these candidate CREs to identify putative regulators of EC2 cells in the pouch. Analogous to our identification of differentially expressed genes between EC1 and EC2 cells (**Figure 2D**), we performed pairwise comparisons between EC1 and EC2 cell populations from pouch, pre-pouch, TI, and AC (**Methods**). We identified 120,569 differentially accessible regions (DARs) in at least one comparison (FDR < 0.01, log fold change > 1) (**Figure 6C, Supplementary Figure 22A**) and 23,650 DARs between EC1 and EC2 cells in the pouch (14,133 up in EC1, 9,517 up in EC2). This observation provided evidence for epigenetic differences between the two enterocyte cell types in the pouch and suggested remarkable epigenetic plasticity within the

epithelial lineages. Of note, 6,892 (72.4%) DARs in pouch EC2 vs EC1 populations were also identified as DARs when comparing EC2 and EC1 cells in AC and TI (**Figure 6C, Supplementary Figure 22A**). Similarly, the comparison of EC1 and EC2 cells in the pouch yielded 11,147 (78.9%) DARs also found when comparing TI to AC enterocytes (**Figure 6C, Supplementary Figure 22A**). This suggests that differences between EC2 and EC1 cells in the pouch were indeed associated with regulatory regions that broadly differentiated AC from TI. However, we also detected 36,848 DARs in total between EC2 cells in the pouch and AC mirroring the results of our DEG analysis between these two tissues (**Supplementary Figure 5D, Supplementary Figure 22A**).

These sets of DARs were enriched for a range of motifs for TFs involved in epithelial differentiation as well as regulation of inflammatory processes (**Supplementary Figure 22B**). To identify candidate transcription factors associated with these enriched motifs, we computed the correlation between integrated RNA transcript levels and motif accessibility within the epithelial lineage (**Methods**) (**Supplementary Figure 22C**) and retained TF-motif pairs with a correlation > 0.5 (activating TF), and correlation < -0.5 (repressive TF)[23]. Motif accessibility (**Figure 6E, Supplementary Figure 22D)** and expression (**Figure 6D, Supplementary Figure 22D**) reflected differential activity in EC1 and EC2 cell types for most TFs.

Among the identified TFs, CDX1 and 2, NFIA, EHF, and KLF5 showed higher expression and motif accessibility in EC2 cells from both Pouch and AC (**Figure 6D, E)**. HNF4G, PPARA, GATA6, JUND were among the TFs associated with EC1 cells in the TI, pouch, and pre-pouch. By comparing accessibility of these TF motifs along pseudotime-based differentiation trajectories from stem cells into EC1 and EC2 cells, respectively, we observed cell-type-specific activation pattern for these TFs during epithelial differentiation (**Figure 6E, Supplementary Figure 24)**. Several of these TFs were independently identified to have ileal and colon-specific activity in a multimodal single-cell atlas of the human intestine[24].

Notably, TF expression and motif enrichment appeared to be more heterogeneous in EC2 pouch cells compared to AC. EC2 pouch cells also maintained activity of several TFs associated with EC1 cells, including BACH1 (**Figure 6E**). Motif enrichment and expression of NFIA was more strongly associated with pouch EC2 highlighting differences in TF activity compared to AC. NFIA was identified in progenitor cells of early colonocytes in our data and by Hickey et al. In contrast to AC EC2, NFIA motif activity remained high in mature pouch EC2 cells.

Together, these TFs represent a key part of the gene regulatory network associated with the two transcriptionally distinct enterocyte populations observed in the pouch epithelium and provide evidence for the activation of a colon-like regulatory network in pouch EC2 cells.

Activation of a colon specific GRN underlies transcriptional identity of EC2 cells in the pouch

To identify putative target genes of identified cell-type-specific TFs in the pouch epithelium and to characterize the regulatory network defining EC2 cells, we linked CREs to genes using co-accessibility of peaks with integrated expression[22] (**Methods**). We linked 31,808 peaks to 6,312 genes across EC1 and EC2 populations identifying a large set of putative targets (**Figure 7A**). We observed a strong enrichment of DARs associated with DEG compared to non-DEGs (**Supplementary Figure 26**). For example, distal regulatory elements linked to the colon-specific marker gene CA2 gained chromatin accessibility in pouch EC2 cells (**Figure 7B).** In contrast, accessibility at CREs linked to ileal marker gene APOA4 was lower in pouch EC2 cells compared to EC1 cells in pouch, pre-pouch and TI (**Supplementary Figure 27**).

We complemented this identification of target genes using FigR to identify domains of regulatory chromatin (DORCs)[25] which might harbor key target genes and facilitate the identification of regulatory factors. Using this metric, we ranked the genes by the number of associated peaks and identified 765 DORC-associated genes within the epithelial lineage. These target genes included both regulatory genes (e.g. ETS2, GATA5, KLF6) and genes associated with cellular function (e.g. OLFM4, MUC2, RBP2) (**Supplementary Table 4**).

To summarize the regulatory mechanisms underlying EC1 and EC2 cells, we constructed a partial gene regulatory network by collating information from motif enrichment, TF expression, co-accessibility links, DORC scores, and differential gene expression (**Methods**, **Supplementary Figure 6C**). This partial regulatory network highlighted that the DEGs in the two pouch enterocytes populations are targeted by groups of overlapping but distinct TFs (**Figure 7C**). Our data therefore suggested that the colon-like transcriptional program in pouch EC2 cells is the result of the activation of a gene regulatory network driven by TFs predominantly associated with the colon. For example, genes associated with colon function (CA2, CD24, CFTR) are induced in pouch EC2 cells compared to EC1 (**Supplementary Figure 4B**l). While we showed that this cell type shares regulatory features (CREs and TFs) with enterocytes from the colon and undergoes similar differentiation dynamics, we note that when compared to enterocytes from the ascending colon the acquisition of a colonic phenotype is incomplete. We identified a number of ileal genes that are still expressed at significant levels in EC2 cells in the pouch compared to colon (e.g., FABP6, APOA4, APOB) (**Supplementary Figure 5D**l) possibly reflecting the different developmental origin of the tissues or the differences between these two tissues.

Discussion

In this study we presented, to our knowledge, the first single cell atlas of ileal pouch mucosa from UC patients. Histological changes associated with UC pouch mucosa are well-documented[6,8]. However, the nature of the cellular changes and their molecular basis remain unclear[6]. Our study contributed to the understanding of these changes in the pouch by measuring the differences in cellular composition and cell-type-specific gene expression in the pouch compared to pre-pouch and control samples from AC and TI.

Our multiomic single-cell approach identified partial colonic metaplasia of epithelial cell types in UC pouches and provided insights into the mechanisms underlying their differentiation. We identified transcriptional regulators specifically associated with colon cell fate in colon-like pouch cells, including CDX1, CDX2[26,27], EHF[28] and NFIA[24]. These TFs were associated with the expression of genes with colon-specific functions including metabolic (CA2) and anti-microbial (CEACAM5) activity. We also observed an increase in the expression of SATB2 in a subset of pouch EC2 cells. SATB2 is necessary for colonocytes differentiation, and its overexpression is sufficient to convert ileal epithelium into colonic epithelium in mice. However, SATB2 was expressed at lower levels and more heterogeneously in pouch EC2 cells compared to AC EC2. In contrast to AC EC2, pouch EC2 cells also maintained expression of a subset of genes with generally ileal-specific gene expression pattern (APOA4, APOB, FABP6) (**Supplementary Table 5**), and appeared less mature compared to AC EC2. In addition, pouch EC2 cells and Goblet 2 cells showed evidence of a transcriptional signature enriched for immune and secretory functions. Similar expression changes in epithelial cells have been observed in chronic inflammatory conditions in human IBD studies and animal models[19,29,30]. Bulk RNA-seq studies observed apparent shifts of gene expression indicative of 'regionalization' in CD[31] and it will be interesting to understand whether similar metaplastic process can be observed.

While these observations suggested a remarkable plasticity of the intestinal epithelium in adult tissues, the precise progenitor cell type of origin for the colon-like metaplastic epithelial cell populations remains unclear. Separate colon-like and ileum-like enterocyte populations were most obvious in mature secretory and absorptive cells of the pouch. However, stem cells and progenitor populations in pouches separated into two populations as well and co-clustered with their respective cell types from ileum and colon.

Changes in gene expression of the pouch mucosa compared to the pre-pouch had previously been observed in bulk RNA-seq data[9,10]. Our study linked these changes to a particular cell population and comparing our cell-type-resolved data to bulk data provided additional evidence for the presence of EC2 cells in the majority of pouches. This analysis also suggested that the presence of EC2 cells alone, was not linked to development of pouchitis. It is possible that the presence of EC2 cells is unrelated to pouchitis and a consequence of establishing pouches. Alternatively, EC2 cells might be necessary for the development of pouchitis but not sufficient (e.g., an additional microbial trigger is needed). We speculate that comparison to individuals with familial adenomatous polyposis (FAP) would help to shed light on the role of EC2 cells. FAP patients frequently undergo the same procedure to reduce risk of colon cancer yet reportedly rarely develop pouchitis[32] , which suggests that the pouch alone does not render patients susceptible to inflammation.

Our study has several limitations. The samples were obtained from a small cross-sectional cohort consisting of six individuals. We therefore focused our analyses on cell types in the epithelial lineage that were well represented and showed a strong and consistent effect across individuals. The study's focus does not imply that changes in other cell types are unimportant but reflect a lack of power to adequately study them. For example, identification of colon-enriched NRG1+ fibroblasts[26] in the pouch might suggest the development of a colon-like niche in the pouch. We observed an increase in IgG plasma B-cells in 3 out of 6 pouches, a change previously observed in UC patients[13]. A previous study profiling immune cell populations in the pouch found significant shifts between uninflamed and inflamed UC pouches[11]. Additionally, our single cell genomics data lacks spatial context which is important to understand cellular niches and cell-cell interactions contributing to the observed cellular changes in the pouch. For example, spatial transcriptome data would provide a measurement of the local distribution of EC1 and EC2 cells within the pouch as well as the make-up of their cellular neighborhoods.

These data serve as a reference for further cellular and molecular studies of IPAA pouches. Our results suggest that UC pouches provide a unique system to study intrinsic and extrinsic cues contributing to epithelial plasticity in the pouch and how modulation of epithelial identity relates to disease risks in the pouch and IBD in general.

Methods

Ethics statement

This study was conducted under the translational core research program protocol (15573A), approved by the Institution Review board of University of Chicago. Licensed gastroenterologists collected biopsy samples and clinical data from patients identified as UC-pouch and non-IBD control subjects. Biopsy sample and patients' demographic data were obtained from all participants with signed informed consent. Samples and demographic data were anonymized.

Patient information

Non-IBD controls were consented and recruited by the time of routine colonoscopy. UC-pouch patients were consented and recruited at follow-up pouchoscopy. Non-IBD controls were patients without a history of IBD, an immune related disease, an ongoing infection in intestines, a history of Familial adenomatous polyposis, a history of primary digestive tract malignancies, a history of metastatic malignancies in digestive tract, and symptoms of digestive disorders. The UC-pouch patients had a clinical diagnosis of ulcerative colitis, have a history of ileal pouch-anal anastomosis and were reported free of active disease via macroscopic assessment from a physician during the pouchoscopy. Demographic data of patients are provided in Table 1.

Tissue biopsy sample collection

All biopsies were collected in locations determined by scoping gastroenterologists. Non-IBD controls had up-to-six biopsies obtained from each location during colonoscopy at two separate anatomic regions (terminal ileum and ascending colon). UC-pouch patients had up-to-six biopsies obtained from each region, at two regions (pre-pouch and pouch), during pouchoscopy. In short, mucosal linings were obtained using biopsy forceps following standardized care. Biopsied mucosal tissues from each region were immersed in Advanced DMEM/F-12 [Thermo Fisher 12634010] in an Eppendorf tube and placed on ice for short-term preservation and transfer. The wait times between biopsy and sample processing were under two hours.

Sample processing

Single cells from mucosal biopsies were acquired using a modified version of previously published protocols [19,33,34]. In brief, intact mucosal tissues were rinsed in 1.5 ml ice-cold PBS [Thermo Fisher 10010023] six times. Mucosal tissues from the same regions were combined and minced by iris scissors. Minced tissues were transferred to 2 ml fresh chelator solution (HBSS Ca/Mg-Free [Thermo Fisher 14175095], 10 mM EGTA [Fisher Scientific 50-255-956], 10 mM HEPES [Thermo Fisher 15630080], and 10% FBS [Gemini Bio 100-106]). The tissues were incubated at 37°C for 15 minutes with end-over-end rotation, then cooled on ice for 10 minutes. The epithelium was separated from lamina propria by mechanical agitation (shaking 20 times). Visualizing epithelial sheets in the solution, the supernatant was collected and spun down at 300xg for 5 minutes at 4°C. The remnant tissue chunks were rinsed in ice-cold PBS and incubated in 2 ml enzyme solution (HBSS Ca/Mg-Free [Thermo Fisher 14175095], 10 mM HEPES [Thermo Fisher 15630080], 200 µg/ml Liberase TM [Millipore Sigma 5401127001] and DNase I [Gold Bio D-301-100]) at 37°C for 20 minutes. The spun-down epithelium was resuspended and incubated in 2 ml TrypLE express [Thermo Fisher 12605010], at 37°C for 5 minutes. Both incubations were applied with end-over-end rotation. Tissues were gently triturated with a P1000 pipette after incubations and no visible tissue chunks remained after trituration. Cells were filtered through a 40 µM cell strainer [Falcon/VWR 21008–949]. After filtration, epithelial cells and cells from lamina propria were combined and spun down at 300xg for 5 minutes at 4°C. Cell pellet from each region were resuspended in 200 µl cold wash buffer (HBSS Ca/Mg-Free [Thermo Fisher 14175095] and 10% FBS [Gemini Bio 100-106]). Cell suspension was then diluted by 2 ml lysis buffer [Miltenyi Biotec 130-094-183] and sat for 3 minutes at room temperature to eliminate red blood cells. Cells in suspensions were spun at 300g for 5 minutes at 4°C. Cell pellets were resuspended in 200 µl suspension buffer (PBS and 0.04% BSA [Miltenyi Biotec 130-091-376]) and assessed for density and viability with Trypan Blue [Thermo Fisher 15250061]. Only cell suspensions with >85% viability were used for downstream single cell assays. The detailed processing is fully described here (https://www.protocols.io/view/preparation-of-single-cell-suspensions-from-human-q26g7bxqklwz/v4).

scRNA library preparation and sequencing

Single cells were loaded onto chip G per the manufacturer's protocol for the Chromium Single Cell 3' Library (V3.1) [10X Genomics 1000121]. In short, single cells were partitioned into Gel Beads in Emulsion (GEMs) in Chromium controller. GEMs were incubated for cell lysis and barcoded reverse transcription, followed by cDNA amplification. The cDNAs were then enzymatically fragmented and indexed. Approximately 8,300 single cells were loaded to a channel with a target recovery for 5,000 cells. Libraries were sequenced on Illumina Novaseq platform.

scATAC library preparation and sequencing

Single nuclei were isolated from single cell suspensions following the manufacturer's protocol for the Chromium Single cell ATAC (v 1.1). Briefly, single nuclei were transposed before loading to chip H. Transposed nuclei were partitioned into GEMS in Chromium controller. GEMs were incubated for barcoding, followed by fragment indexing. Approximately 8,000 single nuclei were transposed and loaded to a channel with a target recovery for 5,000 nuclei. Libraries were sequenced on Illumina Novaseq platform.

scRNA-seq Preprocessing and quality control

scRNA-seq gene expression raw sequencing data were processed using the Cell Ranger software v7.0.1and the 10X human transcriptome GRCh38-2020-A as the reference. The expected number of cells were auto-estimated in cellranger.

SoupX was used with default parameters for the estimation and removal of cell free mRNA contamination for each sample separately[35]. Clustering information from cellranger output was loaded to estimate the contamination fraction. R(V 4.1.0) and Seurat(V 4.2.0) were used to pool single cell counts and for downstream analyses[36]. Cells were filtered for more than 500 UMIs, fewer than 20,000 UMIs, more than 200 unique genes, fewer than 6,000 unique genes, less than 50% of reads mapping to mitochondrial genes and less than 40% of reads mapping to ribosomal genes and a log10GenesPerUMI score higher than 0.7 (to remove contamination with low complexity cell types like red blood cells). Genes expressed in fewer than 120 cells were filtered out (average 5 cells per sample in all 24 samples). Doublet score was calculated using Scrublet (V 0.2.3), and a cut-off score of 0.2 was applied to exclude doublets. The cells with unexpected co-expression of different lineage makers such as CD3D and EPCAM were also excluded as doublets. Gene expression for each cell was normalized and log transformed. Cell cycle scores (G2/M phase, S phase) were calculated using the expression of cell cycle genes listed in Seurat (V 4.2.0). Cell cycle score (the difference between the G2M and S phase scores), the percentage of mitochondrial reads and unique molecular identifiers (UMIs) were regressed out before scaling the data.

scRNA-seq dimensionality reduction, clustering analysis and visualization

We restricted the expression matrix to the 2000 most highly variable genes identified by fitting the mean-variance relationship and high-quality cells noted above as input for principal component analysis (PCA). Principal components were corrected using harmony (V 0.1.0)[37] to remove patient-specific effect signals. Cells were clustered using the Louvain algorithm[38] (resolution 0.3–1.5) for modularity optimization using the kNN graph as input. Cell clusters were visualized in Uniform Manifold Approximation and Projection (UMAP) embedding[39].

The union dataset was divided into different lineages based on clustering analysis and marker gene expression. Cells from epithelial, T cell, B cell, myeloid, stromal lineages were subset for further analysis. For each lineage, we repeated the process of lineage-specific QC, dimensionality reduction, principal component correction, Louvain clustering and UMAP visualization as

described above. Cell identities for cell subtypes in each lineage were assigned based on markers outlined in the following section.

Annotation of scRNA cells

Cell-type clusters were annotated based on marker genes previously reported as markers of intestinal cell populations[12].

Epithelial lineage cells were defined on the basis of EPCAM and FABP1 expression[12]. Stem cells had specific expression of LGR5 and SMOC2. Transit-amplifying cells were classified based on the highest expression of MKI67, TOP2A and TUBA1B. Enterocyte clusters with ileum functional signatures include EC1-1 (SI, GSTA2) and EC1-2 (SI, SLC15A1, APOA4, CUBN). Enterocytes in EC1-1 cluster is at an earlier differentiation stage than EC1-2, with higher OLFM4 expression. Enterocyte clusters with colonic signatures include EC2-1 (CD24, CA2) and EC2-2 (CD24, CA2, CEACAM5). EC2-1 cluster is at an earlier differentiation stage than EC2-2, with higher OLFM4 expression. Goblet clusters were identified based on high MUC2 and TFF3 expression. Goblet2 cluster is a dominating type of goblet cells in the colon with higher BEST2 and KLK1 expression compared to Goblet1. Paneth cells (DEFA5, ITLN2, REG3A) and Microfold-like cells (CCL20) are observed in ileum, pre-pouch and pouch. BEST4 enterocytes, enteroendocrine cells (CHGA, KCNB2, RIMBP2) and tuft cells (POU2F3, FYB1) are identified in all four regions.

T cell lineage cells were defined based on positive CD3D expression. CD4 T cells had expression of CD4, but not CD8A, and vice versa for CD8 T cells. Tissue resident memory T cells were characterized by high CD69 expression. CD4 tissue resident memory T cells were further divided based on *ITGAE* (encodes CD103) expression. $CD103^{-}$CD4 T cells had higher IL7R expression, while $CD103^{+}$CD4 T cells had higher IL17A and IL26 expression. CD8 tissue resident memory T cells include KLRG1 positive and negative clusters, indicating different cytotoxic functions. Central memory T cells had the highest LEF1, TCF7 and CCR7 expression. Regulatory T cells were characterized by high CTLA4 and unique FOXP3, IL2RA expression. Two $TRDC^{high}$ clusters were classified as gamma-delta T cells (ENTPD1, $GZMA^{high}$) and NK cells ($NCR1^{-}KLRF1^{-}CD3D^{-}$) respectively. NK T cells expressed CD3D, FGFBP2, as well as NK genes NCR1, KLRF1. Mucosal-associated invariant T (MAIT) cells showed high SLC4A10 and NCR3 expression. Innate lymphoid cells (ILCs) (PRKG1, PCDH9, AFF3, AREG, IL1R1, IL23R, KIT) are a diverse group of immune cells, but we were not able to further sub-cluster ILCs in this atlas.

B cell lineage cells were defined based on high CD79A and MS4A1 expression, and plasma cell clusters had high XBP1 and IGHA1 expression. BCL6 expression defines germinal center B cells, which is crucial for somatic hypermutation and class switch recombination. Naïve B cells had the highest expression of immunoglobulin heavy chains D (IGHD). Memory B cells were characterized by positive CD27 expression and negative IGHD. Class switch IgA and IgG plasma cells showed expression of the XBP1, and immunoglobulin heavy chains A and G, respectively.

Myeloid lineage cells were defined based on high expression of TYROBP and CPA3. Monocytes and macrophages both are CD14 and C1QA positive. Monocytes had high expression of VCAN, FCN1 and EREG, and macrophages had highest expression of CD163 and MMP12. Among dendritic cells (DCs) we observed cDC1 (CLEC9A) and cDC2 (CD1C), and lymphoid DCs (LAMP3). Mast cells were identified by CPA3, KIT and TPSB2. Neutrophils were classified by high expression of FCGR3B (encodes Fc gamma receptor 3B).

Stromal lineage cells were defined based on high expression of IGFBP7 and COL3A1. Fibroblasts include three main populations, fibroblast1 (ADAMDEC1, CCL11, CCL13), fibroblast2 (NRG1, NPY, PTGS1), fibroblast3 ($SOX6^{high}$). Endothelial cells (PECAM1) include arterial endothelium population (HEY1, EFNB2), venous endothelium population (ACKR1) and lymphatic endothelium population (PROX1, LYVE1, CCL21). Pericytes were identified by high NOTCH3 expression, and a contractile pericyte subset were characterized by high PLN, RERGL and KCNAB1 expression. Myofibroblasts and smooth muscle cells both show high actin (ACTA2) and transgelin (TAGLN) expression, while Myofibroblasts lack smooth muscle marker desmin (DES). Glial cells were characterized by S100B and NRXN1.

Marker gene detection and differential expression analysis

In order to regress out patient specific effects when calculating the differentially expressed genes in enterocytes across regions, we used a linear mixed model (model formula = cluster + (1|Patient_ID)) implemented in variancePartition (V 1.21.1)[40,41] to quantify gene expression attributable to individual patients. We verified that other confounding factors are not contributing a significant amount of variance, so we didn't involve other factors in the linear model. The cell number threshold to create pseudo-bulk cell clusters was set considering the size of the cell type population (enterocyte groups: 200 cells, goblet: 30 cells, stem cell: 30 cells). Differentially expressed genes in different clusters were identified by the 'FindAllMarkers' function (Wilcoxon Rank Sum test with default parameters) in Seurat. Other pairwise differential tests were performed by the 'FindMarkers' function (Wilcoxon Rank Sum test with default parameters).

Gene set enrichment analysis

Gene set categories (H, C2, C5, C7 and C8) implemented in Molecular Signatures Database (MSigDB) were pulled and examined for each list of differentially expressed genes using migdbr package (V 7.5.1). R package 'org.Hs.eg.db' (V 3.14.0) was used to map gene identifiers. Gene set enrichment analyses were performed using the 'clusterProfiler' R package 'GSEA' R function (pvalueCutoff = 0.05, pAdjustMethod = "fdr", minGSSize = 10, maxGSSize = 5000) (Yu et al., 2012).

Enrichment score of signature gene sets

'Area Under Curve' (AUC) was used to calculate the enrichment of a geneset for each cell, as implemented in AUCell (V 1.16.0)[42]. For each cell, the genes are ranked from highest to lowest value. The genes with the same expression value are shuffled. Therefore, genes with expression '0' are randomly sorted at the end of the ranking. The AUC value is representing the fraction of genes from the pathway within the top 5% (aucMaxRank=0.05) of all the genes in our dataset. Since the gene rankings are created individually for each cell, and the scoring method is only based on ranking, AUC score is independent of normalization method or how we subset the group of cells.

Immunostaining of human biopsy

Biopsies matched to the patients assessed in single-cell RNA-seq experiments were obtained from the pathology archives at UChicago Medicine under an IRB-approved protocol. Serial sections were prepared from the paraffin-embedded material. Immunohistochemical staining was

performed as previously described[43]. Briefly, sections were deparaffinized and rehydrated, heat/pressure-treated for 20 min with citrate buffer (pH 6), bleached with 0.3% hydrogen peroxide for 30 minutes, and blocked with horse serum. Primary antibodies were prepared in 0.3% Triton-X and 4% horse serum in phosphate-buffered saline (PBS) and incubated with the tissue overnight at 4C. Primary antibodies used were rabbit anti-APOA4 (1:1,000, Sigma HPA001352), rabbit anti-FABP6 (1:200, Sigma HPA012601), mouse anti-CA2 (Santa Cruz Biotechnology sc-48351), and mouse anti-CEACAM5 (R&D Systems MAB41281). After washing with PBS, sections were exposed to horseradish peroxidase-conjugated secondary antibodies (Immpress kit, Vector Labs) for 2 h at room temperature, and signal was developed for 2-10 minutes with a diaminobenzidine (DAB) kit (Vector Labs). Tissue was counterstained with methyl green prior to dehydration, clearing, and coverslipping. A whole-slide scanner (Olympus) mounted with a 20X objective was used to obtain images of the staining. Analysis was performed with QuPath[44].

Histology evaluation and quantitation of enteritis score

Histological assessment of hematoxylin-and-eosin stained biopsy sections was performed by a clinical pathologist (C.W.) blinded to the slide labels. The scoring method used was based on previous assessment[45] of colonic inflammation and mucosal damage. The score ranged from 0 (no abnormalities) to 3 (severe damage), with intermediate grading defined as follows:

Score	Category	Description
0	Quiescent or Inactive	No active inflammation (no intraepithelial neutrophils)
1	Mild Active	Presence of intraepithelial neutrophils
2	Moderate Active	Presence of crypt abscesses
3	Severe Active	Erosions or ulcers

scATAC analysis workflow preprocessing and quality control

scATAC-seq raw sequencing data were processed using the Cell Ranger ATAC 2.0.0, and the 10X human genome GRCh38-2020-A-2.0.0 as the reference. Further filtering was conducted using ArchR(v.1.0.3). Only high-quality individual cells were kept for downstream analysis (doublet filter ratio = 2.0, minimum fragments per cell = 2000, minimum TSS enrichment score = 6).

Dimensionality reduction and clustering analysis

We used ArchR to convert the fragments file into a 500bp tile matrix of fragment counts and generated a gene score matrix using the default model (model 42) with default parameters, taking into account chromatin accessibility within a gene body as well as proximally and distally from the TSS. We performed latent semantic indexing (LSI) on the tile matrix and retained the top 100 LSI vectors (addIterativeLSI function, 6 iterations including 5 different clustering resolutions: 0.5, 1, 1.5, 2, 2). The first LSI vector was removed because of high correlation with sequencing depth (correlation score > 0.75). The LSI vectors were further corrected using Harmony (v.0.1.0). The remaining 99 LSI vectors were used to create the UMAP (nNeighbors = 30, minDist = 0.2). Cell clusters were identified using the addClusters function with multiple different resolutions.

Annotation of scATAC cells

The scRNA-seq expression matrix was integrated with the scATAC-seq gene score matrix using the 'addGeneIntegrationMatrix' function from ArchR. It identified the corresponding cells across datasets using Seurat's mutual nearest neighbors algorithm, thus transferring the labels from scRNA data to scATAC data. The labels in scATAC data were further curated based on specific marker gene score distribution. A notable exception were TA cells which did not have a clear corresponding cell type in scATAC-seq. We interpreted this as a consequence of the annotation of TA cells based on the expression of cell cycle genes which might not show strong corresponding changes in chromatin accessibility at their genomic loci. We therefore removed TA designation from scATAC-seq and used progenitor/immature instead. In addition, M-cells which represent a small proportion of cells in the scRNA-seq data (< 1.7%) and which are transcriptionally distinguished from enterocytes solely by expression of CCL20 showed a diffuse distribution, so we removed this label.

Accessible chromatin peak calling

We merged cells within each cell group for the generation of pseudo-bulk replicates and then merged these replicates into a single insertion coverage file. The addReproduciblePeakSet function was used to get insertions from coverage files, call peaks, and merge peaks to get a "Union Reproducible Peak Set" (peaksPerCell = 500, maxPeaks = 150000, minCells = 50, extsize = 150, cutOff = 0.05, extendSummits = 250).

Differential accessible region (DAR) analysis

To identify cell-type specific marker peaks, a single-cell insertion count matrix was created using the function addPeakMatrix in ArchR. To perform DA, we used getMarkerFeatures（Wilcoxon rank-sum test, maxCells = 20000). To control for technical variation, "TSSEnrichment" and "log10(nFrags)" were indicated as biases to account for in selecting a matched null group for marker feature identification (FDR < 0.01, log2 fold-change > 1).

Motif annotation and TF motif enrichment

We used 870 human motif matrices from the CisBP[46] database to annotate each of the peaks in the union peakset for each lineage separately by the addMotifAnnotations function. The hypergeometric test was used to assess the enrichment of the number of times a motif overlaps with a given set of peaks, compared to random expectation (adjusted p-value < 0.001). The top 20 transcription factor motifs significantly enriched for each cell type were filtered based on top 3000 broadly expressed genes (ranked by the proportion of cells expressing this gene).

ChromVAR deviation score

We computed background peaks controlling for total accessibility and GC-content by the function addBgdPeaks and used the function addDeviationsMatrix function to calculate the TF ChromVAR activity score. The TFs for each enterocyte group were further further filtered based on positive or negative TF activity-expression correlation across enterocyte cell types.

Trajectory analysis

The trajectory analysis was performed using the 'addTrajectory' function in ArchR by specifying the cluster order in two enterocyte lineages (EC1 lineage: Stem1, EC1-1, EC1-2; EC2 lineage:

Stem2, EC2-1, EC2-2). We further visualized trajectory-dependent changes of TF activity and gene expression after extracting the pseudotime.

Domain of regulatory chromatin analysis

The domains of regulatory chromatin (DORCs) were identified using the standard workflow introduced in FigR (v.0.1.0). The integrated gene expression matrix and peak matrix were used to infer cis-regulatory interactions. Here the integrated gene expression matrix was obtained from the multiomic integration analysis introduced above. The minimum number of interactions (cutoff=7) was used to determine a significant DORC gene.

Gene regulatory network construction

Highly correlated links between gene expression and region accessibility were identified by addPeak2GeneLinks (corCutOff = 0.6), as candidate cis-regulatory elements (cCREs). The core transcription factors were first identified by the TF enrichment analysis of different groups of differentially accessible regions between enterocyte groups. These groups of TFs were then filtered for highly expressed genes. Finally, only the most significantly positively and negatively TF regulators in the same TF family were preserved as nodes in the network. We filtered the gene targets for the top 100 most significantly linked genes, the well-known functional genes (within top 300 most significantly linked genes), and top 100 DORC genes (within top 300 most significantly linked genes).

Data availability

The snRNA-seq and scATAC-seq data anlayzed in this study have been deposited in the GEO repository under accession code GSE242087.

Code availability

Code to reproduce data processing and analysis are available at:

https://github.com/EthanZhaoChem/Code_documentation_2023_IBD_pouch_paper

Supplementary Information

Table 1. Patient medical information

Table 2. Pathway enrichment analysis for multiple contrasts.

Table 3. Differentially expressed genes between pouch EC2 (Goblet2) and AC EC (Goblet).

Table 4. DORC (Domains of regulatory chromatin) genes and the number of associated regions.

Table 5. Differential gene expression analysis for multiple contrasts.

Acknowledgment

This work was enabled through grants from the National Institutes of Health R35GM142986 (to S.P.), RC2DK122394, T32DK007074, P30DK042086 and a grant from the GI Research Foundation (to E.B.C.), and a grant from The Leona M. and Harry B. Helmsley Charitable Trust to the University of Chicago. This work was completed in part with resources provided by the University of Chicago Research Computing Center. We thank Christian Jones for assistance with editing.

Author contributions

A.B. and S.P. conceived the study and supervised this project. R.Z. performed experiments and generated data. C.W. performed histological assessment of tissue sections. C.Y.L. and M.K. performed IHC and microscopy. Y.Z. performed data analysis. S.P. and Y. Z. wrote the manuscript with input from all authors.

Correspondence and requests for materials

Correspondence should be addressed to Anindita Basu (onibasu@uchicago.edu) or Sebastian Pott (spott@uchicago.edu)

Competing interests

The authors declare no competing interests.

References

Figures

Figure 1. Overview for the scRNA-seq profiling of 24 human intestinal biopsies from 4 different anatomical regions. (*A*) Schematic of biopsy sampling strategy for scRNA-seq. We profiled 6 pairs of terminal ileum (TI) and ascending colon (AC) biopsies from 6 healthy individuals, and 6 pairs of prepouch (PP) and pouch (POU) biopsies from 6 different healthy IPAA (ileal pouch-anal anastomosis) patients with ileal pouch. (*B*) Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction of 5 major cell lineages in all 24 samples. Each dot represents a single cell (103,772 cells in total). (*C*) High resolution UMAP showing 47 cell types in 24 biopsies. (*D*) Heatmap of top 10 differentially expressed genes and representative cell marker for each cell type within a lineage. Mean expression in each cell type was scaled as z score in each row across all cell types.

Figure 2. Identification of both ileum-like enterocytes and colon-like enterocytes in pouch. (*A*) UMAP of 59,751 cells in epithelial lineage. There are two different populations of enterocytes: ileum-like EC1 and colon-like EC2. EC1-2 and EC2-2 are more mature than EC1-1 and EC2-1 respectively. (*B*) Epithelial UMAP split based on the four different anatomical

regions. (*C*) Proportions of different cell types in each region. EC1 and EC2 are dominant enterocyte populations in terminal ileum and ascending colon, respectively. Pouch has significant proportions of both EC1 and EC2 enterocytes. Color legends for different cell types are shared in panel A, B and C. (*D*) Enterocytes are split into 5 groups: ascending colon enterocytes, terminal ileum enterocytes, pre-pouch enterocytes, and two populations of pouch enterocytes (EC1 ileum-like population and EC2 colon-like population, as shown in Supplementary Figure 3B). The differential gene tests are performed in multiple contrasts and all DEGs (differentially expressed genes) are summarized in the heatmap. The contrasts and corresponding number of DEGs are illustrated in Supplementary Figure 4A. The DEGs were further grouped into 10 different clusters by the set operations of different contrasts, as summarized in Supplementary Figure 4E. Mean expression in each cell group is shown as z-score in each row. (*E*) Heatmap showing pathway enrichment score in 5 enterocytes groups. The score is scaled across enterocyte groups in each column. (*F*) AUC enrichment score for antimicrobial humoral response and lipid metabolism related pathway activity. Wilcoxon rank-sum test is used for significance test: **** indicates $p < 0.0001$.

Figure 3. Heatmap showing key differentially expressed genes between (A) pouch EC2 and AC EC2, (B) pouch Goblet2 and AC Goblet. These genes were reported in previous single cell UC studies (Smillie et al.). Highlighted genes were reported in more than one study. Pink bars on top are the number of cells in each sample of the corresponding cell type. The RNA expression is scaled in each row among 12 different samples. The last row shows the histological enteritis score of the sample, as also shown in Figure 4B. The purple bar is indicating the differentially expressed genes reported to be up-regulated in (A) UC absorptive lineage or (B) secretory lineage in Smillie et al. The pink bar is indicating the differentially expressed genes reported to be down-regulated in (A) UC absorptive lineage or (B) secretory lineage in Smillie et al. Blue bar is indicating the genes up-regulated in TI (A) enterocytes (B) goblets when compared with AC (A) enterocytes (B) goblets. Green bar is indicating the genes up-regulated in AC (A) enterocytes (B) goblets when compared with TI (A) enterocytes and (B) goblets, respectively. Light yellow bar indicates the genes that are not differentially expressed when comparing TI and AC (A) enterocytes (B) goblets.

Figure 4. Validation of a metaplastic enterocyte population in UC pouches. (*A*) Representative immunohistochemistry images from pre-pouch and pouch biopsies for H&E, FABP6, CEACAM5, CA2 staining. (*B*) Histological enteritis score of pre-pouch and pouch samples. (*C*) Average gene expression in the bulk RNA-seq data from healthy pouch patients. (*D*) Average gene expression in the bulk RNA-seq data from patients with actively inflammatory pouches. The expressions in panel C and D are scaled in each column including prepouch and pouch. (*E*) The deconvolved cell type proportions in healthy pouches from bulk RNA-seq data. (*F*) The deconvolved cell type proportions in inflammatory pouches from bulk RNA-seq data.

Figure 5. scATAC-seq profiling of the same group of 24 biopsies studied in scRNA-seq. (*A*) UMAP and clustering based on single-cell chromatin accessibility identifies 39 distinct cell types from 88,595 cells in all lineages. (*B*) Epithelial lineage UMAP includes all cell types identified in scRNA-seq analysis. (*C*) Epithelial UMAP split for four different anatomical regions. (*D*) Proportions of different cell types in each region. (*E*) Normalized region accessibility in promoter regions of cell type specific marker genes, validating that the scATAC-seq clusters are consistent with scRNA-seq clusters.

Figure 6. Identification of marker peaks and transcription factors (TFs) in each cell type in epithelial lineage. (*A*) Marker peaks for each cell type in epithelial lineage. The peak accessibility is scaled in each column across cell types. (*B*) Transcription factors enriched in each cell type. (*C*) Differentially accessible regions comparing ascending colon enterocytes with terminal ileum enterocytes (top left) and comparing pouch EC2 enterocytes with pouch EC1 enterocytes (top right). The Venn diagram shows the intersection of peaks upregulated in AC vs TI and pouch EC2 vs pouch EC1 (bottom left) and the intersection of peaks upregulated in TI vs AC and pouch EC1 vs pouch EC2 (bottom right). (*D*) Stacked violin plot shows the RNA expression of different TFs in each enterocyte group. (*E*) Heatmap shows the scaled mean ChromVAR TF activity score in each enterocyte group. The score is scaled in each row across enterocyte groups (left). In pouch, the TF activity changes in EC1 lineage and EC2 lineage along simulated differentiation pseudotime (middle). The TF logos are plotted for each TF in each row.

Figure 7. Construction of TF-gene regulatory network in pouch enterocytes by peak-gene linkages based on multiomic analysis. (*A*) 31808 peak-gene links are identified based on the correlation of gene expression and peak accessibility with correlation score cutoff>0.6. (*B*) Peaks highly correlated with CA2 gene expression show higher accessibility in colon and pouch EC2 enterocytes (pointed by arrows). (*C*) TF-gene regulatory network shows that the TFs (red nodes) are dominating different groups of genes in pouch EC1 and EC2. The green edges are showing the strongest TF-gene correlations in pouch EC1. The purple edges are showing the strongest TF-gene correlations in pouch EC2.

Supplementary Figure 1. (A) Metadata of the patients involved in this study. (B) The number of cells in each biopsy sample that have been analyzed after QC in scRNA-seq analysis.

Supplementary Figure 2. (A) scRNA UMAP splitted by regions shows that the largest cross-region differences lie in epithelial lineage. (B) Markers for different cell lineages. The color shows scaled mean expression in each lineage. The dot size represents the proportion of cells expressing that marker.

Supplementary Figure 3. scRNA data. (A) Gene expression UMAPs showing the representative markers for epithelial stem cells (LGR5), transit-amplifying cells (MKI67), early stage enterocytes (OLFM4), ileum-like enterocytes (FABP6), colon-like enterocytes (CEACAM5). (B) Pouch enterocytes are splitted into two populations: pouch EC1 and pouch EC2. (C) Violin plots for FABP6 and CEACAM5 expression in 5 enterocyte groups. (D) Cell proportion in epithelial lineages. The statistical significance is tested by the Wilcoxon rank sum test. ns: $p > 0.05$, *: $p <= 0.05$, **: $p <= 0.01$, ***: $p <= 0.001$, ****: $p <= 0.0001$.

Supplementary Figure 4. scRNA data. (A) Number of differentially expressed genes that are shared across different contrasts. (B) Volcano plot shows the top differentially expressed genes comparing pouch EC2 vs pouch EC1. (C) Volcano plot shows the top differentially expressed genes comparing Stem2 vs pouch Stem1. (D) Volcano plot shows the top differentially expressed genes comparing Goblet2 vs pGoblet1. (E) The DEGs across different enterocyte groups are grouped into 10 different clusters by the set operations of different contrasts.

Supplementary Figure 5. scRNA data. (A) The epithelial cell type proportions in each sample. Volcano plot shows the top differentially expressed genes comparing different enterocyte groups (B) PP vs pouch EC2, (C) TI vs AC, (D) pouch EC2 vs AC, (E) pouch Goblet2 vs AC Goblet.

Supplementary Figure 6. scRNA data. Volcano plot shows the top differentially expressed genes comparing different enterocyte groups (A) PP vs TI, (B) PP vs pouch EC1, (C) all EC2 vs EC1. (D) Number of differentially expressed genes that are shared across contrasts in different cell types.

Supplementary Figure 7. scRNA data. (A) UMAP shows different cell types in T cell lineage. (B) The proportion of different cell types in different regions. (C) Cell types that show proportional differences across regions. (D) The cell type proportions in each sample

Supplementary Figure 8. scRNA data. (A) UMAP shows different cell types in B cell lineage. (B) The proportion of different cell types in different regions. (C) The cell type that shows proportional differences across regions. (D) The cell type proportions in each sample

Supplementary Figure 9. scRNA data. (A) UMAP shows different cell types in myeloid cell lineage. (B) The proportion of different cell types in different regions. (C) The cell type proportions in each sample

Supplementary Figure 10. scRNA data. (A) UMAP shows different cell types in stromal cell lineage. (B) The proportion of different cell types in different regions. (C) (D) Cell types that show proportional differences across regions. (E) The cell type proportions in each sample

Supplementary Figure 11. scRNA data. Markers for different cell types in epithelial lineage. The color shows scaled mean expression in each cell type. The dot size represents the proportion of cells expressing that marker.

Supplementary Figure 12. scRNA data. Markers for different cell types in T cell lineage. The color shows scaled mean expression in each cell type. The dot size represents the proportion of cells expressing that marker.

Supplementary Figure 13. scRNA data. Markers for different cell types in (A) B cell lineage (B) myeloid lineage. The color shows scaled mean expression in each cell type. The dot size represents the proportion of cells expressing that marker.

Supplementary Figure 14. scRNA data. Markers for different cell types in stromal cell lineage. The color shows scaled mean expression in each cell type. The dot size represents the proportion of cells expressing that marker.

Supplementary Figure 15. (*A, B*) Representative immunohistochemistry images from ascending colon and pouch biopsies for FABP6, CEACAM5, CA2 staining. (*C*) The deconvolved cell type proportions in healthy prepouches from bulk RNA-seq data. (*D*) The deconvolved cell type proportions in inflammatory prepouches from bulk RNA-seq data.

Supplementary Figure 16. (A) Gene score of each cell type in scATAC-seq clusters. The markers are the same group of genes as scRNA analysis. (B) Sample wise within lineage cell

type proportion in each region. The cell groups with fewer than 50 cells are labeled with '-' to indicate low confidence value.

Supplementary Figure 17. scATAC data. (A) The number of cells in each biopsy sample that have been analyzed after QC in scATAC-seq analysis. (B) The epithelial cell type proportions in each sample.(C) Gene scores for marker genes of different epithelial subpopulations overlayed onto UMAP of epithelial cells. (D) The number of peaks that were initially called in each cell type.

Supplementary Figure 18. scATAC data. (A) UMAP shows different cell types in T cell lineage. (B) The proportion of different cell types in different regions. (C) The cell type proportions in each sample. (D) Marker peaks for each cell type. (E) Enriched TF motifs in each cell type.

Supplementary Figure 19. scATAC data. (A) UMAP shows different cell types in B cell lineage. (B) The proportion of different cell types in different regions. (C) The cell type proportions in each sample. (D) Marker peaks for each cell type. (E) Enriched TF motifs in each cell type.

Supplementary Figure 20. scATAC data. (A) UMAP shows different cell types in myeloid cell lineage. (B) The proportion of different cell types in different regions. (C) The cell type proportions in each sample. (D) Marker peaks for each cell type. (E) Enriched TF motifs in each cell type.

Supplementary Figure 21. scATAC data. (A) UMAP shows different cell types in stromal cell lineage. (B) The proportion of different cell types in different regions. (C) The cell type proportions in each sample. (D) Marker peaks for each cell type. (E) Enriched TF motifs in each cell type.

Supplementary Figure 22. scATAC data. (A) Number of differentially accessible peaks that are shared across different contrasts. (B) Enriched TF motifs in each differential test. (C) TFs are categorized into activators and repressors. (D) The normalized gene expression and scaled average chromVAR score in 5 enterocyte groups. (E) The logos of repressive TFs.

Supplementary Figure 23. scATAC data. (A) Differentially accessible peaks comparing Stem2 and Stem1 and enriched TF motifs. (B) Differentially accessible peaks comparing Goblet2 and Goblet1 and enriched TF motifs.

Supplementary Figure 24. Multiomic analysis. (A) UMAP of enterocytes and stem cells in pouch. (B) Pseudotime of EC1 differentiation. (C) Pseudotime of EC2 differentiation. (D, E) Gene expression changes along pseudotime in each lineage.

Supplementary Figure 25. Multiomic analysis. (A) UMAP of enterocytes and stem cells in AC and TI. (B) Pseudotime of EC1 differentiation. (C) Pseudotime of EC2 differentiation. (D) In AC and TI, the TF activity changes in EC1 lineage and EC2 lineage along simulated differentiation pseudotime.

Supplementary Figure 26. Multiomic analysis. For the genes that have at least one linked peak, the number of linked peaks is calculated for each gene, and then plotted as a distribution. The genes that are linked to a higher number of peaks are usually DEGs (purple).

Supplementary Figure 27. Multiomic analysis. (A) Peaks highly correlated with APOA4 gene expression show higher accessibility in terminal ileum, prepouch and pouch EC1 enterocytes. (B) Peaks highly correlated with MAF gene expression show higher accessibility in terminal ileum, prepouch and pouch EC1 enterocytes. (C) Peaks highly correlated with SATB2 gene expression show higher accessibility in ascending colon and pouch EC2 enterocytes.

1. Fazio VW, Kiran RP, Remzi FH, et al. Ileal Pouch Anal Anastomosis: Analysis of Outcome and Quality of Life in 3707 Patients. Ann Surg 2013;257:679.

2. Hoda KM, Collins JF, Knigge KL, et al. Predictors of pouchitis after ileal pouch-anal anastomosis: a retrospective review. Dis Colon Rectum 2008;51:554–560.

3. Penna C, Dozois R, Tremaine W, et al. Pouchitis after ileal pouch-anal anastomosis for ulcerative colitis occurs with increased frequency in patients with associated primary sclerosing cholangitis. Gut 1996;38:234–239.

4. Stocchi L, Pemberton JH. Pouch and pouchitis. Gastroenterol Clin North Am 2001;30:223–241.

5. Landy J, Al-Hassi HO, McLaughlin SD, et al. Etiology of pouchitis. Inflamm Bowel Dis 2012;18:1146–1155.

6. Toritani K, Kimura H, Otani M, et al. Inflammatory bowel disease-specific findings are common morphological changes in the ileal pouch with ulcerative colitis. Sci Rep 2022;12:20361.

7. Mitsialis V, Wall S, Liu P, et al. Single-Cell Analyses of Colon and Blood Reveal Distinct Immune Cell Signatures of Ulcerative Colitis and Crohn's Disease. Gastroenterology 2020;159:591-608.e10.

8. Ettorre GM, Pescatori M, Panis Y, et al. Mucosal changes in ileal pouches after restorative proctocolectomy for ulcerative and Crohn's colitis. Dis Colon Rectum 2000;43:1743–1748.

9. Huang Y, Dalal S, Antonopoulos D, et al. Early Transcriptomic Changes in the Ileal Pouch Provide Insight into the Molecular Pathogenesis of Pouchitis and Ulcerative Colitis: Inflamm Bowel Dis 2017:1.

10. Morgan XC, Kabakchiev B, Waldron L, et al. Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. Genome Biol 2015;16:67.

11. Devlin JC, Axelrad J, Hine AM, et al. Single-Cell Transcriptional Survey of Ileal-Anal Pouch Immune Cells From Ulcerative Colitis Patients. Gastroenterology 2021;160:1679–1693.

12. Elmentaite R, Kumasaka N, Roberts K, et al. Cells of the human intestinal tract mapped across space and time. Nature 2021;597:250–255.

13. Boland BS, He Z, Tsai MS, et al. Heterogeneity and clonal relationships of adaptive immune cells in ulcerative colitis revealed by single-cell analyses. Sci Immunol 2020;5:eabb4432.

14. 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;62:967–976.

15. Fukui H, Sekikawa A, Tanaka H, et al. DMBT1 is a novel gene induced by IL-22 in ulcerative colitis. Inflamm Bowel Dis 2011;17:1177–1188.

16. Rosenstiel P, Sina C, End C, et al. Regulation of DMBT1 via NOD2 and TLR4 in Intestinal Epithelial Cells Modulates Bacterial Recognition and Invasion. J Immunol 2007;178:8203–8211.

17. Dotti I, Mora-Buch R, Ferrer-Picón E, et al. Alterations in the epithelial stem cell compartment could contribute to permanent changes in the mucosa of patients with ulcerative colitis. Gut 2017;66:2069–2079.

18. Beelen Granlund A van, Østvik AE, Brenna Ø, et al. REG gene expression in inflamed and healthy colon mucosa explored by in situ hybridisation. Cell Tissue Res 2013;352:639–646.

19. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell 2019;178:714-730.e22.

20. Wang X, Park J, Susztak K, et al. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun 2019;10:380.

21. Buenrostro JD, Wu B, Litzenburger UM, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015;523:486–490.

22. Granja JM, Corces MR, Pierce SE, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet 2021;53:403–411.

23. Domcke S, Hill AJ, Daza RM, et al. A human cell atlas of fetal chromatin accessibility. Science 2020;370:eaba7612.

24. Hickey JW, Becker WR, Nevins SA, et al. Organization of the human intestine at single-cell resolution. Nature 2023;619:572–584.

25. Kartha VK, Duarte FM, Hu Y, et al. Functional inference of gene regulation using single-cell multi-omics. Cell Genomics 2022;2:100166.

26. Yu Q, Kilik U, Holloway EM, et al. Charting human development using a multi-endodermal organ atlas and organoid models. Cell 2021;184:3281-3298.e22.

27. Gu W, Wang H, Huang X, et al. SATB2 preserves colon stem cell identity and mediates ileum-colon conversion via enhancer remodeling. Cell Stem Cell 2022;29:101-115.e10.

28. Luk IY, Jenkins LJ, Schoffer KL, et al. Epithelial de-differentiation triggered by co-ordinate epigenetic inactivation of the EHF and CDX1 transcription factors drives colorectal cancer progression. Cell Death Differ 2022;29:2288–2302.

29. Kong L, Pokatayev V, Lefkovith A, et al. The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon. Immunity 2023;56:444-458.e5.

30. Parian AM, Limketkai BN, Chowdhury R, et al. Serrated Epithelial Change Is Associated With High Rates of Neoplasia in Ulcerative Colitis Patients: A Case-controlled Study and Systematic Review With Meta-analysis. Inflamm Bowel Dis 2021;27:1475–1481.

31. Weiser M, Simon JM, Kochar B, et al. Molecular classification of Crohn's disease reveals two clinically relevant subtypes. Gut 2018;67:36–42.

32. Sriranganathan D, Kilic Y, Nabil Quraishi M, et al. Prevalence of pouchitis in both ulcerative colitis and familial adenomatous polyposis: A systematic review and meta-analysis. Colorectal Dis 2022;24:27–39.

33. Fujii M, Matano M, Toshimitsu K, et al. Human Intestinal Organoids Maintain Self-Renewal Capacity and Cellular Diversity in Niche-Inspired Culture Condition. Cell Stem Cell 2018;23:787-793.e6.

34. Persson EK, Uronen-Hansson H, Semmrich M, et al. IRF4 Transcription-Factor-Dependent CD103+CD11b+ Dendritic Cells Drive Mucosal T Helper 17 Cell Differentiation. Immunity 2013;38:958–969.

35. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. GigaScience 2020;9:giaa151.

36. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184:3573-3587.e29.

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

38. Blondel VD, Guillaume J-L, Lambiotte R, et al. Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008;2008:P10008.

39. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 2018.

40. Hoffman GE, Roussos P. Dream: powerful differential expression analysis for repeated measures designs. Bioinforma Oxf Engl 2021;37:192–201.

41. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics 2016;17:483.

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

43. Liu CY, Girish N, Gomez ML, et al. Transitional Anal Cells Mediate Colonic Re-epithelialization in Colitis. Gastroenterology 2022;162:1975–1989.

44. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep 2017;7:16878.

45. Weber CR, Nalle SC, Tretiakova M, et al. Claudin-1 and claudin-2 expression is elevated in inflammatory bowel disease and may contribute to early neoplastic transformation. Lab Investig J Tech Methods Pathol 2008;88:1110–1120.

46. Weirauch MT, Yang A, Albu M, et al. Determination and inference of eukaryotic transcription factor sequence specificity. Cell 2014;158:1431–1443.

Figure 1

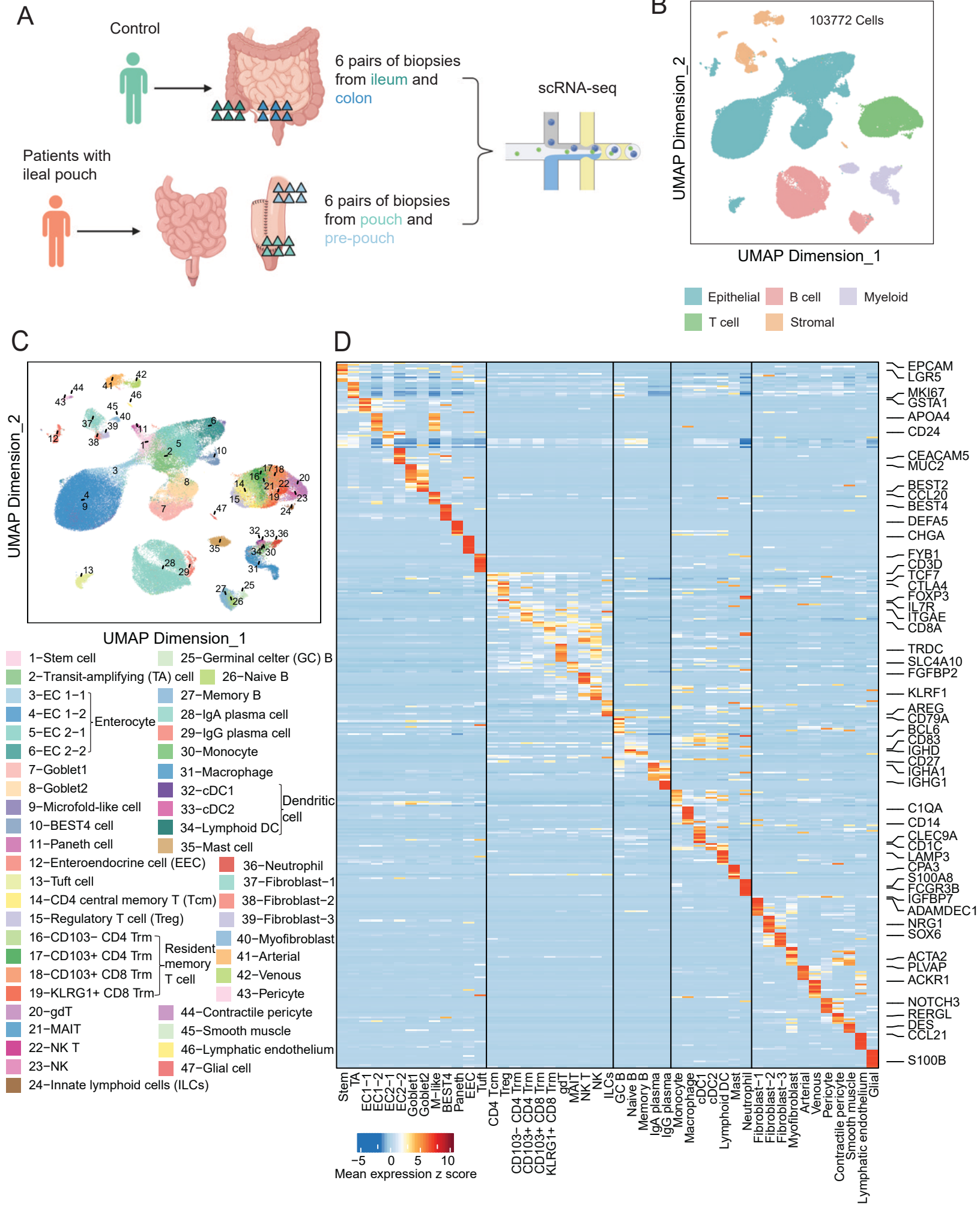

A
Control
6 pairs of biopsies from ileum and colon
Patients with ileal pouch
6 pairs of biopsies from pouch and pre-pouch
scRNA-seq
B
103772 Cells
UMAP Dimension_2
UMAP Dimension_1
Epithelial
B cell
Myeloid
T cell
Stromal
C
UMAP Dimension_2
UMAP Dimension_1
1−Stem cell
2−Transit-amplifying (TA) cell
3−EC 1−1
4−EC 1−2
5−EC 2−1
6−EC 2−2
Enterocyte
7−Goblet1
8−Goblet2
9−Microfold-like cell
10−BEST4 cell
11−Paneth cell
12−Enteroendocrine cell (EEC)
13−Tuft cell
14−CD4 central memory T (Tcm)
15−Regulatory T cell (Treg)
16−CD103− CD4 Trm
17−CD103+ CD4 Trm
18−CD103+ CD8 Trm
19−KLRG1+ CD8 Trm
Resident memory T cell
20−gdT
21−MAIT
22−NK T
23−NK
24−Innate lymphoid cells (ILCs)
25−Germinal celter (GC) B
26−Naive B
27−Memory B
28−IgA plasma cell
29−IgG plasma cell
30−Monocyte
31−Macrophage
32−cDC1
33−cDC2
34−Lymphoid DC
Dendritic cell
35−Mast cell
36−Neutrophil
37−Fibroblast−1
38−Fibroblast−2
39−Fibroblast−3
40−Myofibroblast
41−Arterial
42−Venous
43−Pericyte
44−Contractile pericyte
45−Smooth muscle
46−Lymphatic endothelium
47−Glial cell
D
EPCAM
LGR5
MKI67
GSTA1
APOA4
CD24
CEACAM5
MUC2
BEST2
CCL20
BEST4
DEFA5
CHGA
FYB1
CD3D
TCF7
CTLA4
FOXP3
IL7R
ITGAE
CD8A
TRDC
SLC4A10
FGFBP2
KLRF1
AREG
CD79A
BCL6
CD83
IGHD
CD27
IGHA1
IGHG1
C1QA
CD14
CLEC9A
CD1C
LAMP3
CPA3
S100A8
FCGR3B
IGFBP7
ADAMDEC1
NRG1
SOX6
ACTA2
PLVAP
ACKR1
NOTCH3
RERGL
DES
CCL21
S100B
Stem
TA
EC1−1
EC1−2
EC2−1
EC2−2
Goblet1
Goblet2
M−like
BEST4
Paneth
EEC
Tuft
CD4 Tcm
Treg
CD103− CD4 Trm
CD103+ CD4 Trm
CD103+ CD8 Trm
KLRG1+ CD8 Trm
gdT
MAIT
NK T
NK
ILCs
GC B
Naive B
Memory B
IgA plasma
IgG plasma
Monocyte
Macrophage
cDC1
cDC2
Lymphoid DC
Mast
Neutrophil
Fibroblast−1
Fibroblast−2
Fibroblast−3
Myofibroblast
Arterial
Venous
Pericyte
Contractile pericyte
Smooth muscle
Lymphatic endothelium
Glial
−5
0
5
10
Mean expression z score

Figure 2

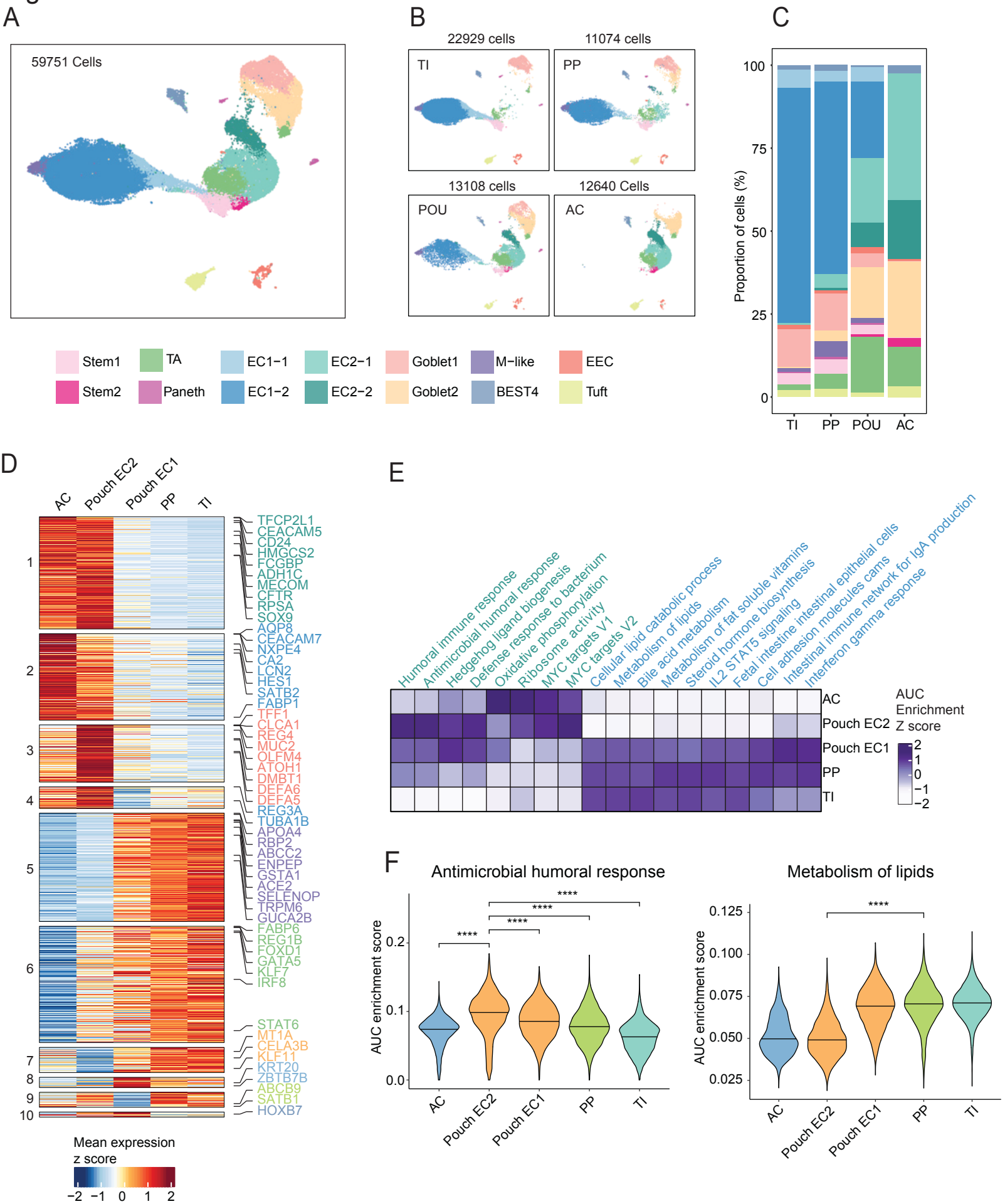

A
59751 Cells
B
22929 cells
TI
11074 cells
PP
13108 cells
POU
12640 Cells
AC
C
Proportion of cells (%)
100
75
50
25
0
TI
PP
POU
AC
Stem1
Stem2
TA
Paneth
EC1−1
EC1−2
EC2−1
EC2−2
Goblet1
Goblet2
M−like
BEST4
EEC
Tuft
D
AC
Pouch EC2
Pouch EC1
PP
TI
1
2
3
4
5
6
7
8
9
10
TFCP2L1
CEACAM5
CD24
HMGCS2
FCGBP
ADH1C
MECOM
CFTR
RPSA
SOX9
AQP8
CEACAM7
NXPE4
CA2
LCN2
HES1
SATB2
FABP1
TFF1
CLCA1
REG4
MUC2
OLFM4
ATOH1
DMBT1
DEFA6
DEFA5
REG3A
TUBA1B
APOA4
RBP2
ABCC2
ENPEP
GSTA1
ACE2
SELENOP
TRPM6
GUCA2B
FABP6
REG1B
FOXD1
GATA5
KLF7
IRF8
STAT6
MT1A
CELA3B
KLF11
KRT20
ZBTB7B
ABCB9
SATB1
HOXB7
Mean expression
z score
−2
−1
0
1
2
E
Humoral immune response
Antimicrobial humoral response
Hedgehog ligand biogenesis
Defense response to bacterium
Oxidative phosphorylation
Ribosome activity
MYC targets V1
MYC targets V2
Cellular lipid catabolic process
Metabolism of lipids
Bile acid metabolism
Metabolism of fat soluble vitamins
Steroid hormone biosynthesis
IL2 STAT5 signaling
Fetal intestine intestinal epithelial cells
Cell adhesion molecules cams
Intestinal immune network for IgA production
Interferon gamma response
AC
Pouch EC2
Pouch EC1
PP
TI
AUC
Enrichment
Z score
2
1
0
−1
−2
F
Antimicrobial humoral response
AUC enrichment score
0.2
0.1
0.0
****
****
****
****
AC
Pouch EC2
Pouch EC1
PP
TI
Metabolism of lipids
AUC enrichment score
0.125
0.100
0.075
0.050
0.025
****
AC
Pouch EC2
Pouch EC1
PP
TI

Figure 3

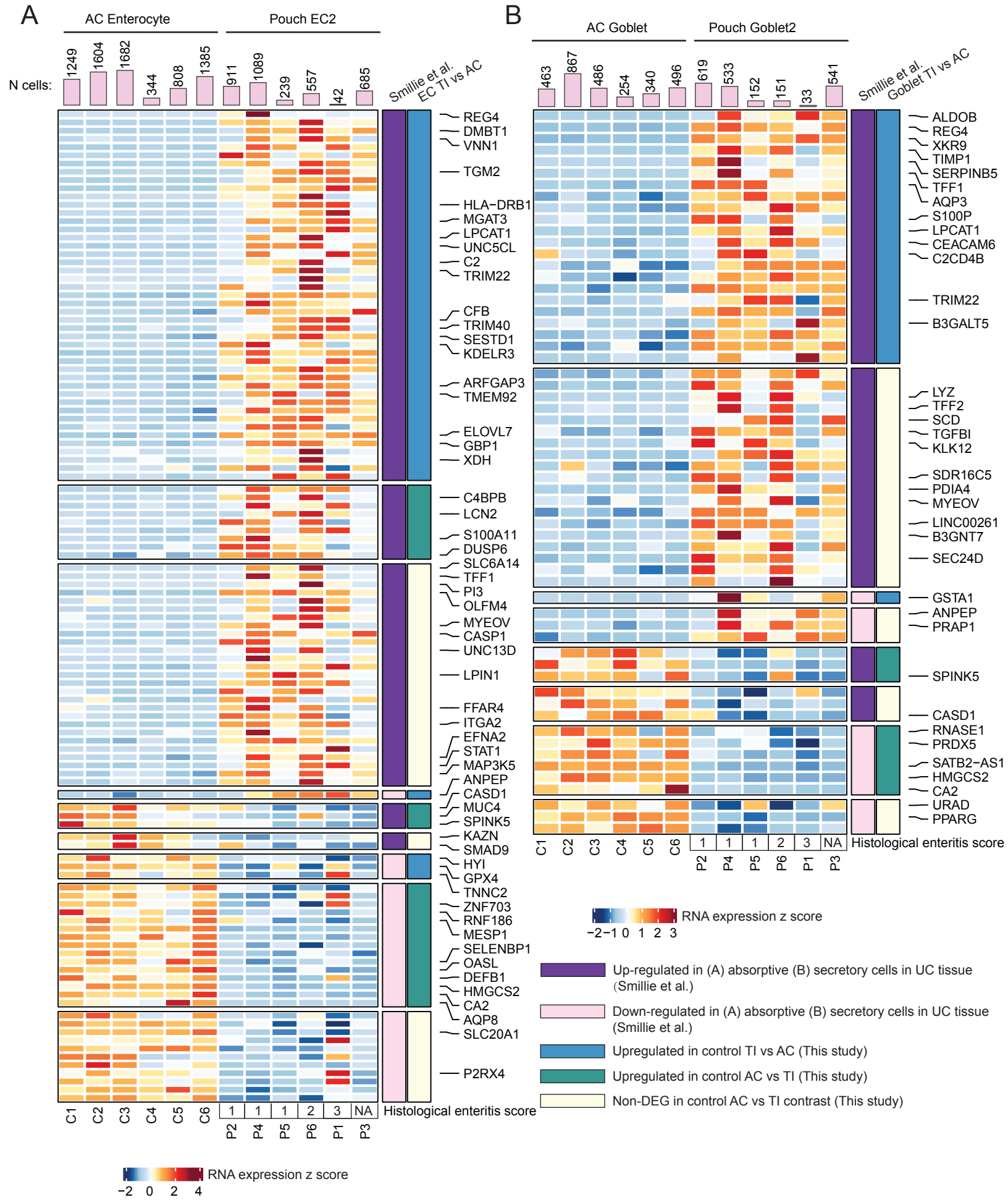

A
AC Enterocyte
Pouch EC2
N cells:
1249
1604
1682
344
808
1385
911
1089
239
557
42
685
Smillie et al. EC TI vs AC
REG4
DMBT1
VNN1
TGM2
HLA-DRB1
MGAT3
LPCAT1
UNC5CL
C2
TRIM22
CFB
TRIM40
SESTD1
KDELR3
ARFGAP3
TMEM92
ELOVL7
GBP1
XDH
C4BPB
LCN2
S100A11
DUSP6
SLC6A14
TFF1
PI3
OLFM4
MYEOV
CASP1
UNC13D
LPIN1
FFAR4
ITGA2
EFNA2
STAT1
MAP3K5
ANPEP
CASD1
MUC4
SPINK5
KAZN
SMAD9
HYI
GPX4
TNNC2
ZNF703
RNF186
MESP1
SELENBP1
OASL
DEFB1
HMGCS2
CA2
AQP8
SLC20A1
P2RX4
C1
C2
C3
C4
C5
C6
1
1
1
2
3
NA
P2
P4
P5
P6
P1
P3
Histological enteritis score
RNA expression z score
−2
0
2
4
B
AC Goblet
Pouch Goblet2
463
867
486
254
340
496
619
533
152
151
33
541
Smillie et al. Goblet TI vs AC
ALDOB
REG4
XKR9
TIMP1
SERPINB5
TFF1
AQP3
S100P
LPCAT1
CEACAM6
C2CD4B
TRIM22
B3GALT5
LYZ
TFF2
SCD
TGFBI
KLK12
SDR16C5
PDIA4
MYEOV
LINC00261
B3GNT7
SEC24D
GSTA1
ANPEP
PRAP1
SPINK5
CASD1
RNASE1
PRDX5
SATB2-AS1
HMGCS2
CA2
URAD
PPARG
C1
C2
C3
C4
C5
C6
1
1
1
2
3
NA
P2
P4
P5
P6
P1
P3
Histological enteritis score
RNA expression z score
−2
−1
0
1
2
3
Up-regulated in (A) absorptive (B) secretory cells in UC tissue (Smillie et al.)
Down-regulated in (A) absorptive (B) secretory cells in UC tissue (Smillie et al.)
Upregulated in control TI vs AC (This study)
Upregulated in control AC vs TI (This study)
Non-DEG in control AC vs TI contrast (This study)

Figure 4

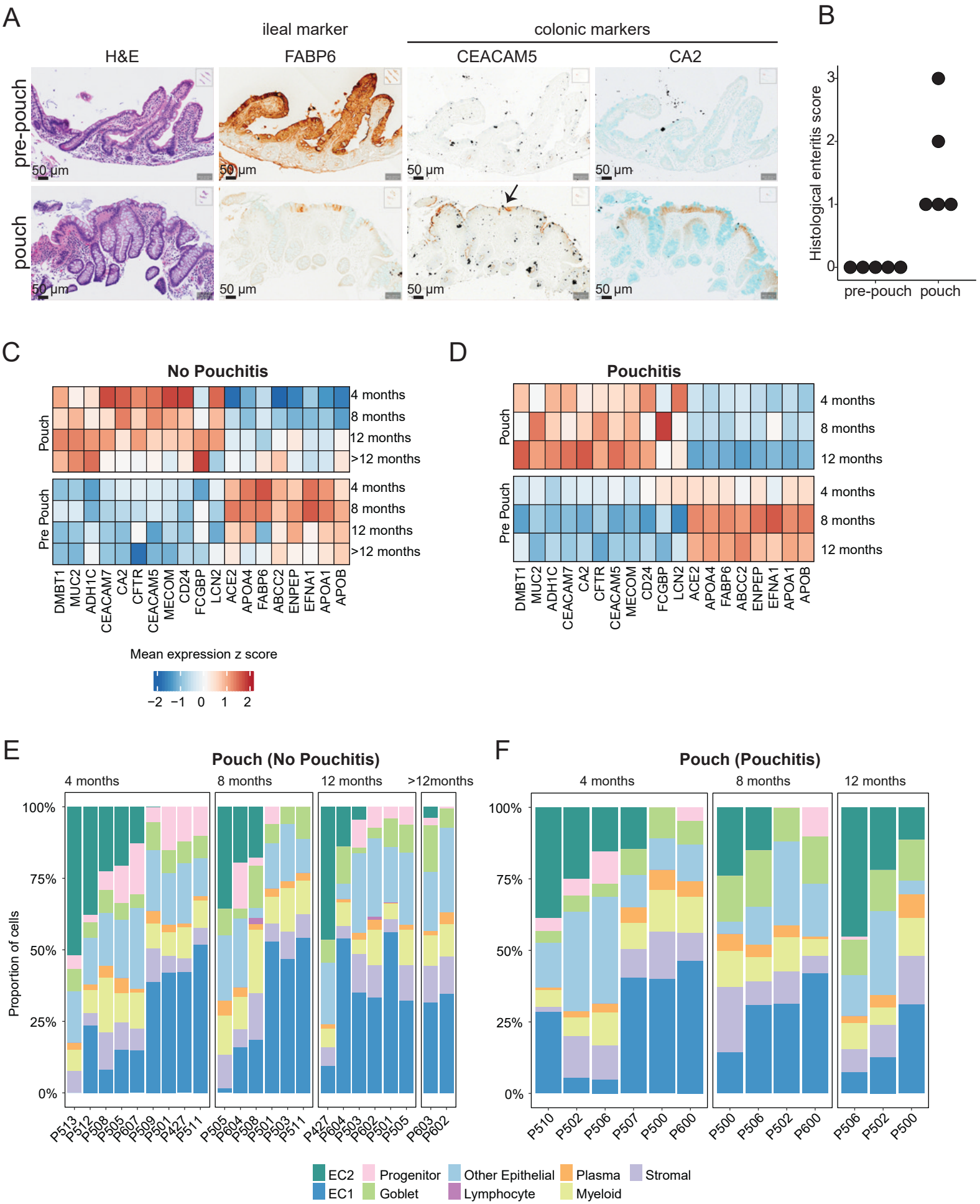

A
ileal marker
colonic markers
H&E
FABP6
CEACAM5
CA2
pre-pouch
pouch
50 µm
B
Histological enteritis score
pre-pouch
pouch
C
No Pouchitis
Pouch
Pre Pouch
4 months
8 months
12 months
>12 months
DMBT1
MUC2
ADH1C
CEACAM7
CA2
CFTR
CEACAM5
MECOM
CD24
FCGBP
LCN2
ACE2
APOA4
FABP6
ABCC2
ENPEP
EFNA1
APOA1
APOB
Mean expression z score
−2 −1 0 1 2
D
Pouchitis
E
Pouch (No Pouchitis)
4 months
8 months
12 months
>12months
Proportion of cells
100%
75%
50%
25%
0%
F
Pouch (Pouchitis)
EC2
EC1
Progenitor
Goblet
Other Epithelial
Lymphocyte
Plasma
Myeloid
Stromal

Figure 5

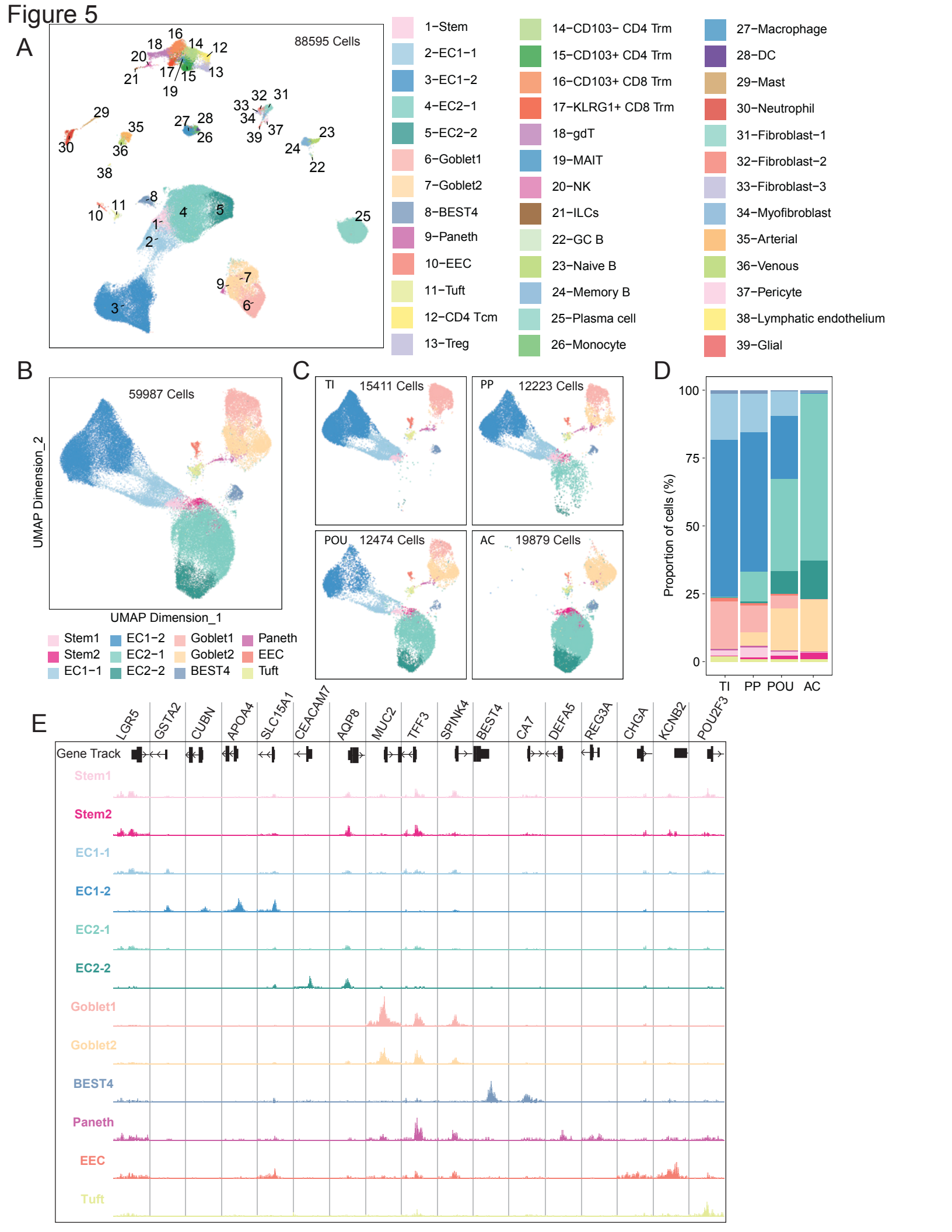

A
88595 Cells
1-Stem
2-EC1-1
3-EC1-2
4-EC2-1
5-EC2-2
6-Goblet1
7-Goblet2
8-BEST4
9-Paneth
10-EEC
11-Tuft
12-CD4 Tcm
13-Treg
14-CD103- CD4 Trm
15-CD103+ CD4 Trm
16-CD103+ CD8 Trm
17-KLRG1+ CD8 Trm
18-gdT
19-MAIT
20-NK
21-ILCs
22-GC B
23-Naive B
24-Memory B
25-Plasma cell
26-Monocyte
27-Macrophage
28-DC
29-Mast
30-Neutrophil
31-Fibroblast-1
32-Fibroblast-2
33-Fibroblast-3
34-Myofibroblast
35-Arterial
36-Venous
37-Pericyte
38-Lymphatic endothelium
39-Glial
B
59987 Cells
UMAP Dimension_2
UMAP Dimension_1
Stem1
Stem2
EC1-1
EC1-2
EC2-1
EC2-2
Goblet1
Goblet2
BEST4
Paneth
EEC
Tuft
C
TI 15411 Cells
PP 12223 Cells
POU 12474 Cells
AC 19879 Cells
D
Proportion of cells (%)
100
75
50
25
0
TI
PP
POU
AC
E
LGR5
GSTA2
CUBN
APOA4
SLC15A1
CEACAM7
AQP8
MUC2
TFF3
SPINK4
BEST4
CA7
DEFA5
REG3A
CHGA
KCNB2
POU2F3
Gene Track
Stem1
Stem2
EC1-1
EC1-2
EC2-1
EC2-2
Goblet1
Goblet2
BEST4
Paneth
EEC
Tuft

Figure 6

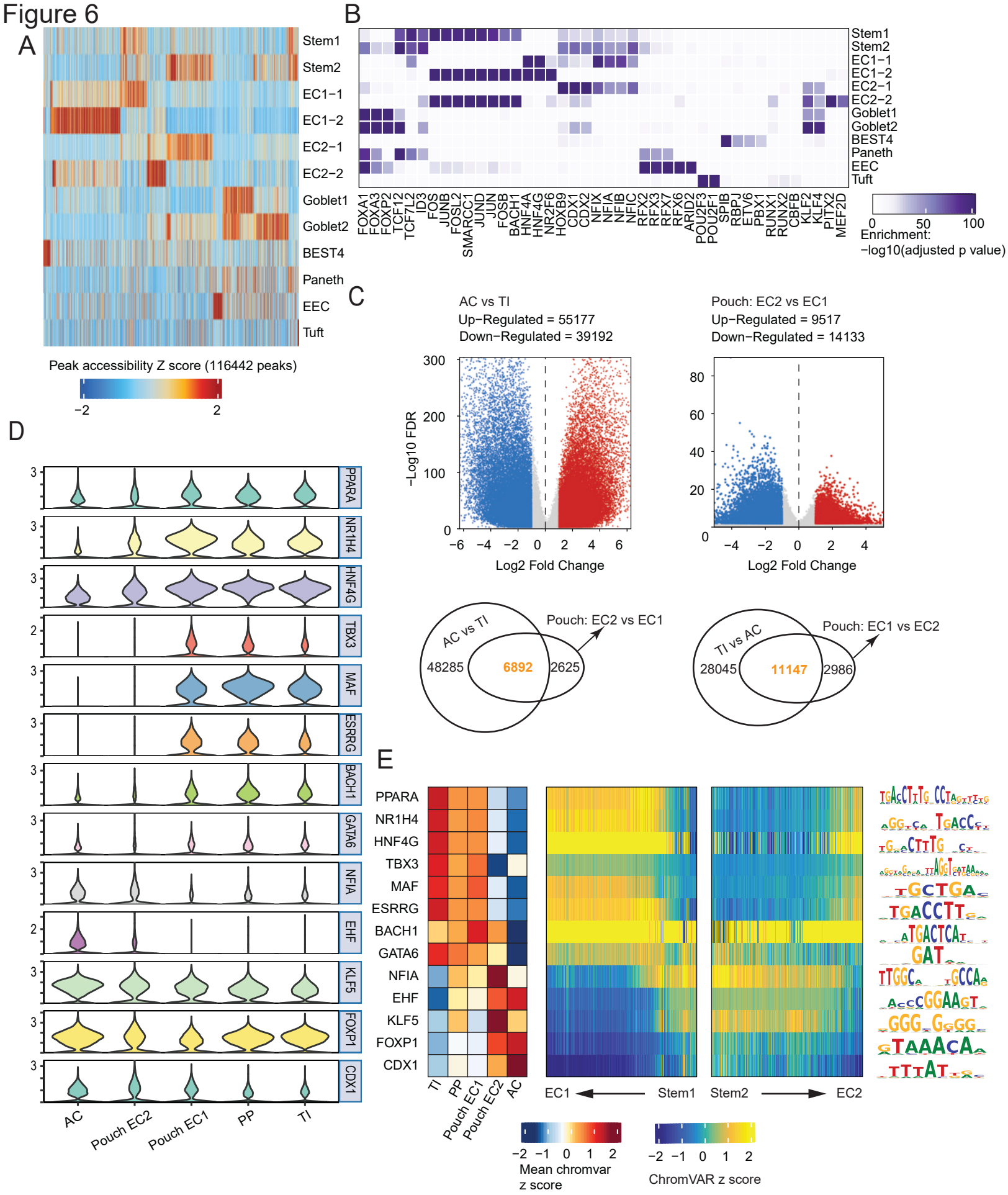
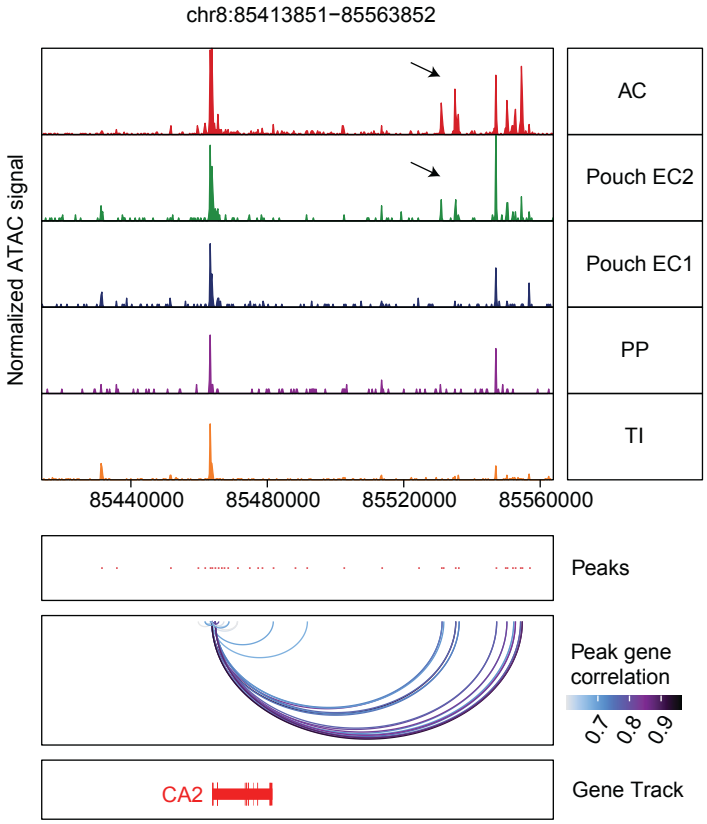

A
Stem1
Stem2
EC1−1
EC1−2
EC2−1
EC2−2
Goblet1
Goblet2
BEST4
Paneth
EEC
Tuft
Peak accessibility Z score (116442 peaks)
−2
2
B
FOXA1 FOXA3 FOXP2 TCF12 TCF7L2 ID3 FOS JUNB FOSL2 SMARCC1 JUND JUN FOSB BACH1 HNF4A HNF4G NR2F6 HOXB9 CDX1 CDX2 NFIX NFIA NFIB NFIC RFX2 RFX3 RFX7 RFX6 ARID2 POU2F3 POU2F1 SPIB RBPJ ETV6 PBX1 RUNX1 RUNX2 CBFB KLF2 KLF4 PITX2 MEF2D
Enrichment: −log10(adjusted p value)
0 50 100
C
AC vs TI
Up−Regulated = 55177
Down−Regulated = 39192
−Log10 FDR
Log2 Fold Change
Pouch: EC2 vs EC1
Up−Regulated = 9517
Down−Regulated = 14133
Log2 Fold Change
AC vs TI
Pouch: EC2 vs EC1
48285
6892
2625
TI vs AC
Pouch: EC1 vs EC2
28045
11147
2986
D
PPARA
NR1H4
HNF4G
TBX3
MAF
ESRRG
BACH1
GATA6
NFIA
EHF
KLF5
FOXP1
CDX1
AC
Pouch EC2
Pouch EC1
PP
TI
E
PPARA
NR1H4
HNF4G
TBX3
MAF
ESRRG
BACH1
GATA6
NFIA
EHF
KLF5
FOXP1
CDX1
TI
PP
Pouch EC1
Pouch EC2
AC
EC1
Stem1
Stem2
EC2
−2 −1 0 1 2
Mean chromvar z score
−2 −1 0 1 2
ChromVAR z score

Figure 7

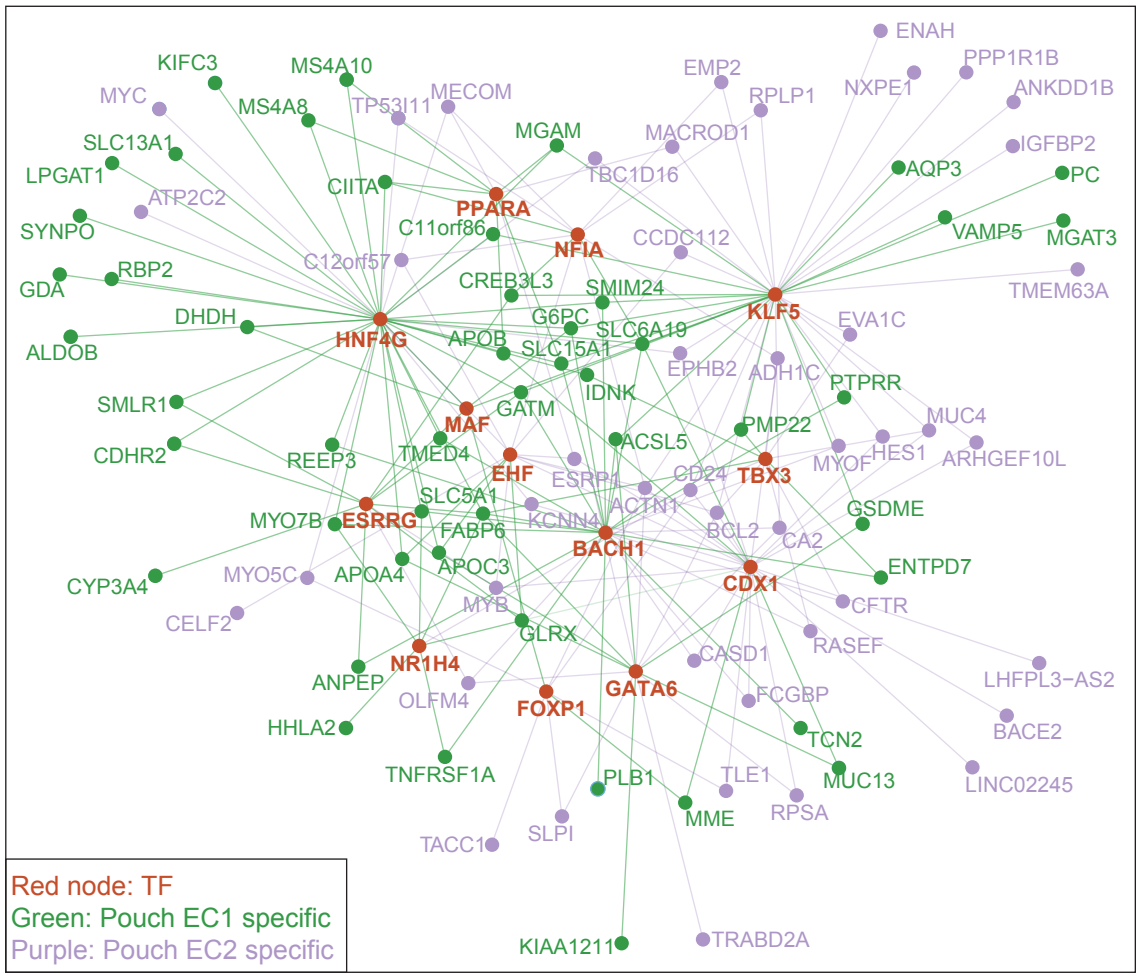
A

Peak accessibility z score
RNA expression z score
31808 peak to gene links
EC1−1
EC2−1
EC1−2
EC2−2
−2
2
−2
2

B


chr8:85413851−85563852
Normalized ATAC signal
AC
Pouch EC2
Pouch EC1
PP
TI
85440000
85480000
85520000
85560000
Peaks
Peak gene correlation
0.7
0.8
0.9
CA2
Gene Track

C


ENAH
PPP1R1B
KIFC3
MS4A10
EMP2
NXPE1
ANKDD1B
MYC
MS4A8
TP53I11
MECOM
RPLP1
MGAM
MACROD1
IGFBP2
SLC13A1
LPGAT1
TBC1D16
AQP3
PC
CIITA
ATP2C2
PPARA
C11orf86
SYNPO
NFIA
CCDC112
VAMP5
MGAT3
C12orf57
RBP2
GDA
CREB3L3
SMIM24
TMEM63A
DHDH
G6PC
KLF5
EVA1C
HNF4G
APOB
SLC6A19
ALDOB
SLC15A1
EPHB2
ADH1C
PTPRR
IDNK
SMLR1
GATM
MAF
MUC4
PMP22
CDHR2
ACSL5
REEP3
TMED4
MYOF
HES1
ARHGEF10L
EHF
ESRP1
CD24
TBX3
SLC5A1
ACTN1
GSDME
MYO7B
ESRRG
KCNN4
FABP6
BCL2
CA2
BACH1
ENTPD7
CYP3A4
MYO5C
APOA4
APOC3
CDX1
MYB
CFTR
CELF2
GLRX
RASEF
CASD1
NR1H4
LHFPL3-AS2
ANPEP
OLFM4
GATA6
FCGBP
FOXP1
HHLA2
TCN2
BACE2
TNFRSF1A
PLB1
TLE1
MUC13
LINC02245
RPSA
MME
SLPI
TACC1
Red node: TF
Green: Pouch EC1 specific
Purple: Pouch EC2 specific
KIAA1211
TRABD2A