

Assessing Cellular and Transcriptional Diversity of Ileal Mucosa Among Treatment-Naïve and Treated Crohn's Disease

Sushma Chowdary Maddipatla, MS,* Vasantha L. Kolachala, PhD,* Suresh Venkateswaran, PhD,* Anne F. Dodd, BS,* Ranjit Singh Pelia, MPH,* Duke Geem, MD, PhD,* Hong Yin, MD,[†] Yutong Sun, MS,[‡] Congmin Xu, PhD,[§] Angela Mo, PhD,[¶] Astrid Kusters, PhD,^{||} Junkai Yang, MS,^{||} Jason D. Matthews, PhD,* Eliver Ghosn, PhD,** Subra Kugathasan, MD,*^{††,‡‡,a}  and Peng Qiu, PhD^{§,a}

From the

^{*}Division of Pediatric Gastroenterology, Department of Pediatrics and Pediatric Research Institute, Children's Healthcare of Atlanta, Emory University School of Medicine, Atlanta, GA, USA;

[†]Department of Pathology, Children's Healthcare of Atlanta, Emory University, Atlanta, GA, USA

[‡]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA;

[§]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA;

[¶]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA;

^{||}Lowance Center for Human Immunology, Division of Immunology and Rheumatology, Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA;

^{**}Emory Vaccine Center, Lowance Center for Human Immunology, Departments of Medicine and Pediatrics, Emory University School of Medicine, Atlanta, GA, USA;

^{††}Genetics and Molecular Biology Program, Emory University School of Medicine, Atlanta, GA, USA; and

^{‡‡}Department of Human Genetics, Emory University, Atlanta, GA, USA.

^aEqual contribution for last authorship.

Address correspondence to: Peng Qiu, PhD, Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, 950 Atlantic Drive, Atlanta, GA 30332, USA (peng.qiu@bme.gatech.edu) or Subra Kugathasan, MD, Division of Pediatric Gastroenterology Emory University School of Medicine and Children's Healthcare of Atlanta, 1760 Haygood Drive, W-427, Atlanta, GA 30322, USA (skugath@emory.edu).

Background: Crohn's disease is a lifelong disease characterized by chronic inflammation of the gastrointestinal tract. Defining the cellular and transcriptional composition of the mucosa at different stages of disease progression is needed for personalized therapy in Crohn's.

Methods: Ileal biopsies were obtained from (1) control subjects (n = 6), (2) treatment-naïve patients (n = 7), and (3) established (n = 14) Crohn's patients along with remission (n = 3) and refractory (n = 11) treatment groups. The biopsies processed using 10x Genomics single cell 5' yielded 139 906 cells. Gene expression count matrices of all samples were analyzed by reciprocal principal component integration, followed by clustering analysis. Manual annotations of the clusters were performed using canonical gene markers. Cell type proportions, differential expression analysis, and gene ontology enrichment were carried out for each cell type.

Results: We identified 3 cellular compartments with 9 epithelial, 1 stromal, and 5 immune cell subtypes. We observed differences in the cellular composition between control, treatment-naïve, and established groups, with the significant changes in the epithelial subtypes of the treatment-naïve patients, including microfold, tuft, goblet, enterocyte, s and BEST4+ cells. Surprisingly, fewer changes in the composition of the immune compartment were observed; however, gene expression in the epithelial and immune compartment was different between Crohn's phenotypes, indicating changes in cellular activity.

Conclusions: Our study identified cellular and transcriptional signatures associated with treatment-naïve Crohn's disease that collectively point to dysfunction of the intestinal barrier with an increase in inflammatory cellular activity. Our analysis also highlights the heterogeneity among patients within the same disease phenotype, shining a new light on personalized treatment responses and strategies.

Key Words: single cell RNA-seq, Crohn's disease, inflammatory bowel disease, ileum, enterocytes

Introduction

Crohn's disease (CD) is characterized by chronic inflammation of the gastrointestinal tract,¹ particularly the ileal or ileocecal locations,² but is distinct from ulcerative colitis (UC), which shows clinical overlap with colonic CD.³ The

incidence of CD in North America is roughly 20 cases per 100 000,⁴ but reaching upward of 200 per 100 000 in some regions of the world. The exact cause of CD is unknown but is thought to be the result of a complex and multifactorial etiology with interplay between lifestyle, genetic predisposition,

Received for publication: August 11, 2022. Editorial Decision: August 15, 2022

© 2022 Crohn's & Colitis Foundation. Published by Oxford University Press on behalf of Crohn's & Colitis Foundation.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Key Messages

- We report a single cell analysis of ileal mucosa from 21 (7 treatment naïve, 14 established) subjects with Crohn's disease and 6 controls, resulting in a total of 139906 cells .
- Compared with established Crohn's disease, the epithelial compartment of the treatment naïve subjects undergoes distinct changes in enterocyte, goblet, BEST4+, microfold and tuft cell composition and behavior.
- Expression of high levels of MHC class II genes by epithelial cells was seen in the ileum of in Crohn's disease regardless of the disease duration.
- Elevated apolipoprotein and goblet cell trefoil factor activity was observed in remission but not in refractory amongst established Crohn's disease subjects.

epigenetics, epithelial barrier disruption, and unbridled immune response to intestinal triggers.⁵⁻⁷ The disease onset can be at any age, but most of the cases are diagnosed in the pediatric population with the peak between 10 and 25 years of age.⁸ Although the clinical course of CD is characterized by relapses and remission, once the disease begins, it typically is a lifelong illness. Clearly, CD has a genetic component, as several genome-wide association studies (GWASs) and follow-up studies demonstrated risk-associated loci,^{4,8} yet most of the 200 or so inflammatory bowel disease (IBD) risk loci fall within noncoding regions, suggesting that more epigenetic and molecular characterization needs to be performed. Because the broadly used therapies like corticosteroids and immunomodulatory drugs are ineffective in controlling the disease, recently developed biologicals including anti-tumor necrosis factor (TNF), anti-adhesion molecules, and anti-interleukin (IL)-12/23 have become frontline targeted therapies for CD.⁹ Although the recent cumulative data suggest that these biologic therapies are better than conventional therapies, they are only effective in about 30% to 50% of the patients. Unfortunately, biologic-treated patients, particularly those on anti-TNF therapy, have an increased risk for opportunistic infections and malignancies such as lymphomas and skin carcinomas.¹⁰ Thus, new and improved targeted therapies in CD are an unmet need, and expansion of available treatment options requires a shift toward a precision medicine approach. Furthermore, primary and secondary nonresponse (refractory) to biologics is increasingly common, but understanding the mechanism for nonresponse requires more understanding of the cellular and transcriptional variability across different CD subjects. There has been limited success from clinical and biological phenotyping in identifying novel biomarkers as companion diagnostics to guide precision CD treatment.

To date, several bulk biopsy transcriptomic studies have been performed for IBD; however, those studies were limited to using whole tissue for analysis of phenotypic associations and pathways. Therefore, recent studies have aimed at enabling precision medicine in IBD by using new technologies like single-cell RNA sequencing (scRNA-seq) and mass cytometry (CyTOF) to predict therapy response and disease activity.⁷ One recent study aimed to identify pathogenic cellular modules that were weakly, although significantly, associated with resistance to anti-TNF therapy.⁷ The presence

of these immune/macrophage cell modules suggests a spatial and functional relationship between underlying cell types. Because those studies used surgically resected tissue that is the end stage of a chronic inflammatory process, it is vital that colonoscopically retrieved mucosal biopsies are used to study the different locations, stages, and course of CD. Application of a high multidimensional scRNA-seq approach provides unprecedented resolution of cell lineages to characterize the transcriptomic heterogeneity of cell types and their functional states in human tissues. A deeper understanding in cellular heterogeneity and specificity at the single-cell level can be achieved if the mucosa is studied at diagnosis when patients are treatment-naïve and may provide additional insights for ileum-specific inflammation during early CD when compared with established CD patients.

Thus, in this study, we have used scRNA-seq of colonoscopically retrieved terminal ileal mucosal biopsies from 27 subjects across 3 different phenotypes, including non-IBD control subjects, treatment-naïve CD (TN_CD) patients, and established CD (Es_CD) patients. Es_CD patients allowed us to study the mucosa in those who had healing (remission) and in those that did not (refractory). Our analysis revealed distinct cellular and transcriptional complexities within the ileal mucosa of TN_CD patients, allowing identification of both cell types and cell type-specific genes that exhibited changes across different patient phenotypes, and further providing a glimpse into mucosal differences at the cellular and transcriptional level between 2 treatment response groups, remission, and refractory in Es_CD patients, all of which can be used to formulate new hypotheses toward personalized treatment strategies for different stages of CD.

Methods**Study Population**

Terminal ileal biopsy samples were collected from 27 patients who were undergoing clinically indicated colonoscopy at the Division of Pediatric Gastroenterology at Children's Healthcare of Atlanta and Emory University. Written informed consent was obtained before sample collection. The study group includes control subjects (n = 6), TN_CD patients (patients that had not yet received treatment at the time of diagnosis; n = 7), and Es_CD patients (patients that were diagnosed earlier and received treatment for at least 1 year; n = 14). Endoscopic (Simple Endoscopic Score for Crohn Disease in CD) and histological activity (presence or absence of histological inflammation) were documented for all the subjects. Depending on a patients' response to the treatment, the Es_CD group was divided into remission (who responded to treatment with evidence of mucosal healing; n = 3) and refractory (who did not respond and had active inflammation with colonoscopy and histology; n = 11) groups ([Supplementary Table 1](#)).

Tissue Processing

Once the biopsy was collected from the patient through endoscopy, it was immediately transported from the endoscopy suite to the lab in complete RPMI media with 10% fetal bovine serum and fully processed within 2 hours of collection. The biopsies were first minced with micro-scissors on ice in Hank's Balanced Salt Solution buffer containing 2 mM EDTA without Ca²⁺/Mg²⁺. The minced pieces were vortexed gently

and allowed to settle for 2 minutes on ice. The supernatant containing some of the loosely associated immune cells and other single cells released by the mincing process (primarily lamina propria) were removed and the sedimented minced tissue pieces were digested on ice with *Bacillus licheniformis* protease for 30 minutes. Samples were periodically vortexed gently and placed back on ice during enzymatic digestion. For each patient sample, the lamina propria fraction was mixed with the enzyme digested epithelial fraction 1:1 for noninflamed mucosa and 1:4 for inflamed mucosa (minimizing overloading with immune cells). The ability to only capture a few thousand cells per patient by the scRNA-seq required us to decrease the immune signature in inflamed samples by mixing higher ratios of the fractioned epithelial compartment to the final cell dilution with the immune compartment before single cell encapsulation. The cells were then passed through a 40- and 20- μ m filter, washed, and finally resuspended at the appropriate cell density for loading onto the 10x Chromium Controller (~1000-1200 cells/ μ L; 10x Genomics, Pleasanton, CA, USA).

10x Library Preparation and Sequencing

Tissue-derived single cells were loaded onto the 10x Chromium Controller targeting 10 000 cells. Single cell capture, barcoding, gel bead in emulsion room temperature, cleanup, complementary DNA amplification, and library construction were performed according to the manufacturers' instructions using Chromium Single Cell A Chip Kit (cat no: PN-1000152; 10x Genomics) and Chromium 5' Library and Gel Bead Kit (v1 chemistry; cat no: PN-1000006; 10x Genomics). Final library pools were sequenced in 4 batches on Illumina's NovaSeq 6000 (Illumina, San Diego, CA, USA) instrument at Georgia Tech Genomics Core or Yerkes Genomic Core with a targeted sequencing depth of 50 000 reads/cell.

10x Genomics Computational Analysis and Quantification

Cell Ranger v6.0 (10x Genomics) was used to generate gene expression count matrices for every sample, after sequence alignment to Grch38 human reference genome using STAR¹¹ to obtain Unique Molecular Identifier counts for each gene per cell. Based on Unique Molecular Identifier counts and cell barcodes, numbers of genes and cells for each sample were estimated. Matrices for each sample were analyzed by Seurat v4.1 (<https://satijalab.org/seurat/>). Because the intestine has high cell turnover, it is important to consider all the cells, including the cells that are under stress. Therefore, the quality control threshold for mitochondrial percentage was not set for the analysis, as intestinal epithelium undergoes constant regeneration and apoptosis. In addition, we observed a small population of epithelial cells showing very low expression of immune markers and vice versa, but it did not affect the overall cluster annotations.

Clustering Analysis

The Seurat v4.1 reciprocal principal component analysis integration¹² workflow was applied to cluster cells from all the samples. The workflow started with log normalization of the data and highly variable feature selection independently for each sample. Features repeatedly selected as variable genes across the samples formed the integrated feature list for scaling and principal component analysis (PCA). This

integration workflow corrected for batch effects in the data. PCA was performed on the integrated dataset to reduce the data into 30 dimensions (ie, top 30 principal components). The PCA space was used to generate cell clusters and generate Uniform Manifold Approximation and Projection (UMAP) to visualize the cell clusters.

Marker Genes and Annotations of Clusters

After cell clustering, differential gene expression (DGE) analysis was performed between each cell cluster against all other cells to obtain marker genes for interpreting the cell clusters. The differential expression analysis was performed on the uncorrected data before integration, but after library size normalization with scaling factor of 10 000 and log transformation. The Wilcoxon rank sum test was used to identify and rank differentially expressed genes, using the Seurat FindMarker function. Cell clusters were annotated by comparing the marker genes derived from our dataset with publicly available databases and the marker genes published by Martin et al⁷ and Wang et al.¹³ The marker genes were manually examined by color-coding the UMAP visualization, allowing for confirmation of appropriate expression patterns across cell clusters and that they agreed with the cell type annotations.

Differential Gene Expression Across CD Phenotypes

Differential gene expression analysis was performed on every cell type from each pair of CD phenotypes (TN_CD vs control groups; TN_CD vs Es_CD groups; Es_CD vs control groups), using the Wilcoxon rank sum test implemented by the Seurat FindMarker function (min.pct = 0.1, log.threshold = 0.25). We used the same settings to identify differentially expressed genes between remission and refractory in Es_CD patients. These analyses were performed for each cell cluster separately. Seurat's BuildClusterTree function was used to construct phylogenetic trees that represent distances between the average cells for each epithelial (**Supplementary Figure 1A**) and immune (**Supplementary Figure 1B**) subtype.

Ontology Enrichment Analysis

Gene Ontology (GO) enrichment analysis was performed using the cell type-specific differentially expressed genes between the CD phenotypes with P value <.05, using the R package cluster profiler.¹⁴ The results were viewed using the cluster profiler dotplot function and cowplot package.

Results

TN_CD Mucosa Shows Disrupted Epithelial Subtype Proportions While B Cells Are Decreased in Es_CD

First, we sought to define the differences in cellular composition of the terminal ileum in TN_CD and Es_CD patients in comparison with non-IBD control subjects, and to better understand the association of different cell types with disease phenotypes. To this end, we generated 139 906 single-cell transcriptomes from control and CD ileal mucosal biopsies. The scRNA-seq data went through quality control (**Supplementary Figure 2A-C**) and integration pipelines (**Supplementary Figure 2D, E**) to account for batch effects. The integrated data were clustered and visualized by UMAP

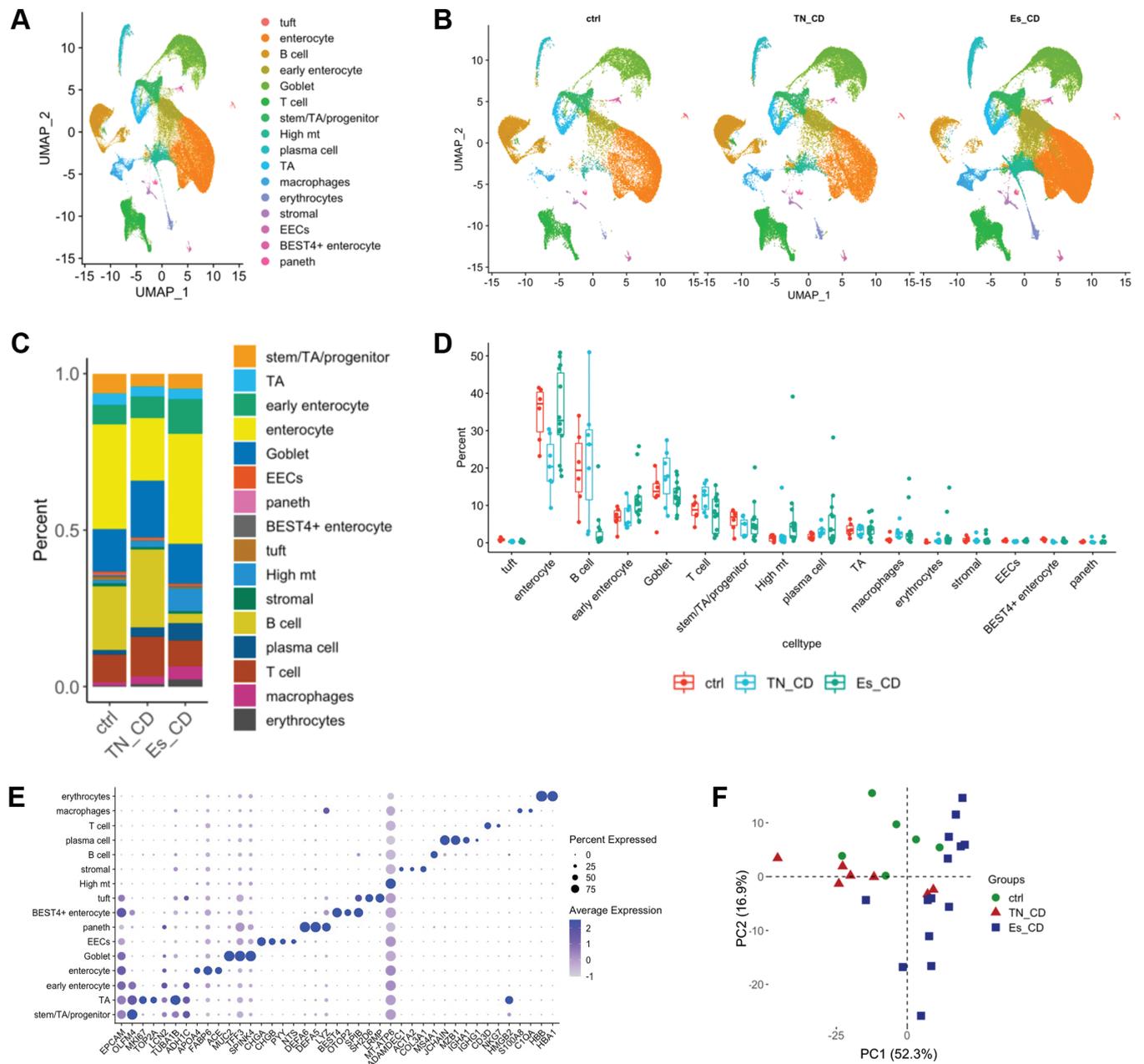


Figure 1. Human mucosal terminal ileal tissue heterogeneity. A, Uniform Manifold Approximation and Projection (UMAP) representation colored by cell types. B, Split UMAP based on Crohn's disease (CD) phenotypes of control (ctrl) subjects, treatment-naïve CD (TN_CD) patients, and established CD (Es_CD) patients. C, Bar plot showing the cell type proportion in each phenotype group. D, Box plot showing the cell type proportion differences among phenotype groups (green: control subjects; red: TN_CD patients; violet: Es_CD patients). E, Dot plot representing marker genes for each cell type. F, Principal component (PC) analysis of cell type proportions, showing that the 3 phenotype groups can be separated based on cell type proportions. EEC, enteroendocrine cell; TA, transit amplifying.

(Figure 1A). When the UMAP visualization was stratified by patient phenotype groups (Figure 1B), a high degree of overlap between control subjects and CD patients can be observed. We performed DGE between the cell clusters to identify marker genes that can be used to annotate the cell clusters with known cell types, based on known cell type signature genes from published literature.^{7,13} We were able to manually annotate 9 epithelial cell types, 5 immune cell types, and 1 stromal cell type (Supplementary Figure 3A). Epithelial cells were subdivided into 9 cell types with the shared expression of *EPCAM*, *KRT8*, and *KRT18* expression in both control subjects and CD patients. The genes

used to define each subcluster of epithelial cells were stem/transit-amplifying (TA)/progenitor cells (*OLFM4*, *LGR5*), TA cells (*OLFM4*, *MKI67*, and *TOP2A*), early enterocyte cells (*OLFM4*, *LCN2*, and *ADH1C*), enterocyte cells (*APOA4* and *FABP6*), Goblet cells (*MUC2*, *TFF3*, and *SPINK4*), enteroendocrine cells (EECs) (*CHGA*, *CHGB*, *PYY*, and *NTS*), Paneth cells (*DEFA5*, *DEFA6*, and *LYZ*), BEST4+ enterocyte cells (*BEST4*, *OTOP2*, and *SPIB*),¹⁵ and tuft cells (*SPIB*, *SH2D6*, and *LRMP*). For the immune compartment defined by *PTPRC* expression, 5 subtypes were defined by the following gene expression patterns: B cells (*MS4A1*), T cells (*CD3D* and *NKG7*), plasma cells

(*JCHAIN*, *MZB1*, *IGHA1*, and *IGHG1*), macrophages (*S100A8*, *C1QA*, and *C1QB*), and erythrocytes (*HBB* and *HBA1*). The stromal cells were defined by *ADAMDEC1*, *ACTA2*, and *COL3A1* expression. In this global analysis, at the level of cell type, with the exception of enterocytes that are significantly reduced in TN_CD patients ($P < .05$), and B cells that are significantly reduced in Es_CD patients compared with control subjects,¹⁶ most of the detected cell types were found in similar proportions between the 3 patient phenotype groups ($P < .05$) (Figure 1C and D). Less prominent cells such as BEST4+ enterocytes and tuft cells were also detected, demonstrating the robustness of our annotation strategy (Figure 1A and E). Principal components extracted on cell type proportions demonstrated unique differences in cellular composition among the patient phenotype groups, while also showing patient-to-patient variabilities within the same phenotypic group (Figure 1F). These data pointed toward changes in cellular subtypes of TN_CD and Es_CD patients that might have diagnostic and alternative treatments potential.

Increased Secretory and Reduced Absorptive Epithelial Cells in TN_CD

To gain more insight into the cellular differences of Ileal mucosa that appeared to be taking place in the epithelium ($n = 88\ 913$) (defined in Figure 1), deeper analysis was performed on the epithelial cells without the immune and stromal cells, thus increasing epithelial cell cluster granularity. A separate analysis without interference of immune and stromal cell genes allowed the pipeline to more clearly identify highly variable genes and principal components driven by the variability among epithelial cell types. In doing so, we identified low-abundant M cells (Figure 2A) and defined each cluster based on its unique gene expression patterns (Figure 2B). The 13 epithelial subtypes (Supplementary Figure 3B) included differentiated absorptive enterocytes, secretory goblet, EECs, pH-sensing absorptive BEST4+ enterocytes, chemosensory tuft cells, defensive secretory Paneth cells, low-abundant M cells, early enterocytes, proliferating cells, and 4 clusters of undifferentiated cells enriched for early enterocytes, along with TA, or stem/TA signatures (Figure 2A and B). The early enterocyte cluster was likely grouped based on cell cycle genes and precursor lineage genes for absorptive and secretory epithelial cell subtypes. The low-abundant M cells were missed in the first round of clustering analysis when immune and stromal cells were included.

Remarkably, the abundance of the M cells, along with several other epithelial subtypes including tuft, goblet, TA, and BEST4+ cells, appears to be altered in the mucosa of the TN_CD group compared with non-IBD control group ($P = .035$) (Figure 2C). Compared with the results from (Figure 1) when all the cell types were considered, here in the epithelial-focused clustering analysis (Figure 2), the enterocytes in TN_CD patients were significantly reduced in proportions compared with control subjects ($P = .0047$) and Es_CD patients ($P = .016$). Both TN_CD and Es_CD patients had significantly decreased BEST4+ enterocyte ($P = .0047$ and $.0087$) and tuft ($P = .035$ and $.0066$) cells compared with control subjects. TN_CD patients also had goblet cells in a higher abundance when ($P = .008$ and $.00012$) compared with Es_CD patients and control subjects (Figure 2C). PCA on epithelial cellular composition across

patient phenotype groups (Supplementary Figure 4) showed that cell type changes were more pronounced in TN_CD patients compared with Es_CD patients and control subjects (Figure 2D). These data highlight differences in the epithelial cell subtype proportions related to disease and treatment and point to cellular microenvironments of the epithelium that are altered during the early stages of untreated CD.

Decreased B Cells and Heterogeneity in Inflammatory Macrophages in Es_CD

To gain insight into the cellular diversity and behavior of the immune compartment, we subclustered and analyzed the immune cells using a separate analysis without the epithelial and stromal cells ($n = 43\ 648$) (defined in Figure 1A). In this analysis, we identified 9 immune cell subtypes (Figure 3A). The identified cell subtypes included B cells, plasma cells, T cells, erythrocytes, and a cluster showing mixed marker genes that we call unmatched cells (Supplementary Figure 3C). We were able to identify CD8+ T-cells and natural killer cells and distinguish inflammatory macrophages from tissue-resident macrophages (Figure 3B). Es_CD patients had significantly lower B cell proportions compared with TN_CD patients and control subjects ($P = .0097$ and $.0064$) (Figure 3C). Es_CD patients interestingly showed more unmatched cells when compared with TN_CD patients ($P = .031$) and control subjects ($P = .00021$) (Figure 3C). Inflammatory macrophages were found to be more abundant in Es_CD and TN_CD patient mucosa when compared with control subjects ($P = 5.2 \times 10^{-5}$ and $.022$), but the increase in this type of macrophages were primarily contributed by a few patients, highlighting the patient-to-patient heterogeneity for some cell type proportions (Supplementary Figure 5). PCA analysis showed the Es_CD group with strong differences in cell type composition compared with the TN_CD and control groups (Figure 3D), which was likely caused by the decrease in their B cell populations. Taken together, we found a remarkable similarity in immune cell-type abundance, suggesting that the disease may be more related to the behavior of these cells than their proportions.

Cell Type-Specific Expression Changes in Epithelial Cell Subtypes

To gain a closer insight into the behavior of the different mucosal cell types, we performed their cluster-specific DGE analysis. Here, we aim to identify cell type-specific gene expression changes between the CD phenotype groups compared with control subjects in each epithelial cell type (Supplementary Table 2). This differential expression analysis found 18 genes highlighted in Figure 4A. Interestingly, both untreated (TN_CD) and treated (Es_CD) patients expressed more major histocompatibility complex (MHC) class II protein complex binding genes, such as HLA-DP and HLA-DR, in most of the epithelial subtypes except EECs, Paneth cells, and tuft cells when compared with control subjects, indicating a change in epithelial behavior as it begins to display more foreign antigens on its surface.¹⁷ Likewise, both CD groups also showed MHC class I protein binding gene *HLA-B* expressed in almost all epithelial subtypes except EECs and Paneth cells when compared with control subjects. Additionally, *CD74*, a receptor linked to MHC class II membrane transport,¹⁸ inflammation, and intestinal mucosal healing,¹⁹ had increased *CD74* transcript

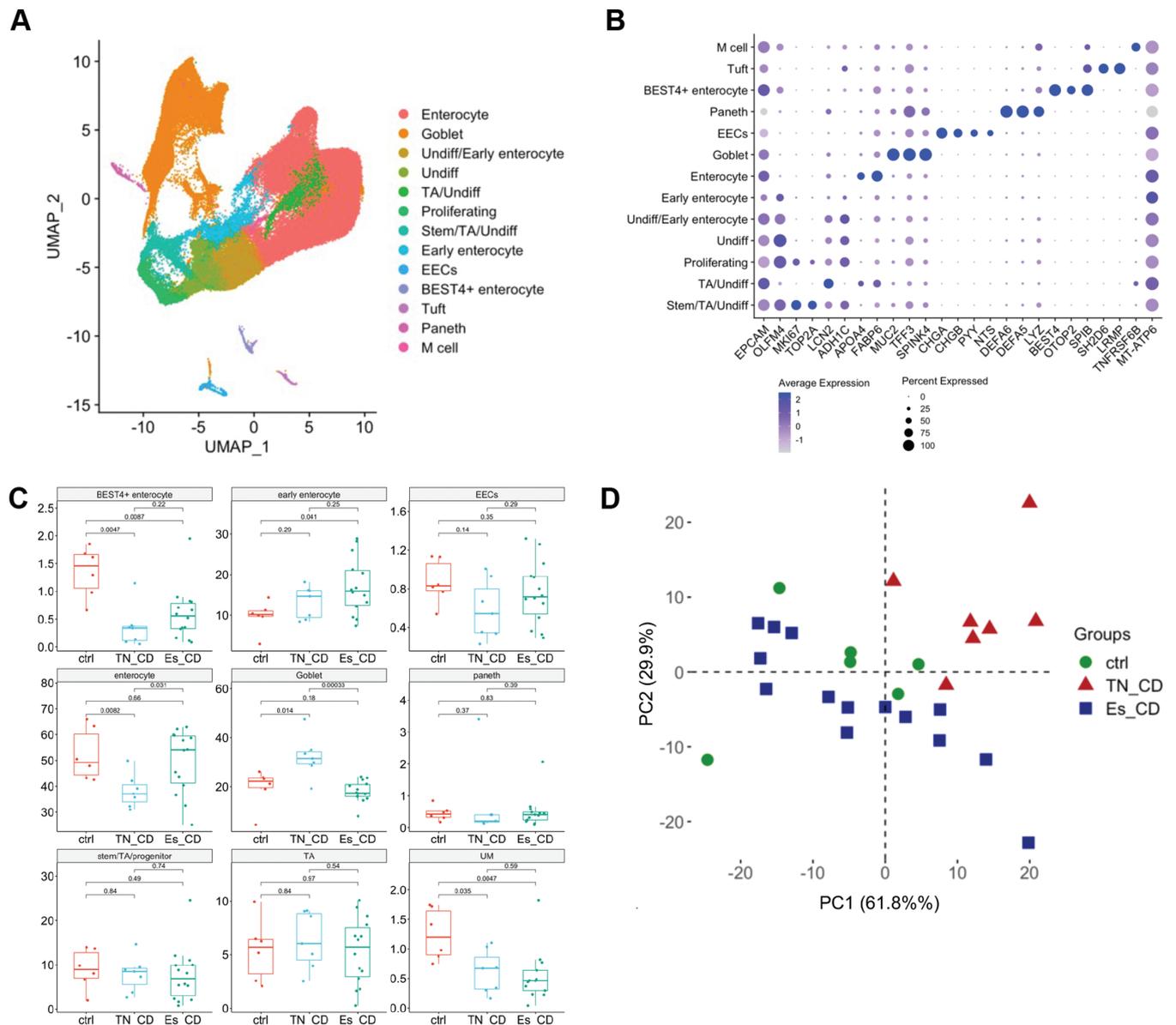


Figure 2. Epithelial subclustering. A, Uniform Manifold Approximation and Projection (UMAP) representation of epithelial cells, colored by 13 cell types. B, Dot plot representing marker genes for each epithelial cell type. C, Box plot showing the cell type proportion differences among the phenotype groups. D, Principal component (PC) analysis of cell type proportions, showing the variations of epithelial cell type compositions among the phenotype groups. ctrl, control; EEC, enteroendocrine cell; Es_CD, established Crohn's disease; TA, transit amplifying; TN_CD, treatment-naïve Crohn's disease.

levels in most epithelial subtypes from Es_CD and TN_CD patients (Figure 4A). Furthermore, the M cells showed expression of the GWAS-implicated IBD risk gene *CCL20* along with TA/undifferentiated cells especially in TN_CD.²⁰ When we analyzed the differences in cellular composition between remission and refractory mucosa, surprisingly, the epithelial cell-type compositions between these groups appeared to be similar (Figure 4B). However, DGE analysis for each epithelial cell type between remission and refractory patients within the established CD group showed significant transcriptional differences (Figure 4C). It indicates that indeed in some instances it is not the composition of the cell types per se but more likely their activity, ie, gene transcription, that is contributing to the disease and progression. In fact, enterocytes and TA/undifferentiated cell types from patients in remission have increased levels of *LCT*, *APOA1*, *APOA4*, and *RBP2*

transcripts when compared with refractory cells. Goblet cells from the mucosa of patients in remission showed more expression of *TFF1* when compared with refractory cells. When the differentially expressed genes between CD patients and control subjects were used to perform GO enrichment to identify biological processes involved in each subtype (Figure 4D), the predominant biological activity highlighted by this analysis indicated changes in antigen presentation, reactive oxygen production, and other immune functions, processes that point to decreases in epithelial barrier function and an increase in exposure of the lamina propria to luminal contents.

Cell Type-Specific Expression Changes in Immune Cell Subtypes

The behavior of the immune compartment was also studied by DGE analysis for every immune cell subtype of each

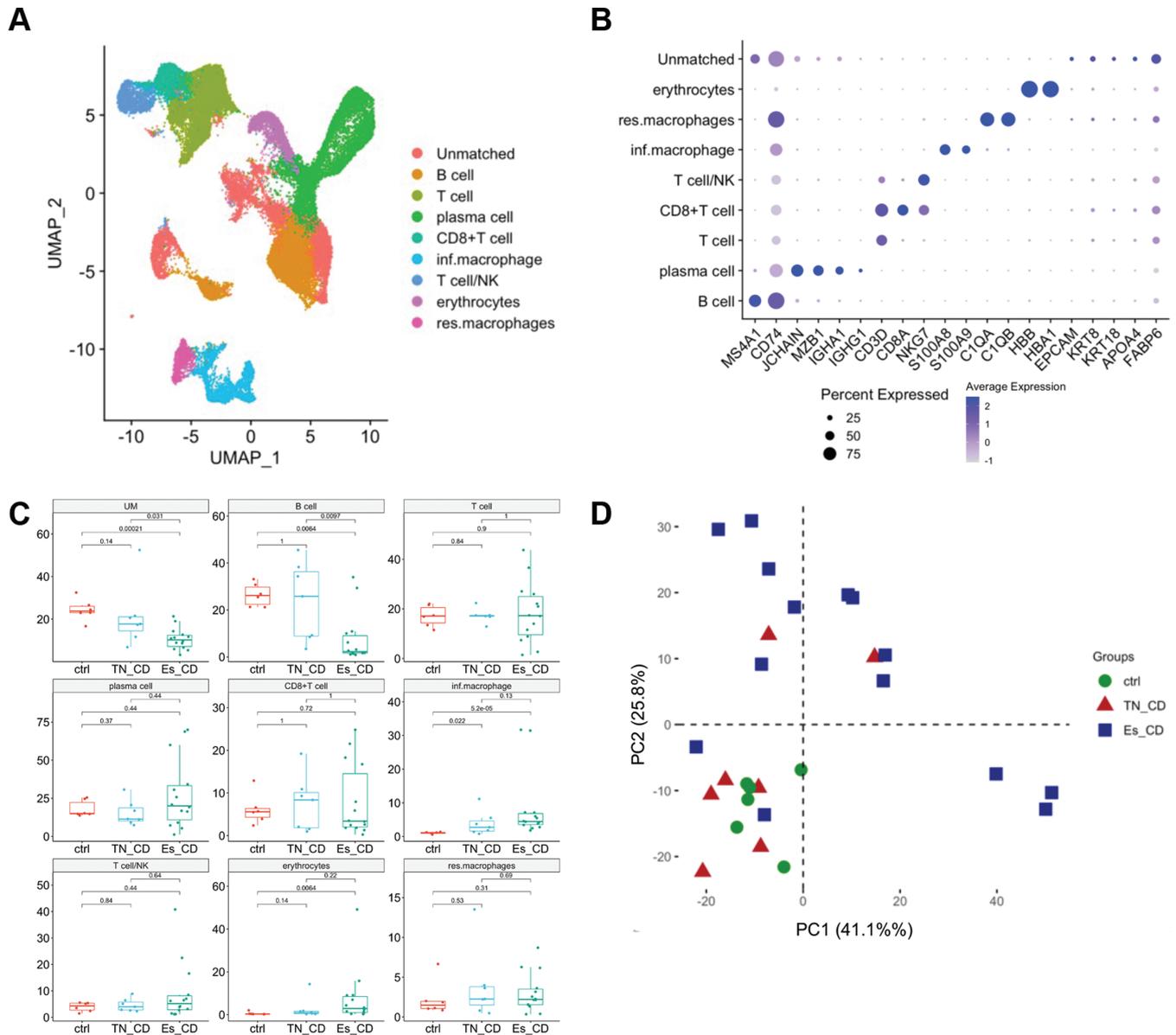


Figure 3. Immune subclustering. A, Uniform Manifold Approximation and Projection (UMAP) representation of immune cells, colored by 11 immune subclusters. B, Dot plot representing marker genes for each immune cell type. C, Box plot showing the cell type proportion differences among the phenotype groups. D, Principal component (PC) analysis of cell type proportions. ctrl, control; EEC, enteroendocrine cell; Es_CD, established Crohn's disease; NK, natural killer; TA, transit amplifying; TN_CD, treatment-naïve Crohn's disease.

phenotype pair (Supplementary Table 3). Here, B cells also showed changes in MHC class II expression like that of inflammatory macrophages and plasma cells, while B and T cells showed higher expression of *BTG1* in TN_CD patients compared with non-IBD control subjects and Es_CD patients (Figure 5A). In the erythrocytes, *ALAS2* expression was higher in TN_CD and Es_CD patients, while inflammatory macrophages showed higher expression of *HLA-DR* genes in control subjects and TN_CD patients when compared with Es_CD patients. However, in a few patients, inflammatory macrophages in Es_CD and TN_CD patients showed higher expression of inflammatory markers *S100A8*, *IL1B*, *SOD2*, and *SAT1* that was not detected in resident macrophages. Differential expression and GO analysis was performed separately on remission and refractory CD treatment groups, and found no significant proportional differences between

immune subtypes in these 2 groups (Figure 5B). However, we did find transcriptional differences (Figure 5C), further exemplifying the “behavior over abundance” hypothesis our data have helped to develop. For example, *HLA-DRB5* expression was increased in remission of inflammatory macrophages and refractory of resident macrophages, whereas *CCL5* expression was higher in T/natural killer cells in the remission group. GO analysis of the genes differentially expressed between CD patients and control subjects showed a strong metabolic signature along with changes in MHC class II activity, along with numerous other processes related to immune function (Figure 5D). Taken together, a clear change in the activity of the immune compartment in the ileal mucosa is taking place during CD despite observing only minor changes in the relative abundance of different immune subtypes.

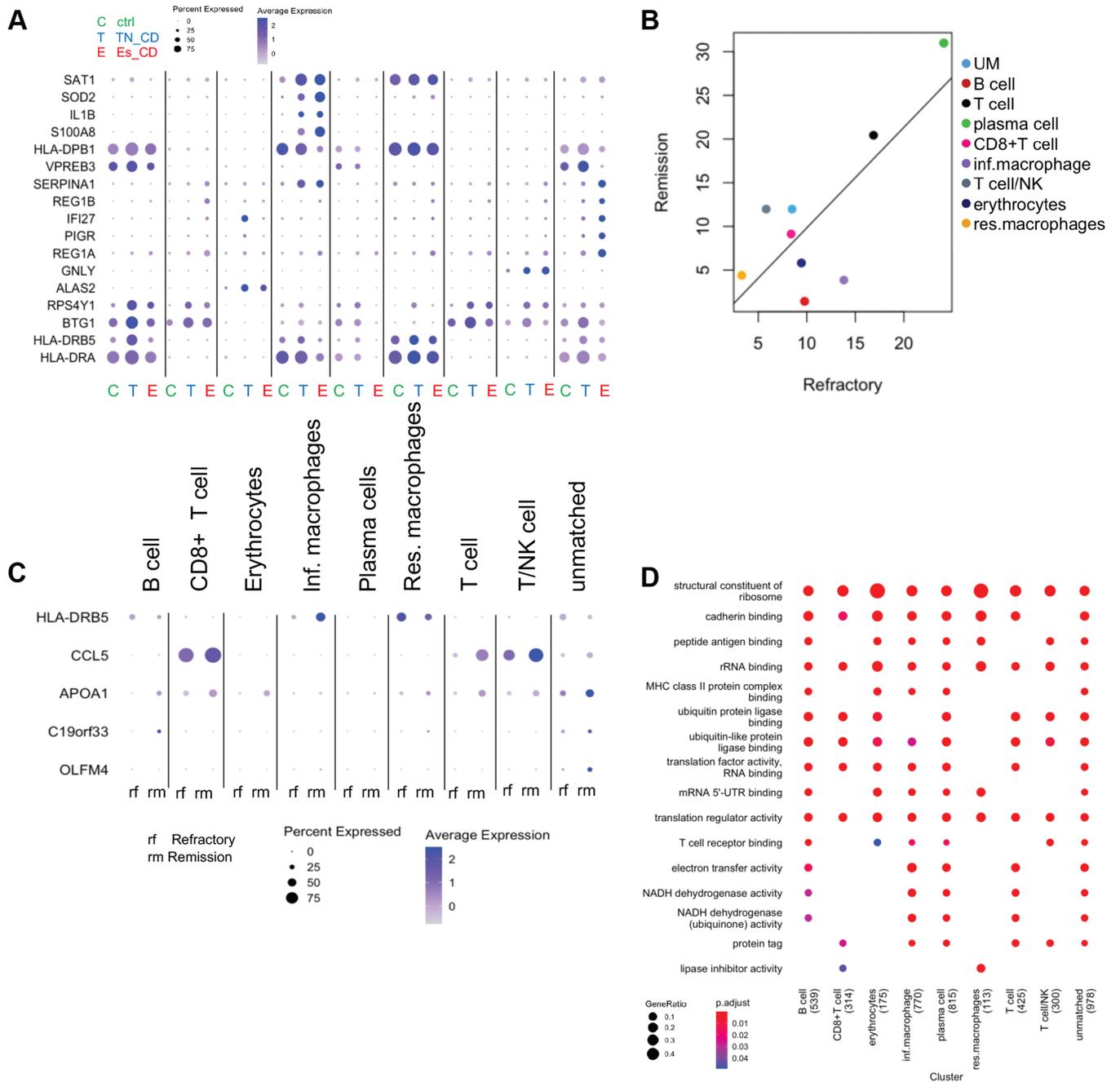


Figure 5. Immune cell type-specific transcriptional changes in Crohn's phenotypes. A, Dot plot showing differentially expressed genes in immune cell types between the phenotype groups. B, Scatter plot showing the immune cell type proportions between remission and refractory are correlated. C, Dot plot showing the differentially expressed genes in immune cell types between remission and refractory. D, Dot plot showing GO enrichment of biological processes in immune cell types. ctrl, control; EEC, enteroendocrine cell; Es_CD, established Crohn's disease; inf., inflammatory; MHC, major histocompatibility complex; mRNA, messenger RNA; NK, natural killer; res., resident; rRNA, ribosomal RNA; TA, transit amplifying; TN_CD, treatment-naïve Crohn's disease.

during CD and exposure of the underlying lamina propria to luminal contents that promote inflammatory activity in the surrounding mucosa.

We have identified both unique and patient-specific signals mapping to the epithelium of TN_CD. The alterations we found in epithelial abundance in TN_CD determined by scRNA-seq is supported by years of clinical and histological findings pointing toward degradation of the epithelium coincident with leaky barrier function and is a hallmark feature of newly diagnosed, particularly inflamed CD.²² Loss of the

epithelial barrier leads to exposure of the lamina propria to an influx of luminal contents that initiate immune responses that we also detected in the form of MHC class II expression, a process that can take place in both immune and nonimmune cells.^{23,24} An earlier study also showed that intestinal M cells with immune surveillance activity were induced during inflammation²⁵ in colonic mucosa, supporting our observations in the ileum. Although we are not the first to show the importance of BEST4+ enterocytes^{13,15} and tuft cells²⁶ to intestinal homeostasis and disease or changes in enterocytes and

goblet cell levels in CD,²⁷ we are for the first time, with the largest number of cells mapped from the terminal ileum in such a phenotypically diverse cohort to date, characterizing the global level of disruption in these cell subtypes and their activity within the TN_CD mucosa.

Although demonstrated by Suzuki et al,²⁸ we did not find modifications in the stem cell compartment of the epithelium related to disease activities. However, we did observe significant increases in TA/undifferentiated cells along with the *CDKN2A* senescence marker (data not shown) in the CD group compared with non-IBD control subjects, which might be related to the stem and TA/undifferentiated cell activity shown by the studies of Wang et al,²⁹ whereby senescence of TA cells were no longer able to proliferate into new enterocytes. In line with those findings, disruption of the proliferative capacity of the TA cells would likely result in loss of differentiated enterocyte abundance and contribute to epithelial erosion, villi disruption and decreased nutrition absorption.

The effects of therapy can be observed in our scRNA-seq data from the established cohort. Supported by results from Haberman et al, who showed the association of *APOA4* with better IBD outcome^{30,31} in pediatric CD patients, our analysis showed that the remission group had increased levels of *LCT*, *APOA1*, *APOA4*, and *RBP2* expression in enterocytes and TA/undifferentiated cell types when compared with treatment-refractory patients. Other links to lipid metabolism, including those to *APOA1* and IBD, have been established, with some therapeutic potential,³¹ suggesting that disruption of the epithelial subtypes during CD impacts the absorptive and metabolic processes these cells normally perform, and may be linked to some of the nutritional differences and peripheral effects observed during CD. Moreover, alterations in goblet cell behavior were also observed between remission and refractory patients, a finding supported by Hensel et al³² that demonstrated increased mRNA expression levels of *TFF1*, *MUC1*, and *HGF* in pediatric IBD during clinical remission. Thus, focusing on lipid metabolism and goblet cell function may be an important target in helping patients who are refractory to treatment. It should also be noted that where TN_CD patients had decreased epithelial numbers, the established patients, many of whom were refractory to anti-TNF therapy, showed similar epithelial profiles to non-IBD control subjects, indicating at least some form of mucosal healing after treatment.

Except for a few CD patients having increased levels of inflammatory macrophages in our cohort, along with an overall decrease in B cell populations, no significant changes in the abundance of cells in the immune compartment were detected for CD patients as a whole. While changes in mucosal B cell abundance were not surprising,³³ we were surprised to not find more cellular compositional changes in the immune compartment during CD, especially for those in the TN_CD group. For example, the absence of macrophages having significantly higher expression of inflammatory markers such as *S100A8*,³⁴ *IL1B*,³⁵ *SOD2*,³⁶ and *SAT1* in most of our diseased patients suggested alternative mechanisms contributing to disease other than just changes in cell subtype abundance or an influx of inflammatory cells. This has led us to consider a new hypothesis that in some instances it is the cellular activity involving gene transcription, protein function, and cellular crosstalk that is playing a larger role in disease onset and progression than it is a change in cellular composition.

In support of this new hypothesis, our analysis showed higher levels of a well-studied IBD chemokine *CCL20* in M cells of TN_CD patients and found increases in *CCL5* production in T and natural killer cells. Recent work from our group using patient-derived intestinal organoids has begun to show the importance of the epithelium in CD and its role in producing soluble molecules that can impact the behavior of surrounding cell types, including those in the immune compartment.^{5,37} Studies of Kaser et al³⁸ and Smillie et al²⁰ showed that the *CCR6* receptor and its ligand *CCL20* were upregulated in both CD and UC in 114 patients and is a GWAS-implicated risk gene²⁰ for UC. Furthermore, our analysis showed that epithelial subclusters such as enterocytes, M cells, and TA/undifferentiated cells had significant increases in *HLA-DP*, *HLA-DR*, and *HLA-B* in CD patients with a fold change >2. This pattern of expression by the epithelium is directly involved in antigen presentation pathways and processing,³⁹ indicating that these cells have been stimulated by microbes and other luminal contents have been allowed to cross the damaged epithelial barrier and are displaying foreign antigens to prompt an immune response. Alternatively, changes in the microbiome of some of these patients may also precipitate alterations in MHC class II expression. Nonetheless, our single-cell data have enabled us to further substantiate the idea that it could be the crosstalk across cell subtypes and the modulation of cellular behavior that is more important to disease in some cases than changes in cellular abundance.

Conclusions

In summary, to better understand the differences in cell type and gene expression profiles between treatment-naïve patients and established patients with therapy, we used the revolutionary technology of scRNA-seq and subsequent bioinformatics to analyze 27 patients, generating single-cell transcriptomic data for ~130 000 cells that were annotated into epithelial,^{13,15,26} immune,⁷ and stromal⁴⁰ cell subtypes to reveal an overwhelming change in the mucosal epithelium of TN_CD along with numerous changes in immune and epithelial activity in distinct disease states. Our analysis showed cell type abundance and transcriptional heterogeneity that was unique to the individual and in some cases the disease that can help us give new direction in designing personalized therapies and possible explanations for treatment failure. Our study was not without limitations, however, in that we only captured a subset of the genes expressed by these cells, the spatial organization of these cells was lost upon tissue dissociation, and we did not look at protein levels or surface expression on the cells. Despite these limitations, we have shown an important role of epithelial dysfunction during the early stages of CD, and future studies using spatial transcriptomics on mucosal tissue sections along with CITE-seq that couples cells surface protein detection with sequencing will be used to overcome these limitations and further refine and granulate our analysis.

Supplementary Data

Supplementary data is available at *Inflammatory Bowel Diseases* online.

Acknowledgments

We are grateful to Murugadas Anbazhagan and Greg Gibson for their support and helpful comments. We acknowledge the Genomic Cores at Yerkes Non-Human Primate Research Center at Emory University (National Institutes of Health Grant Nos. P51 OD011132 and S10 OD026799) and the Parker H. Petit Institute for Bioengineering and Bioscience at the Georgia Institute of Technology for the single-cell library sequences. The authors confirm all relevant ethical guidelines were followed, and ethics committees/necessary Emory University Institutional Review Board approvals were obtained. The review board approved the protocol. All necessary patient/participant written consent has been obtained.

Funding

This work was supported by the Leona M. and Harry B. Helmsley Charitable Trust (G-2007-04028 and G-2204-05197) as well as the National Science Foundation (CCF2007029). P.Q. is an ISAC Marylou Ingram Scholar and a Wallace H. Coulter Distinguished Faculty Fellow. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability

The scRNA-seq data for the 27 patient's terminal ileal biopsies included in this study have been deposited in Gene Expression Omnibus and is accessible through Gene Expression Omnibus accession GSE202052.

References

- Kugathasan S, Denson LA, Walters TD, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet*. 2017;389:1710–1718.
- Atreya R, Siegmund B. Location is important: differentiation between ileal and colonic Crohn's disease. *Nat Rev Gastroenterol Hepatol*. 2021;18:544–558.
- Lang-Schwarz C, Angeloni M, Agaimy A, et al. Validation of the "Inflammatory Bowel Disease-Distribution, Chronicity, Activity [IBD-DCA] Score" for ulcerative colitis and Crohn's disease. *J Crohns Colitis*. 2021;15:1621–1630.
- Gajendran M, Loganathan P, Catinella AP, Hashash JG. A comprehensive review and update on Crohn's disease. *Dis Mon*. 2018;64:20–57.
- Niklinska-Schirtz BJ, Venkateswaran S, Anbazhagan M, et al. Ileal derived organoids from Crohn's disease patients show unique transcriptomic and secretomic signatures. *Cell Mol Gastroenterol Hepatol*. 2021;12:1267–1280.
- Pelia R, Venkateswaran S, Matthews JD, et al. Profiling non-coding RNA levels with clinical classifiers in pediatric Crohn's disease. *BMC Med Genomics*. 2021;14:194.
- Martin JC, Chang C, Boschetti G, et al. Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. *Cell*. 2019;178:1493–1508.e20.
- Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol*. 2015;12:205–217.
- Agrawal M, Spencer EA, Colombel JF, Ungaro RC. Approach to the management of recently diagnosed inflammatory bowel disease patients: a user's guide for adult and pediatric gastroenterologists. *Gastroenterology*. 2021;161:47–65.
- Axelrad JE, Lichtiger S, Yajnik V. Inflammatory bowel disease and cancer: the role of inflammation, immunosuppression, and cancer treatment. *World J Gastroenterol*. 2016;22:4794–4801.
- Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
- Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573–3587.e29.
- Wang Y, Song W, Wang J, et al. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J Exp Med*. 2020;217:e20191130.
- Yu G, Wang LG, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–287.
- Parikh K, Antanaviciute A, Fawcner-Corbett D, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature*. 2019;567:49–55.
- Roda G, Sartini A, Zambon E, et al. Intestinal epithelial cells in inflammatory bowel diseases. *World J Gastroenterol*. 2010;16:4264–4271.
- Wosen JE, Mukhopadhyay D, Macaubas C, Mellins ED. Epithelial MHC class II expression and its role in antigen presentation in the gastrointestinal and respiratory tracts. *Front Immunol*. 2018;9:2144.
- Schroder B. The multifaceted roles of the invariant chain CD74—More than just a chaperone. *Biochim Biophys Acta*. 2016;1863:1269–1281.
- Farr L, Ghosh S, Jiang N, et al. CD74 signaling links inflammation to intestinal epithelial cell regeneration and promotes mucosal healing. *Cell Mol Gastroenterol Hepatol*. 2020;10:101–112.
- 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.
- Carraro G, Langerman J, Sabri S, et al. Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell states and composition. *Nat Med*. 2021;27:806–814.
- Soderholm JD, Streutker C, Yang PC, et al. Increased epithelial uptake of protein antigens in the ileum of Crohn's disease mediated by tumour necrosis factor alpha. *Gut*. 2004;53:1817–1824.
- Koyama M, Mukhopadhyay P, Schuster IS, et al. MHC class II antigen presentation by the intestinal epithelium initiates graft-versus-host disease and is influenced by the microbiota. *Immunity*. 2019;51:885–898.e7.
- Biton M, Haber AL, Rogel N, et al. T helper cell cytokines modulate intestinal stem cell renewal and differentiation. *Cell*. 2018;175:1307–1320.e22.
- Bennett KM, Parnell EA, Sanscartier C, et al. Induction of colonic M cells during intestinal inflammation. *Am J Pathol*. 2016;186:1166–1179.
- Elmentaite R, Kumasaka N, Roberts K, et al. Cells of the human intestinal tract mapped across space and time. *Nature*. 2021;597:250–255.
- Elmentaite R, Ross ADB, Roberts K, et al. Single-cell sequencing of developing human gut reveals transcriptional links to childhood Crohn's disease. *Dev Cell*. 2020;55:771–783.e5.
- Suzuki K, Murano T, Shimizu H, et al. Single cell analysis of Crohn's disease patient-derived small intestinal organoids reveals disease activity-dependent modification of stem cell properties. *J Gastroenterol*. 2018;53:1035–1047.
- Wang X, Bootsma H, Kroese F, Dijkstra G, Pringle S. Senescent stem and transient amplifying cells in Crohn's disease intestine. *Inflamm Bowel Dis*. 2020;26:e8–e9.
- Meriwether D, Sulaiman D, Volpe C, et al. Apolipoprotein A-I mimetics mitigate intestinal inflammation in COX2-dependent inflammatory bowel disease model. *J Clin Invest*. 2019;129:3670–3685.

31. Haberman Y, Tickle TL, Dexheimer PJ, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest.* 2014;124:3617–3633.
32. Hensel KO, Boland V, Postberg J, et al. Differential expression of mucosal trefoil factors and mucins in pediatric inflammatory bowel diseases. *Sci Rep.* 2014;4:7343.
33. Castro-Dopico T, Colombel JF, Mehandru S. Targeting B cells for inflammatory bowel disease treatment: back to the future. *Curr Opin Pharmacol.* 2020;55:90–98.
34. Palmer NP, Silvester JA, Lee JJ, et al. Concordance between gene expression in peripheral whole blood and colonic tissue in children with inflammatory bowel disease. *PLoS One.* 2019;14:e0222952.
35. 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.
36. Ho GT, Aird RE, Liu B, et al. MDR1 deficiency impairs mitochondrial homeostasis and promotes intestinal inflammation. *Mucosal Immunol.* 2018;11:120–130.
37. Cottle CM, Lipat A, Patel M, et al. Complexity of secretory chemokines in human intestinal organoid cultures ex vivo. *Gastro Hep Adv.* 2022;1:457–460.
38. Kaser A, Ludwiczek O, Holzmann S, et al. Increased expression of CCL20 in human inflammatory bowel disease. *J Clin Immunol.* 2004;24:74–85.
39. Lin XP, Almqvist N, Telemo E. Human small intestinal epithelial cells constitutively express the key elements for antigen processing and the production of exosomes. *Blood Cells Mol Dis.* 2005;35:122–128.
40. Kinchen J, Chen HH, Parikh K, et al. Structural remodeling of the human colonic mesenchyme in inflammatory bowel disease. *Cell.* 2018;175:372–386.e17.