Title: Single-cell profiling of environmental enteropathy reveals signatures of epithelial remodeling and immune activation in severe disease

Authors: Conner Kummerlowe[1,2,3,4], Thomas Wallach[5], Simutanyi Mwakamui[6], Travis K. Hughes[2,3,4], Nolawit Mulugeta[2,3,4], Victor Mudenda[6], Ellen Besa[6], Kanekwa Zyambo[6], Ira Fleming[2,3,4], Marko Vukovic[2,3,4], Ben A. Doran[2,3,4,8], Juliet Tongue Bramante[3], Amiko M. Uchida[9,10,11], John J. Garber[9,11], Jose Ordovas-Montanes[4,8,12,13], Zev Gartner[14], Alex K. Shalek[1,2,3,4,12,13*†], Paul Kelly[6,7*†]

Affiliations:

[1] Program in Computational and Systems Biology, Massachusetts Institute of Technology; Cambridge, MA, 02139, USA.

[2] Institute for Medical Engineering and Science (IMES), Department of Chemistry, and Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology; Cambridge, MA, 02139, USA.

[3] Ragon Institute of MGH, MIT, and Harvard; Cambridge, MA, 02139, USA.

[4] Broad Institute of MIT and Harvard; Cambridge, MA, 02142, USA.

[5] SUNY Downstate Health Sciences University; Department of Pediatrics, Brooklyn, NY, 11203, USA.

[6] Tropical Gastroenterology and Nutrition group, University of Zambia School of Medicine; Lusaka, Zambia.

[7] Blizard Institute, Queen Mary University of London; London E1 2AT, United Kingdom.

[8] Division of Gastroenterology, Hepatology, and Nutrition, Boston Children's Hospital; Boston, MA 02115, USA.

[9] Gastrointestinal Unit, Department of Medicine, Massachusetts General Hospital; Boston, MA, 02114, USA.

[10] Cancer Immunology and Virology, Dana Farber Cancer Institute; Boston, MA, 02115, USA.

[11] Department of Medicine, Harvard Medical School; Boston MA, 02115, USA.

[12] Program in Immunology, Harvard Medical School; Boston, MA, 02115, USA.

[13] Harvard Stem Cell Institute; Cambridge, MA, 02138, USA.

[14] University of California San Francisco; San Francisco, CA, 94185 USA.

*Corresponding authors. Emails: shalek@mit.edu (A.K.S.); m.p.kelly@qmul.ac.uk (P.K.)

† These authors contributed equally to this work

One Sentence Summary: Environmental enteropathy (EE) is a highly prevalent condition of the small intestine that is thought to be a primary cause of most global growth-stunting cases, and a key contributing factor to childhood malnutrition and diminished oral vaccine responses. Using single-cell RNA sequencing (scRNA-seq), we characterize, at unprecedented resolution, the pathophysiology of EE to derive new insights into the epithelial and immune correlates of disease severity, suggesting new therapeutic targets for future investigation.

Abstract: Environmental enteropathy (EE) is a subclinical condition of the small intestine that is highly prevalent in low- and middle-income countries. It is thought to be a primary cause of most global growth-stunting cases and a key contributing factor to childhood malnutrition and diminished oral vaccine responses. While EE has been shown to be the by-product of recurrent enteric infection, to date, its full pathophysiology remains unclear. Here, we mapped the cellular and molecular correlates of EE severity by performing high-throughput single-cell RNA-sequencing on 33 small intestinal biopsies from 11 adults with EE from Lusaka, Zambia (8 HIV-negative, 3 HIV-positive) and 6 adults without EE in Boston, USA. Using the resulting cellular atlas, we scored existing bulk-transcriptomic signatures of reduced villus height and decreased plasma LPS levels in EE, finding that these signatures may be driven by an increased abundance of surface mucosal cells–a gastric-like subset previously implicated in epithelial repair in the gastrointestinal tract. In addition, we identified several cell subsets whose fractional abundances associated with histological determined EE severity, small intestinal region, and HIV infection. Furthermore, by comparing distal duodenal EE samples with those from two U.S. control cohorts, we identified broadly decreased epithelial proliferative signaling, lower fractional abundances of goblet cells, and a T cell subset highly expressing a transcriptional signature of tissue-resident memory cells but with increased pro-inflammatory cytokine expression in EE. Altogether, our work illuminates the epithelial and immune correlates of EE severity and provides new molecular targets for intervention.

Main Text:

INTRODUCTION

Environmental enteropathy (EE) is a subclinical condition of the small intestine that is driven by enteropathogen exposure through environmental contamination *(1, 2)*. Also referred to as Environmental Enteric Dysfunction (EED), EE impacts millions of children and adults around the world. It is associated with stunted growth, neurocognitive impairment, reduced oral vaccine efficacy, and life-long increased risk of metabolic syndrome *(1, 3, 4)*.Water, sanitation, and hygiene (WASH) interventions for preventing EE have proven ineffective, and ongoing work is assessing alternative therapeutic interventions such as antibiotics, anti-inflammatory therapeutics, and dietary supplementation *(5)*. However, development of effective treatments has been hindered by a limited understanding of the underlying mechanisms of EE.

Studies of the tissue biology of EE have been practically limited by operational constraints. Obtaining small intestinal biopsies from patients with EE requires esophago-gastroduodenoscopy (EGD)—a procedure that, while generally safe, carries increased risk to perform in undernourished pediatric populations. Accordingly, novel approaches and initial mechanistic studies using invasive techniques should ethically be completed in affected adult populations before profiling pediatric patients. While the largest global health consequences of EE are seen in children, this condition also negatively impacts adult quality of life via its associated increased intestinal permeability, which can lead to systemic inflammation, increased risk of metabolic diseases, and reduced intestinal absorption *(1, 6)*. Given the environmental nature of enteric infections, finding unaffected controls for studies of EE in affected populations is similarly challenging. Thus, EE is often contextualized to health by either comparing intermediate EE with severe EE *(7)* or by comparing EE patients to control cohorts in the United States or the United Kingdom *(8)*. The validity of these international comparisons is supported by the environmental nature of EE and the resolution of EE in Peace Corps volunteers upon repatriation to the United States *(9)*.

Pathologically, EE in the proximal small intestine is continuous (not patchy), does not extend beyond the mucosa, and is characterized by reduced villus height, increased villus fusion, greater crypt depth, and increased microbial translocation *(10)*. However, in a study of Zambian children with EE and non-responsive growth stunting over time, reduced villus height was associated with decreased circulating levels of LPS (a measure of microbial translocation) *(11)*. In agreement, a

bulk transcriptomic study of Zambia children with enteropathy due to severe acute malnutrition (SAM) showed that many genes differentially upregulated in biopsies with reduced villus height were downregulated in biopsies from participants with high circulating levels of LPS *(7)*. These studies raise the intriguing possibility that EE is an adaptive response to potentially lethal enteropathogen exposure that comes at the cost of reduced absorptive capacity and thus impaired growth. However, a detailed picture of the cellular changes underpinning these processes is currently missing.

Histological analysis of EE has revealed increased abundance of lymphocytes (rather than granulocytes), abnormalities of secretory cells, reduced goblet cell numbers, and abnormal Paneth cell morphology *(8, 12)*. Low plasma levels of tryptophan in children with growth stunting *(13)* and the amelioration of villus blunting in Zambian adults given tryptophan, glutamine, and leucine supplementation suggest that amino acid deficiency may play a role in epithelial remodeling in EE *(14)*. Bulk transcriptomic studies of EE duodenal biopsies, meanwhile, have revealed increased expression of NADPH oxidases, CXC chemokines, mucin genes, matrix metalloprotease genes, interferon stimulated genes, and antimicrobial genes including *LCN2, DUOX2*, and *DUOXA2 (15, 16)*. In addition, immunohistochemistry staining of the DUOX2 protein has been validated as a marker distinguishing tissues from Bangladeshi children with EE from control samples obtained from North American children with healthy tissue or with celiac disease *(16)*. Furthermore, transcriptomic analysis of feces from Malawian children has illustrated increased innate immune activity and interferon signaling, as well as reduced expression of mucins and pro-proliferative genes (including EGFR and MAPK7), in severe EE *(17)*. However, previous work has lacked the single-cell resolution required to localize these changes to specific epithelial and immune cell subsets in the small intestine.

High-throughput single-cell genomic profiling holds transformative promise for understanding the cellular populations, phenotypic states, and signaling changes that underlie EE *(18)*. In the small intestine, it is challenging to use bulk level measurements to resolve the states of rarer cell types (such as intestinal stem cells, enteroendocrine cells, or tissue resident lymphocytes) that play crucial roles in tissue maintenance *(19)*. In contrast, single-cell RNA-sequencing (scRNA-seq) can be used to comprehensively profile the distinct cellular subsets that compose complex tissues,

thereby enabling analysis of rare yet influential cell subsets, and discovery of altered cell states and signaling. Illustratively, the unprecedented cellular resolution provided by scRNA-seq has been leveraged to identify novel features of human epithelial inflammation in the skin*(20)*, the nasal passages *(21)*, and the large intestine *(22)*. While scRNA-seq has also been used to study the healthy human small intestine *(23)*, to our knowledge, it has yet to be leveraged to understand inflammation in this organ.

Here, we applied the Seq-Well platform for massively-parallel scRNA-seq *(24)* to profile small intestinal biopsies from 11 adults from a community in Zambia where EE is known to be ubiquitous *(25)*. Across these individuals, we profiled 27 biopsies spanning 3 small intestinal regions, HIV-positive and HIV-negative patients, and a range of histological EE severity scores. This provided capacity to define the cellular subsets associated with these covariates. In addition, by comparing EE biopsies with those from two control groups in the USA, we found broadly decreased epithelial proliferative signaling, and lower abundances of goblet cells and a T cell subset highly expressing a transcriptional signature of tissue-resident memory cells as well as pro-inflammatory cytokines in EE. Altogether, our data provide new insight into epithelial remodeling and immune signaling in EE, suggesting several novel therapeutic targets for further investigation.

RESULTS

scRNA-seq atlas of the proximal small intestine with and without EE

We collected 27 small intestinal biopsies from 11 Zambian volunteers with EE (8 HIV-negative, 3 HIV-positive). For all 11, we profiled duodenal bulb and distal duodenum; for a subset, we also collected jejunal samples (**Table S1, S2**). Biopsies displayed varying levels of EE severity, with villus height:crypt depth ratios ranging from 0.64:1 to 2.38:1, compared to a normal ratio of 3:1 or greater (**Fig. 1A**, **Table S3**) *(8)*. Across these samples, we sought to identify the cellular correlates of intestinal region, HIV infection, and EE histological severity. However, due to both the widespread prevalence of EE in Zambia and a lack of existing screening methods to identify patients with EE, we could not obtain control biopsies from patients without EE in Zambia. Thus, we chose to distinguish the cellular and molecular features of EE biology by comparing to samples from patients in Boston, USA where EE can safely be assumed not to occur, and to mitigate the possibility of unexpected disease effects we included two control groups. We profiled distal duodenum samples from 3 patients undergoing screening for eosinophilic esophagitis (EoE) and

3 patients undergoing duodenal resection for pancreatic malignancy (**Table S1, S2**). Among the former, pathology did not report eosinophilic gastroenteritis for any patients screened for EoE, and revealed that one patient did not have EoE; duodenal resections from pancreatic cancer patients meanwhile were taken from uninvolved tissue.

Iterative clustering identifies cell subsets across heterogenous patient conditions

In total, we obtained 23,545 high quality single-cell transcriptomes across 33 samples (**Fig. 1B**). After data pre-processing (see **supplementary methods**), UMAP visualization of the entire dataset revealed differences in cellular distribution by patient cohort, intestinal region, and HIV status (**Fig. 1c**). In order to rigorously identify cellular subsets across these biological covariates, we adopted an existing pipeline for iterative clustering of cell subsets *(26)*, annotating the resulting clusters by known lineage markers and, in rare cases, merging together subsets across batches that co-expressed known lineage markers (see **supplementary methods**). This revealed that the major cell types (epithelial cells, T & NK cells, B cells, myeloid & stromal cells) were represented in almost all biopsies, indicating, along with more standard QC metrics (**Fig. S1-S4**) consistent sample quality (**Fig. 1D**). We then identified 48 granular cellular subsets that varied in abundance across samples (**Fig. 1E, Fig. S1-S4, Table S4**) *(19, 22, 23)*. These 48 subsets included: 23 epithelial subsets – inclusive of 3 stem cell subsets, 1 subset highly expressing *BEST4*, 1 secretory precursor subset, 1 goblet cell subset, 1 Paneth cell subset, 1 tuft cell subset, 6 enteroendocrine subsets, 1 surface mucosal cell subset (high expression of *TFF1* and *MUC5AC*), 1 mucosal neck cell subset (high expression of *TFF2* and *MUC6*), one foveolar precursor subset, and 6 enterocyte subsets, two of which highly expressed *ANXA2* (a marker of intestinal epithelial cells engaged wound healing *(27)*) and either *FABP1* or *PTMA* (markers of dedifferentiating intestinal cells in a murine model system *(28)*); 16 T & NK cellular subsets – consisting of 5 *CD8*^{hi} T cell, 2 *CD4*^{hi} T cell, 2 γδ T cell, 3 NK cell, 1 ILC3, and 3 T cell subsets respectively defined by interferon genes, mitochondrial genes, and cell cycle genes; 6 myeloid subsets – inclusive of macrophages, *CD14* monocytes, *CD16* monocytes, dendritic cells, mast cells, and eosinophils; 2 B cell subsets –plasma cells and *MS4A1* (i.e. CD20) expressing B cells; and, one stromal cell type, endothelial cells.

Comparison to past bulk RNA-seq EE signatures implicates surface mucosal cells expressing *DUOX2* in epithelial remodeling in EE

When identifying cell subsets, we noticed a striking similarity between the marker genes of surface mucosal cells (a cell subset most commonly found in the distal stomach that highly expresses *TFF1* and *MUC5AC* and has been implicated in the mucosal injury response across the gastrointestinal tract) *(29)* and existing bulk transcriptomic gene signatures characterizing EE samples with reduced villus height and decreased plasma LPS levels in Zambian children with SAM (**Fig. S5A**) *(7)*. Scoring all subsets on these signatures, we found that surface mucosal cells were significantly enriched for both signatures relative to all other cell subsets ($p < 10^{15}$, Wilcoxon test) (**Fig. 2A, B, Fig. S5B, C**). This suggests that a higher abundance of this cell subset may have driven these previously derived signatures. To more closely examine this possibility, we visualized the overlap between the genes from these signatures and the marker genes of all subsets (**Fig. 2C, D**). This revealed that the inferred enrichment of surface mucosal cells in this these signatures is driven by high expression of known surface mucosal cell markers *(30)*, as well as the antimicrobial genes *LCN2, DUOXA2,* and *DUOX2*. These last three genes were also highly upregulated in EE relative to controls in a recent bulk transcriptomic study of Bangladeshi children *(16)*. That study demonstrated that immunohistochemical staining of DUOX2 protein in the villus epithelium distinguished EE biopsies from controls. Given that reduced villus height in EE has been shown to associate with decreased microbial translocation over time *(11)*, it is remarkable that surface mucosal cells scored highly for signatures of both reduced villus height and decreased microbial translocation (as measured by decreased plasma LPS concentrations). This suggests that this cell subset may be associated with adaptive shortening of the villus in response to enteropathogen exposure.

Next, we sought to identify the epithelial subsets whose fractional abundances correlated with those of surface mucosal cells in our data. While the *DUOX2* signature was present in distal duodenal samples in past bulk transcriptomic studies, in our data, surface mucosal cells expressing *DUOX2* predominantly came from duodenal bulb samples of patients with EE (**Fig. S2C**). Thus, we focused our analysis on the duodenal bulb. We correlated the fractional abundances of all subsets in these samples and hierarchically clustered the results (**Fig. 2E**). Surface mucosal cell fractional abundances were significantly correlated (Permutation testing, p value < 0.05) with those of cycling stem cells (Stem cycling) and the Ent *ANXA2 PTMA* subset which highly expressed marker genes of dedifferentiating enterocytes (*PTMA*) *(28)* and wound associated

epithelial cells (*ANXA2*) *(27)* in murine studies. In addition, surface mucosal cell fractional abundances were weakly correlated with those of the Ent *ANXA2 FABP1 CD55* subset (which also highly expressed marker genes of dedifferentiation and wound healing) (Permutation testing, p value < 0.1). This suggests that surface mucosal cells are associated with a wound healing-like epithelial state.

To further characterize the Ent *ANXA2 PTMA* and Ent *ANXA2 FABP1 CD55* subsets that co-occurred with surface mucosal cells, we inferred differentiation trajectories for the epithelial cells in our dataset with PAGA *(31)* – an unbiased algorithm for estimating cell trajectories based on cell subset similarity (**Fig. 2F**). These two epithelial subsets and surface mucosal cells all lay in between mature enterocyte and stem cell subsets in the inferred differentiation hierarchy. In agreement, upon running the RNA velocity package Velocyto, these subsets again lay in between the enterocyte and stem cell subsets (**Fig. S5D, E**) *(32)*. Together, these results suggest that the presence of these wound healing-like epithelial subsets and *DUOX2* expressing surface mucosal cells in EE is associated with a shift towards an intermediate-like epithelial phenotype. This is consistent with the fact that surface mucosal cells highly expressed an array of transcripts encoding mucins and antimicrobial proteins, including *MUC5AC,* whose protein product has been shown to colocalize with *H. pylori* in gastric infections *(33)*, and *DPCR1* (also known as *MUCL3*), which has been shown to be a transcriptomic marker of wound associated epithelial cells in murine studies (**Table S4**) *(34)*. A shift towards this intermediate wound healing-like phenotype in response to enteropathogen mediated damage would provide a mechanism to explain the previous observation *(11)* that, over time, villus shortening is associated with decreased microbial translocation in EE.

Mapping the cellular correlates of intestinal region and disease severity in HIV-negative EE

Next, we sought to leverage the single-cell resolution of our scRNA-seq atlas to examine previously undescribed yet fundamental variation in cellular subsets associated with small intestinal region and histologically determined EE severity. To avoid any confounding impacts of HIV infection biology, we focused this analysis on the eight HIV-negative participants with EE in our study.

First, we identified cell subsets whose fractional abundances robustly shifted across intestinal region within HIV-negative EE patients (see **supplementary methods, Fig. 3A**). Duodenal bulb samples were enriched for surface mucosal cells, mucosal neck cells, and intermediate wound healing-like enterocytes highly expressing *ANXA2, FABP1* and *CD55*, as well as three T cell subsets expressing markers of immune activation (*IL17A, CXCR4*, *GZMA*) *(35, 36)*. On the other hand, distal duodenal samples were enriched for enterocytes, goblet cells, and stem cells highly expressing *OLFM4*. Due to a limited number of jejunal samples in our study, no cell subsets were significantly and robustly enriched in jejunal samples. These changes reveal that the duodenal bulb contains a more gastric-like transcriptional program than the distal duodenum in EE. This suggests variation in the cellular features of EE along the proximal small intestine.

Second, we sought to identify correlates of histologically determined EE severity across the entire small intestine. For eleven biopsies with matched histology and severity scores (**Table S5**) *(8)*, we used a Dirichlet regression model to assess how severity associated with cell subset composition and regress out small intestinal region-specific effects (**Fig. 3B,C**). As Dirichlet regression accounts for compositional dependencies in sequencing data, significant subsets do not always display the greatest absolute Pearson correlations with total EE severity, as these were generally observed in rarer subsets (e.g., tuft cells) whose relative abundances were likely to be highly influenced by random sampling or by shifts in the abundances of other subsets. Across small intestinal regions, in the epithelium, greater severity was significantly associated with lower fractional abundances of mature enterocytes and stem cells highly expressing *OLFM4*, as well as higher fractional abundances of immature enterocytes. This indicates that more severe EE may lead to a lower abundance of progenitor and mature cell subsets and a higher abundance of intermediate-like cell subsets. In the immune compartment, greater severity associated with a higher abundance of two T cell subsets expressing markers associated with inflammation in the intestine (*GZMA and CD6*) *(35)* and one T cell subset with high expression of *MALAT1* and the lowest median number of UMIs of all T cell subsets, which together suggest that this subset may represent low quality pre-apoptotic cells (**Fig. S2D**) *(37)*. In line with past histological findings in children with EE in the Gambia *(38)*, plasma cells abundances decreased with EE severity. Altogether, these results suggest that severe EE is characterized by an intermediate-like epithelial phenotype and provides new resolution into the lymphocyte subsets associated with EE.

Identifying the cellular features associated with both HIV infection and EE

Next, we sought to characterize the impact of highly active antiretroviral treated HIV infection on EE pathology. First, we examined the subset of the EE dataset (16 biopsies from 8 patients including all 3 HIV-positive patients) for which we obtained matched histological images, and found that HIV-positive samples displayed higher EE severity than HIV-negative samples (Wilcoxon test, p=0.034) (**Fig. S6A**). Next, we examined robust shifts in cell subset fractional abundances with HIV infection (**Fig. S6B, C**). Significant changes included known features of HIV biology, such as decreased fractional abundances of *CD4*^{hi} T cells *(39)* and increased fractional abundances of γδ T cells highly expressing the HIV co-receptor *CXCR4 (40)*. In addition, within duodenal bulb samples, HIV pathology associated with increased fractional abundances of the intermediate wound healing-like enterocytes highly expressing *ANXA2, FABP1 and CD55* (**Fig. S6D**). This suggested that HIV pathology may also contribute to the presence of these subsets within the duodenal bulb. Altogether, these results suggest that HIV infection has a distinct influence on EE pathology.

Inflammation is accompanied by decreased proliferative signaling in the EE epithelium

We next sought to identify features that distinguished EE small intestine samples from those without EE. To accomplish this, we compared the HIV-negative EE distal duodenal samples with two groups of matched small intestinal samples from Boston, USA: distal duodenal samples from patients referred to endoscopy for suspected EoE and distal duodenal samples from surgically resected uninvolved tissue from patients with pancreatic cancer. To identify distinguishing attributes of EE, we focused on biological features that: 1) were differentially regulated in EE compared to both outgroups and 2) displayed the same direction of regulation when comparing EE samples to each outgroup individually.

Applying this framework to the epithelium, we identified shifts in cell subset proportions, including reduced goblet cell fractional abundances in EE (**Fig. 4A**), consistent with previous reports *(8)*. Stem cells expressing *OLFM4* and enterocytes co-expressing *APOA4* and *ALPI* displayed were enriched in EE relative to controls, despite the fact that they displayed lower fractional abundances in severe EE relative to intermediate EE. One potential explanation for the

lack of consistency in these comparisons may be that enterocytes highly expressing *PCK1* constituted a large proportion of epithelial cells in the resection tissues yet was rarely observed in EoE or EE samples.

Accompanying these changes, epithelial compartment-wide and cell-type specific differential expression between EE and control cohorts revealed upregulation of genes (*PIGR, DMBT1, CCL25*) involved in mucosal defense mechanisms (antibody transport, NOD2 signaling, and lymphocyte recruitment, respectively) (**Fig. 4B,C**, **Table S6**) *(41–43)*. EE samples also displayed compartment-wide upregulation of *CTNNB1*, a key component of Wnt/ ß-catenin signaling, and compartment- and cell-type specific downregulation of genes (*KLF4, ATF3, FOS,* and *JUN*) involved in epithelial proliferation, IL-22 signaling in the intestine, and goblet cell development *(43–46)*. Consistent with this, gene set enrichment analysis (GSEA) displayed reduced activity of pro-proliferative MAPK and oxidative phosphorylation signaling in EE (**Fig. 4D, Table S7).** GSEA also revealed increased enrichment of the reactome signature for response of EIF2AAK4 and GCN2 to amino acid deficiency in the epithelial cells of EE relative to the US cohorts. As amino acid deficiency has been shown to restrict intestinal epithelial proliferation *(47)*, this finding of decreased proliferative signaling and amino acid deficiency response may provide a link connecting low-quality diets in EE endemic areas and the villus regenerative effect of amino acid supplementation in EE populations *(14)*. In addition, goblet cells from patients with EE upregulated markers for lower crypt goblet cells (*LYZ, REG3A*, and ribosomal genes), suggesting that EE goblet cells have a more immature phenotype and that the lower relative abundances of goblet cells in EE may result from impaired goblet cell maturation *(48)*. This was consistent with an observed upregulation of *GATA4* in mature enterocytes expressing *APOA* and *ALPI*, as this transcription factor has been shown to bias the intestinal epithelium towards enterocyte differentiation and away from goblet cell differentiation *(49)*. In sum, our findings suggest that inflammation in EE is associated with a hypoproliferative state in the epithelium that is accompanied by reduced goblet cell differentiation. This stands in stark contrast to previously characterized inflammation of the small intestine in Crohn's disease and celiac disease, where a hyperproliferative epithelium is induced by inflammation *(50, 51)*.

To further interrogate potential changes in proliferative capacity, we scored all stem cells in our

data on a module of genes annotated as upregulated in cycling human cells *(52)* (which included *MKI67,* cyclins, CDKs, topoisomerases, and polymerases) and found that this module score was significantly reduced in the EE stem cell compartment compared to the US cohorts (**Fig. 4E**). This further implicated altered proliferation and differentiation in EE relative to control cohorts. We extended these findings by inferring signaling activity in EE with PROGENy *(53)*, which inferred increased stress responsive pathway activation (JAK/STAT, TNFa, p53) and decreased proliferative pathway activation (EGFR, MAPK, PI3K) in *OLFM4* expressing stem cells (**Fig. 4F, Table S8**). We next inferred upstream transcription factor activity in EE with DoRothEA *(54)*, which found reduced activation of ATF2 and ATF4 broadly across the epithelium, which further implicated reduced IL-22 signaling*(50)* and has been shown to alter goblet cell development *(55)* (**Fig. S7A, Table S9**). In addition, STAT6, a key regulator of secretory lineage differentiation, was predicted to have lower activity in *OLFM4* expressing stem cells and immature enterocytes *(56)*. This is consistent with predicted upregulation of MYC in secretory precursors and goblet cells, which negatively regulates differentiation towards goblet cells *(57)*. Furthermore, predicted upregulation of GATA4 and GATA6 in *OLFM4* expressing stem cells and mature enterocytes, respectively, suggested that the loss of goblet cell differentiation may occur due to a polarization towards enterocyte differentiation *(49)*.

Altogether these predicted changes indicate altered epithelial differentiation and reduced proliferation of stem cells, the latter of which is surprising as T cell driven enteropathies such as celiac disease have been shown to induce a hyperproliferative state in the epithelium *(58)*. Thus, we wondered if EE severity within the Zambia cohort would negatively correlate with epithelial proliferation (like in EE relative to the US cohorts) or positive correlate with epithelial proliferation (like in other T cell driven enteropathies). To investigate this question, we correlated PROGENy pathway scores and proliferation gene module scores with histological EE severity scores in the stem cell subsets from HIV-negative patients in the Zambia cohort (**Fig. S7B**). Across all cells from stem cell subsets, the proliferation module weakly positively correlated with EE severity. This suggested that although EE patients as a whole display lower levels of epithelial proliferation, more severe EE leads to relatively higher epithelial proliferation than less severe EE.

Analysis of the immune compartment in EE reveals increased IFNγ signaling and reduced

fractional abundances of ILC3s and T cells expressing markers of tissue resident memory cells

Finally, we compared the immune cell compartment between EE and the US cohorts. Consistent with the within EE cohort comparison, cell proportion analysis revealed that, relative to control cohorts, EE samples were enriched for *CD8*^{hi} T cells highly expressing the lncRNA *MALAT1* and γδ T cells highly expressing *GZMA* (**Fig. 5A**). Consistent with previous findings *(38)*, plasma cells were increased in EE relative to non-EE cohorts despite exhibiting decreased abundances with increased severity within the EE cohort (**Fig. S8**). NK cells highly expressing *GZMA* and ILC3s were decreased in EE relative to the US cohorts. Notably, ILC3s have been shown to be a major source of IL-22 in the intestine *(59)* and γδ T cells have been shown to reduce IL-22 production by ILC3s in mice fed a low protein diet *(60)*. Thus, the relative increase in γδ T cells highly expressing *GZMA* and the relative decrease in ILC3s in EE relative to the U.S. cohorts may reflect an immune response to a low protein diet in EE patients that could lead to decreased IL-22 production.

We next identified changes in gene expression in the T cell compartment in EE relative to the US cohorts and found that genes involved in immune activation (*B2M, HLA*) were upregulated whereas downregulated genes included *TXNIP* (a negative regulator of T cell glucose uptake and proliferation *(61)*), *TAOK1* (a negative regulator of IL-17 *(62)*) and *IL-7R* (which has been implicated in the formation of memory T cells in response to infection *(63)*) (**Fig. 5B**). Differential expression between EE and the US cohorts within each cell type revealed that the vast majority of significantly differentially expressed genes within specific cell types came from T *CD8 CD69*^{hi} cells, which displayed upregulation of effector-like genes (*IFNG, HLA-DR*, CCL5, IL32, SLAMF7*) in EE and a downregulation of genes (*IL7R* and *CXCR4*) implicated in memory T cell formation in response to infection *(63–65)* (**Fig. 5C, Table S6**). Furthermore, as *CD69* is a potential marker for T cell tissue residency, we sought to ask if the T *CD8 CD69*^{hi} subset represented tissue resident cells. Scoring each T cell subset by a previously published tissue resident memory gene signature, we found that the T *CD8 CD69*^{hi} cluster scored highest (adjusted $p = 3.11e^{-64}$, one sided Wilcoxon test) (**Fig. 5D**) *(66)*. This suggests that this subset- which out of all subsets displayed the most differentially expressed genes between EE and the US cohorts– consisted of tissue resident cells. To further characterize this subset, we scored all cells from this

subset on a set of activation signatures for T cells in scRNA-seq data *(67)*. Relative to the US cohorts, EE T *CD8 CD69*^{hi} cells scored higher for T cells signatures for proliferation, cytokine production, and cytotoxicity (**Fig. 5E**). Together, these results suggested that, relative to the US cohorts, the tissue resident T cell niche in EE is more activated and polarized towards production of pro-inflammatory cytokines, which may contribute to the pathogenesis of EE.

To better understand the relationship between our observations in the epithelial and immune compartments in EE, we sought to identify potential ligand-receptor interactions between cellular subsets using NicheNet (**Fig. S9**) *(68)*. In addition to nominating putative drivers of immune epithelial-interactions in EE, this analysis recapitulated our previous observations that relative to the US cohorts, EE is characterized by increased IFNγ signaling in the epithelium stemming from IFNγ production by T cells, especially the T *CD8 CD69*^{hi} cells. These results further suggest that IFNγ is a critical regulator of EE pathology and may be a valuable avenue for therapeutics looking to reverse pathogenesis.

DISCUSSION

Environmental enteropathy is a subclinical condition of the small intestine that contributes to the continuing ill-health and entrenched poverty of millions of people around the world. Here, by utilizing Seq-Well–a transformative single-cell transcriptomics platform that can be applied in modestly equipped laboratories–we generated the first scRNA-seq dataset profiling the upper human small intestine in order to characterize EE. In the resultant single-cell atlas, we identified a cell subset–surface mucosal cells–which highly expressed *DUOX2* and whose gene expression pattern matched past bulk-transcriptomic signatures of reduced villus height and reduced plasma LPS concentrations in EE. In addition, our atlas revealed regional cell subset fractional abundance differences between the duodenal bulb and distal duodenum, demonstrated that more severe EE is characterized by a shift towards a more intermediate-like epithelial phenotype, provided enhanced resolution of the T cell subsets found in EE, and suggested that HIV enteropathy is associated with distinct cell subsets. In addition, by comparing distal duodenal samples from HIV-negative EE patients to distal duodenal samples from two U.S. cohorts, we found that EE is characterized by reduced proliferative signaling in the epithelium, and reduced fractional abundances of goblet cells and a T cell subset highly expressing a transcriptional signature of tissue-resident memory cells

with increased expression of pro-inflammatory cytokine genes. Altogether, our single-cell atlas and analyses re-contextualizes past bulk-transcriptomic studies of EE and identify previously unappreciated correlates of EE pathology.

The presence of surface mucosal cells highly expressing *DUOX2* in EE in our atlas may reflect adaptive remodeling of the epithelium into an intermediate-like wound healing state. One possible means of forming this state may be enterocyte dedifferentiation, which would be consistent with immunohistochemistry staining of EE samples from Bangladeshi children that found DUOX2+ cells at the villus tip *(16)*. While surface mucosal cells have previously been implicated in the mucosal injury response in pyloric metaplasia and in the Ulcer-associated cell lineage (UACL), these conditions are also associated with the presence of mucosal neck cells (high expression of *TFF2* and *MUC6*), whereas mucosal neck cells did not appear to be associated with a wound healing response in our data *(29)*. This may explain why our data suggests surface mucosal cells are a feature of injury response in EE yet past histological analyses of EE did not find any evidence of foveolar metaplasia *(8)*. Furthermore, the presence of this wound healing-like state associated with surface mucosal cells may occur at the expense of absorptive capacity, as absorptive enterocytes are reprogrammed into surface mucosal cells engaged in wound healing. This, in turn, may provide a mechanism that explains why, over time, reduced nutrient absorption in EE is associated with decreased microbial translocation *(69)*. However, this phenomenon may not be unique to EE as a previous bulk transcriptomic study has shown a similar tradeoff between *DUOX2* and *APOA4* (a mature enterocyte marker gene) in pediatric Crohn's disease relative to health *(70)*.

Examination of the other subsets in our atlas provided us the capacity to describe variation in EE due to intestinal region, histological severity, and HIV infection. However, this analysis lacked the power to detect significant changes in cell subsets with low fractional abundance and future work is needed to determine if shifts in rarer cell populations are associated with EE severity. In addition, profiling of pediatric patients will be required to understand the generality of these observations and link cellular responses to EE sequalae observed in children. Nonetheless, the data presented here provides a valuable resource for hypothesis generation for future studies of EE.

Strikingly, comparison of EE to the US cohorts suggested that the EE epithelium is characterized by decreased epithelial proliferation and found decreased abundance of ILC3s–a major producer of IL-22 (a positive regulator of intestinal epithelial proliferation) *(59)*–and increased abundances of γδ T cells– which have been shown to negatively regulate IL-22 production by ILC3s in mice fed a low protein diet *(60)*. Our findings are consistent with past work suggesting decreased proliferation in EE, including a study of fecal transcripts in Malawian children with EE that showed decreased abundances of transcripts in the EGFR and MAPK pathways in the feces of kids with EE relative to healthy kids *(17)*, as well as mouse studies showing that during *Cryptosporidium* infection protein malnutrition leads to reduced turnover and decreased expression of cell cycle genes in intestinal epithelial cells *(71)*. Consistently, reduced epithelial proliferation in EE relative to the US cohorts was associated with GSEA enrichment of a response to amino acid deficiency. This may explain the reduced relative abundances of goblet cells in EE relative to the US cohorts in our data, as the metabolism of tryptophan–an essential amino acid– has been shown to induce the differentiation of stem cells into goblet cells in mice *(72)*. This is consistent with work showing bulk transcriptomic evidence for decreased tryptophan metabolism in Bangladeshi children with EE *(16)*. Together with recent work showing that amino acid supplementation ameliorates villus blunting in adults with EE *(14)*, our results suggest that amino acid deficiency in the small intestine of adults with EE may lead to hypoproliferative signaling in the epithelium characterized by reduced stem cell proliferation and goblet cell abundance.

However, in contrast to the decreased proliferation in EE patients relative to the US cohorts, within HIV-negative Zambian adults with EE, we found that EE histological severity was associated with increased proliferation across all stem cell subsets. This is consistent with past bulk RNA-sequencing of Bangladeshi children with EE, which suggested that more severe EE leads to more proliferation than less severe EE, but lacked the single-cell resolution to demonstrate that this phenomenon was occurring in epithelial stem cells *(16)*. Thus, when comparing EE to US cohorts, EE appears to be negatively associated with stem cell proliferation, whereas when comparing within EE cohorts, more severe EE appears to be positive associated with stem cell proliferation. We believe that this discrepancy is due to the interplay of malnutrition and infection. We propose that the reduced dietary quality (and markedly diminished dietary diversity) in the population of Zambian adults we studied imposes a proliferative constraint on enteropathy. If this is correct, the

whole population has reduced stem cell turnover relative to intestinal homeostasis, but some individuals have more proliferation than others in response to infective and inflammatory drivers. However, we acknowledge that dissecting out the different processes induced by reduced dietary quality and pathogen exposure is very difficult, and experimental approaches to this are required.

One intriguing aspect of the observed hypoproliferative state in EE relative to US cohorts is that it stands in stark contrast to the hyperproliferative signaling observed in Crohn's disease *(50, 73)*. This is of particular interest as limited evidence suggests lower rates of IBD in countries where EE is endemic *(74)*. Thus, we propose that decreased intestinal epithelial proliferation in EE may be protective against inflammation-induced hyperproliferative states in the intestine such as Crohn's disease. This hypothesis is at an early stage of development and will require validation in other patient cohorts as well as animal or *in vitro* model studies mapping the causal factors driving reduced proliferative signaling. If validated, this result would 1) nominate genes associated with proliferation in IBD (such as *ATF3*) as potential therapeutic targets for modulating proliferation in EE and 2) call for future work investigating whether therapeutics inspired by the causal factors of EE may help to ameliorate IBD.

In addition to identifying changes in the epithelium in EE, our study provides new resolution of the immune subsets that associate with inflammation in EE. In our cross cohort comparisons, we found lower relative abundances of T cells expressing a transcriptional signature of tissue-resident memory T cells and higher relative abundances of plasma cells, suggesting that, similar to recent observations in IBD, EE may be characterized by an imbalance in cell mediated vs humoral immunity *(75)*. In addition, this tissue-resident memory-like subset highly expressed genes for inflammatory cytokines (including interferon-γ), suggesting that although they are present in a lower abundance in EE, these cells may be chronically activated and contribute to pathology. This, in turn, may lead to immune exhaustion and impaired response to new immune stimuli, which may contribute a hindered response to oral vaccines *(76)*. In addition, while interferon-γ is often viewed as a pro-inflammatory cytokine in acute inflammation, numerous studies have demonstrated that in chronic inflammation, interferon-γ can produce tolerogenic effects *(77)*. While a tolerogenic role for interferon-γ would be consistent with reduced oral vaccine efficacy in EE, future work is needed to delineate the specific mechanisms and impacts of interferon-γ signaling in the small

intestine with EE. Altogether, these identified immune changes provide clues about the mechanisms underlying the lack of oral vaccine efficacy in EE, though future work is needed to corroborate these findings in other cohorts and validate them in animal models.

It is important to recognize that our study has several inherent limitations. We were not able to include a non-EE control group of age-matched adults in Lusaka, Zambia. Thus, we cannot rule out the possibility that the observed differences between EE patients and the U.S. cohorts are due to ethnic differences or another unobserved variable that differed between these patient populations. In addition, our findings are primarily correlative due to the associative nature of measuring mRNA expression and the difficulties associated with mechanistic follow up validation in humans. Future studies of protein deficiency, tryptophan deficiency, and repletion will be necessary for understanding the significance of these results. In addition, the inflamed small intestinal epithelium and lamina propria are highly heterogenous environments containing several rare cell types such as tuft cells, Paneth cells, and a variety of adaptive and innate immune subsets which we did not have sufficient power to analyze in greater detail. Additionally, we did not screen EE patients for celiac disease (which displays a highly similar histological phenotype to EE) and we cannot completely rule out the possibility that some patients may have had celiac disease, however, we note that in Zambia the staple diet is maize (a gluten free food) and that past studies of EE in Zambian adults have found no evidence of celiac disease*(78)*. Furthermore, pediatric EE may display different aspects of biology than EE in adults, which calls for future studies in pediatric cohorts. In addition, as EE is an endemic condition in low- and middle-income countries across the globe, it will be necessary to validate that our results hold in cohorts with EE from geographic locations other than Zambia.

Examining our work as a whole, a potential picture of how malnutrition and enteropathogens interact to generate EE pathology emerges. Amino acid deficiency due to a low-quality diet may lead to reduced epithelial proliferation and differentiation towards goblet cells, which in turn would lead to a significant diminution of the host ability to maintain anti-microbial defense of the mucosa and thus may permit increased pathogen-mediated damage of the enterocyte. Increased enteropathogen damage may then lead to epithelial remodeling towards an intermediate wound healing-like phenotype associated with the presence of surface mucosal cells; further work on this

intriguing subset is urgently required. In addition, enteropathogen mediated damage would further exaggerate the pathogen-induced interferon-γ response and may explain the observed increased interferon-γ response in *CD8 CD69*^{hi} T cells in our data. Strikingly, metabolites of the essential amino acid tryptophan have been also been shown to induce IL-22 production by ILC3s in the intestine, and thus protein malnutrition may explain the lower relative abundances of ILC3s and the reduced evidence of IL-22 signaling in the epithelium in EE relative to control cohorts *(79)*. The combined effects of these changes may culminate in reduced IL-22 signaling, thereby promoting a more dormant epithelium, further reducing goblet differentiation, and thereby weakening the host ability to maintain anti-microbial defense. This would represent a vicious cycle of disease that may explain why the effect of combined malnutrition and enteropathogen exposure is far more deleterious than the effect of either factor alone. Together, these findings nominate several therapeutic axes for inducing healthy epithelial proliferation and immune efficacy in patients with EE and potentially other disorders of altered intestinal inflammation and epithelial function like IBD.

MATERIALS AND METHODS

Study design

Adult men and women, all volunteers, were recruited from a disadvantaged community in Lusaka, Zambia, in which we have carried out previous studies of environmental enteropathy *(11, 15)*. All volunteers gave written, fully informed consent. The study was approved by the University of Zambia Biomedical Research Ethics Committee (reference 006-11-15, dated 12th January 2016). From July 2018 to August 2018, volunteers underwent endoscopy with a Pentax 2990i gastroscope or a Pentax VSB2990i enteroscope, in the endoscopy unit of the University Teaching Hospital, Lusaka, under sedation with diazepam and pethidine. Duodenal tissue was collected from eosinophilic esophagitis (EoE) patients undergoing surveillance gastroscopy at Massachusetts General Hospital (MGH), Boston. Informed consent was obtained from EoE patients under a protocol approved by MGH. Resection samples were obtained from patients undergoing duodenal resection for pancreatic cancer (but in whom no local spread was apparent) in accordance with MGH IRB guidance under Mass General Brigham Protocol 2010P000632,

Biopsy handling and tissue digestion

Two biopsies from the patients with environmental enteropathy were collected into normal saline, then orientated under a dissecting microscope, fixed in buffered formal saline, and processed to 3μm sections for haematoxylin/eosin staining. These sections were scanned using an Olympus VS120 scanning microscope, measured for average villus height (VH) and crypt depth (CD) , and scored using a new scoring system for environmental enteropathy, which captures information about villus and crypt architecture, lamina propria and epithelial inflammation, epithelial injury and detachment, and abnormalities of Paneth and goblet cells *(8)*. Biopsies from EE, EoE, and resection patients were dissociated into single-cell suspensions using a modified version of a previously published protocol *(22)*. Further detail for tissue digestion is provided in the supplementary methods.

Single-cell RNA-sequencing with Seq-Well S^3:

Please refer to the supplementary methods for further detail. Briefly, single-cells were loaded onto a functionalized polydimethylsiloxane (PDMS) array preloaded with ~80,000 uniquely-barcoded mRNA capture beads (Chemgenes; MACOSKO-2011-10), and sequencing libraries were obtained following the protocol described in Hughes et al. *(20)* Arrays were sequenced on an Illumina Next-Seq. Sequencing read alignment and demultiplexing was performed on the cumulus platform *(80)* using snapshot 6 of the Drop-seq pipeline previously described in *(52)*, resulting in a cell barcode by UMI digital gene expression (DGE) matrix. QC was performed to remove low quality cell barcodes and doublet cells. To identify cell subsets, we adopted an existing pipeline for automated iterative clustering of single-cell data (pipeline described in detail in supplementary methods) that has been shown to identify batch effects without collapsing distinct rare cell types *(26, 81)*. To facilitate visualization and presentation of the data, final cell subsets obtained from automated iterative clustering and manual annotation with lineage markers were grouped into 1) epithelial 2) T & NK cells and 3) B cells, myeloid cells, and stromal cells. Cells from each group were then normalized with SCTransform and PCA and UMAP embeddings were calculated for visualization. Furthermore, final cell types were input to the FindAllMarkers function in Seurat with default parameters and a Wilcoxon test, producing marker genes for each final cell type. All subsets were scored on gene signatures of reduced villus height and decreased and plasma LPS signatures from Chama et al. *(7)* using the AddModuleScore function in Seurat and a Wilcoxon test was used to assess the significance of the difference of the mean module scores in surface mucosal cells relative

to all other subsets. Epithelial trajectories were inferred with PAGA *(31)* in Scanpy *(82)* run with default parameters. Prior to analysis in Scanpy, data was normalized in Seurat with SCTransform. Prior to running PAGA, the nearest neighbor graph was calculated in Scanpy with neighbors = ceiling(0.5(number of epithelial cells)^0.5)). RNA velocity trajectories were inferred with velocyto using default parameters *(32)*.

Analyses examining variation within samples from the EE cohort

To assess the epithelial subsets associated with surface mucosal cells in duodenal bulb samples from EE patients, we calculated the Pearson correlation between the fractional abundances of all epithelial subsets within these subsets. We then hierarchically clustered the resulting correlations using the k-means approach implemented in the ComplexHeatMap package in R.

Changes in the relative abundances of cell subsets by differing HIV infection status and small intestinal region were detected by a leave-one-out approach in order to avoid identifying patient specific effects. For each comparison (HIV positive vs negative, intestinal region A vs other intestinal regions, etc.), across all patients, a Fisher's exact test was used to identify subsets with significant changes in relative abundance (adjusted p value < 0.05). Then this analysis was repeated with each patient left out and only cell subsets that were significant in all analyses with patients left out were included in the final results.

Cell subsets significantly associated with histological severity from the Zambian cohort were identified by running Dirichlet Regression, which allows for testing for differences in cell subsets along a continuous dependent variable, with the R package DirichletReg. Further details are provided in the supplementary methods.

Comparison of distal duodenal samples from HIV-negative EE and U.S. control cohorts

Distal duodenal samples from patients with HIV-negative EE were compared to distal duodenal samples from HIV-negative EE patients to distal duodenal samples from patients with EoE and from tissue resections from patients with pancreatic cancer. In all analyses for this comparison, we sought to identify biological features driven by variation in EE biology relative to the two control cohorts (as opposed to identifying biology that distinguished only one cohort from EE). Thus, in all subsequent analyses for this comparison we required that results pass the following two criteria: 1) Result significant when comparing EE vs both control cohorts 2) Result displayed the same

direction of change between EE and each control cohort. Following this approach, we conducted compositional testing, differential expression, PROGENy, DoRothEA, GSEA, and NicheNet analyses. Further details for each of these analyses in provided in the supplementary methods.

Supplementary Materials

Materials and Methods

Fig. S1: Characterization of epithelial cell subsets revealed by single-cell RNA-sequencing.

Fig. S2: Characterization of T and NK cell subsets revealed by single-cell RNA-sequencing.

Fig. S3: Characterization of B cell, myeloid, and stromal subsets revealed by single-cell RNA-sequencing.

Fig. S4: Number of genes and UMIs per cell across samples

Fig. S5: Characterization of surface mucosal and dedifferentiation-like subsets

Fig. S6: Variation in EE biology associated with HIV infection

Fig. S7: Additional characterization of epithelial signaling changes

Fig. S8: Cell subset relative abundance and differential expression changes in B cells and myeloid cells between EE and controls

Fig. S9: NicheNet analysis of cell-cell signaling in EE relative to controls

Table S1: Clinical characteristics of donors of intestinal biopsy samples

Table S2: Patient cohort, intestinal region, and HIV infection status of participants studied

Table S3: Villus morphometry for EE distal duodenal samples

Table S4: Marker genes for all subsets

Table S5: Histological severity scores for EE biopsies with matched histology

Table S6: Genes differentially expressed in HIV- EE distal duodenal samples relative to samples from U.S cohorts

Table S7: KEGG, REACTOME, and PID pathways differentially enriched in HIV- EE distal duodenal samples relative to samples from U.S cohorts

Table S8: PROGENY pathway activation scores differentially regulated in HIV- EE distal duodenal samples relative to samples from U.S cohorts

Table S9: Transcription factors with upstream activity predictions from DoRothEA differentially scored in HIV- EE distal duodenal samples relative to samples from U.S cohorts

References and Notes

1. R. L. Guerrant, M. D. Deboer, S. R. Moore, R. J. Scharf, A. A. M. Lima, The impoverished gut - A triple burden of diarrhoea, stunting and chronic disease, *Nat. Rev. Gastroenterol. Hepatol.* **10**, 220–229 (2013).
2. R. Y. Chen, V. L. Kung, S. Das, M. S. Hossain, M. C. Hibberd, J. Guruge, M. Mahfuz, S. M. K. N. Begum, M. M. Rahman, S. M. Fahim, M. A. Gazi, R. Haque, S. A. Sarker, R. N. Mazumder, B. Di Luccia, K. Ahsan, E. Kennedy, J. Santiago-Borges, D. A. Rodionov, S. A. Leyn, A. L. Osterman, M. J. Barratt, T. Ahmed, J. I. Gordon, Duodenal Microbiota in Stunted Undernourished Children with Enteropathy, *N. Engl. J. Med.* **383**, 321–333 (2020).
3. J. A. Church, E. P. K. Parker, M. N. Kosek, G. Kang, N. C. Grassly, P. Kelly, A. J. Prendergast, Exploring the relationship between environmental enteric dysfunction and oral vaccine responses, *Future Microbiol.* **13**, 1055–1070 (2018).
4. R. D. Semba, M. Shardell, I. Trehan, R. Moaddel, K. M. Maleta, M. I. Ordiz, K. Kraemer, M. Khadeer, L. Ferrucci, M. J. Manary, Metabolic alterations in children with environmental enteric dysfunction, *Sci. Rep.* **6**, 1–9 (2016).
5. K. D. Tickell, H. E. Atlas, J. L. Walson, Environmental enteric dysfunction: A review of potential mechanisms, consequences and management strategies*BMC Med.* **17**, 181 (2019).
6. R. J. Crane, K. D. J. Jones, J. A. Berkley, Environmental enteric dysfunction: An overview, *Food Nutr. Bull.* **36**, S76–S87 (2015).
7. M. Chama, B. C. Amadi, K. Chandwe, K. Zyambo, E. Besa, N. Shaikh, I. M. Ndao, P. I. Tarr, C. Storer, R. Head, P. Kelly, Transcriptomic analysis of enteropathy in Zambian children with severe acute malnutrition, *EBioMedicine* **45**, 456–463 (2019).
8. T. C. Liu, K. Vanbuskirk, S. A. Ali, M. P. Kelly, L. R. Holtz, O. H. Yilmaz, K. Sadiq, N. Iqbal, B. Amadi, S. Syed, T. Ahmed, S. Moore, I. M. Ndao, M. H. Isaacs, J. D. Pfeifer, H. Atlas, P. I. Tarr, D. M. Denno, C. A. Moskaluk, A novel histological index for evaluation of environmental enteric dysfunction identifies geographic-specific features of enteropathy among children with suboptimal growth, *PLoS Negl. Trop. Dis.* **14**, 1–21 (2020).
9. G. T. Keusch, I. H. Rosenberg, D. M. Denno, C. Duggan, R. L. Guerrant, J. V. Lavery, P. I. Tarr, H. D. Ward, R. E. Black, J. P. Nataro, E. T. Ryan, Z. A. Bhutta, H. Coovadia, A. Lima, B. Ramakrishna, A. K. M. Zaidi, D. C. Hay Burgess, T. Brewer, Implications of acquired environmental enteric dysfunction for growth and stunting in infants and children living in low-and middle-income countries, *Food Nutr. Bull.* **34**, 357–364 (2013).
10. M. N. Kosek, T. Ahmed, Z. A. Bhutta, L. Caulfield, R. L. Guerrant, E. Houpt, G. Kang, G. Lee, A. A. M. Lima, B. J. J. McCormick, J. Platts-Mills, J. C. Seidman, R. R. Blank, M. Gottlieb, S. L. Knobler, D. R. Lang, M. A. Miller, K. H. Tountas, W. Checkley, C. J. Mason, L. E. Murray-Kolb, W. A. Petri, P. Bessong, R. Haque, S. John, E. R. Mduma, R. B. Oriá, P. S. Shrestha, S. K. Shrestha, E. Svensen, A. K. M. Zaidi, C. B. Abreu, A. M. Acosta, I. Ahmed, A. M. Shamsir Ahmed, A. Ali, R. Ambikapathi, L. Barrett, A. Bauck, E. Bayyo, L. Bodhidatta, A. Bose, J. Daniel Carreon, R. K. Chandyo, V. Charu, H. Costa, R. Dillingham, A. Di Moura, V. Doan, J. Q. Filho, J. Graham, C. Hoest, I. Hossain, M. Islam, M. Steffi Jennifer, S. Kaki, B. Koshy, Á. M. Leite, N. L. Lima, B. L. L. Maciel, M. Mahfuz, C. Mahopo, A. Maphula, M. McGrath, A. Mohale, M. Moraes, F. S. Mota, J. Muliyil, R. Mvungi, G. Nayyar, E. Nyathi, M. P. Olortegui, R. Oria, A. O. Vasquez, W. K. Pan, J. Pascal, C. L. Patil, L. Pendergast, S. R. Pinedo, S. Psaki, M. V. Raghava, K. Ramanujam, M. Rasheed, Z. A. Rasmussen, S. A. Richard, A. Rose, R. Roshan, B. Schaefer, R. Scharf, S. L. Sharma, B. Shrestha, R. Shrestha, S. Simons, A. M. Soares, R. M. S. Mota, S. Soofi, T. Strand, F. Tofail, R. J. Thomas, A. Turab, M. Ulak, V. Wang, L. Yarrot, P. P. Yori, D. Alam, C. Amour, C. B. Chavez, S. Babji, R. R. de Burga, J. T. Flores, J. Gratz, A. T. George, D. Hariraju, A. Havt, P. Karunakaran, R. P. Lazarus, I. F. Lima, D. Mondal, P. H. Q. S. Medeiros, R. Nshama, J. Quetz, S. Qureshi, S. Raju, A. Ramachandran, R. Ramadas, A. Catharine Ross, M. S. Salas, A. Samie, K. Schulze, E. Shanmuga Sundaram, B. M. Swema, D. R. Trigoso, Causal Pathways from Enteropathogens to Environmental Enteropathy: Findings from the MAL-ED Birth Cohort Study, *EBioMedicine* **18**, 109–117 (2017).
11. B. Amadi, K. Zyambo, K. Chandwe, E. Besa, C. Mulenga, S. Mwakamui, S. Siyumbwa, S. Croft, R. Banda, M. Chipunza, K. Chifunda, L. Kazhila, K. Vanbuskirk, P. Kelly, Adaptation of the small intestine to microbial enteropathogens in Zambian children with stunting, *Nat. Microbiol.* , doi:10.1038/s41564-020-00849-w.
12. W. Dhaliwal, M. Bajaj-Elliott, P. Kelly, Intestinal defensin gene expression in human populations, *Mol. Immunol.* **40**, 469–475 (2003).
13. M. N. Kosek, E. Mduma, P. S. Kosek, G. O. Lee, E. Svensen, W. K. Y. Pan, M. P. Olortegui, J. H. Bream, C. Patil, C. R. Asayag, G. M. Sanchez, L. E. Caulfield, J. Gratz, P. P. Yori, Plasma tryptophan and the kynurenine-

tryptophan ratio are associated with the acquisition of statural growth deficits and oral vaccine underperformance in populations with environmental enteropathy, *Am. J. Trop. Med. Hyg.* **95**, 928–937 (2016).
14. J. Louis-Auguste, E. Besa, K. Zyambo, D. Munkombwe, R. Banda, T. Banda, A. Watson, J. Mayneris-Perxachs, J. Swann, P. Kelly, Tryptophan, glutamine, leucine, and micronutrient supplementation improves environmental enteropathy in Zambian adults: A randomized controlled trial, *Am. J. Clin. Nutr.* **110**, 1240–1252 (2019).
15. P. Kelly, E. Besa, K. Zyambo, J. Louis-Auguste, J. Lees, T. Banda, R. Soko, R. Banda, B. Amadi, A. Watson, Endomicroscopic and Transcriptomic Analysis of Impaired Barrier Function and Malabsorption in Environmental Enteropathy, *PLoS Negl. Trop. Dis.* **10**, 1–18 (2016).
16. Y. Haberman, N. T. Iqbal, S. Ghandikota, I. Mallawaarachchi, T. Braun, P. J. Dexheimer, N. Rahman, R. Hadar, K. Sadiq, Z. Ahmad, R. Idress, J. Iqbal, S. Ahmed, A. Hotwani, F. Umrani, L. Ehsan, G. Medlock, S. Syed, C. Moskaluk, J. Z. Ma, A. G. Jegga, S. R. Moore, S. A. Ali, L. A. Denson, Mucosal Genomics Implicate Lymphocyte Activation and Lipid Metabolism in Refractory Environmental Enteric Dysfunction, *Gastroenterology* (2021), doi:10.1053/j.gastro.2021.01.221.
17. J. Yu, M. I. Ordiz, J. Stauber, N. Shaikh, I. Trehan, E. Barnell, R. D. Head, K. Maleta, P. I. Tarr, M. J. Manary, Environmental Enteric Dysfunction Includes a Broad Spectrum of Inflammatory Responses and Epithelial Repair Processes, *CMGH* **2**, 158-174.e1 (2016).
18. S. M. Prakadan, A. K. Shalek, D. A. Weitz, Scaling by shrinking: Empowering single-cell "omics" with microfluidic devices, *Nat. Rev. Genet.* **18**, 345–361 (2017).
19. A. L. Haber, M. Biton, N. Rogel, R. H. Herbst, K. Shekhar, C. Smillie, G. Burgin, T. M. Delorey, M. R. Howitt, Y. Katz, I. Tirosh, S. Beyaz, D. Dionne, M. Zhang, R. Raychowdhury, W. S. Garrett, O. Rozenblatt-Rosen, H. N. Shi, O. Yilmaz, R. J. Xavier, A. Regev, A single-cell survey of the small intestinal epithelium., *Nature* **551**, 333–339 (2017).
20. T. K. Hughes, M. H. Wadsworth, T. M. Gierahn, T. Do, D. Weiss, P. R. Andrade, F. Ma, B. J. de Andrade Silva, S. Shao, L. C. Tsoi, J. Ordovas-Montanes, J. E. Gudjonsson, R. L. Modlin, J. C. Love, A. K. Shalek, Second-Strand Synthesis-Based Massively Parallel scRNA-Seq Reveals Cellular States and Molecular Features of Human Inflammatory Skin Pathologies, *Immunity* **53**, 878-894.e7 (2020).
21. J. Ordovas-Montanes, D. F. Dwyer, S. K. Nyquist, K. M. Buchheit, M. Vukovic, C. Deb, M. H. Wadsworth, T. K. Hughes, S. W. Kazer, E. Yoshimoto, K. N. Cahill, N. Bhattacharyya, H. R. Katz, B. Berger, T. M. Laidlaw, J. A. Boyce, N. A. Barrett, A. K. Shalek, Allergic inflammatory memory in human respiratory epithelial progenitor cells*Nature* **560**, 649–654 (2018).
22. C. S. Smillie, M. Biton, J. Ordovas-Montanes, K. M. Sullivan, G. Burgin, D. B. Graham, R. H. Herbst, N. Rogel, M. Slyper, J. Waldman, M. Sud, E. Andrews, G. Velonias, A. L. Haber, K. Jagadeesh, S. Vickovic, J. Yao, C. Stevens, D. Dionne, L. T. Nguyen, A.-C. Villani, M. Hofree, E. A. Creasey, H. Huang, O. Rozenblatt-Rosen, J. J. Garber, H. Khalili, A. N. Desch, M. J. Daly, A. N. Ananthakrishnan, A. K. Shalek, R. J. Xavier, A. Regev, Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis., *Cell* **178**, 714-730.e22 (2019).
23. X. Han, Z. Zhou, L. Fei, H. Sun, R. Wang, Y. Chen, H. Chen, J. Wang, H. Tang, W. Ge, Y. Zhou, F. Ye, M. Jiang, J. Wu, Y. Xiao, X. Jia, T. Zhang, X. Ma, Q. Zhang, X. Bai, S. Lai, C. Yu, L. Zhu, R. Lin, Y. Gao, M. Wang, Y. Wu, J. Zhang, R. Zhan, S. Zhu, H. Hu, C. Wang, M. Chen, H. Huang, T. Liang, J. Chen, W. Wang, D. Zhang, G. Guo, Construction of a human cell landscape at single-cell level, *Nature* **581**, 303–309 (2020).
24. T. M. Gierahn, M. H. Wadsworth, T. K. Hughes, B. D. Bryson, A. Butler, R. Satija, S. Fortune, J. Christopher Love, A. K. Shalek, Seq-Well: Portable, low-cost rna sequencing of single cells at high throughput, *Nat. Methods* **14**, 395–398 (2017).
25. P. Kelly, I. Menzies, R. Crane, I. Zulu, C. Nickols, R. Feakins, J. Mwansa, V. Mudenda, M. Katubulushi, S. Greenwald, M. Farthing, Responses of small intestinal architecture and function over time to environmental factors in a tropical population, *Am. J. Trop. Med. Hyg.* **70**, 412–419 (2004).
26. K. Shekhar, S. W. Lapan, I. E. Whitney, N. M. Tran, E. Z. Macosko, M. Kowalczyk, X. Adiconis, J. Z. Levin, J. Nemesh, M. Goldman, S. A. McCarroll, C. L. Cepko, A. Regev, J. R. Sanes, Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics, *Cell* **166**, 1308-1323.e30 (2016).
27. C. R. Rankin, R. S. Hilgarth, G. Leoni, M. Kwon, K. A. Den Beste, C. A. Parkos, A. Nusrat, Annexin A2 regulates β1 integrin internalization and intestinal epithelial cell migration, *J. Biol. Chem.* **288**, 15229–15239 (2013).
28. P. W. Tetteh, O. Basak, H. F. Farin, K. Wiebrands, K. Kretzschmar, H. Begthel, M. Van Den Born, J. Korving, F. De Sauvage, J. H. Van Es, A. Van Oudenaarden, H. Clevers, Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters, *Cell Stem Cell* **18**, 203–213 (2016).
29. J. R. Goldenring, Pyloric metaplasia, pseudopyloric metaplasia, ulcer-associated cell lineage and spasmolytic polypeptide-expressing metaplasia: reparative lineages in the gastrointestinal mucosa, *J. Pathol.* **245**, 132–137 (2018).

30. W. Hoffmann, TFF2, a MUC6-binding lectin stabilizing the gastric mucus barrier and more (Review), *Int. J. Oncol.* **47**, 806–816 (2015).
31. F. A. Wolf, F. K. Hamey, M. Plass, J. Solana, J. S. Dahlin, B. Göttgens, N. Rajewsky, L. Simon, F. J. Theis, PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells, *Genome Biol.* **20**, 1–9 (2019).
32. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells, *Nature* **560**, 494–498 (2018).
33. G. R. Van Den Brink, K. M. A. J. Tytgat, R. W. M. Van Der Hulst, C. M. Van Der Loos, A. W. C. Einerhand, H. A. Büller, J. Dekker, H pylori colocalises with MUC5AC in the human stomach, *Gut* **46**, 601–607 (2000).
34. H. Miyoshi, K. L. VanDussen, N. P. Malvin, S. H. Ryu, Y. Wang, N. M. Sonnek, C. Lai, T. S. Stappenbeck, Prostaglandin E2 promotes intestinal repair through an adaptive cellular response of the epithelium, *EMBO J.* **36**, 5–24 (2017).
35. S. Velaga, S. N. Ukena, U. Dringenberg, C. Alter, J. Pardo, O. Kershaw, A. Franzke, Granzyme a is required for regulatory T-cell mediated prevention of gastrointestinal graft-versus-host disease, *PLoS One* **10**, 1–9 (2015).
36. N. P. Zimmerman, R. A. Vongsa, S. L. Faherty, N. H. Salzman, M. B. Dwinell, Targeted intestinal epithelial deletion of the chemokine receptor CXCR4 reveals important roles for extracellular-regulated kinase-1/2 in restitution, *Lab. Investig.* **91**, 1040–1055 (2011).
37. E. Madissoon, A. Wilbrey-Clark, R. J. Miragaia, K. Saeb-Parsy, K. T. Mahbubani, N. Georgakopoulos, P. Harding, K. Polanski, N. Huang, K. Nowicki-Osuch, R. C. Fitzgerald, K. W. Loudon, J. R. Ferdinand, M. R. Clatworthy, A. Tsingene, S. Van Dongen, M. Dabrowska, M. Patel, M. J. T. Stubbington, S. A. Teichmann, O. Stegle, K. B. Meyer, ScRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation, *Genome Biol.* **21**, 1–16 (2019).
38. D. I. Campbell, S. H. Murch, M. Elia, P. B. Sullivan, M. S. Sanyang, B. Jobarteh, P. G. Lunn, Chronic T cell-mediated enteropathy in rural West African children: Relationship with nutritional status and small bowel function, *Pediatr. Res.* **54**, 306–311 (2003).
39. A. A. Okoye, L. J. Picker, CD4+ T-Cell Depletion In Hiv Infection: Mechanisms Of Immunological Failure, *Immunol. Rev.* **254**, 54–64 (2013).
40. L. Werner, H. Elad, E. Brazowski, H. Tulchinsky, S. Vigodman, U. Kopylov, Z. Halpern, H. Guzner-Gur, I. Dotan, Reciprocal regulation of CXCR4 and CXCR7 in intestinal mucosal homeostasis and inflammatory bowel disease, *J. Leukoc. Biol.* **90**, 583–590 (2011).
41. D. J. Campbell, E. C. Butcher, Intestinal attraction: CCL25 functions in effector lymphocyte recruitment to the small intestine, *J. Clin. Invest.* **110**, 1079–1081 (2002).
42. H. Turula, C. E. Wobus, The role of the polymeric immunoglobulin receptor and secretory immunoglobulins during mucosal infection and immunity, *Viruses* **10**, 1–15 (2018).
43. P. Rosenstiel, C. Sina, C. End, M. Renner, S. Lyer, A. Till, S. Hellmig, S. Nikolaus, U. R. Fölsch, B. Helmke, F. Autschbach, P. Schirmacher, P. Kioschis, M. Hafner, A. Poustka, J. Mollenhauer, S. Schreiber, Regulation of DMBT1 via NOD2 and TLR4 in Intestinal Epithelial Cells Modulates Bacterial Recognition and Invasion , *J. Immunol.* **178**, 8203–8211 (2007).
44. D. Glal, J. N. Sudhakar, H. H. Lu, M. C. Liu, H. Y. Chiang, Y. C. Liu, C. F. Cheng, J. W. Shui, ATF3 sustains IL-22-induced STAT3 phosphorylation to maintain mucosal immunity through inhibiting phosphatases, *Front. Immunol.* **9** (2018), doi:10.3389/fimmu.2018.02522.
45. J. Zhou, B. A. Edgar, M. Boutros, ATF3 acts as a rheostat to control JNK signalling during intestinal regeneration, *Nat. Commun.* **8** (2017), doi:10.1038/ncomms14289.
46. J. P. Katz, N. Perreault, B. G. Goldstein, C. S. Lee, P. A. Labosky, V. W. Yang, K. H. Kaestner, The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon, *Development* **129**, 2619–2628 (2002).
47. Y. Zhu, G. Lin, Z. Dai, T. Zhou, T. Li, T. Yuan, Z. Wu, G. Wu, J. Wang, L-Glutamine deprivation induces autophagy and alters the mTOR and MAPK signaling pathways in porcine intestinal epithelial cells, *Amino Acids* **47**, 2185–2197 (2015).
48. R. Manco, I. Averbukh, Z. Porat, K. B. Halpern, I. Amit, S. Itzkovitz, Clump sequencing exposes the spatial expression programs of intestinal secretory cells, *bioRxiv* , 2020.08.05.237917 (2020).
49. E. M. Walker, C. A. Thompson, M. A. Battle, GATA4 and GATA6 regulate intestinal epithelial cytodifferentiation during development, *Dev. Biol.* **392**, 283–294 (2014).
50. N. Powell, E. Pantazi, P. Pavlidis, A. Tsakmaki, K. Li, F. Yang, A. Parker, C. Pin, D. Cozzetto, D. Minns, E.

Stolarczyk, S. Saveljeva, R. Mohamed, P. Lavender, B. Afzali, J. Digby-Bell, T. Tjir-Li, A. Kaser, J. Friedman, T. T. Macdonald, G. A. Bewick, G. M. Lord, Interleukin-22 orchestrates a pathological endoplasmic reticulum stress response transcriptional programme in colonic epithelial cells, *Gut* **69**, 578–590 (2020).
51. K. Juuti-Uusitalo, M. Mäki, H. Kainulainen, J. Isola, K. Kaukinen, Gluten affects epithelial differentiation-associated genes in small intestinal mucosa of coeliac patients, *Clin. Exp. Immunol.* **150**, 294–305 (2007).
52. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets, *Cell* **161**, 1202–1214 (2015).
53. M. Schubert, B. Klinger, M. Klünemann, A. Sieber, F. Uhlitz, S. Sauer, M. J. Garnett, N. Blüthgen, J. Saez-Rodriguez, Perturbation-response genes reveal signaling footprints in cancer gene expression, *Nat. Commun.* **9** (2018), doi:10.1038/s41467-017-02391-6.
54. L. Garcia-Alonso, M. M. Ibrahim, D. Turei, J. Saez-Rodriguez, Benchmark and integration of resources for the estimation of human transcription factor activities, *bioRxiv* , 1363–1375 (2018).
55. B. J. Meijer, F. P. Giugliano, B. Baan, J. H. M. van der Meer, S. Meisner, M. van Roest, P. J. Koelink, R. J. de Boer, N. Jones, W. Breitwieser, N. N. van der Wel, M. E. Wildenberg, G. R. van den Brink, J. Heijmans, V. Muncan, ATF2 and ATF7 Are Critical Mediators of Intestinal Epithelial Repair, *Cmgh* **10**, 23–42 (2020).
56. C. Schubart, B. Krljanac, M. Otte, C. Symowski, E. Martini, C. Günther, C. Becker, C. Daniel, D. Voehringer, Selective expression of constitutively activated STAT6 in intestinal epithelial cells promotes differentiation of secretory cells and protection against helminths, *Mucosal Immunol.* **12**, 413–424 (2019).
57. A. J. Finch, L. Soucek, M. R. Junttila, L. B. Swigart, G. I. Evan, Acute Overexpression of Myc in Intestinal Epithelium Recapitulates Some but Not All the Changes Elicited by Wnt/β-Catenin Pathway Activation, *Mol. Cell. Biol.* **29**, 5306–5315 (2009).
58. T. T. MacDonald, J. Spencer, Evidence that activated mucosal T cells play a role in the pathogenesis of enteropathy in human small intestine, *J. Exp. Med.* **167**, 1341–1349 (1988).
59. A. R. Victor, A. P. Nalin, W. Dong, S. McClory, M. Wei, C. Mao, R. D. Kladney, Y. Youssef, W. K. Chan, E. L. Briercheck, T. Hughes, S. D. Scoville, J. R. Pitarresi, C. Chen, S. Manz, L.-C. Wu, J. Zhang, M. C. Ostrowski, A. G. Freud, G. W. Leone, M. A. Caligiuri, J. Yu, IL-18 Drives ILC3 Proliferation and Promotes IL-22 Production via NF-κB, *J. Immunol.* **199**, 2333–2342 (2017).
60. Z. A. Sullivan, W. Khoury-Hanold, J. Lim, C. Smillie, M. Biton, B. S. Reis, R. K. Zwick, S. D. Pope, K. Israni-Winger, R. Parsa, N. H. Philip, S. Rashed, N. Palm, A. Wang, D. Mucida, A. Regev, R. Medzhitov, γδ T cells regulate the intestinal response to nutrient sensing., *Science* **371** (2021), doi:10.1126/science.aba8310.
61. J. Muri, S. Heer, M. Matsushita, L. Pohlmeier, L. Tortola, T. Fuhrer, M. Conrad, N. Zamboni, J. Kisielow, M. Kopf, The thioredoxin-1 system is essential for fueling DNA synthesis during T-cell metabolic reprogramming and proliferation, *Nat. Commun.* **9**, 1–16 (2018).
62. Z. Zhang, Z. Tang, X. Ma, K. Sun, L. Fan, J. Fang, J. Pan, X. Wang, H. An, J. Zhou, TAOK1 negatively regulates IL-17-mediated signaling and inflammation, *Cell. Mol. Immunol.* **15**, 794–802 (2018).
63. A. Chandele, N. S. Joshi, J. Zhu, W. E. Paul, W. J. Leonard, S. M. Kaech, Formation of IL-7Rα high and IL-7Rα low CD8 T Cells during Infection Is Regulated by the Opposing Functions of GABPα and Gfi-1 , *J. Immunol.* **180**, 5309–5319 (2008).
64. J. Chaix, S. A. Nish, W.-H. W. Lin, N. J. Rothman, L. Ding, E. J. Wherry, S. L. Reiner, Cutting Edge: CXCR4 Is Critical for CD8 + Memory T Cell Homeostatic Self-Renewal but Not Rechallenge Self-Renewal , *J. Immunol.* **193**, 1013–1016 (2014).
65. L. Loyal, S. Warth, K. Jürchott, F. Mölder, C. Nikolaou, N. Babel, M. Nienen, S. Durlanik, R. Stark, B. Kruse, M. Frentsch, R. Sabat, K. Wolk, A. Thiel, SLAMF7 and IL-6R define distinct cytotoxic versus helper memory CD8+ T cells, *Nat. Commun.* **11** (2020), doi:10.1038/s41467-020-19002-6.
66. B. V. Kumar, W. Ma, M. Miron, T. Granot, R. S. Guyer, D. J. Carpenter, T. Senda, X. Sun, S. H. Ho, H. Lerner, A. L. Friedman, Y. Shen, D. L. Farber, Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites, *Cell Rep.* **20**, 2921–2934 (2017).
67. P. A. Szabo, H. M. Levitin, M. Miron, M. E. Snyder, T. Senda, J. Yuan, Y. L. Cheng, E. C. Bush, P. Dogra, P. Thapa, D. L. Farber, P. A. Sims, Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease, *Nat. Commun.* **10**, 1–16 (2019).
68. R. Browaeys, W. Saelens, Y. Saeys, NicheNet: modeling intercellular communication by linking ligands to target genes, *Nat. Methods* **17**, 159–162 (2020).
69. B. Amadi, Adaptation of the small intestine to microbial enteropathogens in Zambian children with stunting (2021).
70. Y. Haberman, T. L. Tickle, P. J. Dexheimer, M. O. Kim, D. Tang, R. Karns, R. N. Baldassano, J. D. Noe, J.

Rosh, J. Markowitz, M. B. Heyman, A. M. Griffiths, W. V. Crandall, D. R. Mack, S. S. Baker, C. Huttenhower, D. J. Keljo, J. S. Hyams, S. Kugathasan, T. D. Walters, B. Aronow, R. J. Xavier, D. Gevers, L. A. Denson, Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature, *J. Clin. Invest.* **124**, 3617–3633 (2014).
71. J. Liu, D. T. Bolick, G. L. Kolling, Z. Fu, R. L. Guerrant, Protein malnutrition impairs intestinal epithelial cell turnover, a potential mechanism of increased cryptosporidiosis in a murine model, *Infect. Immun.* **84**, 3542–3549 (2016).
72. D. M. Alvarado, B. Chen, M. Iticovici, A. I. Thaker, N. Dai, K. L. VanDussen, N. Shaikh, C. K. Lim, G. J. Guillemin, P. I. Tarr, M. A. Ciorba, Epithelial Indoleamine 2,3-Dioxygenase 1 Modulates Aryl Hydrocarbon Receptor and Notch Signaling to Increase Differentiation of Secretory Cells and Alter Mucus-Associated Microbiota, *Gastroenterology* **157**, 1093-1108.e11 (2019).
73. M. Pierdomenico, L. Stronati, M. Costanzo, R. Vitali, G. Di Nardo, F. Nuti, S. Oliva, S. Cucchiara, A. Negroni, New insights into the pathogenesis of inflammatory bowel disease: Transcription factors analysis in bioptic tissues from pediatric patients, *J. Pediatr. Gastroenterol. Nutr.* **52**, 271–279 (2011).
74. P. Hodges, P. Kelly, Inflammatory bowel disease in Africa: what is the current state of knowledge?, *Int. Health* **12**, 222–230 (2020).
75. A. Noble, L. Durant, L. Hoyles, A. L. McCartney, R. Man, J. Segal, S. P. Costello, P. Hendy, D. Reddi, S. Bouri, D. N. F. Lim, T. Pring, M. J. O'Connor, P. Datt, A. Wilson, N. Arebi, A. Akbar, A. L. Hart, S. R. Carding, S. C. Knight, Deficient Resident Memory T Cell and CD8 T Cell Response to Commensals in Inflammatory Bowel Disease, *J. Crohn's Colitis* **14**, 525–537 (2020).
76. E. J. Wherry, M. Kurachi, Molecular and cellular insights into T cell exhaustion, *Nat. Rev. Immunol.* **15**, 486–499 (2015).
77. P. Rožman, U. Švajger, The tolerogenic role of IFN-γ*Cytokine Growth Factor Rev.* **41**, 40–53 (2018).
78. B. Amadi, E. Besa, K. Zyambo, P. Kaonga, J. Louis-Auguste, K. Chandwe, P. I. Tarr, D. M. Denno, J. P. Nataro, W. Faubion, A. Sailer, S. Yeruva, T. Brantner, J. Murray, A. J. Prendergast, J. R. Turner, P. Kelly, Impaired Barrier Function and Autoantibody Generation in Malnutrition Enteropathy in Zambia, *EBioMedicine* **22**, 191–199 (2017).
79. T. Zelante, R. G. Iannitti, C. Cunha, A. DeLuca, G. Giovannini, G. Pieraccini, R. Zecchi, C. D'Angelo, C. Massi-Benedetti, F. Fallarino, A. Carvalho, P. Puccetti, L. Romani, Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22, *Immunity* **39**, 372–385 (2013).
80. K. C. Observatory, I. Diseases, M. G. Hospital, M. G. Hospital, Cumulus: a cloud-based data analysis framework for large-scale single-cell and single-nucleus RNA-seq, *bioRxiv* (2019).
81. Y. R. Peng, K. Shekhar, W. Yan, D. Herrmann, A. Sappington, G. S. Bryman, T. van Zyl, M. T. H. Do, A. Regev, J. R. Sanes, Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina, *Cell* **176**, 1222-1237.e22 (2019).
82. F. A. Wolf, P. Angerer, F. J. Theis, Open Access S CANPY : large-scale single-cell gene expression data analysis, , 1–5 (2018).
83. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data, *Cell Syst.* **8**, 281-291.e9 (2019).
84. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data, *Cell* **177**, 1888-1902.e21 (2019).
85. A. E. Moor, Y. Harnik, S. Ben-Moshe, E. E. Massasa, M. Rozenberg, R. Eilam, K. Bahar Halpern, S. Itzkovitz, Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis, *Cell* **175**, 1156-1167.e15 (2018).
86. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, *Nat. Biotechnol.* **36**, 411–420 (2018).
87. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression, *Genome Biol.* **20**, 1–15 (2019).
88. N. S. Kurd, Z. He, T. L. Louis, J. J. Milner, K. D. Omilusik, W. Jin, M. S. Tsai, C. E. Widjaja, J. N. Kanbar, J. G. Olvera, T. Tysl, L. K. Quezada, B. S. Boland, W. J. Huang, C. Murre, A. W. Goldrath, G. W. Yeo, J. T. Chang, Early precursors and molecular determinants of tissue-resident memory CD8+ T lymphocytes revealed by single-cell RNA sequencing, *Sci. Immunol.* **5**, 16–19 (2020).

Acknowledgements

We thank the participants who participated in this study; B. Mead, S. Nguyen, A. Rubin, and R. Xavier for insightful feedback and copy editing; Chola Mulenga for histological processing and morphometry; Rose Banda for recruitment of the Zambian volunteers and sample collection; and Joyce Sibwani and Rose Soko for expert endoscopy nursing. Figure 1b and the graphical abstract were generated with Biorender.com.

Funding

This work was supported, in part, by grants to PK from Barts & The London Charity (G-000907) and from the Bill & Melinda Gates Foundation (OPP1066118). A.K.S. was supported was supported, in part, by the Searle Scholars Program, the Beckman Young Investigator Program, Sloan Fellowship in Chemistry, the NIH (5U24AI118672), and the Bill and Melinda Gates Foundation. T.W. was supported by the NIH (DK007762). J.O.M is a New York Stem Cell Foundation – Robertson Investigator. J.O.M was supported by the Richard and Susan Smith Family Foundation, the HHMI Damon Runyon Cancer Research Foundation Fellowship (DRG-2274-16), the AGA Research Foundation's AGA-Takeda Pharmaceuticals Research Scholar Award in IBD – AGA2020-13-01, the HDDC Pilot and Feasibility P30 DK034854, the Food Allergy Science Initiative, and The New York Stem Cell Foundation. T.K.H. is supported by the NIH (F30-AI143160-01A1).

Author Contributions

C.K., T.W. T.K.H., Z.G., A.K.S., and P.K. conceived and designed experiments; C.K. performed tissue dissociation with assistance from T.K.H. J.O.M., and M.V.; C.K., T.K.H., T.W., S.M., I.F., and M.V. performed scRNA-seq; C.K. and N.M. analyzed the single-cell data; B.D. and J.O.M. developed code for iterative tiered clustering; V.M., E.B., and K.Z. performed histological analyses; J.T.B. coordinated duodenal resection samples; A.M.U. and J.J.G. coordinated distal duodenal samples from patients with EoE; C.K., T.W., T.K.H., J.O.M., Z.G., A.K.S., and P.K. interpreted the data; C.K., T.W., A.K.S., and P.K. wrote the manuscript with input from all authors.

Competing interests

J.O.M. reports compensation for consulting services with Cellarity and Hovione. A.K.S. reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, Repertoire Immune Medicines, Hovione, Third Rock Ventures, Ochre Bio, and Dahlia Biosciences unrelated to this work. A.K.S. has received research support from Merck, Novartis, Leo Pharma, Janssen, the Bill and Melinda Gates Foundation, the Moore Foundation, the Pew-Stewart Trust, Fondation MIT, the Chan Zuckerberg Initiative, Novo Nordisk and the FDA unrelated to this work. T.K.H. has served as a consultant and holds equity in nference, inc. The remaining authors disclose no conflicts.

Data availability

For all samples, the expression matrix and associated metadata can be visualized and downloaded from the Alexandria Project, a Bill and Melinda Gates Foundation-funded portal (part of the Single-Cell Portal hosted by The Broad Institute of MIT and Harvard): https://singlecell.broadinstitute.org/single_cell/study/SCP1307. The raw data for samples from patients in the EE and EoE cohorts will available in the Gene Expression Omnibus (GEO) at accession number GSE168883 The raw data for the duodenal resection samples will be available by contacting the authors and establishing a Data Use Agreement.

Abbreviations

Crypt depth (CD), Environmental enteropathy (EE), Eosinophilic esophagitis (EoE), Severe acute malnutrition (SAM), single-cell RNA-sequencing (scRNA-seq), Villus height (VH)

Figures

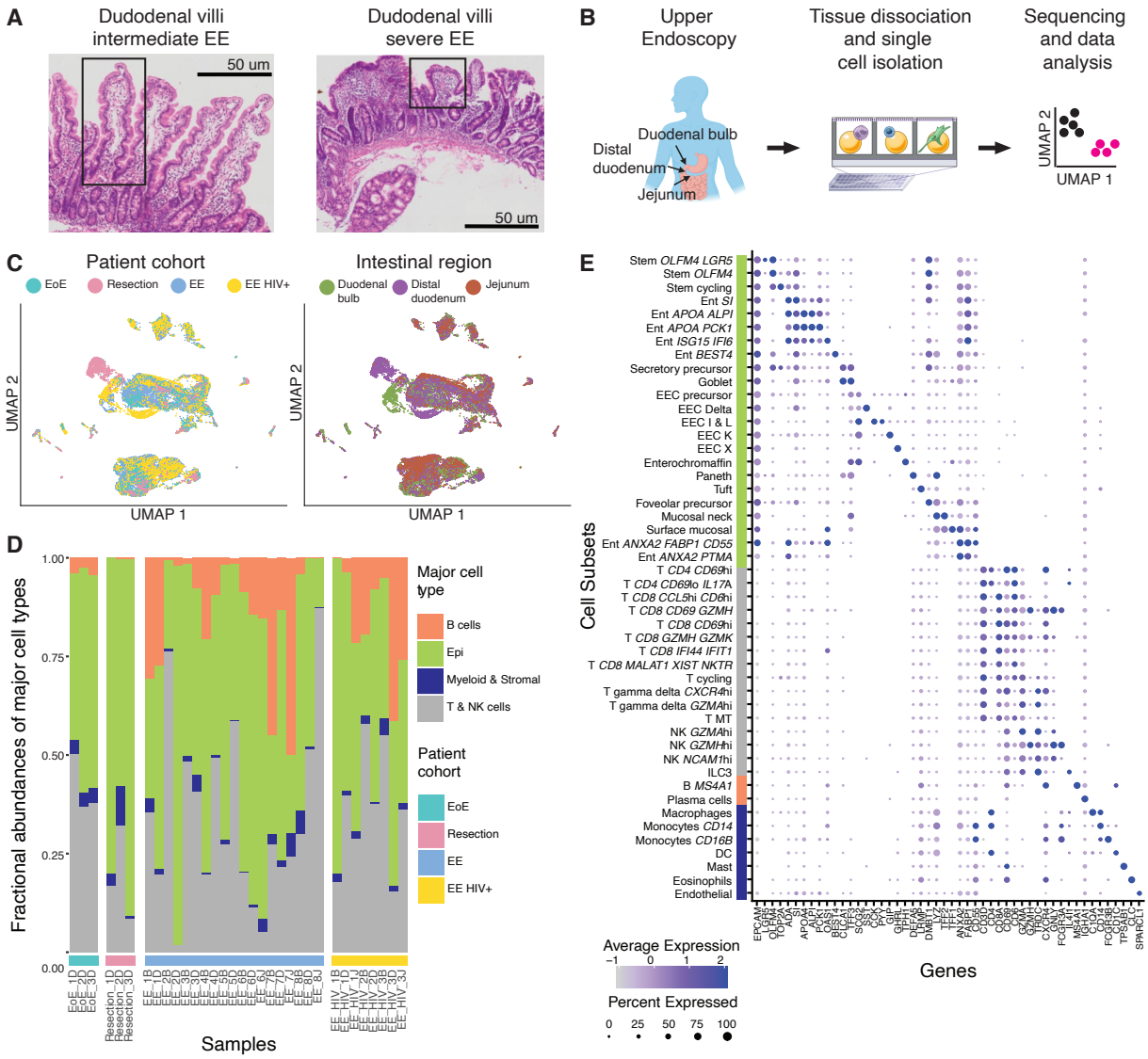

A
Dudodenal villi intermediate EE
50 um
Dudodenal villi severe EE
50 um
B
Upper Endoscopy
Tissue dissociation and single cell isolation
Sequencing and data analysis
Duodenal bulb
Distal duodenum
Jejunum
UMAP 2
UMAP 1
C
Patient cohort
EoE
Resection
EE
EE HIV+
Intestinal region
Duodenal bulb
Distal duodenum
Jejunum
UMAP 2
UMAP 1
D
Fractional abundances of major cell types
Major cell type
B cells
Epi
Myeloid & Stromal
T & NK cells
Patient cohort
EoE
Resection
EE
EE HIV+
Samples
E
Cell Subsets
Stem OLFM4 LGR5
Stem OLFM4
Stem cycling
Ent SI
Ent APOA ALPI
Ent APOA PCK1
Ent ISG15 IFI6
Ent BEST4
Secretory precursor
Goblet
EEC precursor
EEC Delta
EEC I & L
EEC K
EEC X
Enterochromaffin
Paneth
Tuft
Foveolar precursor
Mucosal neck
Surface mucosal
Ent ANXA2 FABP1 CD55
Ent ANXA2 PTMA
T CD4 CD69hi
T CD4 CD69lo IL17A
T CD8 CCL5hi CD6hi
T CD8 CD69 GZMH
T CD8 CD69hi
T CD8 GZMH GZMK
T CD8 IFI44 IFIT1
T CD8 MALAT1 XIST NKTR
T cycling
T gamma delta CXCR4hi
T gamma delta GZMAhi
T MT
NK GZMAhi
NK GZMHhi
NK NCAM1hi
ILC3
B MS4A1
Plasma cells
Macrophages
Monocytes CD14
Monocytes CD16B
DC
Mast
Eosinophils
Endothelial
Genes
Average Expression
Percent Expressed

Figure 1: Single-cell atlas of the upper small intestine with and without EE.

A, H&E imaging of the small intestine with intermediate and severe EE. Intestinal villi with normal morphology (left image) and with severe blunting (right image) are highlighted in boxes.

B, Experimental workflow: small intestinal biopsies from the duodenal bulb, distal duodenum, and jejunum were obtained via endoscopy or tissue resection, dissociated into cells, loaded onto a Seq-Well array, processed for single-cell sequencing, and analyzed.

C, UMAP visualization by patient cohort and intestinal region for all 23,545 cells from 33 samples and 17 patients.

D, Fractional abundances of major cell types amongst all single cells captured per sample.

E, Expression of marker genes for final cell subsets. Dot size represents the fraction of a cell subset (rows) expressing a given gene (columns). Dot hue represents the scaled average expression by gene column. For clarity, dots for genes expressed in 5% or less of cells within a given subset are not shown.

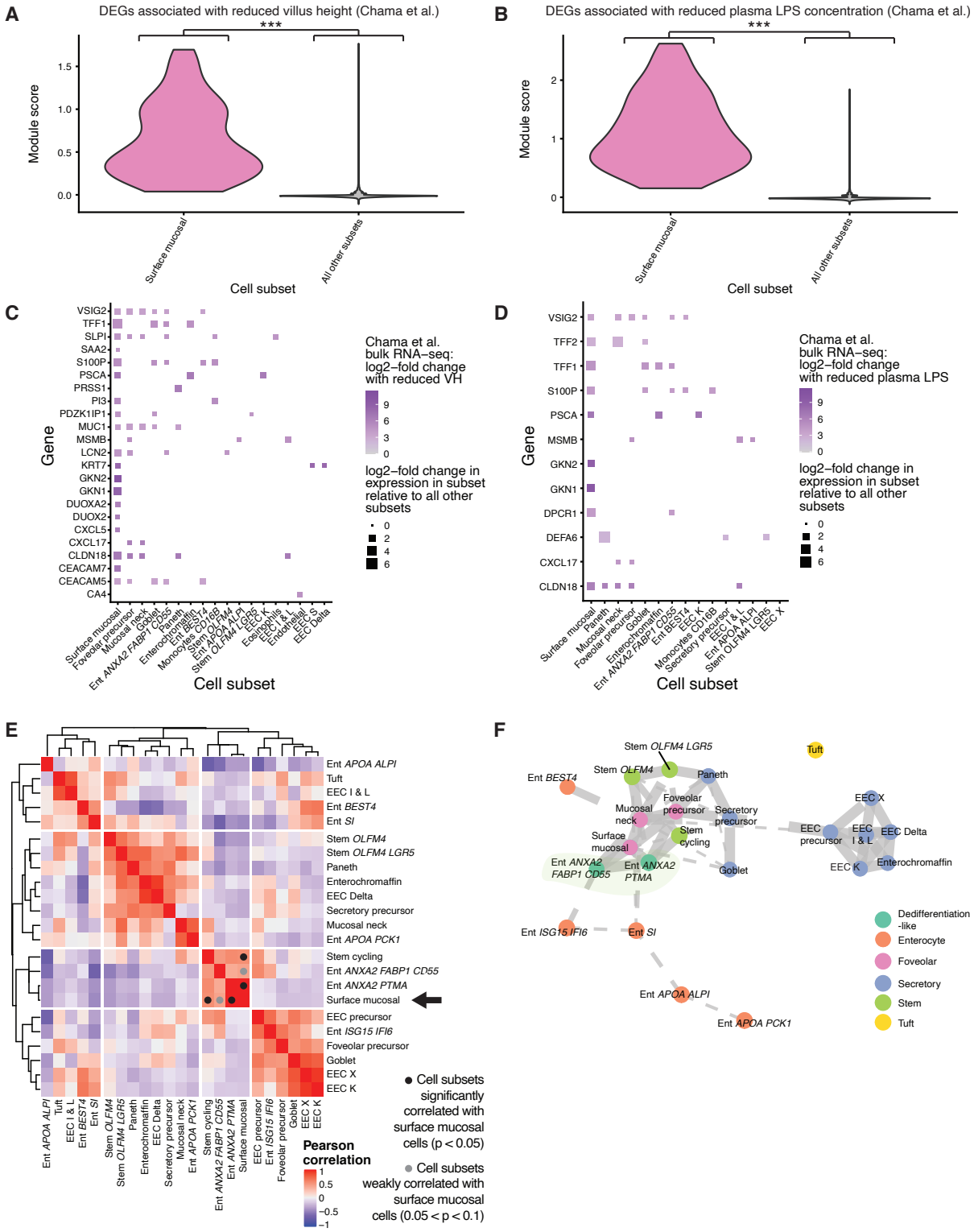

A
DEGs associated with reduced villus height (Chama et al.)
***
Module score
1.5
1.0
0.5
0.0
Surface mucosal
All other subsets
Cell subset
B
DEGs associated with reduced plasma LPS concentration (Chama et al.)
***
Module score
2
1
0
Surface mucosal
All other subsets
Cell subset
C
Gene
VSIG2
TFF1
SLPI
SAA2
S100P
PSCA
PRSS1
PI3
PDZK1IP1
MUC1
MSMB
LCN2
KRT7
GKN2
GKN1
DUOXA2
DUOX2
CXCL5
CXCL17
CLDN18
CEACAM7
CEACAM5
CA4
Surface mucosal
Foveolar precursor
Mucosal neck
Goblet
Ent ANXA2 FABP1 CD55
Enterochromaffin
Paneth
Ent BEST4
Monocytes CD16B
Stem OLFM4
Ent APOA ALPI
Stem OLFM4 LGR5
EEC K
Eosinophils
EEC I & L
Endothelial
EEC S
EEC Delta
Cell subset
Chama et al.
bulk RNA–seq:
log2–fold change
with reduced VH
9
6
3
0
log2–fold change in
expression in subset
relative to all other
subsets
0
2
4
6
D
Gene
VSIG2
TFF2
TFF1
S100P
PSCA
MSMB
GKN2
GKN1
DPCR1
DEFA6
CXCL17
CLDN18
Surface mucosal
Paneth
Mucosal neck
Foveolar precursor
Goblet
Enterochromaffin
Ent ANXA2 FABP1 CD55
Ent BEST4
EEC K
Monocytes CD16B
Secretory precursor
EEC I & L
Ent APOA ALPI
Stem OLFM4 LGR5
EEC X
Cell subset
Chama et al.
bulk RNA–seq:
log2–fold change
with reduced plasma LPS
9
6
3
0
log2–fold change in
expression in subset
relative to all other
subsets
0
2
4
6
E
Ent APOA ALPI
Tuft
EEC I & L
Ent BEST4
Ent SI
Stem OLFM4
Stem OLFM4 LGR5
Paneth
Enterochromaffin
EEC Delta
Secretory precursor
Mucosal neck
Ent APOA PCK1
Stem cycling
Ent ANXA2 FABP1 CD55
Ent ANXA2 PTMA
Surface mucosal
EEC precursor
Ent ISG15 IFI6
Foveolar precursor
Goblet
EEC X
EEC K
Pearson
correlation
1
0.5
0
−0.5
−1
● Cell subsets
significantly
correlated with
surface mucosal
cells (p < 0.05)
● Cell subsets
weakly correlated with
surface mucosal
cells (0.05 < p < 0.1)
F
Stem OLFM4 LGR5
Tuft
Ent BEST4
Stem OLFM4
Paneth
Foveolar
precursor
Secretory
precursor
Mucosal
neck
Stem
cycling
Surface
mucosal
Ent ANXA2
FABP1 CD55
Ent ANXA2
PTMA
Goblet
EEC X
EEC
precursor
EEC
I & L
EEC Delta
EEC K
Enterochromaffin
Ent ISG15 IFI6
Ent SI
Ent APOA ALPI
Ent APOA PCK1
Dedifferentiation
-like
Enterocyte
Foveolar
Secretory
Stem
Tuft

Figure 2: Comparison to past bulk RNA-sequencing EE signatures implicates surface mucosal cells expressing DUOX2 in epithelial remodelling in EE

A, Violin plot of a module score generated from genes differentially upregulated in bulk RNA-sequencing of samples with EE and reduced villus height (VH) in Chama et al.*(7)* Surface mucosal cells were enriched for this signature relative to all other subsets (***, $p < 0.001$, Wilcoxon test)

B, Violin plot of a module score generated from genes differentially upregulated in bulk RNA-sequencing of samples with EE and decreased plasma LPS concentrations in Chama et al.*(7)* Surface mucosal cells were enriched for this signature relative to all other subsets (***, $p < 0.001$, Wilcoxon test)

C, Dot plot of cell subset marker genes that overlapped with the genes used to generate the module scores in panel a

D, Dot plot of cell subset marker genes that overlapped with the genes used to generate the module scores in panel b.

E, Hierarchically clustered heatmap of the Pearson correlations between the fractional abundances of epithelial cells within duodenal bulb samples from patients with EE. Cell subsets significantly correlated with surface mucosal cells are highlighted with a black circle (Permutation testing, $p < 0.05$). Cell subsets weakly correlated with surface mucosal cells are highlighted with a grey circle (Permutation testing, $0.05 < p < 0.1$).

F, Partition-based graph abstraction (PAGA) trajectory visualization of epithelial subsets.

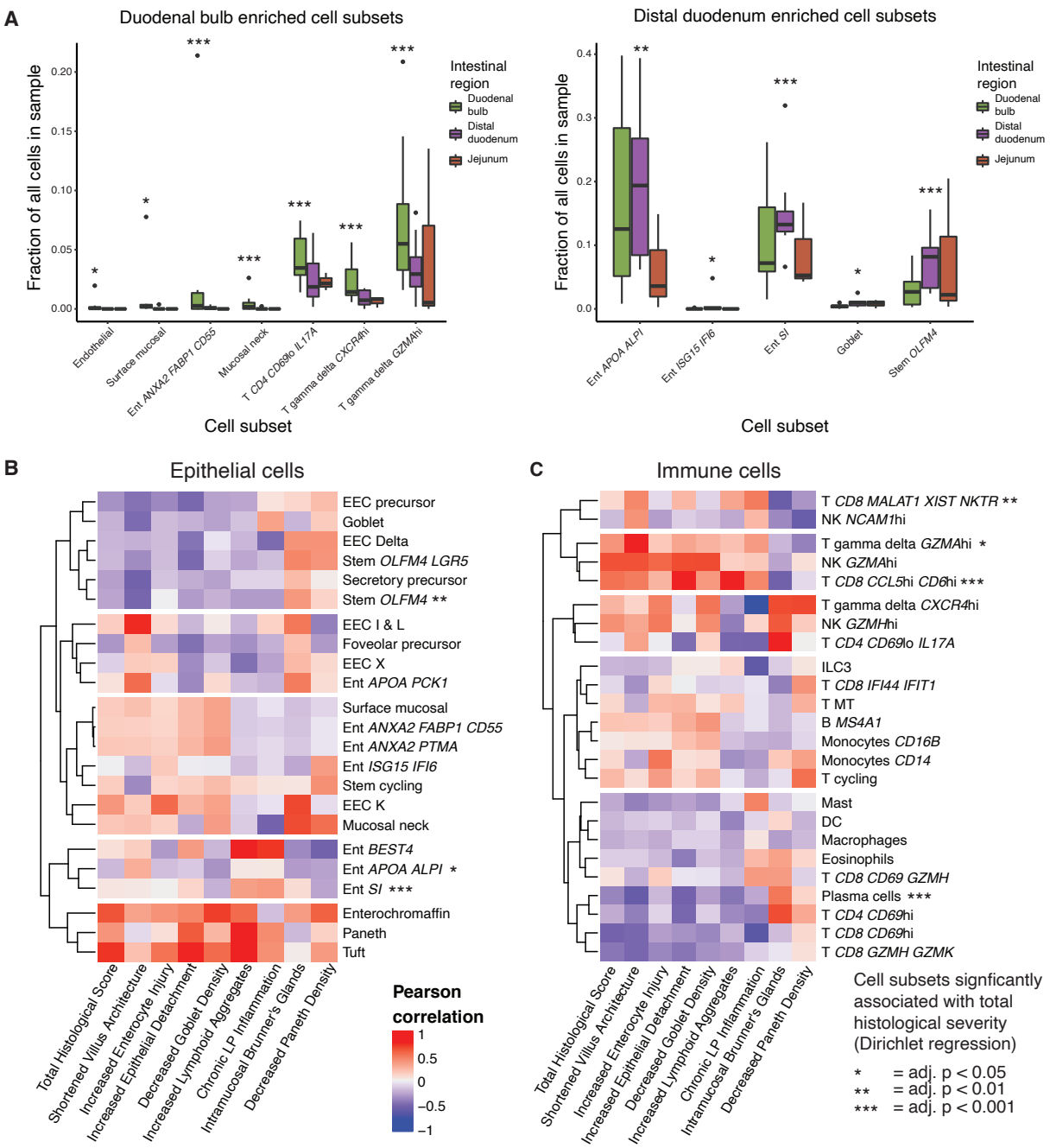

A
Duodenal bulb enriched cell subsets
Distal duodenum enriched cell subsets
Fraction of all cells in sample
Intestinal region
Duodenal bulb
Distal duodenum
Jejunum
Endothelial
Surface mucosal
Ent ANXA2 FABP1 CD55
Mucosal neck
T CD4 CD69lo IL17A
T gamma delta CXCR4hi
T gamma delta GZMAhi
Ent APOA ALPI
Ent ISG15 IFI6
Ent SI
Goblet
Stem OLFM4
Cell subset
B
Epithelial cells
EEC precursor
Goblet
EEC Delta
Stem OLFM4 LGR5
Secretory precursor
Stem OLFM4 **
EEC I & L
Foveolar precursor
EEC X
Ent APOA PCK1
Surface mucosal
Ent ANXA2 FABP1 CD55
Ent ANXA2 PTMA
Ent ISG15 IFI6
Stem cycling
EEC K
Mucosal neck
Ent BEST4
Ent APOA ALPI *
Ent SI ***
Enterochromaffin
Paneth
Tuft
Total Histological Score
Shortened Villus Architecture
Increased Enterocyte Injury
Increased Epithelial Detachment
Decreased Goblet Density
Increased Lymphoid Aggregates
Chronic LP Inflammation
Intramucosal Brunner's Glands
Decreased Paneth Density
Pearson correlation
1
0.5
0
−0.5
−1
C
Immune cells
T CD8 MALAT1 XIST NKTR **
NK NCAM1hi
T gamma delta GZMAhi *
NK GZMAhi
T CD8 CCL5hi CD6hi ***
T gamma delta CXCR4hi
NK GZMHhi
T CD4 CD69lo IL17A
ILC3
T CD8 IFI44 IFIT1
T MT
B MS4A1
Monocytes CD16B
Monocytes CD14
T cycling
Mast
DC
Macrophages
Eosinophils
T CD8 CD69 GZMH
Plasma cells ***
T CD4 CD69hi
T CD8 CD69hi
T CD8 GZMH GZMK
Cell subsets signficantly associated with total histological severity (Dirichlet regression)
* = adj. p < 0.05
** = adj. p < 0.01
*** = adj. p < 0.001

Figure 3: Identifying cell subsets associated with intestinal region and histologically determined EE severity within HIV-negative EE patients.

A, Cell subsets enriched in duodenal bulb and distal duodenal samples from HIV-negative EE patients (* adj. < 0.05, ** adj. p < 0.01, *** adj. p < 0.001). No subsets were enriched in jejunal samples from HIV-negative EE patients.

B, Hierarchically clustered heatmap of HIV-negative EE epithelial cell subset relative abundance Pearson correlations with component scores of the total EE histological severity score. In order to account for compositional dependencies in the data, significant relationships were determined with Dirichlet regression (*adj. p < 0.05, **adj. p < 0.01, ***adj. p < 0.001).

C, Hierarchically clustered heatmap of HIV-negative EE immune cell subset relative abundance Pearson correlations with component scores of the total EE histological severity score. Significant relationships determined with Dirichlet regression (* adj. p < 0.05, ** adj. p < 0.01, *** adj. p value < 0.001).

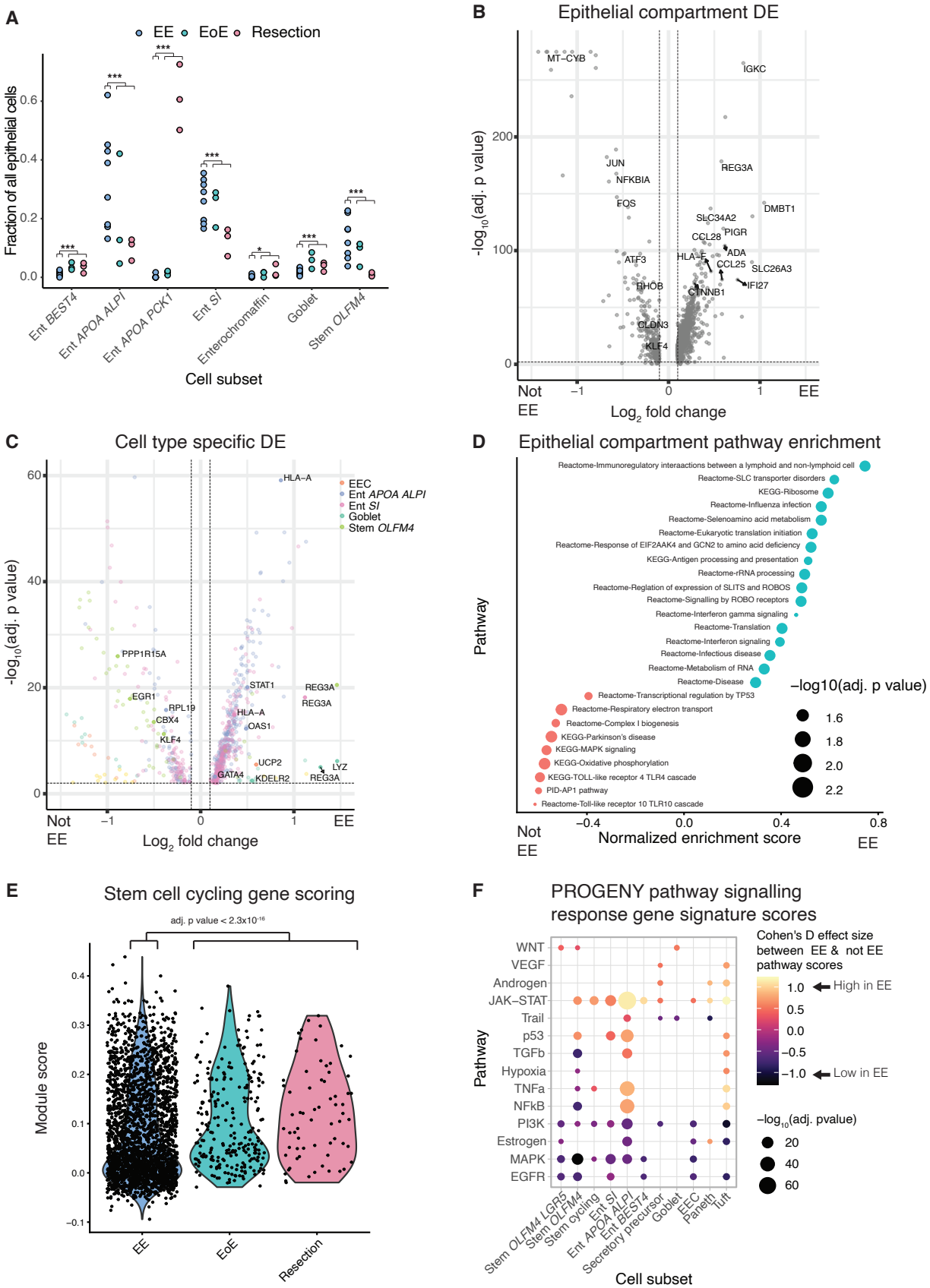

A
EE
EoE
Resection
Fraction of all epithelial cells
Ent BEST4
Ent APOA ALPI
Ent APOA PCK1
Ent SI
Enterochromaffin
Goblet
Stem OLFM4
Cell subset
B
Epithelial compartment DE
-log10(adj. p value)
MT−CYB
IGKC
JUN
NFKBIA
FOS
REG3A
DMBT1
SLC34A2
PIGR
CCL28
ADA
HLA−F
CCL25
SLC26A3
IFI27
ATF3
RHOB
CTNNB1
CLDN3
KLF4
Not EE
EE
Log2 fold change
C
Cell type specific DE
EEC
Ent APOA ALPI
Ent SI
Goblet
Stem OLFM4
HLA−A
PPP1R15A
STAT1
REG3A
EGR1
RPL19
CBX4
KLF4
OAS1
UCP2
LYZ
GATA4
KDELR2
D
Epithelial compartment pathway enrichment
Reactome-Immunoregulatory interaactions between a lymphoid and non-lymphoid cell
Reactome-SLC transporter disorders
KEGG-Ribosome
Reactome-Influenza infection
Reactome-Selenoamino acid metabolism
Reactome-Eukaryotic translation initiation
Reactome-Response of EIF2AAK4 and GCN2 to amino acid deficiency
KEGG-Antigen processing and presentation
Reactome-rRNA processing
Reactome-Reglation of expression of SLITS and ROBOS
Reactome-Signalling by ROBO receptors
Reactome-Interferon gamma signaling
Reactome-Translation
Reactome-Interferon signaling
Reactome-Infectious disease
Reactome-Metabolism of RNA
Reactome-Disease
Reactome-Transcriptional regulation by TP53
Reactome-Respiratory electron transport
Reactome-Complex I biogenesis
KEGG-Parkinson's disease
KEGG-MAPK signaling
KEGG-Oxidative phosphorylation
KEGG-TOLL-like receptor 4 TLR4 cascade
PID-AP1 pathway
Reactome-Toll-like receptor 10 TLR10 cascade
Pathway
−log10(adj. p value)
Normalized enrichment score
E
Stem cell cycling gene scoring
adj. p value < 2.3x10-16
Module score
EE
EoE
Resection
F
PROGENY pathway signalling response gene signature scores
WNT
VEGF
Androgen
JAK−STAT
Trail
p53
TGFb
Hypoxia
TNFa
NFkB
PI3K
Estrogen
MAPK
EGFR
Pathway
Cohen's D effect size between EE & not EE pathway scores
High in EE
Low in EE
−log10(adj. pvalue)
Stem OLFM4 LGR5
Stem OLFM4
Stem cycling
Ent SI
Ent APOA ALPI
Ent BEST4
Secretory precursor
Goblet
EEC
Paneth
Tuft
Cell subset

Figure 4: The epithelium of EE is characterized by increased interferon response and decreased proliferative signaling.

A, Cell subsets with significant shifts in relative abundances between EE and control cohorts. (Fisher's exact test, * adj. $p < 0.05$, ** adj. $p < 0.01$, *** adj. $p < 0.001$).

B, Genes differentially expressed in the epithelial compartment in EE relative to controls. Horizontal and vertical dashed lines respectively refer to an adjusted p value threshold of 0.01 and a log fold change threshold of 0.1.

C, Genes differentially expressed in EE relative to controls within specific cellular subsets. Horizontal and vertical dashed lines respectively refer to an adjusted p value threshold of 0.01 and a log fold change threshold of 0.1.

D, Gene set enrichment analysis of genes upregulated in epithelial compartment cells

E, Module scores for cell cycle genes in EE and control cohorts in all cells from stem cell subsets. EE stem cells displayed lower module scores (adj. $p < 2.3*10^{-16}$. Cohen's D effect size = 0.47)

F, PROGENy pathway prediction scores for epithelial cells in EE relative to controls

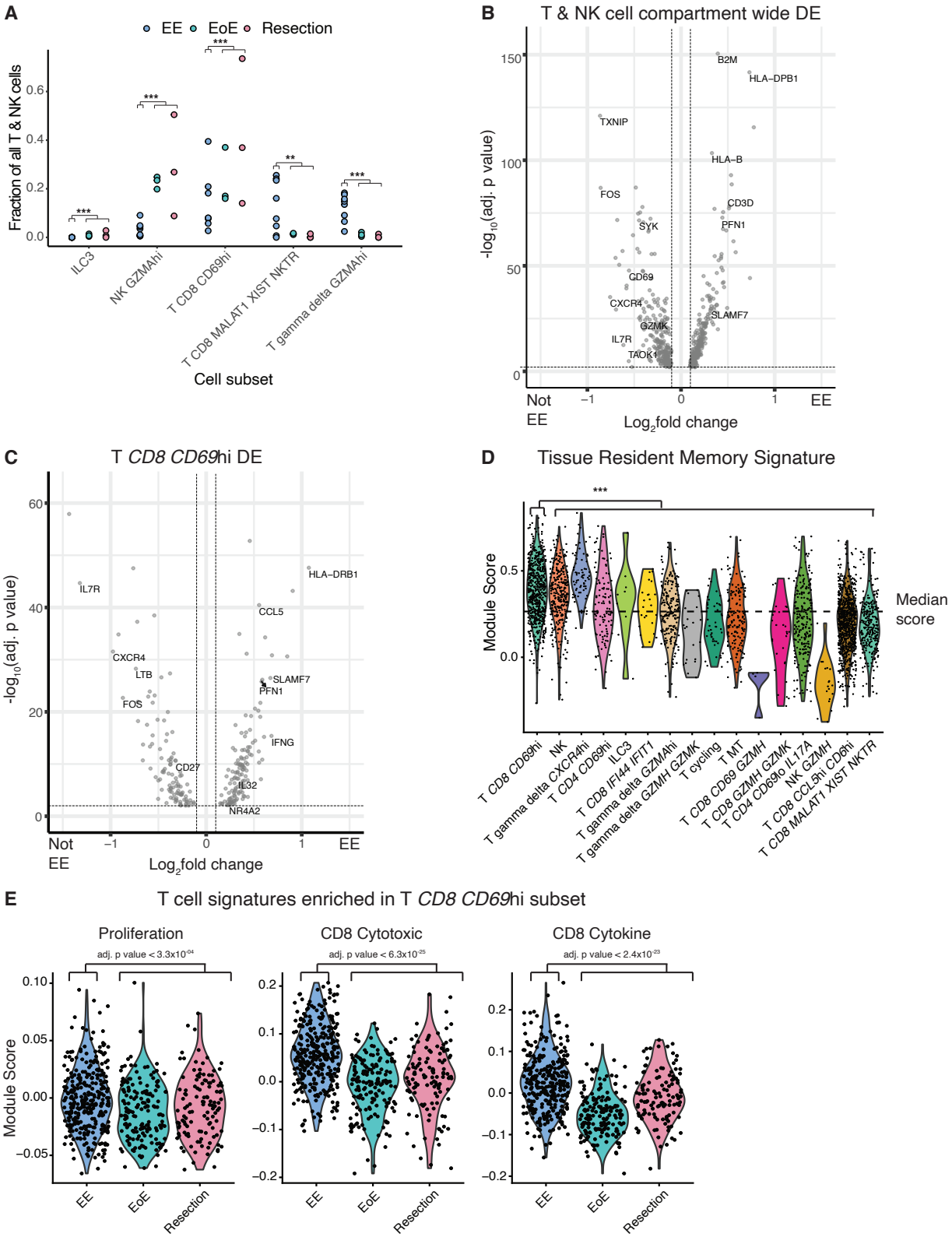

A
EE
EoE
Resection
Fraction of all T & NK cells
0.0
0.2
0.4
0.6
***
**
ILC3
NK GZMAhi
T CD8 CD69hi
T CD8 MALAT1 XIST NKTR
T gamma delta GZMAhi
Cell subset
B
T & NK cell compartment wide DE
-log10(adj. p value)
150
100
50
0
B2M
HLA−DPB1
TXNIP
HLA−B
FOS
CD3D
SYK
PFN1
CD69
CXCR4
SLAMF7
GZMK
IL7R
TAOK1
Not EE
−1
0
1
EE
Log2fold change
C
T CD8 CD69hi DE
-log10(adj. p value)
60
40
20
0
IL7R
HLA−DRB1
CCL5
CXCR4
LTB
SLAMF7
PFN1
FOS
IFNG
CD27
IL32
NR4A2
Not EE
−1
0
1
EE
Log2fold change
D
Tissue Resident Memory Signature
***
Module Score
0.5
0.0
Median score
T CD8 CD69hi
NK
T gamma delta CXCR4hi
T CD4 CD69hi
ILC3
T CD8 IFI44 IFIT1
T gamma delta GZMAhi
T gamma delta GZMH GZMK
T cycling
T MT
T CD8 CD69 GZMH
T CD8 GZMH GZMK
T CD8 GZMH IL17A
T CD4 CD69lo
NK GZMH
T CD8 CCL5hi CD6hi
T CD8 MALAT1 XIST NKTR
E
T cell signatures enriched in T CD8 CD69hi subset
Proliferation
adj. p value < 3.3x10-04
CD8 Cytotoxic
adj. p value < 6.3x10-25
CD8 Cytokine
adj. p value < 2.4x10-23
Module Score
0.10
0.05
0.00
−0.05
0.2
0.1
0.0
−0.1
−0.2
EE
EoE
Resection

Figure 5: EE is associated with a shift towards activated T cell phenotypes

A, Cell subsets with significant shifts in relative abundances between EE and control cohorts (Fischer's exact test, * adj. $p < 0.05$, ** adj. $p < 0.01$, *** adj. $p < 0.001$).

B, Genes differentially expressed in the T & NK cell compartment in EE relative to controls. Horizontal and vertical dashed lines respectively refer to an adjusted p value threshold of 0.01 and a log fold change threshold of 0.1.

C, Genes differentially expressed in in EE relative to controls within specific cellular subsets. Horizontal and vertical dashed lines respectively refer to an adjusted p value threshold of 0.01 and a log fold change threshold of 0.1.

D, Module scores for a tissue resident memory T cell signature in EE relative to controls.

E, T cell activation signatures enriched in T *CD8 CD69*^{hi} cells from EE patients relative to controls. Proliferation: adj. $p = 3.2 * 10^{-04}$, Cohen's D effect size =0.32. CD8 Cytotoxic: adj. $p = 6.2 * 10^{-25}$, Cohen's D effect size =0.95. CD8 Cytokine: adj. $p = 2.3 * 10^{-23}$, Cohen's D effect size =0.90

Supplementary Figures

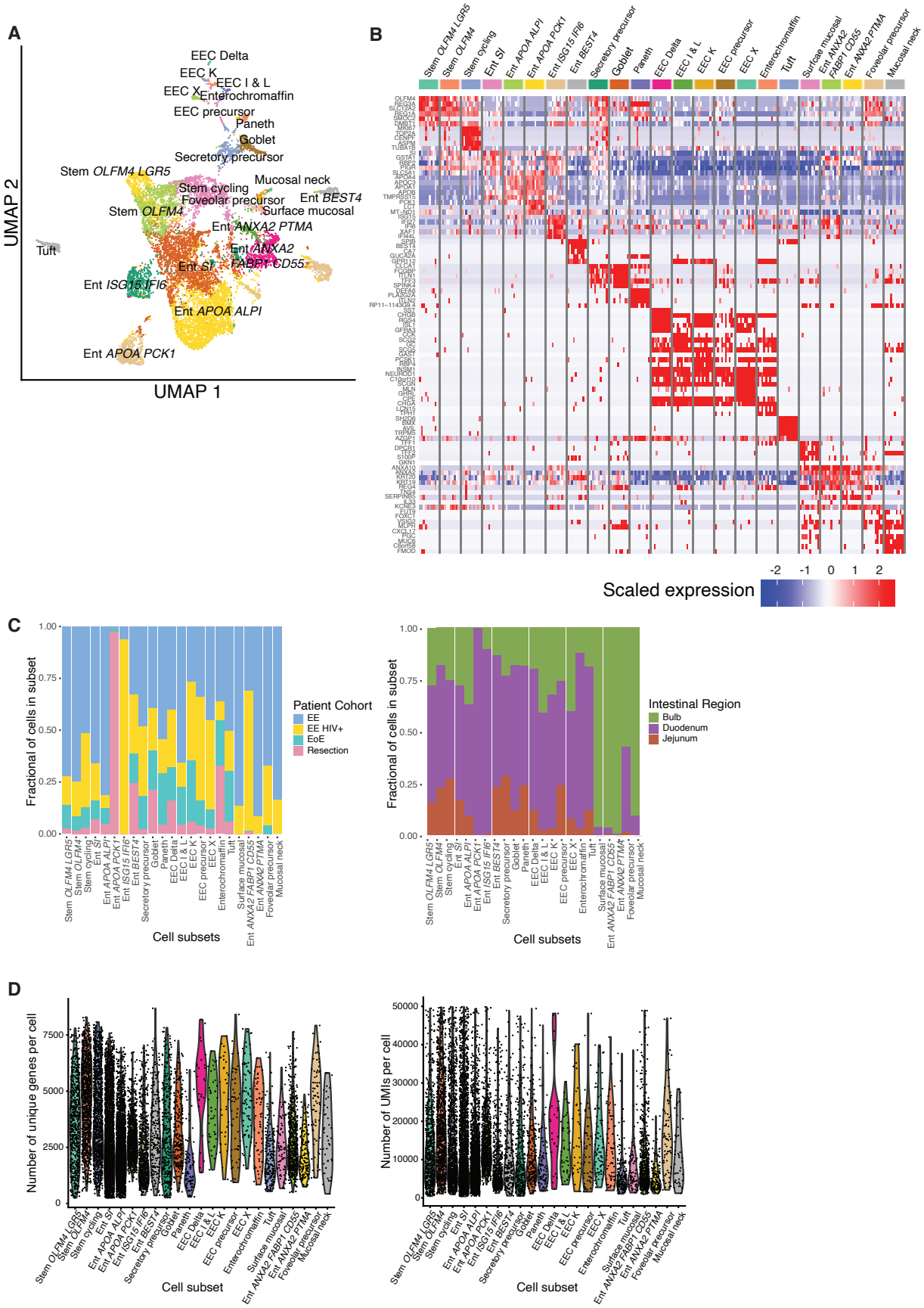

A
UMAP 2
UMAP 1
EEC Delta
EEC K
EEC I & L
EEC X
Enterochromaffin
EEC precursor
Paneth
Goblet
Secretory precursor
Stem OLFM4 LGR5
Mucosal neck
Stem cycling
Foveolar precursor
Ent BEST4
Stem OLFM4
Surface mucosal
Ent ANXA2 PTMA
Tuft
Ent ANXA2
Ent SI
FABP1 CD55
Ent ISG15 IFI6
Ent APOA ALPI
Ent APOA PCK1
B
Stem OLFM4 LGR5
Stem OLFM4
Stem cycling
Ent SI
Ent APOA ALPI
Ent APOA PCK1
Ent ISG15 IFI6
Ent BEST4
Secretory precursor
Goblet
Paneth
EEC Delta
EEC I & L
EEC K
EEC precursor
EEC X
Enterochromaffin
Tuft
Surface mucosal
Ent ANXA2 FABP1 CD55
Ent ANXA2 PTMA
Foveolar precursor
Mucosal neck
Scaled expression
-2
-1
0
1
2
C
Fractional of cells in subset
Cell subsets
Patient Cohort
EE
EE HIV+
EoE
Resection
Intestinal Region
Bulb
Duodenum
Jejunum
D
Number of unique genes per cell
Number of UMIs per cell
Cell subset

Figure S1: Characterization of epithelial cell subsets revealed by single-cell RNA-sequencing.

A, UMAP visualization of epithelial subsets

B, Heatmap of the top five marker genes (Wilcoxon test) of each epithelial subset

C, Fraction of cells in each subset from each patient cohort; from HIV-positive vs. HIV-negative patients; and from each intestinal region.

D, Violin plots by cell subset of the number of unique molecular identifiers (UMIs) per cell and the number of unique genes detected per cell.

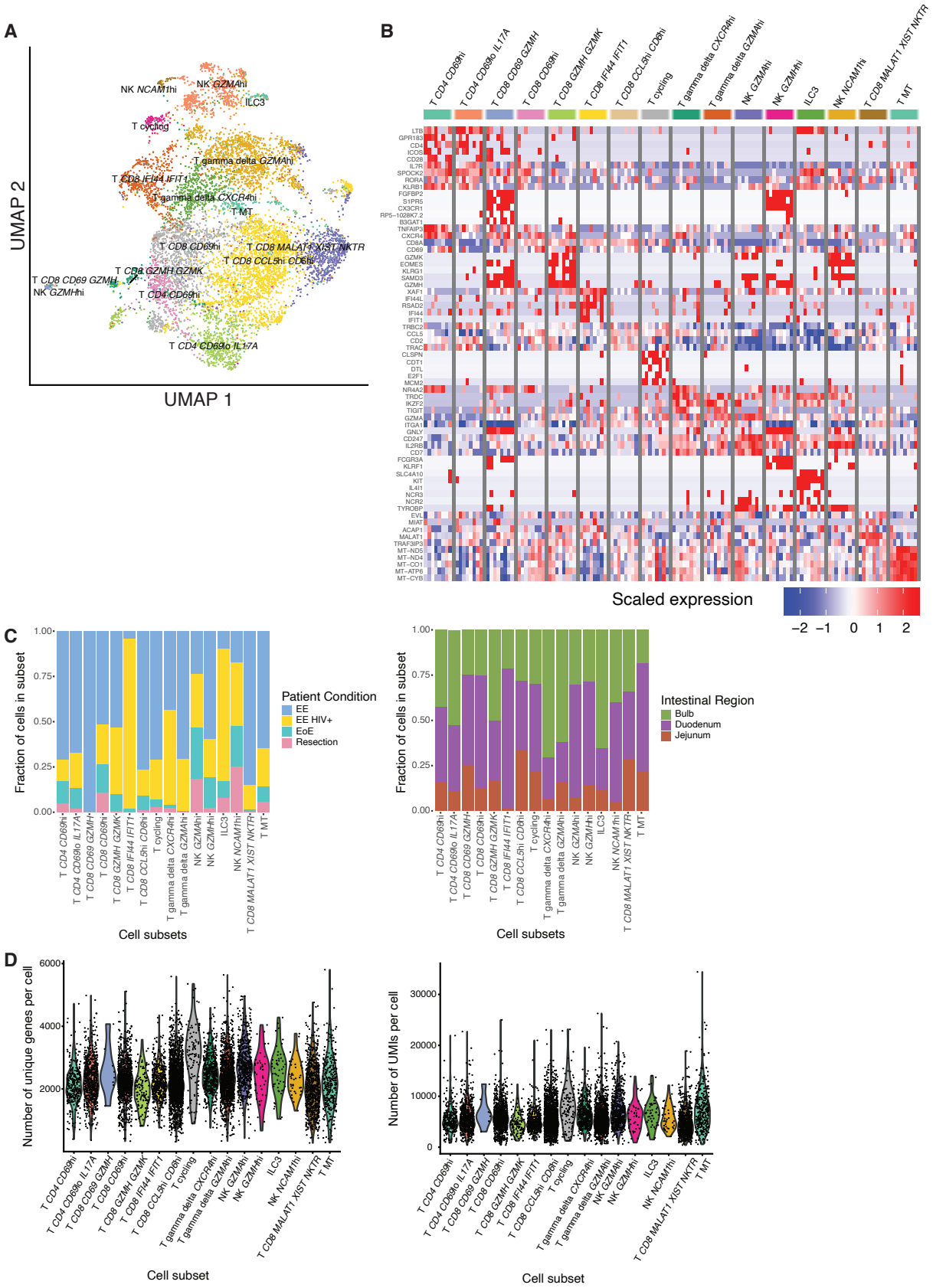

A
UMAP 2
UMAP 1
NK NCAM1hi
NK GZMAhi
ILC3
T cycling
T gamma delta GZMAhi
T CD8 IFI44 IFIT1
T gamma delta CXCR4hi
T MT
T CD8 CD69hi
T CD8 MALAT1+ XIST NKTR
T CD8 CCL5hi CD8hi
T CD8 GZMH GZMK
T CD8 CD69 GZMH
NK GZMHhi
T CD4 CD69hi
T CD4 CD69lo IL17A
B
Scaled expression
−2 −1 0 1 2
C
Fraction of cells in subset
Patient Condition
EE
EE HIV+
EoE
Resection
Cell subsets
Intestinal Region
Bulb
Duodenum
Jejunum
D
Number of unique genes per cell
Number of UMIs per cell
Cell subset

Figure S2: Characterization of T and NK cell subsets revealed by single-cell RNA-sequencing.

A, UMAP visualization of epithelial subsets

B, Heatmap of the top five marker genes (Wilcoxon test) of each T and NK cell subset

C, Fraction of cells in each subset from each patient cohort; from HIV-positive vs. HIV-negative patients; and from each intestinal region.

D, Violin plots by cell subset of the number of unique molecular identifiers (UMIs) per cell and the number of unique genes detected per cell.

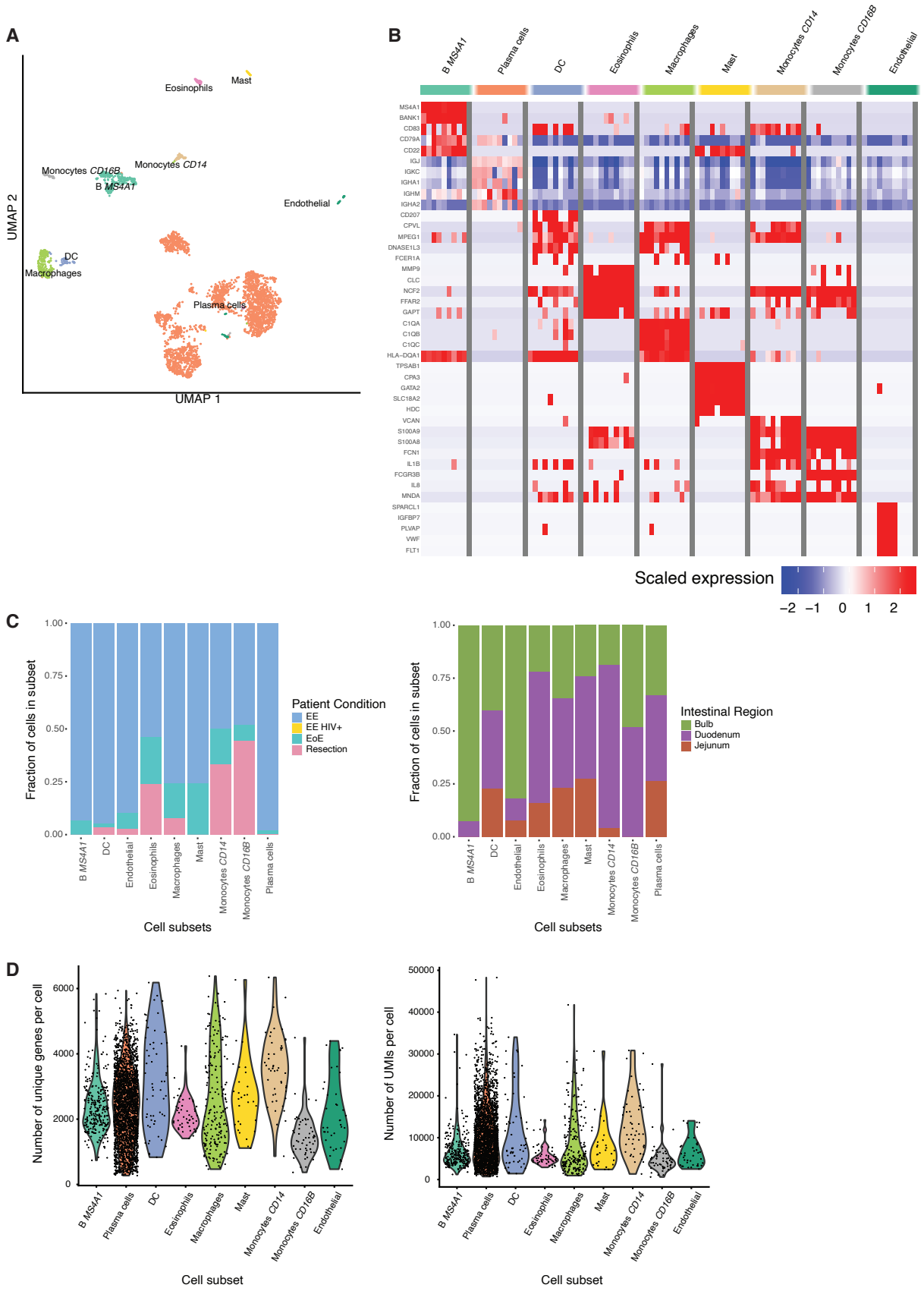

A
Eosinophils
Mast
Monocytes CD16B
Monocytes CD14
B MS4A1
Endothelial
DC
Macrophages
Plasma cells
UMAP 2
UMAP 1
B
B MS4A1
Plasma cells
DC
Eosinophils
Macrophages
Mast
Monocytes CD14
Monocytes CD16B
Endothelial
MS4A1
BANK1
CD83
CD79A
CD22
IGJ
IGKC
IGHA1
IGHM
IGHA2
CD207
CPVL
MPEG1
DNASE1L3
FCER1A
MMP9
CLC
NCF2
FFAR2
GAPT
C1QA
C1QB
C1QC
HLA-DQA1
TPSAB1
CPA3
GATA2
SLC18A2
HDC
VCAN
S100A9
S100A8
FCN1
IL1B
FCGR3B
IL8
MNDA
SPARCL1
IGFBP7
PLVAP
VWF
FLT1
Scaled expression
−2 −1 0 1 2
C
Fraction of cells in subset
1.00
0.75
0.50
0.25
0.00
Patient Condition
EE
EE HIV+
EoE
Resection
Cell subsets
Intestinal Region
Bulb
Duodenum
Jejunum
D
Number of unique genes per cell
6000
4000
2000
0
Number of UMIs per cell
50000
40000
30000
20000
10000
0
Cell subset

Figure S3: Characterization of B cell, myeloid, and stromal subsets revealed by single-cell RNA-sequencing.

A, UMAP visualization of epithelial subsets

B, Heatmap of the top five marker genes (Wilcoxon test) of each B cell, myeloid, and stromal cell subset

C, Fraction of cells in each subset from each patient cohort; from HIV-positive vs. HIV-negative patients; and from each intestinal region.

D, Violin plots by cell subset of the number of unique molecular identifiers (UMIs) per cell and the number of unique genes detected per cell.

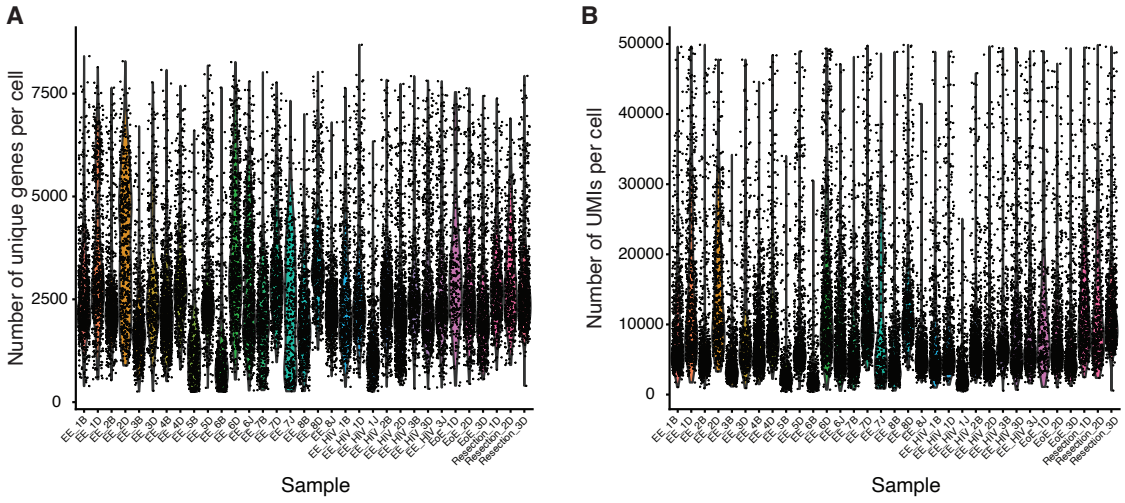

A
Number of unique genes per cell
7500
5000
2500
0
Sample
B
Number of UMIs per cell
50000
40000
30000
20000
10000
0
Sample

Figure S4: Number of genes and UMIs per cell across samples

A, Violin plot of number of genes per cell across samples

B, Violin plot of number of UMIs per cell across samples

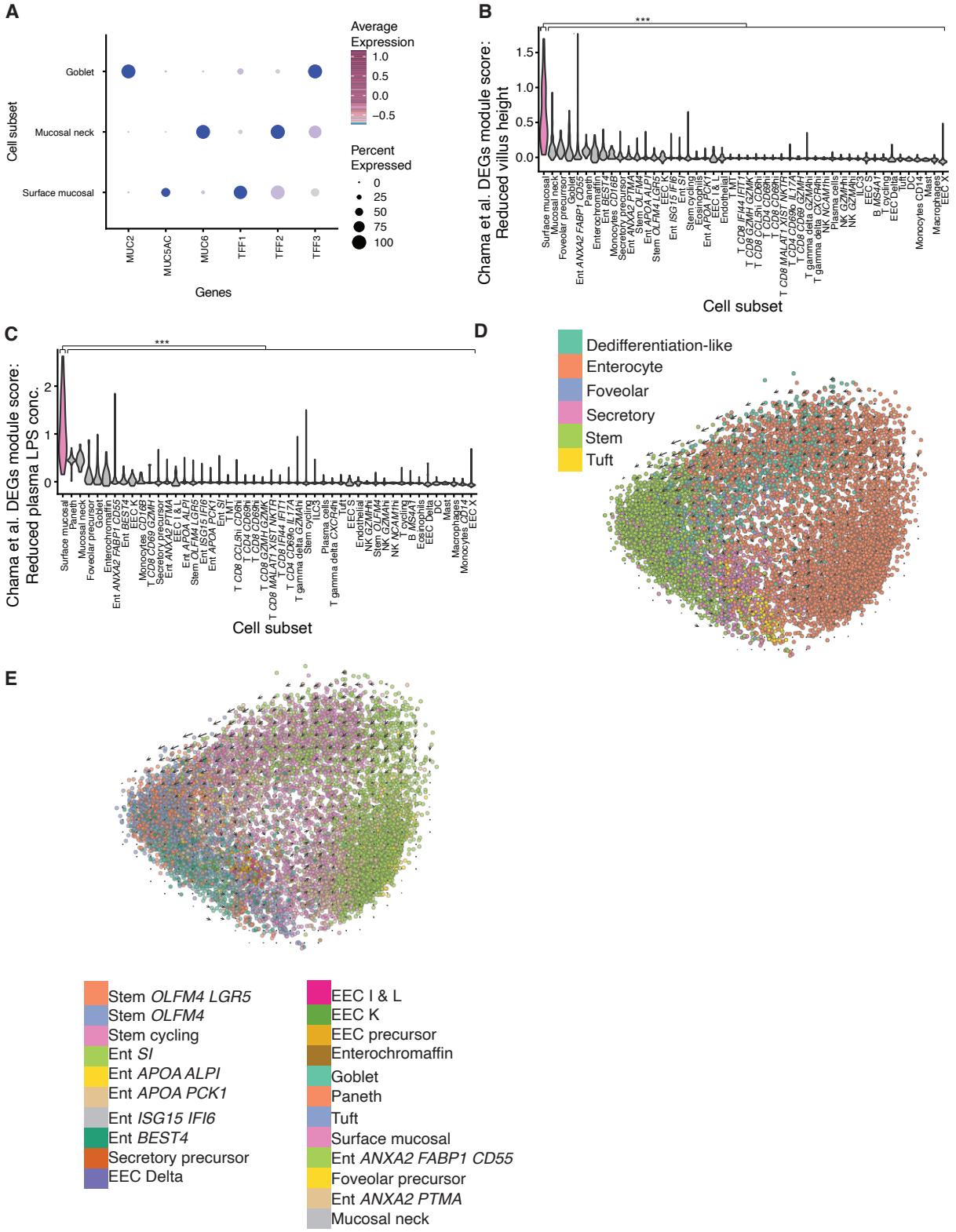

A
Average Expression
1.0
0.5
0.0
−0.5
Percent Expressed
0
25
50
75
100
Cell subset
Goblet
Mucosal neck
Surface mucosal
MUC2
MUC5AC
MUC6
TFF1
TFF2
TFF3
Genes
B
Chama et al. DEGs module score: Reduced villus height
***
Cell subset
C
Chama et al. DEGs module score: Reduced plasma LPS conc.
***
Cell subset
D
Dedifferentiation-like
Enterocyte
Foveolar
Secretory
Stem
Tuft
E
Stem OLFM4 LGR5
Stem OLFM4
Stem cycling
Ent SI
Ent APOA ALPI
Ent APOA PCK1
Ent ISG15 IFI6
Ent BEST4
Secretory precursor
EEC Delta
EEC I & L
EEC K
EEC precursor
Enterochromaffin
Goblet
Paneth
Tuft
Surface mucosal
Ent ANXA2 FABP1 CD55
Foveolar precursor
Ent ANXA2 PTMA
Mucosal neck

Figure S5: Characterization of surface mucosal and dedifferentiation-like subsets

A, Expression of mucin and trefoil factor genes distinguishing Surface mucosal, mucosal neck, and goblet cell subsets. Dot size represents the fraction of a cell subset (rows) expressing a given gene (columns). Dot hue represents the scaled average expression by gene column.

B, Violin plot scoring all subsets on a module score generated from genes differentially upregulated in bulk RNA-sequencing of samples with EE and reduced villus height (VH) in Chama et al. Surface mucosal cells were enriched for this signature relative to all other subsets (***, $p < 0.001$, Wilcoxon test)

C, Violin plot scoring all subsets on a module score generated from genes differentially upregulated in bulk RNA-sequencing of samples with EE and decreased plasma LPS concentrations in Chama et al. Surface mucosal cells were enriched for this signature relative to all other subsets (***, $p < 0.001$, Wilcoxon test)

D, Velocyto results grouped by epithelial differentiation trajectory identified in Figure 2e

E, Velocyto results grouped by epithelial subset

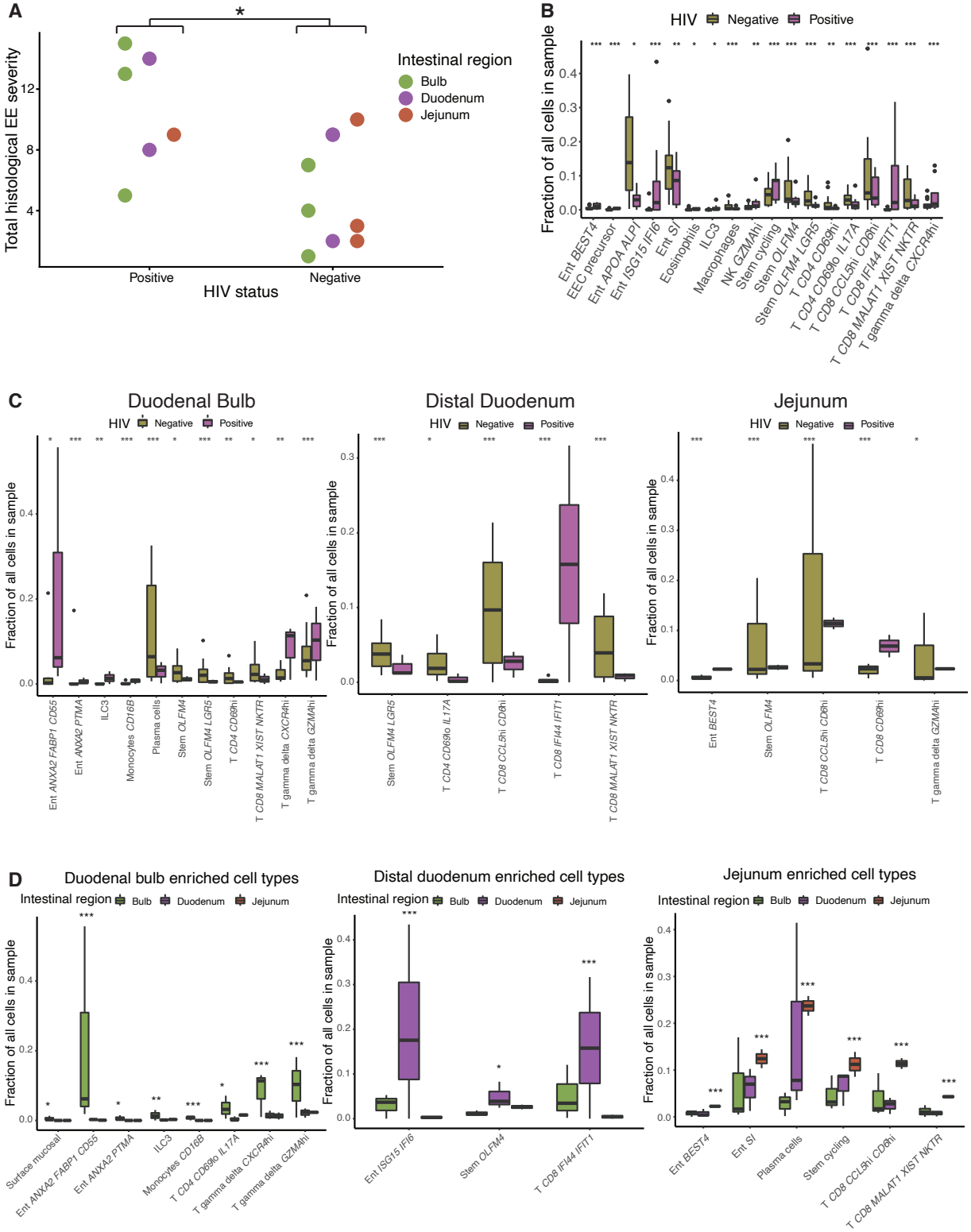

A
Total histological EE severity
HIV status
Positive
Negative
Intestinal region
Bulb
Duodenum
Jejunum
B
HIV
Negative
Positive
Fraction of all cells in sample
Ent BEST4
EEC precursor
Ent APOA ALPI
Ent ISG15 IFI6
Ent SI
Eosinophils
ILC3
Macrophages
NK GZMAhi
Stem cycling
Stem OLFM4
Stem OLFM4 LGR5
T CD4 CD69hi
T CD4 CD69lo IL17A
T CD8 CCL5hi CD6hi
T CD8 IFI44 IFIT1
T CD8 MALAT1 XIST NKTR
T gamma delta CXCR4hi
C
Duodenal Bulb
Distal Duodenum
Jejunum
Ent ANXA2 FABP1 CD55
Ent ANXA2 PTMA
ILC3
Monocytes CD16B
Plasma cells
Stem OLFM4
Stem OLFM4 LGR5
T CD4 CD69hi
T CD8 MALAT1 XIST NKTR
T gamma delta CXCR4hi
T gamma delta GZMAhi
T CD4 CD69lo IL17A
T CD8 CCL5hi CD6hi
T CD8 IFI44 IFIT1
Ent BEST4
D
Duodenal bulb enriched cell types
Distal duodenum enriched cell types
Jejunum enriched cell types
Intestinal region
Bulb
Duodenum
Jejunum
Surface mucosal
Ent ANXA2 FABP1 CD55
Ent ANXA2 PTMA
ILC3
Monocytes CD16B
T CD4 CD69lo IL17A
T gamma delta CXCR4hi
T gamma delta GZMAhi
Ent ISG15 IFI6
Stem OLFM4
T CD8 IFI44 IFIT1
Ent BEST4
Ent SI
Plasma cells
Stem cycling
T CD8 CCL5hi CD6hi
T CD8 MALAT1 XIST NKTR

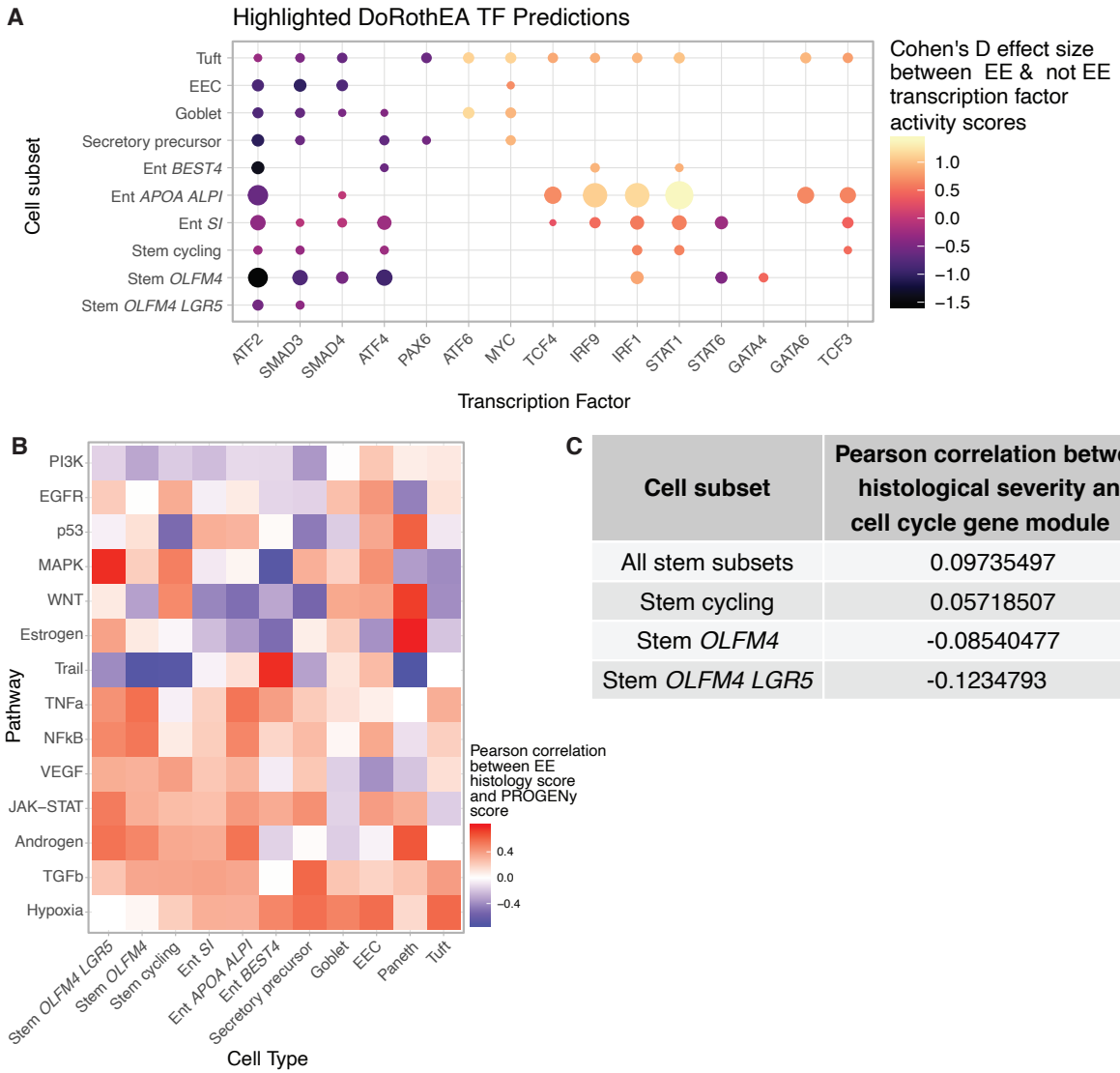
Figure S6: Variation in EE biology associated with HIV infection

A, EE HIV-positive samples had significantly higher total EE histological severity scores than EE HIV-negative samples (Wilcoxon test, $p < 0.05$).

B, Cell subsets with a significant change in fractional abundance between HIV-positive and HIV-negative samples across all EE samples (*adj. $p < 0.05$, **adj. $p < 0.01$, ***adj. $p < 0.001$).

C, Cell subsets with a significant change in fractional abundance between HIV-positive and HIV-negative samples within EE samples from each intestinal region (*adj. $p < 0.05$, **adj. $p < 0.01$, ***adj. $p < 0.001$).

D, Within only HIV-positive EE samples, cell subsets with a significant change increase in fractional abundance in each intestinal region (*adj. $p < 0.05$, **adj. $p < 0.01$, ***adj. $p < 0.001$).



C

Cell subset	Pearson correlation between histological severity and cell cycle gene module
All stem subsets	0.09735497
Stem cycling	0.05718507
Stem *OLFM4*	-0.08540477
Stem *OLFM4 LGR5*	-0.1234793

Figure S7: Additional characterization of epithelial signaling changes

A, Transcription factors with DoRothEA predicted activity varying between EE and controls (adj. $p < 0.05$).

B, Heatmap of Pearson correlations between PROGENY pathway scores and EE histological severity among EE patients.

C, Pearson correlation between EE histological severity and proliferation module score within stem cell subsets from EE patients.

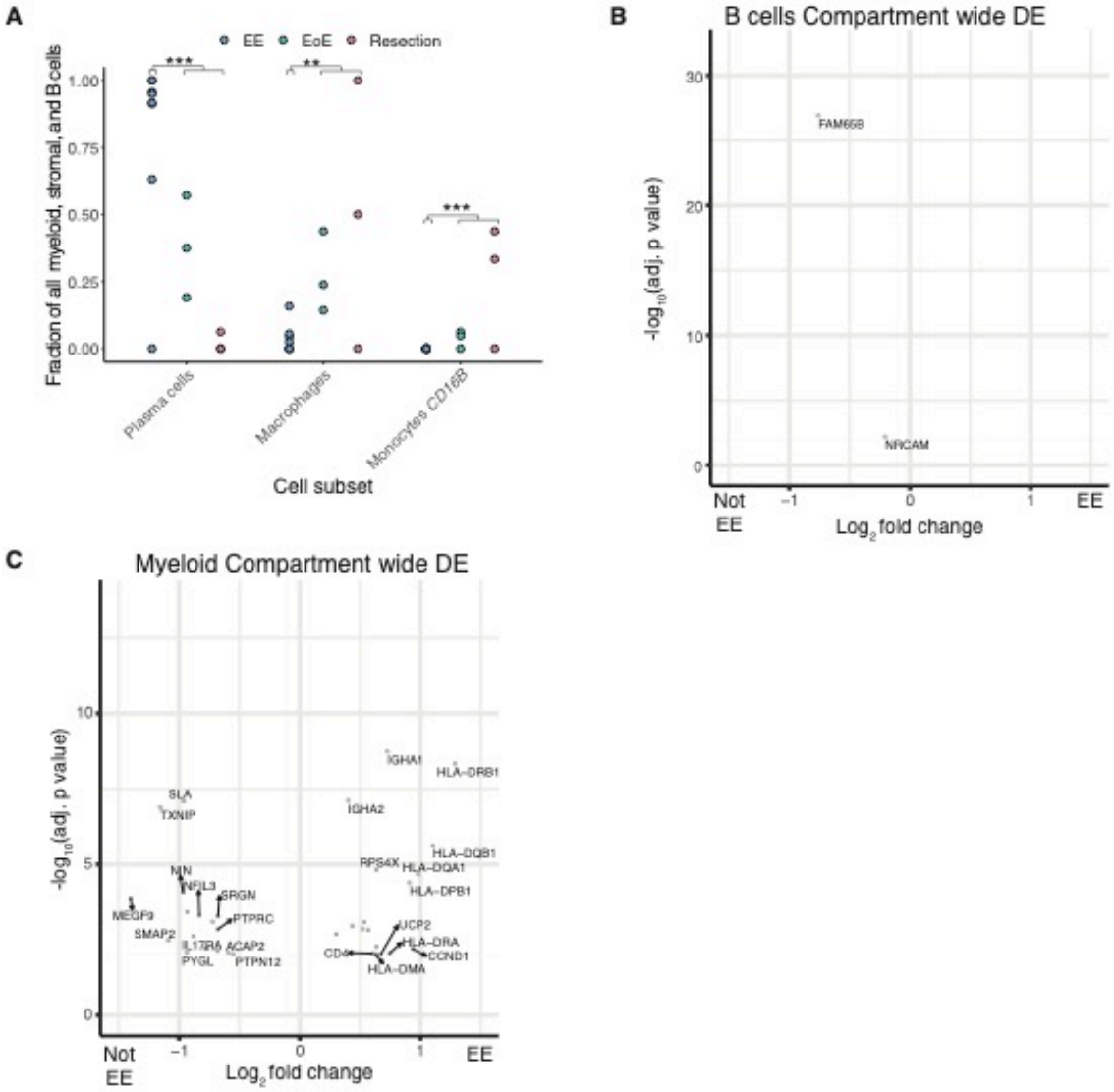

A
EE
EoE
Resection
Fraction of all myeloid, stromal, and B cells
1.00
0.75
0.50
0.25
0.00
***
**
***
Plasma cells
Macrophages
Monocytes CD16B
Cell subset
B
B cells Compartment wide DE
-log10(adj. p value)
30
20
10
0
FAM65B
NRCAM
Not EE
−1
0
1
EE
Log2 fold change
C
Myeloid Compartment wide DE
-log10(adj. p value)
10
5
0
IGHA1
HLA-DRB1
SLA
TXNIP
IGHA2
HLA-DQB1
RPS4X
HLA-DQA1
NN
NFIL3
SRGN
HLA-DPB1
MEGP9
PTPRC
SMAP2
UCP2
IL17RA
ACAP2
HLA-DRA
PYGL
PTPN12
CD4
CCND1
HLA-DMA
Not EE
−1
0
1
EE
Log2 fold change

Figure S8: Cell subset relative abundance and differential expression changes in B cells and myeloid cells between EE and controls

A, Cell subsets with significant shifts in relative abundances between EE and control cohorts (Fischer's exact test, *adj. $p < 0.05$, **adj. $p < 0.01$, ***adj. $p < 0.001$).

B, Genes differentially expressed in all B cells in EE relative to controls.

C, Genes differentially expressed in all myeloid cells in EE relative to controls.

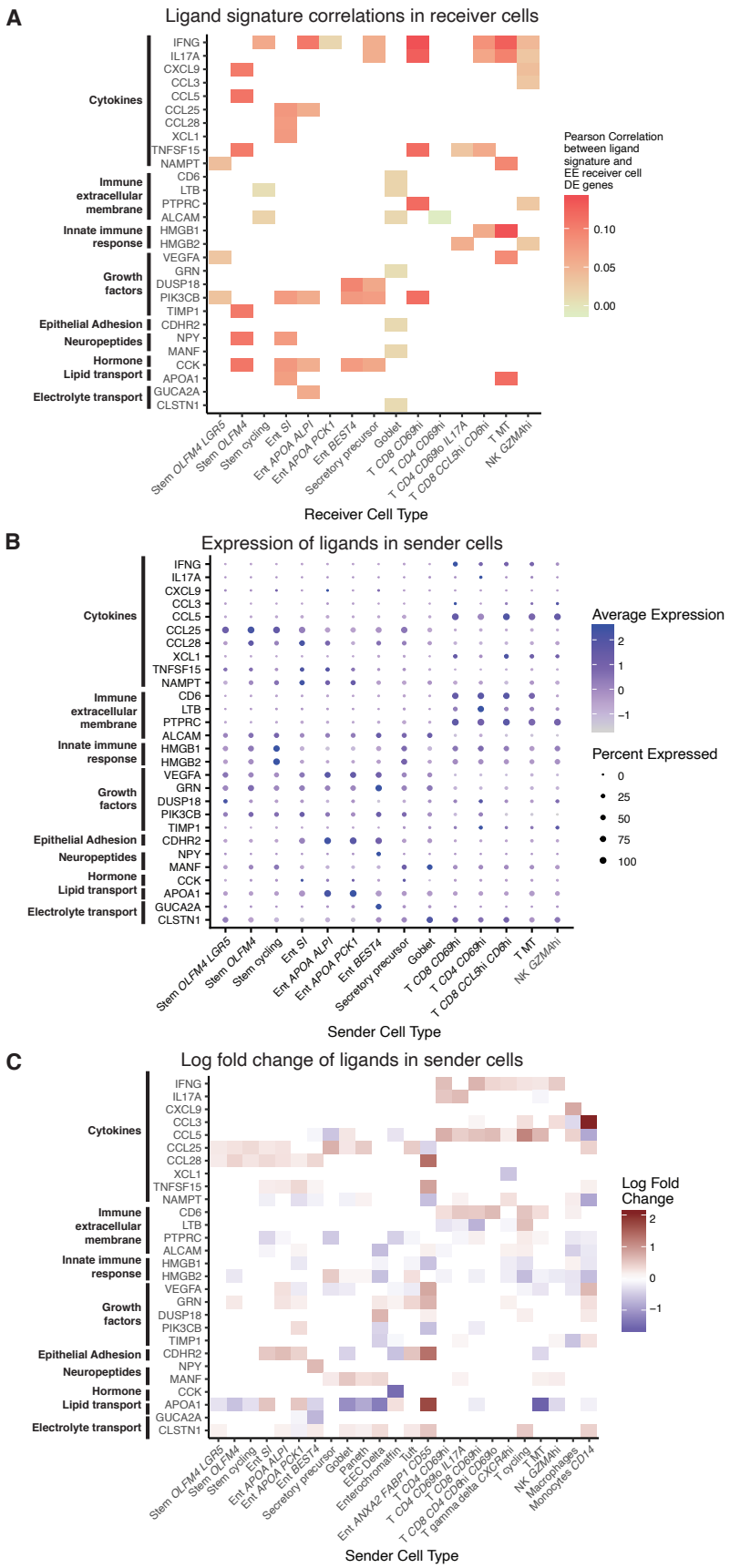

A
Ligand signature correlations in receiver cells
Cytokines
Immune extracellular membrane
Innate immune response
Growth factors
Epithelial Adhesion
Neuropeptides
Hormone
Lipid transport
Electrolyte transport
IFNG
IL17A
CXCL9
CCL3
CCL5
CCL25
CCL28
XCL1
TNFSF15
NAMPT
CD6
LTB
PTPRC
ALCAM
HMGB1
HMGB2
VEGFA
GRN
DUSP18
PIK3CB
TIMP1
CDHR2
NPY
MANF
CCK
APOA1
GUCA2A
CLSTN1
Stem OLFM4 LGR5
Stem OLFM4
Stem cycling
Ent SI
Ent APOA ALPI
Ent APOA PCK1
Ent BEST4
Secretory precursor
Goblet
T CD8 CD69hi
T CD4 CD69hi
T CD4 CD69lo IL17A
T CD8 CCL5hi CD6hi
T MT
NK GZMAhi
Receiver Cell Type
Pearson Correlation between ligand signature and EE receiver cell DE genes
0.10
0.05
0.00
B
Expression of ligands in sender cells
Average Expression
2
1
0
−1
Percent Expressed
0
25
50
75
100
Sender Cell Type
C
Log fold change of ligands in sender cells
Log Fold Change
2
1
0
−1
Paneth
EEC Delta
Enterochromaffin
Tuft
Ent ANXA2 FABP1 CD55
T CD4 CD69lo IL17A
T CD8 CD69hi
T CD8 CD4 CD69hi CD69lo
T gamma delta CXCR4hi
T cycling
Macrophages
Monocytes CD14

Figure S9: NicheNet analysis of cell-cell signaling in EE relative to controls

A, Pearson correlation between ligand NicheNet ligand response signatures and the differential gene expression signature in each subset.

B, Expression levels of the genes for NicheNet ligands in all cell subsets.

C, Log-fold change of the NicheNet ligands in all cell subsets.

SUPPLEMENTARY MATERIALS AND METHODS

Tissue digestion for scRNA-seq

Single-cell suspensions were obtained using a modified version of a previously published protocol*(22)*, described below in detail. Biopsies were rinsed in 30 mL of ice cold HBSS, then transferred to 10 mL of epithelial cell solution (HBSS, 10 mM EDTA, Pen / Strep, 10 mM HEPES, 2% FCS) and incubated for 15 minutes at 37 degrees Celsius at 120 rpm. Samples were transferred to sit on ice for 10 minutes and then vortexed for 10 seconds, after which the tissue was rinsed in 30 mL HBSS before being transferred to1 mL of epithelial digestion mix (2% FCS, 10 mM HEPES. Pen/Strep, 20 ug/ml gentamicin, 100 ug/mL liberase TM, 100 ug/mL DNAse I) in a 1.5 mL Eppendorf tube and spun down at 800g for 2 minutes and resuspended in TrypLE express enzyme [ThermoFisher 12604013] for 5 minutes at 37 degrees Celsius followed by gentle trituration with a P1000 pipette and spun down at 800 g for 2 minutes. Next, after aspiration, the pellet was resuspended in ACK lysis buffer [ThermoFisher A1049201] and put on ice for 3 minutes. Cells were then spun down at 800g for 2 minutes and resuspended in 1 mL of epithelial cell solution and placed on ice for 3 minutes before triturating with a P1000 pipette. Cells were then filtered into a new Eppendorf tube through a 40 uM cell filter [Falcon/VMR 21008-949]. The epithelial cell fraction–which passed through the filter– was spun down at 800g for 2 minutes and then resuspended in 200 mL and put on ice, while the tissue remaining on the filter was added to 5 mL of enzymatic digestion mix at 37 degrees Celsius for 30 minutes at 120 rpm, after which it was quenched with 80 uL of 0.5M EDTA) and placed on ice for 5 minutes. Samples were typically fully dissociated at this step and after gentle trituration with a P1000 pipette filtered through a new 40 uM strainer into a new 50 mL conical tube and rinsed with HBSS to 30 mL total volume, and spun down at 400g for 10 minutes. All but 1 mL was then aspirated off of the 50 mL conical, after which the pellet was resuspended in this 1 mL, which was then transferred to a 1.5 mL Eppendorf. This was spun down at 800g for 2 minutes and then resuspended in 1 mL of ACK lysis butter and put on ice for 3 minutes. Cells were then spun down at 800 g for 2 minutes and resuspended in 200 mL of epithelial cell solution and placed on ice, produced the lamina propria cell fraction. The lamina propria and epithelial fractions were then counted using a hemocytometer.

Single-cell RNA-seq with Seq-Well S^3 experimental details

We used Seq-Well S^3 (Ref. *(20)*), for our single-cell profiling, for which full methods are available on the Shalek Lab website (www.shaleklab.com). Briefly, 16,000 cells (8,000 from the epithelial fraction from tissue digestion and 8,000 from the lamina propria fraction from tissue digestion) were loaded onto a functionalized polydimethylsiloxane (PDMS) array preloaded with ~80,000 uniquely-barcoded mRNA capture beads (Chemgenes; MACOSKO-2011-10). The array was then sealed with a hydroxylated polycarbonate membrane with pore sizes of 10 nm, thereby enabling buffer exchange (and thereby cellular lysis and mRNA transcript hybridization to beads) while retaining biological molecules confined within each well. Beads were then removed from each well, reverse transcribed, treated with Exonuclease I (New England Biolabs; M0293M), and mixed with 0.1 M NaOH for 5 minutes at room temperature to denature the mRNA-cDNA hybrid product on the bead. Second strand synthesis was then performed with a PCR mastermix (40 uL 5x maxima RT buffer, 80 uL 30% PEG8000 solution, 20uL 10 mM DNTPS, 2uL 1 mM dn-SMART oligo, 5 uL Klenow Exo-, and 53 ul of DI ultrapure water) which was incubated with the beads for 1 hour at 37 ºC with end over end rotation. PCR amplification was then preformed using KAPA HiFi PCR Mix (Kapa Biosystems KK2602). After PCR amplification whole transcriptome products were isolated via two rounds of SPRI purification using Ampure Spri beads (Beckman Coulter, Inc.) at both 0.6x and 0.8x volumetric ratio. Sequencing libraries were then generated using the Nextera Tagmentation method on a total of 800 pg of pooled cDNA library per sample. This product was then purified via two rounds of SPRI at 0.6 and 0.8x ratios, producing library size with an average distribution of 500-750 basepairs, as determined using the Aligent hsD1000 Screen Tape System (Aligent Genomics). Arrays were sequenced on an Illumina Next-Seq at the Ragon Institute. The read structure was paired end, with read 1 starting at a custom read 1 primer containing 21 bases with a 12 bp cell barcode and an 8 bp unique molecular identifier (UMI) and read 2 consisting of 50 bases containing transcript information. Sequencing read alignment and demultiplexing was performed on the cumulus platform*(80)* using snapshot 6 of the Drop-seq pipeline previously described in*(52)*, resulting in a cell barcode by UMI digital gene expression (DGE) matrix.

Data QC and clustering overview

Prior to clustering, the DGE matrices were filtered to remove cellular barcodes with less than 250 unique genes, with more than 50,000 UMIs, with over 50% of UMIs mapping to mitochondrial

genes, and identified as doublets via the Scrublet algorithm*(83)* (which was ran separately for each Seq-Well array). Due to the multiple intestinal regions sampled and multiple patient cohorts in this study, different samples contained different distributions of cell types. In particular, unique subsets of cells were present in duodenal bulb samples, in samples from HIV-positive individuals, and in samples from uninvolved tissue from patients with pancreatic cancer. We tested the Seurat V3 integration approach*(84)* to see if we could correct for batch effects between samples while preserving differences driven by biology (i.e. differences driven by the region of the tissue sampled or the disease state of the patient). Initial testing revealed that integration incorrectly merged biologically distinct cell types in the duodenal bulb with secretory lineage cells from the second part of the duodenum and the jejunum. Furthermore, integration removed differences in clustering driven by HIV status that were seen without integration, indicating that integration may have been over correcting the dataset and minimizing biological differences driven by disease biology. Thus, rather than running an integration method for batch correction, we adopted an existing pipeline for automated iterative clustering of single-cell data (pipeline described in detail in the next methods section) that has been shown to identify batch effects without collapsing distinct rare cell types*(26, 81)*.

Workflow for iterative clustering

1. We ran the clustering pipeline on the full filtered DGE matrix and manually identified base clusters with co-expression of known mutually exclusive cellular lineage markers and removed these base clusters that correspond to doublet populations.
2. We ran the clustering pipeline separately on the epithelial, T and NK cell, B cell, myeloid, and stromal cellular compartments, generating a hierarchical tree of clusters for each compartment.
3. Known lineage specific genes*(19, 22, 28)* were used to annotate cell types on the clustering tree. The genes used to annotate each subset are available alongside the genes differentially upregulated in each subset in Supplementary Table 4. If sub clusters of a cluster did not have lineage defining genes as marker genes, then the cell type designation was made at the parent cluster. In over 95%, cell type designations were made at the first or second level of clustering, and lower tiers of clusters (representing transcriptional differences within a cell type) were not annotated as cell types. In some cases, multiple sub clusters expressed

marker genes for one cell type and were all given that cell type designation. For example, within an epithelial cluster composed of secretory lineage cells, two clusters that highly expressed *TFF3* and *CLCA1* were merged together as one final goblet cell type.

4. Finally, in one instance, we performed batch correction by identifying populations of cells with high expression of known lineage markers that formed separate sub clusters in the hierarchical clustering tree due to transcriptional differences between disease states and merging the clusters with common high expression of a lineage marker back together. This correction was only carried out on stem cells highly expressing *OLFM4* and *LGR5*, where the clustering pipeline split a tier 1 cluster of stem cells by disease specific expression changes in the AP-1 signalling pathway at tier 2 before splitting by the co-expressed stem marker genes *LGR5* and *SMOC2* at tier 3, producing two separate *LGR5*^{hi} sub clusters at tier 3 with differing levels of AP-1 signalling activation. These two sub clusters were merged together as the Stem *OLFM4*^{hi} *LGR5*^{hi} cell type. Furthermore, although *APOA1* and *APOA4* are markers for mature enterocytes near the villus tips*(85)*, we chose to not merge a cluster of enterocytes predominantly from pancreatic cancer resection samples that exhibited high expression of *APOA1* and *APOA4* with a different cluster of enterocytes expressing *APOA1* and *APOA4* as enterocyte sub types lie on a transcriptional gradient and these two clusters displayed distinctly different expression of *ALPI*–another marker for mature villus enterocytes*(85)*– which raised the possibility that these enterocyte clusters were biologically distinct. We thus did not merge these cell types.

Iterative clustering pipeline

Starting with our filtered DGE matrix, cells were iteratively clustered with the following unbiased pipeline built using the Seurat R pacakge*(86)* and adopted from*(26, 81)*. For each starting cluster, normalization and variable gene selection were conducted with SCtransform*(87)*, then dimensional reduction was carried out by running PCA. For starting clusters with less than 500 cells, the number of PCs was chosen with the JackStraw function in Seurat, while for runtime considerations, for starting clusters with more than 500 cells, the number of PCs was chosen as the elbow on the PCA variance explained plot. Cellular neighbours were calculated using a k.param of 0.5*(number of cells)^0.5. Next clustering was carried out at 40 resolutions evenly spaced between 0.2 to 0.8, and the average silhouette width across cells was calculated for each resolution.

After calculating a histogram of the average silhouette scores with (num resolutions)/1.2 bins, the resolution in the top bin of silhouette scores with the smallest average silhouette score in that bin was chosen as the final silhouette score, thereby creating a bias towards under clustering the data. Next, marker genes were calculated for each sub cluster generated with the final resolution using the Wilcoxon test with pre-test thresholds that each gene must have an average log fold change of at least 0.025 between clusters and be expressed in at least 0.2 percent of cells in the cluster that it is a marker for. For each sub cluster, this clustering pipeline was iteratively repeated until sub clusters were reached that did not have at least 10 marker genes with at least an adjusted p value of 0.01 and an average log fold change of 0.2. Altogether, this pipeline generated a hierarchical tree of clusters which then used in conjunction with known lineage markers to annotate cell types as described above.

Details on module scores for reduced villus height and reduced plasma LPS associated genes

Gene signatures for reduced villus height and decreased and plasma LPS signatures in Chama et al.*(7)* were obtained by aggregating all genes positively differentially expressed in these conditions in this study. Module scores were generated with the AddModuleScore function in Seurat with default parameters. Cell subsets were then scored for the modules, a Wilcoxon test was used to assess significance, and effect sizes were calculated with Cohen's D.

Identifying cell subsets varying with histological EE severity across all HIV-negative patients in the Zambian cohort

Cell subsets significantly associated with histological severity from the Zambian cohort were identified by running Dirichlet Regression, which allows for testing for differences in cell subsets along a continuous dependent variable, with the R package DirichletReg. Across all HIV-negative samples, relative abundances were regressed against the total EE histological severity score and small intestinal region, and an adjusted p value was generated for the association between EE severity and the relative abundance of each subset. Other cell subsets co-varying with the significantly associated subsets from Dirichlet regression were visualized by calculating the Pearson correlation between all subset relative abundances across patients and then hierarchically clustering the resulting correlations with Ward's method using the ComplexHeatmap R package. To identify cell subsets significantly associated with surface mucosal cells, sample labels were

permuted 10,000 times and the Pearson correlations of each subset with surface mucosal cells were recalculated to form null distributions for the association between each cell type and surface mucosal cells. Original Pearson correlation values within the top 5% of values of the respective null distribution for each cell subset were designated as significant correlations.

Further details for analyses identifying biological features that distinguish HIV-negative EE from U.S. control cohorts

Compositional testing: Cell subset differences were calculated using the leave-one-out approach described to identify changes in HIV status and intestinal region.

Differential expression: Differentially expressed genes for each cell subset and for the epithelial and immune compartments as a whole were found with a Wilcoxon test implemented in the Seurat FindAllMarkers function with the minimum log-fold change threshold set to 0.1 and the minimum percent of cells within a subset expressing a gene set to 0.025.

PROGENy: PROGENy scores for signalling pathway activities were calculated using the progeny R package using 500 genes to generate the model matrix and all other parameters set to default as suggested by the tutorial vignette for applying PROGENy to scRNA-seq data.

DoRothEA: DoRothEA scores for upstream transcription factor activities were calculated using the dorothea R package following the parameters suggested in the Bioconductor vignette for applying DoRothEA to scRNA-seq data.

Module scoring: To identify changes in proliferative capacity in stem cells in EE relative to stem cells in the U.S. control cohorts, we examined a list of genes previously identified as upregulated in cycling human cells*(52)*. To score for a signature of tissue residency in T and NK cells, we used a previously identified signature*(88)* for tissue resident memory T cells consisting of the genes *CD69, ITGAE, ITGA1, IL2, IL10, CXCR6, CXCL13, KCNK5, RGS1, CRTAM, DUSP6, PDCD1, and IL23R.* T cell signatures for activation were obtained from a previous scRNA-seq atlas of T cell activation*(67)*. For all gene signatures, module scores were generated with the

AddModuleScore function in Seurat with default parameters. Cell subsets were then scored for the modules, a Wilcoxon test was used to assess significance, and effect sizes were calculated with Cohen's D.

Predicting ligand-receptor interactions with NicheNet

NicheNet signatures for changes in ligand-receptor interactions between EE and the control cohorts were generated from the differentially expressed genes generated as described above. All other parameters in NicheNet were kept the same as in the NicheNet tutorial "Perform NicheNet analysis starting from a Seurat object".