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

Authors: Conner Kummerlowe[1,2,3,4], , Simutanyi Mwakamui[5], Travis K. Hughes[2,3,4], Nolawit Mulugeta[2,3,4], Victor Mudenda[5], Ellen Besa[5], Kanekwa Zyambo[5], Jessica E. S. Shay[2,6], Ira Fleming[2,3,4], Marko Vukovic[2,3,4], Ben A. Doran[2,3,4,7], Toby P. Aicher[2,3,4], Marc H. Wadsworth II[2,3,4], Juliet Tongue Bramante[3], Amiko M. Uchida[7,8,9], Rabiah Fardoos[10], Osaretin E. Asowata[10], Nicholas Herbert[10], Ömer H. Yilmaz[2,11], Henrik N. Kløverpris[10], John J. Garber[7,9], Jose Ordovas-Montanes[3,4,7,12,13], Zev Gartner[14], Thomas Wallach[15*†], Alex K. Shalek[1,2,3,4,11,12*†], Paul Kelly[6,16*†]

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

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

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

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

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

[10]Africa Health Research Institute, Durban, 4001, South Africa

[11]Department of Pathology, MGH, 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.

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

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

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

† These senior authors contributed equally to this work

One Sentence Summary: Using single-cell genomic profiling, we characterize the epithelial and immune correlates of environmental enteropathy (EE).

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 key contributing factor to childhood malnutrition, growth-stunting, and diminished oral vaccine responses. While EE has been shown to be the by-product of recurrent enteric infection, its full pathophysiology remains unclear. Here, we mapped the cellular and molecular correlates of EE by performing high-throughput single-cell RNA-sequencing on 33 small intestinal biopsies from 11 adults with EE in Lusaka, Zambia (8 HIV-negative, 3 HIV-positive), 6 adults without EE in Boston, USA, and 2 adults in Durban, South Africa, which we complemented with published data from 3 additional individuals from the same clinical site. By leveraging these data to reanalyze previously-defined bulk-transcriptomic signatures of reduced villus height and decreased plasma LPS levels in EE, we found 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 determined cell subsets whose fractional abundances associate with EE severity, small intestinal region, and HIV infection. Furthermore, by comparing duodenal EE samples with those from three control cohorts, we identified dysregulated WNT and MAPK signaling in the EE epithelium and increased pro-inflammatory cytokine gene expression in a T cell subset highly expressing a transcriptional signature of tissue-resident memory cells in the EE cohort. Altogether, our work illuminates epithelial and immune correlates of EE, and nominates cellular and molecular targets for intervention.

Main Text:

INTRODUCTION

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

Pathologically, EE in the proximal small intestine is characterized by reduced villus height, greater crypt depth, and increased microbial translocation *(1, 2, 6)*. However, in a study of Zambian children with EE and non-responsive growth stunting over time, reduced villus height was associated with decreased microbial translocation *(7)*. A bulk transcriptomic study of Zambia children with enteropathy showed that genes differentially upregulated in biopsies with reduced villus height were downregulated in biopsies from participants increased microbial translocation *(8)*. These studies suggest that EE is an adaptive response to potentially lethal enteropathogen exposure that comes at the cost of reduced absorptive capacity and impaired growth.

Histological analyses of EE have revealed increased abundance of lymphocytes, reduced goblet cell numbers, and altered Paneth cell morphology *(6)*. Low plasma levels of tryptophan in children with growth stunting *(9)* and the amelioration of villus blunting in Zambian adults given amino acid supplementation suggest that amino acid deficiency plays a role in EE *(10)*. Bulk transcriptomic studies of EE duodenal biopsies have revealed increased expression of NADPH oxidases, chemokines, mucins, matrix metalloproteases, interferon stimulated genes, and antimicrobial genes including *LCN2, DUOX2*, and *DUOXA2 (11, 12)*. While previous work has lacked the single-cell resolution required to localize these changes to specific epithelial and immune cell subsets, application of single-cell RNA-sequencing (scRNA-seq) could help to resolve comprehensively the cellular and molecular changes that underlie EE pathophysiology *(13)*.

Here, we applied the Seq-Well S^3 platform for massively-parallel scRNA-seq *(14)* to characterize small intestinal biopsies from 11 adults from a community in Zambia where EE is known to be ubiquitous *(15)*. Across these individuals, we profiled 27 biopsies spanning 3 small intestinal regions, HIV-positive and HIV-negative patients, and a range of EE severity scores. In addition, by comparing EE biopsies with those from control groups from South Africa and the USA, we found that EE was associated with upregulated WNT and downregulated MAPK signaling within the epithelium, as well as increased pro-inflammatory cytokine gene expression in a T cell subset highly expressing a transcriptional signature of tissue-resident memory cells. Altogether, our data provide insight into the epithelial and immune correlates of EE, suggesting several therapeutic targets for further investigation.

RESULTS

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

We collected 27 small intestinal biopsies from 11 Zambian volunteers with varying levels of EE severity (8 HIV-negative, 3 HIV-positive) (**Fig. 1A**, **Table S3**). For all 11, we profiled the duodenal bulb and distal duodenum; for a subset, we also collected jejunal samples (**Table S1, S2**). Due to the widespread prevalence of EE in Zambia and a lack of existing screening methods to identify patients without EE, we could not obtain control biopsies from participants without EE in Zambia. Thus, as controls, we profiled samples from 5 adults recruited from a gastrointestinal unit in Durban, South Africa (2 of which we profiled and 3 for which data were publicly available) *(16, 17)* as well as two cohorts of patients in Boston, USA where EE can safely be assumed not to occur (**Table S1, S2**). We note that EE is often contextualized to health by either comparing intermediate EE with severe EE *(8)*, or by comparing EE patients to control cohorts in the United States or the United Kingdom *(6)*. 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 *(18)*.

In total, we analyzed 26,556 high quality single-cell transcriptomes across 38 samples from the Zambian, U.S., and South African cohorts (**Fig. 1B**). After data pre-processing, UMAP

visualization of the entire dataset revealed differences in cellular distribution by patient cohort, intestinal region, and HIV status (**Fig. 1c**). To identify cellular subsets, we applied an existing pipeline for iterative clustering of cell subsets to the Zambian and U.S. datasets *(19)*. Next, we integrated this data with the South African data to account for potential batch effects *(20)*. These analyses revealed that the expected major cell types were represented in almost all biopsies and that epithelial cells were the most abundant major cell type (**Fig. 1D**). Along with more standard QC metrics (**Fig. S1-S4**), this indicated consistent sample quality. In total, we identified 48 detailed cellular subsets that varied in abundance across samples (**Fig. 1E, Fig. S1-S4, Table S4**) *(13, 21, 22)*.

Surface mucosal cells expressing *DUOX2* in the EE epithelium

When identifying cell subsets, we noticed a similarity between the marker genes of surface mucosal cells (a cell subset most commonly found in the distal stomach) and existing gene signatures of reduced villus height and decreased plasma LPS levels in EE (**Fig. S5A**) *(8, 23)*. Relative to all other cell subsets, surface mucosal cells were significantly enriched for both gene signatures ($p < 10^{15}$, Wilcoxon test) (**Fig. 2A, B, Fig. S5B, C**). In addition, the genes in these signatures overlapped with surface mucosal cell marker genes and three antimicrobial genes (*DUOX2, DUOXA2, LCN2)* recently identified as histological markers of EE (**Fig. 2C, D**) *(12, 23)*. Thus, our data suggest these bulk gene signatures may have been driven by an increase in surface mucosal cell abundance. In our data, the vast majority of surface mucosal cells came from duodenal bulb samples of patients with EE (**Fig. S2C**). In these tissues, surface mucosal cell fractional abundances were significantly correlated (Permutation test, $p < 0.05$) with those of the stem cycling subset and the Ent *ANXA2 PTMA* subset which highly expressed marker genes of dedifferentiating enterocytes (*PTMA*) *(24)* and wound associated epithelial cells (*ANXA2*) (**Fig. 2E**) *(25)*. To further examine potential relationships between these cell subsets that co-occurred with surface mucosal cells, we inferred differentiation trajectories for the epithelial cells in our dataset with PAGA (**Fig. 2F**) *(26)* The wound healing-like epithelial subsets and surface mucosal cells all lay in between mature enterocyte and stem cell subsets in the inferred differentiation hierarchy. Running the RNA velocity package Velocyto produced similar results (**Fig. S5D, E**) *(27)*. In addition, immunohistochemical staining revealed that DUOX2 localized to the villus tip in blunted villi (**Fig. 2G, Fig. S6)**. Together, these results suggest that surface mucosal cells occur

at the villus tip in EE and are associated with the presence of intermediate wound healing-like cell populations. Furthermore, we observed that surface mucosal cells uniquely expressed *MUC5AC*– a marker of *H. pylori* infection (**Fig. S5A, Table S4**) *(28)*. Applying the metagenomic classification tool Kraken 2, we found that relative to the control cohort samples, 6 samples from 4 participants in the EE cohort contained significantly higher levels of *H. pylori* mapping reads (**Fig. S5**). These samples were predominantly from the duodenal bulb (**Fig. S5G**). Thus, the presence of surface mucosal cells in EE may be associated with *H. pylori* infection.

Cellular correlates of intestinal region, disease severity, and HIV infection in EE

We next identified cell subsets whose fractional abundance shifted across intestinal region within HIV-negative EE patients (**Fig. 3A**). Duodenal bulb samples were enriched for surface mucosal cells, mucosal neck cells, and enterocytes highly expressing *ANXA2, FABP1,* and *CD55*, as well as three T cell subsets expressing markers of immune activation (*IL17A, CXCR4*, *GZMA*) *(29, 30)*. Distal duodenal samples were enriched for enterocytes, goblet cells, and stem cells highly expressing *OLFM4*.

We then mapped the correlates of histologically determined EE severity in eleven biopsies (**Fig. 3B,C; Supplementary Methods; Table S5**) *(6)*. In the epithelium, greater severity was associated with lower fractional abundances of the mature enterocyte and stem *OLFM4* subsets, as well as higher fractional abundances of immature enterocytes. 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*) *(29)* 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**) *(31)*. In line with past findings, plasma cells abundances decreased with EE severity *(32)*. These results suggest that severe EE is associated with an intermediate-like epithelial phenotype and inflammatory lymphocyte subsets.

Next, we sought to characterize the impact of antiretroviral-treated HIV infection on EE pathology. We found that HIV-positive samples displayed higher EE severity than HIV-negative samples (Wilcoxon test, $p = 0.034$) (**Fig. S7A**). Examining the shifts in cell subset fractional abundances

with HIV infection, we found known features of HIV biology including decreased fractional abundances of *CD4*^{hi} T cells and increased fractional abundances of γδ T cells highly expressing the HIV co-receptor *CXCR4* (**Fig. S7B, C**) *(16)*. In addition, within duodenal bulb samples, HIV pathology associated with increased fractional abundances of enterocytes highly expressing *ANXA2, FABP1,* and *CD55,* suggesting that HIV pathology may contribute to the presence of this wound healing-like subset within the duodenal bulb (**Fig. S7D**).

Epithelial correlates of EE

We next sought to identify features that distinguished HIV-negative EE distal duodenal samples from matched samples from participants in South Africa and the USA. While histology was not available for the South African dataset, H&E staining of a duodenal biopsy from a separate patient at the same clinical site revealed features of EE, including villus blunting, goblet cell depletion, and Paneth cell depletion, but no signs of inflammation (**Fig. S8A**). To investigate whether samples from this site displayed features of EE, we performed a pairwise comparison of the fractional abundances of all cell subsets between the three geographical locations in this study (**Fig. S8B-D**). Relative to the U.S. cohorts, both the Zambian and South African cohorts displayed two characteristic features of EE: reduced goblet cell and increased plasma cell fractional abundances *(6)*. However, plasma cells and T cell subsets expressing markers of inflammation (*IL17A*, *GZMA*) were increased in fractional abundance in the Zambia EE cohort relative to the South African cohort *(29)*, suggesting that not all features of EE were present in the South African samples. Thus, we took two approaches to identify the distinguishing attributes of confirmed EE in the Zambian cohort. First, we compared the Zambian cohort to all control cohorts. Then, we compared the Zambian cohort with confirmed EE to only the U.S. cohorts in case our previous analysis was confounded by potential features of EE in the South African cohort.

Comparing the fractional abundances of epithelial cells from patients with confirmed EE with those from all other cohorts, we found an enrichment of stem *OLFM4* cells, foveolar precursor cells, and enterocytes co-expressing *APOA4* and *ALPI* in EE, as well as reduced fractional abundance of EEC K cells (**Fig. 4A**). Differential expression analysis between epithelial cells in EE and control cohorts revealed compartment-wide upregulation of genes (*PIGR, CCL25*) involved in antibody transport and lymphocyte recruitment, among others (**Fig. 4B,C Table S6**)

(33, 34). EE epithelial cells also highly expressed *CTNNB1*, a key component of WNT/ß-catenin signaling. In agreement, PROGENy analysis suggested increased WNT signaling in all three stem cell subsets and decreased MAPK signaling in cycling stem cells and stem *OLFM4* cells in EE (**Fig. 4D)** *(35)*. To help corroborate and extend these findings, we immunohistochemically stained Zambian EE and U.S. control samples for ß-catenin. We found that ß-catenin stained at higher intensity in the EE epithelium (**Fig. S9**). In addition, tuft cells in EE highly expressed *ALOX5AP* (which is involved in inflammation via leukotriene biosynthesis *(36)*) and *SERPINA1* (which encodes α-1-antitrypsin (AAT), a biomarker of epithelial damage in EE *(5)*) (**Fig. 4C**). Finally, comparison of our data to past intestinal scRNA-seq datasets revealed no evidence for "colonification" of the small intestine in EE and showed limited overlap between the genes upregulated in EE and ulcerative colitis (UC). (**Fig. S10**).

Immune correlates of EE

Next, we compared cell proportions of immune cells between the Zambian cohort with confirmed EE and all control cohorts. This revealed that EE samples were enriched for *CD8^{hi}* T cells highly expressing *MALAT1* and γδ T cells highly expressing *GZMA* (**Fig. 5A**). Consistent with previous findings, plasma cells were increased in EE relative to non-EE cohorts (**Fig. S10C**) *(32)*. Conducting differential expression analyses, we found that the majority of significant gene expression changes in EE immune cells occurred within the T cell compartment (**Fig. 5B, Fig. S11A-C**). Out of all T cell subsets, T *CD8 CD69*^{hi} cells displayed the most differentially expressed genes between EE and controls, including upregulation of effector-like genes (*IFNG, CCL5, IL32*) in EE and downregulation of genes (*IL7R*, *CXCR4*) promoting memory T cell formation after infection *(37, 38)* (**Fig. 5C, Table S6**). As *CD69* is a potential marker for T cell tissue residency, we scored all T cells subsets on a gene signature of tissue resident memory T cells and found that the T *CD8 CD69*^{hi} subset scored highest (adjusted $p = 4.98*10^{-78}$, one sided Wilcoxon test) (**Fig. 5D**) *(39)*. Scoring all T *CD8 CD69*^{hi} cells on T cell activation signatures, we found that relative to controls, cells in this subset from EE scored higher for signatures of cytotoxicity and cytokine production (**Fig. 5E**) *(40)*. In agreement, immunohistochemical staining revealed more cells positive for Granzyme B in EE samples relative to controls (**Fig. S12**). Nominating putative ligand-receptor interactions between cellular subsets using NicheNet, we found potentially increased IFNγ signaling in the EE epithelium stemming from IFNγ production by T cells, especially the T

CD8 CD69^{hi} cells. (**Fig. S13**) *(41)*. Altogether, our data reveal immune correlates of EE that may contribute to pathogenesis and reduced oral vaccine efficacy.

Evidence of reduced epithelial proliferation in EE relative to the U.S. cohorts

Finally, we compared only the Zambian EE and U.S. cohorts. Differential expression analysis revealed that EE samples displayed compartment-wide downregulation of genes (*KLF4, ATF3, FOS,* and *JUN*) involved in epithelial proliferation, IL-22 signaling, and goblet cell development (**Fig. 6A**) *(42–44)*. Consistently, the EE samples displayed lower fractional abundances of goblet cells and ILC3s (producers of IL-22), as well as higher fractional abundances of γδ T cells (negative regulators of IL-22 production in mice fed a low protein diet) (**Fig. S8B)** *(45, 46)*. Furthermore, gene set enrichment analysis (GSEA) revealed enrichment of the Reactome signature for response of EIF2AAK4 and GCN2 to amino acid deficiency in the epithelial cells (**Fig. 6B, Table S8**). In addition, goblet cells from patients with confirmed EE upregulated markers for lower crypt goblet cells suggesting that EE goblet cells have a more immature phenotype (**Fig. 6C**) *(47)*. Furthermore, EE stem cells scored significantly lower on a gene signature of cycling human cells, and displayed lower PROGENY scores for the pro-proliferative EGFR, MAPK, and PI3K pathways (**Fig. 6D-E, Table S8**) *(48)*. Upstream transcription factor activity inference with DoRothEA revealed reduced activation of ATF2 and ATF4 broadly across the epithelium, consistent with reduced IL-22 signaling (**Fig. S14A, Table S9)** *(49)*. In sum, our results suggest that relative to the U.S. controls, the EE epithelium is characterized by reduced proliferation, IL-22 signaling, and goblet cell development. However, within the EE cohort, EE severity scores correlated with cycling scores in stem cells (**Fig. S1B**). Intriguingly, this suggested that although EE patients as a whole display lower levels of epithelial proliferation relative to U.S. controls, more severe EE leads to relatively higher epithelial proliferation than less severe EE.

DISCUSSION

Here, we profiled EE with the Seq-Well S^3 platform for scRNA-seq. We thereby identified a cell subset – surface mucosal cells – which highly expressed *DUOX2* and whose gene expression pattern matched existing bulk gene signatures of reduced villus height and reduced plasma LPS concentrations in EE. In addition, our dataset revealed variations in cell subset fractional

abundance by small intestinal region, HIV infection, and EE severity, as well as epithelial and immune subsets differing between EE and control samples. Altogether, our work re-contextualizes past bulk-transcriptomic studies of EE and maps the cellular correlates of EE pathology.

The presence of surface mucosal cells highly expressing *DUOX2* in EE may reflect remodeling of the epithelium into an intermediate wound healing-like state. Dedifferentiation of mature cells could facilitate repair of the epithelial barrier and reduce microbial translocation at the expense of reducing surface area and absorptive capacity, explaining why reduced nutrient absorption in EE is associated with decreased microbial translocation *(7)*. This process may be due, in part, to *H. pylori:* as *H. pylori* gastritis has been shown often to be associated with duodenal colonization in children *(50)*, *H. pylori* infection may explain why previous studies have found high levels of *DUOX2* transcripts in the distal duodenum of some children with EE, whereas our study found *DUOX2* expressing surface mucosal cells in the duodenal bulb (which is closest to the stomach where most *H. pylori* infection occurs) but not the distal duodenum of adults with EE.

Comparison to all control cohorts illuminated the epithelial and immune cell correlates of EE. Increased abundances of immature epithelial cell subsets, increased WNT/ß-catenin signaling, and decreased MAPK signaling suggested that the EE epithelium is biased towards an immature phenotype. Furthermore, Tuft cells upregulated genes involved in promoting and responding to inflammation *(51)*. In addition, we found lower relative abundances of T cells expressing a transcriptional signature of tissue-resident memory T cells, but those cells had elevated expression of inflammatory cytokines (including IFNγ) in EE, suggesting that while these cells are present in a lower abundance in EE, they may be chronically activated. This, in turn, may lead to immune exhaustion and impaired responses to new immune stimuli, which may contribute a hindered response to oral vaccines *(52)*. While IFNγ is often viewed as a pro-inflammatory cytokine in acute inflammation, numerous studies have demonstrated that in chronic inflammation, IFNγ can produce tolerogenic effects, which would be consistent with reduced oral vaccine efficacy in EE *(53)*.

Additionally, we compared EE to only the U.S. cohorts to account for potential confounding features of moderate EE in the South African cohort. Relative to the U.S. cohorts, the EE

epithelium was characterized by decreased epithelial proliferation and changes in cell subset fractional abundances, consistent with decreased IL-22 signaling *(45, 46, 54)*. This is in line with work that found decreased abundances of transcripts from pro-proliferative pathways in the feces of Malawian children with EE *(55)*, as well as work showing that during *Cryptosporidium* infection protein malnutrition leads to reduced turnover of intestinal epithelial cells *(56)*. Consistently, reduced epithelial proliferation in EE relative to the U.S. cohorts was associated with GSEA enrichment of a response to amino acid deficiency and reduced goblet cell abundances–whose differentiation can be induced by tryptophan *(57)*. This agrees with work demonstrating decreased tryptophan metabolism in Pakistani children with EE and work showing that amino acid supplementation ameliorates villus blunting in adults with EE *(10, 12)*. Thus, amino acid deficiency may lead to hypoproliferative signaling in the EE epithelium characterized by reduced stem cell proliferation and goblet cell abundance. One intriguing aspect of this hypoproliferative signaling is that it stands in stark contrast to the hyperproliferative signaling observed in Crohn's disease, which is of particular interest as limited evidence suggests lower rates of IBD in countries where EE is endemic *(16, 49)*.

However, when comparing within the EE cohort, more severe EE was positively associated with stem cell proliferation. This may be due to the interplay of malnutrition and infection. The reduced dietary quality in the population of Zambian adults we studied may impose a proliferative constraint on enteropathy, leading to reduced stem cell turnover relative to intestinal homeostasis. However, individuals with more severe EE may have more proliferation than others in response to infective and inflammatory drivers. This discrepancy between our within and across country analyses is congruent with past work which found similar discrepancies, highlighting the necessity of comparing to an outgroup to fully understand the pathophysiology of EE *(32, 58)*. Follow-up mechanistic studies are needed to clarify the role that malnutrition and infection play in epithelial proliferation in EE.

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. and South African cohorts are due to unobserved variables that differed between these patient populations,

especially the high burden of enteropathogens in tropical settings. 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. This is further limited by the lack of available tissue for histological analysis of samples from the South African cohort. As we saw few stromal cells in our scRNA-seq dataset, our tissue dissociation was likely biased against this subset and future work will be needed to characterize these cells in EE. In addition, the inflamed small intestinal epithelium and lamina propria are highly heterogenous environments containing several relatively rare cell types such as Paneth cells, and a variety of immune subsets which we did not have sufficient power to analyze in great detail. Furthermore, the HIV status of the participants from the U.S. cohorts was not determined. Additionally, we did not screen EE patients for Celiac disease 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 *(59)*. Also, pediatric EE may differ from EE in adults, which calls for future studies in pediatric cohorts. Finally, as EE is an endemic condition in low- and middle-income countries across the globe, it will be necessary to validate our results in cohorts with EE from geographic locations other than Zambia.

Examining our work as a whole, a potential picture of EE pathogenesis emerges. Relative amino acid deficiency due to a low-quality diet may lead to reduced epithelial proliferation and differentiation towards goblet cells, which would diminish anti-microbial mucosal defense, leading to increased pathogen-mediated damage of the enterocyte. This may then lead to epithelial remodeling towards an intermediate wound healing-like phenotype associated with the presence of surface mucosal cells. In addition, enteropathogen mediated damage would further exaggerate the pathogen-induced IFNγ response and may explain the observed pro-inflammatory polarization of *CD8 CD69*^{hi} T cells in our data. Together, these findings nominate several therapeutic axes for inducing healthy epithelial proliferation and immune efficacy in EE.

MATERIALS AND METHODS

Study design

Adult volunteers were recruited from a disadvantaged community in Lusaka, Zambia, in which we have carried out previous studies of environmental enteropathy *(7, 11)*. 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. Informed consent was obtained from participants recruited into this study at the Inkosi Albert Luthuli Central Hospital in Durban, South African. No randomization or blinding was done in this study. No power analyses were conducted due to the observational nature of this study and the lack of pre-existing scRNA-seq datasets of EE. The number of samples used is presented in the figures, supplementary figures, and supplementary tables.

Biopsy handling, immunohistochemical staining, and tissue digestion

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 for EE severity using a recently published methodology *(6)*. Duodenal bulb samples from EE patients were stained for DUOX2 protein and distal duodenal samples from EE patients and from normal tissue obtained from Mass General Brigham were stained for ß-catenin and GZMB protein; for more details see supplementary methods. Biopsies from EE, EoE, resection, and South African patients were dissociated into single-cell suspensions using a modified version of a previously published protocol *(21)*. For further detail see the supplementary methods.

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

Please refer to the supplementary methods for further detail. Briefly, the epithelial and lamina propria layers of the biopsies were dissociated into single-cell suspensions. Then, single-cells were loaded onto a functionalized polydimethylsiloxane (PDMS) array preloaded with uniquely-barcoded mRNA capture beads (Chemgenes; MACOSKO-2011-10), and sequencing libraries were obtained and sequenced on an Illumina Next-Seq. Sequencing read alignment and demultiplexing was performed on the cumulus platform using snapshot 6 of the Drop-seq pipeline *(48)*, resulting in a cell barcode by UMI digital gene expression (DGE) matrix. QC was performed to remove low quality cell barcodes and doublet cells. Bam files from sequencing were classified with Kraken2 to find metagenomic mapping reads. To identify cell subsets, we adopted an existing pipeline for automated iterative clustering of single-cell data that has been shown to identify batch effects without collapsing distinct rare cell types *(19)*. We applied this pipeline to the data from the EE and U.S. cohorts. To correct for batch effects between data collected in different laboratories, we integrated our data with the South African dataset *(20)*. All subsets were scored on gene signatures of reduced villus height and decreased and plasma LPS signatures from Chama et al. *(8)* using the AddModuleScore function in Seurat and a Wilcoxon test was used to assess significance. Epithelial trajectories were inferred with PAGA *(26)*. RNA velocity trajectories were inferred with velocyto *(27)*.

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 with the ComplexHeatMap R package. 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 using a Fisher's exact test. Cell subsets significantly associated with histological EE severity were identified by running Dirichlet Regression. Further details are provided in the supplementary methods.

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

Distal duodenal samples from HIV-negative patients iwth EE were compared to matched samples from control cohorts. We sought to identify biological features driven by variation in EE biology relative to the control cohorts (as opposed to identifying biology that distinguished only one cohort from EE). Thus, in all subsequent analyses we required that results pass the following two criteria: 1) Result significant when comparing EE vs all control cohorts 2) Result displayed the same direction of change between EE and each control cohort. Further details are provided in the supplementary methods.

Statistical analysis

The statistical test used for each comparison is denoted in the corresponding figure legend. Tests were conducted in R, and a Benjamini-Hochberg adjusted p value of 0.05 was used for significance.

Supplementary Materials

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

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 EE relative to control cohorts.

Table S7. KEGG, REACTOME, and PID pathway analysis.

Table S8. PROGENY pathway activation scores.

Table S9. Transcription factors with upstream activity predictions from DoRothEA.

References (60-62)

References and Notes

1. 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).
2. 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).
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. 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).
7. 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.
8. 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).
9. 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).
10. 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).
11. 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).
12. 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.
13. 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).
14. 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).
15. 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).
16. R. Fardoos, O. E. Asowata, N. Herbert, S. K. Nyquist, Y. Zungu, A. Singh, A. Ngoepe, I. M. Mbano, N. Mthabela, D. Ramjit, F. Karim, W. Kuhn, F. G. Madela, V. T. Manzini, F. Anderson, B. Berger, T. H. Pers, A. K. Shalek, A. Leslie, H. Kløverpris, HIV infection drives interferon signalling within intestinal SARS-CoV-2 target cells, *JCI Insight* (2021), doi:10.1172/jci.insight.148920.
17. O. E. Asowata, A. Singh, A. Ngoepe, N. Herbert, R. Fardoos, K. Reddy, Y. Zungu, F. Nene, N. Mthabela, D. Ramjit, F. Karim, K. Govender, T. Ndung'u, J. Z. Porterfield, J. H. Adamson, F. G. Madela, V. T. Manzini, F. Anderson, A. Leslie, H. N. Kløverpris, Irreversible depletion of intestinal CD4+ T-cells is associated with T-cell activation during chronic HIV infection, *JCI Insight* (2021), doi:10.1172/jci.insight.146162.
18. 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).
19. H. B. Zheng, B. A. Doran, K. Kimler, A. Yu, V. Tkachev, V. Niederlova, K. Cribbin, R. Fleming, B. Bratrude, K. Betz, L. Cagnin, C. McGuckin, P. Keskula, A. Albanese, M. Sacta, J. de Sousa Casal, F. Taliaferro, M. Ford, L. Ambartsumyan, D. L. Suskind, D. Lee, G. Deutsch, X. Deng, L. V Collen, V. Mitsialis, S. B. Snapper, G. Wahbeh, A. K. Shalek, J. Ordovas-Montanes, L. S. Kean, A Treatment-Na{\"\i}ve Cellular Atlas of Pediatric Crohn{\textquoteright}s Disease Predicts Disease Severity and Therapeutic Response, *medRxiv* (2021), doi:10.1101/2021.09.17.21263540.
20. 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).
21. 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).
22. 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).
23. 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).
24. 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).
25. 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).
26. 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).
27. 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).
28. 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).
29. 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).
30. 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).
31. 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).
32. 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).
33. 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).
34. 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).
35. 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.
36. S. Ualiyeva, E. Lemire, E. C. Aviles, C. Wong, A. A. Boyd, J. Lai, T. Liu, I. Matsumoto, N. A. Barrett, J. A. Boyce, A. L. Haber, L. G. Bankova, Tuft cell-produced cysteinyl leukotrienes and IL-25 synergistically initiate lung type 2 inflammation., *Sci. Immunol.* **6**, eabj0474 (2021).
37. 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).
38. 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).
39. 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).
40. 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).
41. R. Browaeys, W. Saelens, Y. Saeys, NicheNet: modeling intercellular communication by linking ligands to target genes, *Nat. Methods* **17**, 159–162 (2020).
42. 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.
43. 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.
44. 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).
45. 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).
46. 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.
47. 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).
48. 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).
49. 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).
50. M. Bonamico, P. Mariani, F. M. Magliocca, V. Petrozza, M. Montuori, C. Pezzella, I. Luzzi, F. Carpino, Helicobacter pylori duodenal colonization in children., *Acta Paediatr.* **86**, 356–360 (1997).
51. R. A. F. Dixon, R. E. Diehl, E. Opast, E. Rands, D. K. Millerll, Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis, *Nature* **343**, 282–284 (1990).
52. E. J. Wherry, M. Kurachi, Molecular and cellular insights into T cell exhaustion, *Nat. Rev. Immunol.* **15**, 486–499 (2015).
53. P. Rožman, U. Švajger, The tolerogenic role of IFN-γ*Cytokine Growth Factor Rev.* **41**, 40–53 (2018).
54. M. E. Keir, T. Yi, T. T. Lu, N. Ghilardi, The role of IL-22 in intestinal health and disease, *J. Exp. Med.* **217**, 1–9 (2020).
55. 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).
56. 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).
57. 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).
58. P. Kelly, M. Bajaj-Elliott, M. Katubulushi, I. Zulu, R. Poulsom, R. A. Feldman, C. L. Bevins, W. Dhaliwal, Reduced gene expression of intestinal α-defensins predicts diarrhea in a cohort of African adults, *J. Infect. Dis.* **193**, 1464–1470 (2006).
59. 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).
60. 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).
61. 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).
62. R. Elmentaite, N. Kumasaka, K. Roberts, A. Fleming, E. Dann, H. W. King, V. Kleshchevnikov, M. Dabrowska, S. Pritchard, L. Bolt, S. F. Vieira, L. Mamanova, N. Huang, F. Perrone, I. Goh Kai'En, S. N. Lisgo, M. Katan, S. Leonard, T. R. W. Oliver, C. E. Hook, K. Nayak, L. S. Campos, C. Domínguez Conde, E. Stephenson, J. Engelbert, R. A. Botting, K. Polanski, S. van Dongen, M. Patel, M. D. Morgan, J. C. Marioni, O. A. Bayraktar, K. B. Meyer, X. He, R. A. Barker, H. H. Uhlig, K. T. Mahbubani, K. Saeb-Parsy, M. Zilbauer, M. R. Clatworthy, M. Haniffa, K. R. James, S. A. Teichmann, Cells of the human intestinal tract mapped across space and time. Nature. 597, 250–255 (2021).
63. C. Kummerlowe. (2022). Code for single-cell profiling of environmental enteropathy reveals signatures of epithelial remodeling and immune activation. Zenodo. https://doi.org/10.5281/zenodo.6703873

Acknowledgements

We thank the participants 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 was 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), the Bill and Melinda Gates Foundation, and the Ragon Institute. T.W. was supported by the NIH (DK007762). 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, the Leona M. and Harry B. Helmsley Charitable Trust, The Pew Charitable Trusts Biomedical Scholars and The New York Stem Cell Foundation. T.K.H. was supported by the NIH (F30-AI143160-01A1). J.E.S.S. was supported by the NIH (F32DK128872).

Author Contributions

C.K., T.W. T.K.H., Z.G., H.NK., O.H.Y. A.K.S., and P.K. designed experiments; C.K. performed tissue dissociation with assistance from T.K.H. J.O.M., M.V, R.F, O.A., and N.H..; C.K., T.K.H., T.W., S.M., I.F., M.V., R.F. T.A.P., and M.HW. performed scRNA-seq; C.K. and N.M. analyzed the single-cell data; B.A.D. and J.O.M. developed code for iterative tiered clustering; V.M., E.B., and K.Z. performed histological analyses; J.E.S.S. performed immunohistochemical staining 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., 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, FL82, Empress Therapeutics, Relation Therapeutics, Senda Biosciences, IntrECate biotherapeutics, 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, Foundation 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

All data associated with this study are in the paper or supplementary materials. All scRNA-seq data is available at 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 the EE and EoE cohorts is available at Gene Expression Omnibus (GEO) accession number GSE168883. The raw data for the resection samples is available by contacting the authors and establishing a Data Use Agreement. Code is available at https://doi.org/10.5281/zenodo.6703873 (*63*)

Abbreviations

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

Figures

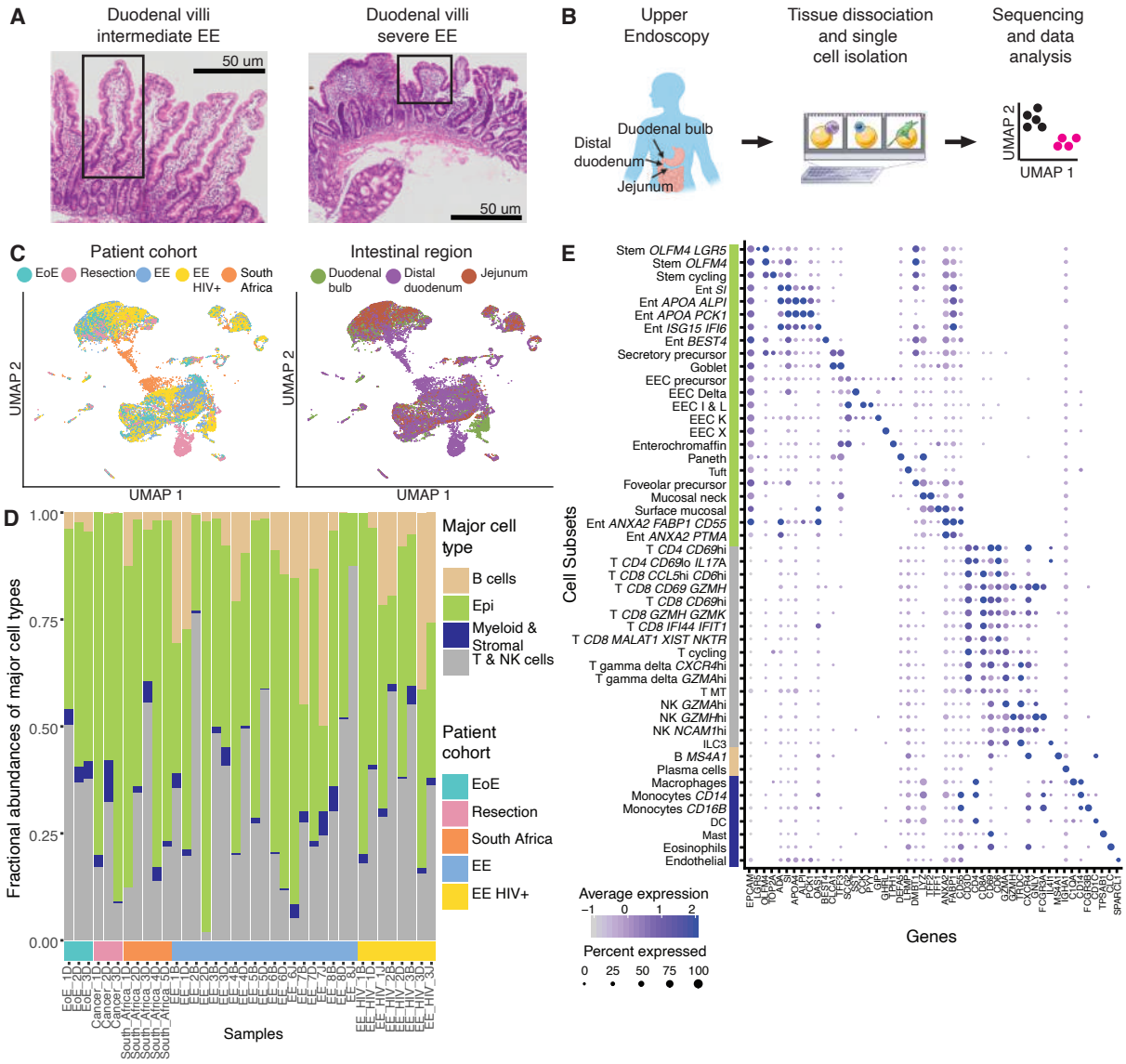


Figure 1: Single-cell RNA-sequencing of the small intestine with and without EE.

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 26,556 high quality cells from 38 samples and 22 patients.

D, Fractional abundances of major cell types amongst all single cells analyzed 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.

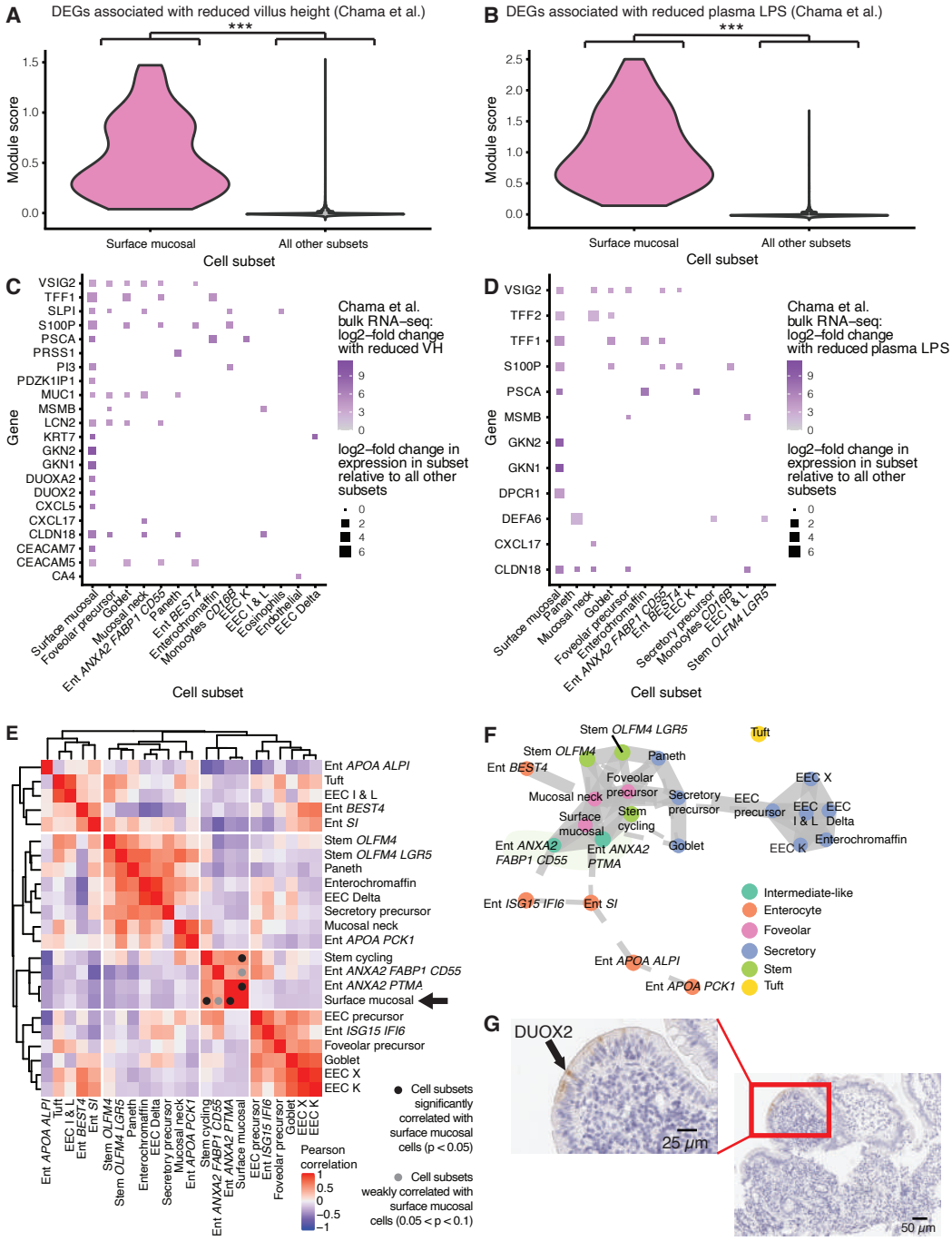


Figure 2: Surface mucosal cells uniquely express *DUOX2* and are associated with a wound healing-like phenotype

A, Violin plot of a module score generated from genes upregulated in EE samples with reduced villus height (VH) in Chama et al. *(8)* (***, $p < 0.001$; Wilcoxon test).

B, Violin plot of a module score generated from genes upregulated in EE samples wwith decreased plasma LPS concentrations in Chama et al. *(8)* (***, $p < 0.001$; 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 ($0.05 < p < 0.1$; Permutation testing).

F, PAGA trajectory visualization of epithelial subsets.

G, H&E (purple) and immunohistochemical staining for DUOX2 protein (brown) on an EE biopsy from the duodenal bulb.

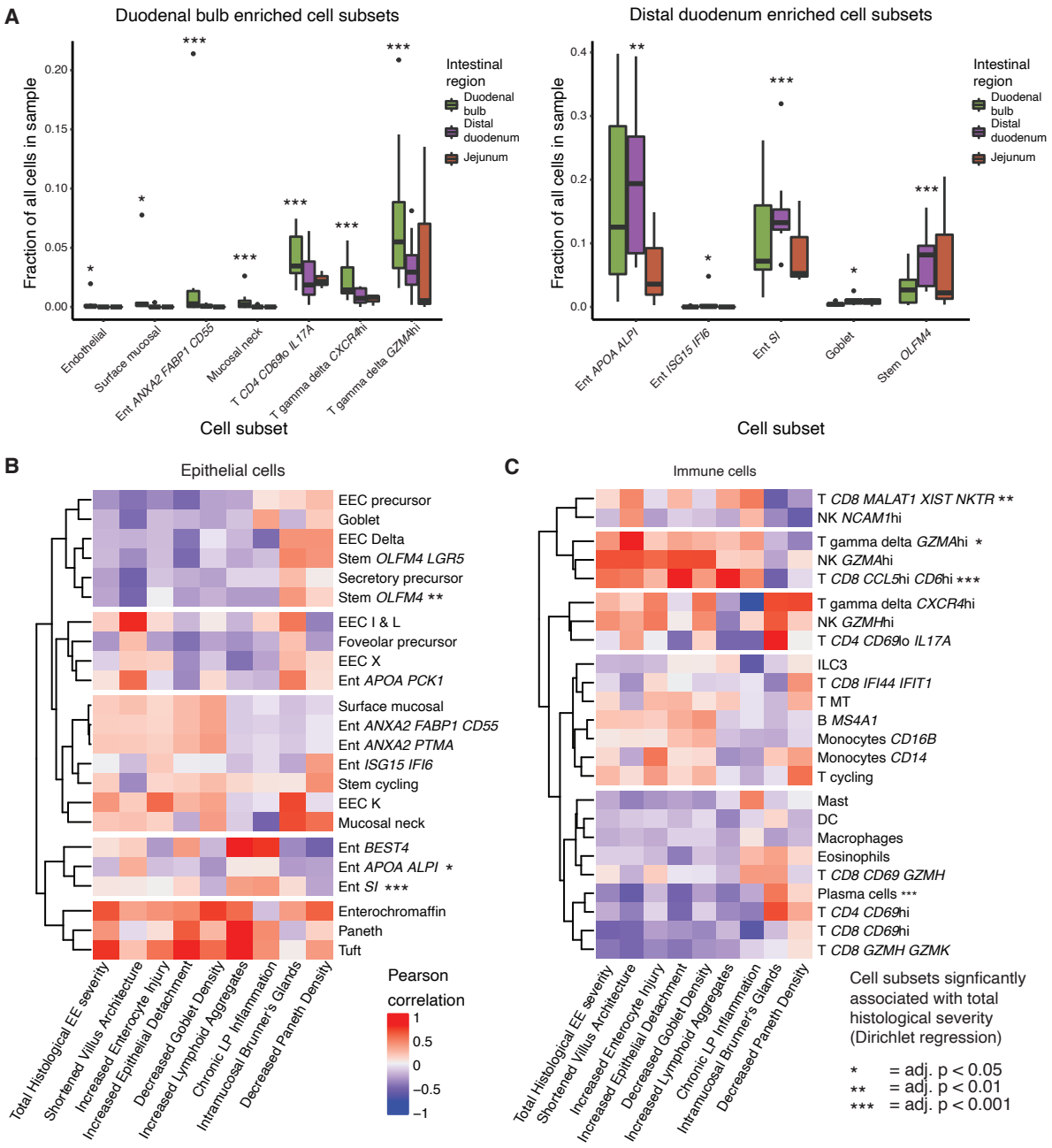


Figure 3: Cell subsets associated with intestinal region and histologically determined EE severity

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, Fisher's exact test).

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 (*, adj. p < 0.05; **, adj. p < 0.01,; ***, adj. p < 0.001; Dirichlet regression).

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. (*, adj. p < 0.05; **, adj. p < 0.01; ***, adj. p value < 0.001; Dirichlet regression).

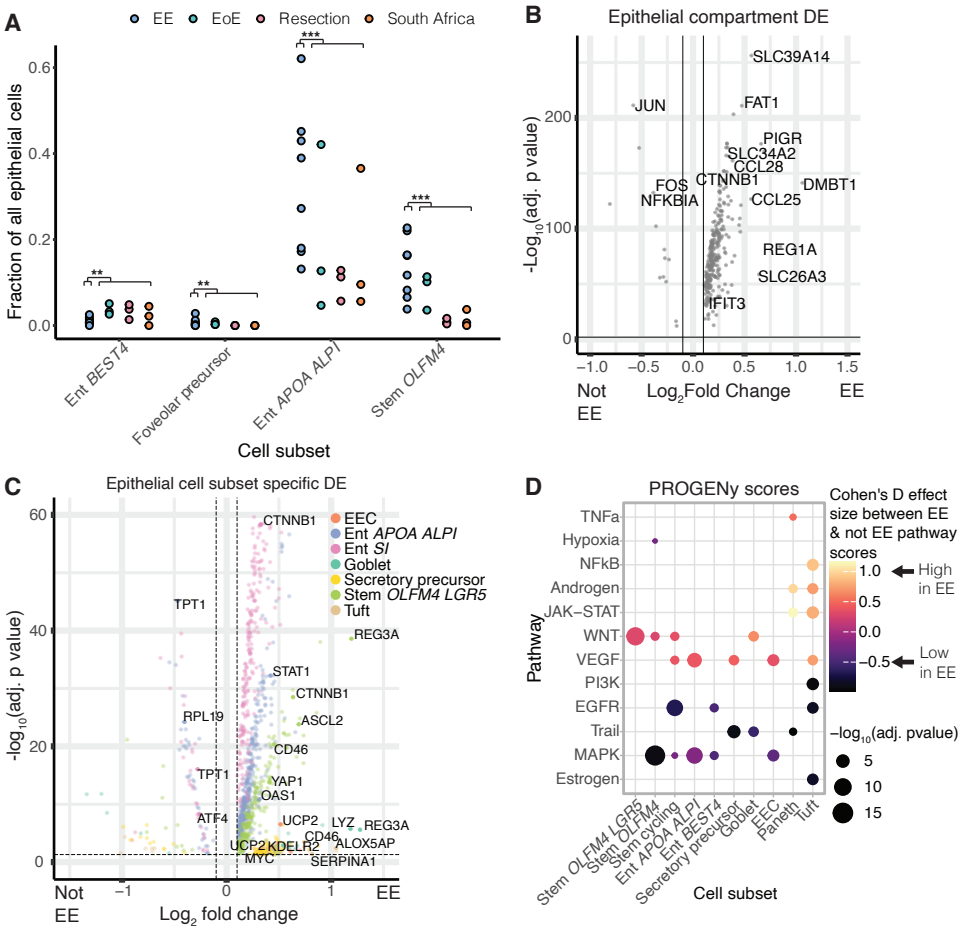


Figure 4: The epithelium of EE is characterized by increased WNT signaling, and decreased MAPK signaling.

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

B, Genes differentially expressed in the epithelial compartment in EE relative to all control cohorts. 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 all control cohorts 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, PROGENy pathway prediction scores for epithelial cells in EE relative to controls

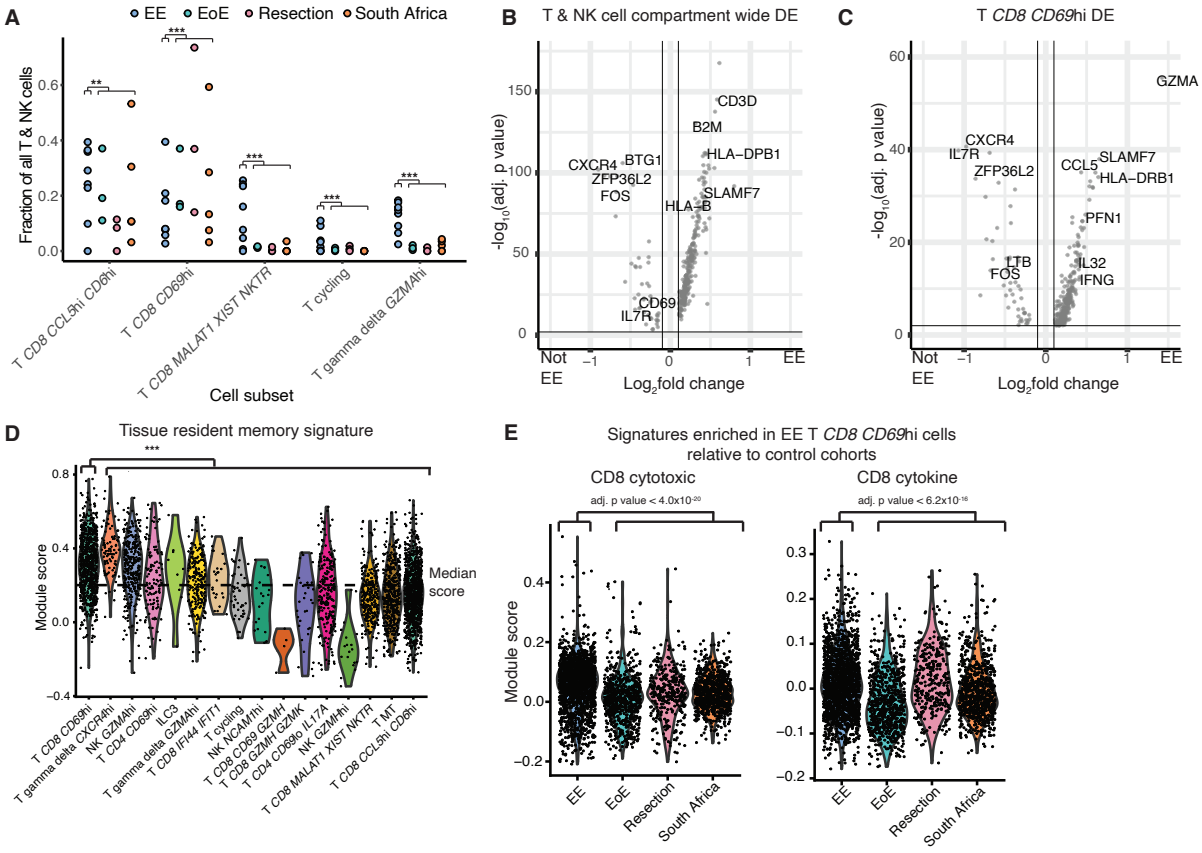


Figure 5: EE is associated with a shift towards activated T cell phenotypes

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

B, Genes differentially expressed in the T & NK cell compartment in EE relative to all control cohorts. 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 all control cohorts 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 all control cohorts.

E, T cell activation signatures enriched in T *CD8 CD69*^{hi} cells from EE patients relative to all control cohorts. 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

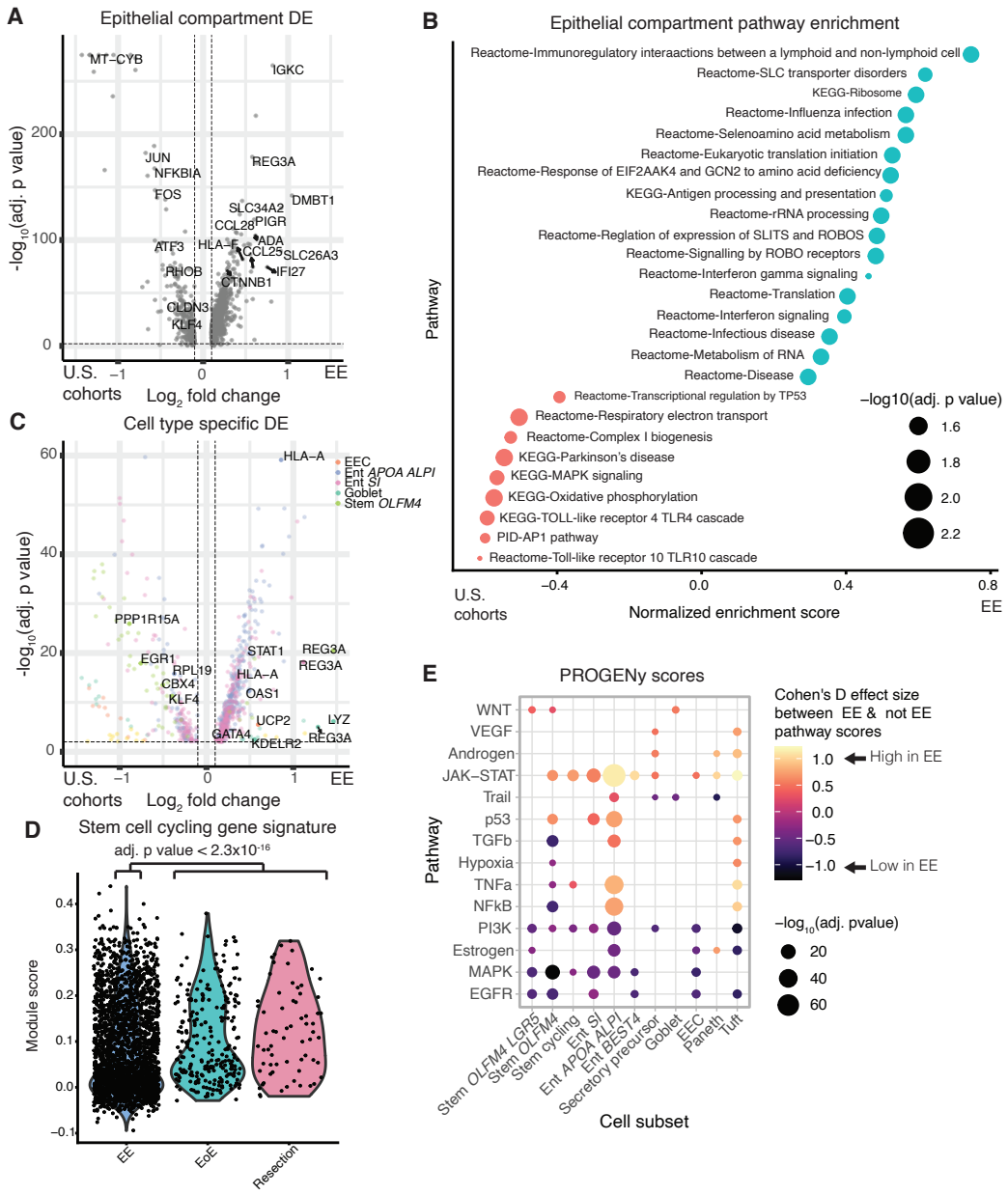


Figure 6: Evidence of reduced proliferation in the EE epithelium relative to U.S. cohorts

A, Genes differentially expressed in the epithelial compartment in EE relative to the U.S. cohorts. 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.

B, Gene set enrichment analysis of genes upregulated in epithelial compartment cells in EE relative to the U.S. cohorts.

C, Genes differentially expressed in EE relative to the U.S. cohorts 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 cell cycle genes in EE and U.S. cohorts in all cells from stem cell subsets.

E, PROGENy pathway prediction scores for epithelial cells in EE relative to U.S. cohorts.

SUPPLEMENTARY MATERIALS

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, 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 base pairs, 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 Terra platform (https://app.terra.bio) using snapshot 6 of the Drop-seq pipeline 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 *(60)* (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 *(21)* 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 *(20)*.

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 *(15,22)* 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 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 *(61)*, 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– 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 package and adopted from *(20)*. For each starting cluster, normalization and variable gene selection were conducted with SCtransform 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 neighbors 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. The final 48 cell types 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 *(26)*) and either *FABP1* or *PTMA* (markers of dedifferentiating intestinal cells in a murine model system *(25)*); 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.

Analyses characterizing surface mucosal cells and associated cell populations

Gene signatures for reduced villus height and decreased and plasma LPS signatures in Chama et al. *(6)* 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. Prior to PAGA analysis, nearest neighbor graph was calculated in Scanpy with neighbors = ceiling(0.5(number of epithelial cells)^0.5)). PAGA and Velcyto analyses were otherwise ran with default parameters.

Identifying cell subsets with significant changes in relative abundance by HIV infection status and small intestinal region

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 using a Fisher's exact test. 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). We then repeated this analysis with each patient left out and only cell subsets that were significant in all analyses were included in the final results.

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*(10)*. To score for a signature of tissue residency in T and NK cells, we used a previously identified signature*(11)* 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*(12)*. 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.

Comparison with large intestine gene signature from the cut cell at atlas.: Intestinal scRNA-seq data from adults was downloaded from https://www.gutcellatlas.org and the Find Markers function with default parameters in Seurat was used to find genes upregulated in the large intestine vs the small intestine (*62*).

Comparison with Ulcerative colitis gene signature from Smillie et al (*21*): A gene signature comparing the inflamed colon with ulcerative colitis was obtained from the supplemental information in Smillie et al.

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”.

Metagenomic mapping of H. pylori reads with Kraken2

Bam files from sequencing were classified using Kraken2 with the following Terra pipeline from the Broad institute: https://dockstore.org/workflows/github.com/broadinstitute/viral-pipelines/classify_kraken2:master?tab=files. The following inputs were used. Kraken2_db_tgz: "gs://pathogen-public-dbs/v1/kraken2-broad-20200505.tar.zst”. Krona_taxonomy_db_tgz: "gs://pathogen-public-dbs/v1/krona.taxonomy-20200505.tab.zst”. The outputted results for the samples were then aggregated across all bacterial species using the merge_metagenomics pipeline on the Terra platform from the Broad institute: https://viral-pipelines.readthedocs.io/en/latest/merge_metagenomics.htmls

SUPPLEMENTARY FIGURES

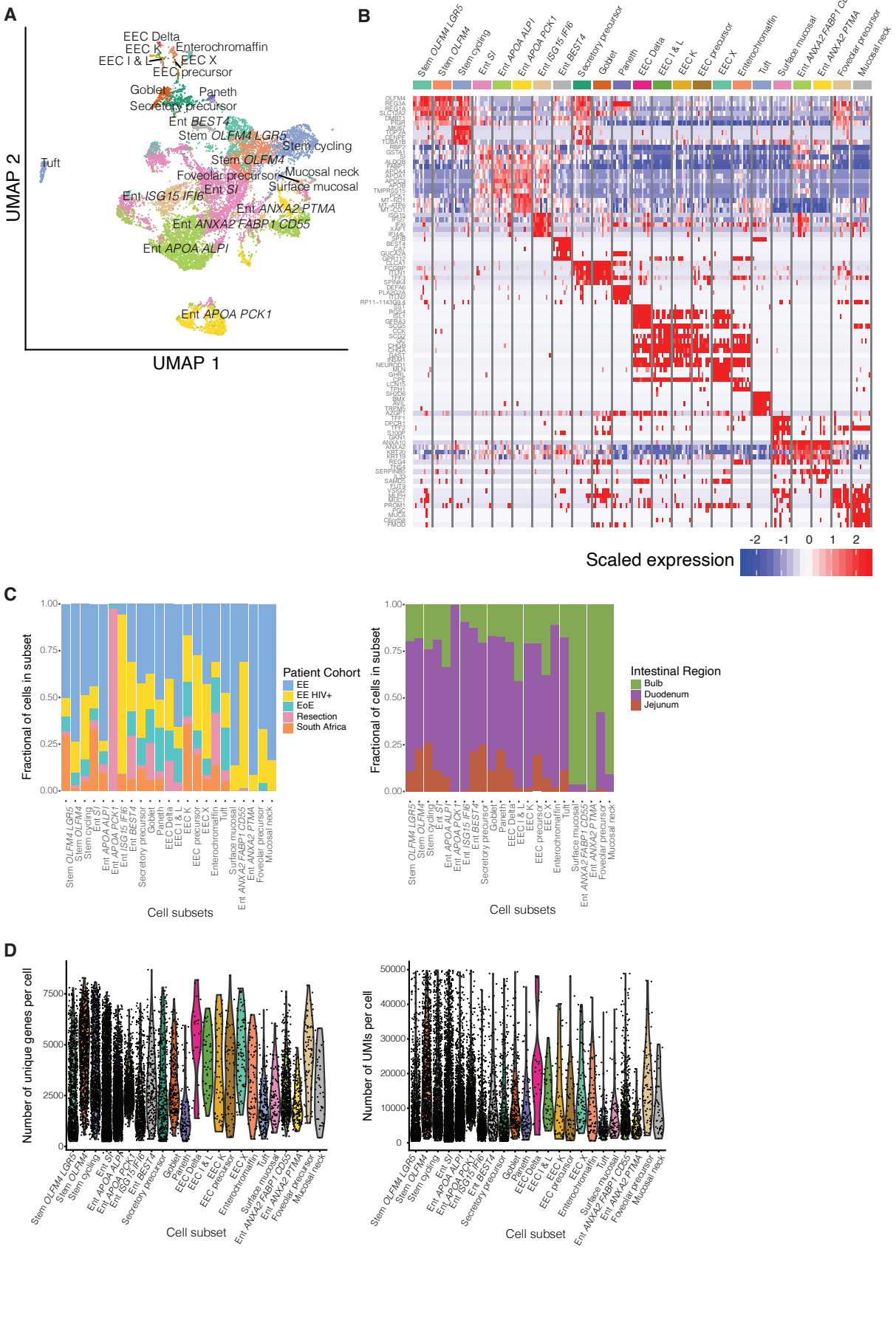

A
EEC Delta
EEC I
Enterochromaffin
EEC I & L
EEC X
EEC precursor
Goblet
Paneth
Secretory precursor
Ent BEST4
Stem OLFM4 LGR5
Stem cycling
Stem OLFM4
Tuft
Foveolar precursor
Mucosal neck
Ent SI
Surface mucosal
Ent ISG15 IFI6
Ent ANXA2 PTMA
Ent ANXA2 FABP1 CD55
Ent APOA ALPI
Ent APOA PCK1
UMAP 2
UMAP 1
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
-2
-1
0
1
2
Scaled expression
C
Fractional of cells in subset
1.00
0.75
0.50
0.25
0.00
Patient Cohort
EE
EE HIV+
EoE
Resection
South Africa
Cell subsets
Intestinal Region
Bulb
Duodenum
Jejunum
D
Number of unique genes per cell
7500
5000
2500
0
Cell subset
Number of UMIs per cell
50000
40000
30000
20000
10000
0
Cell subset

Fig. S1. Characterization of epithelial subsets.

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.

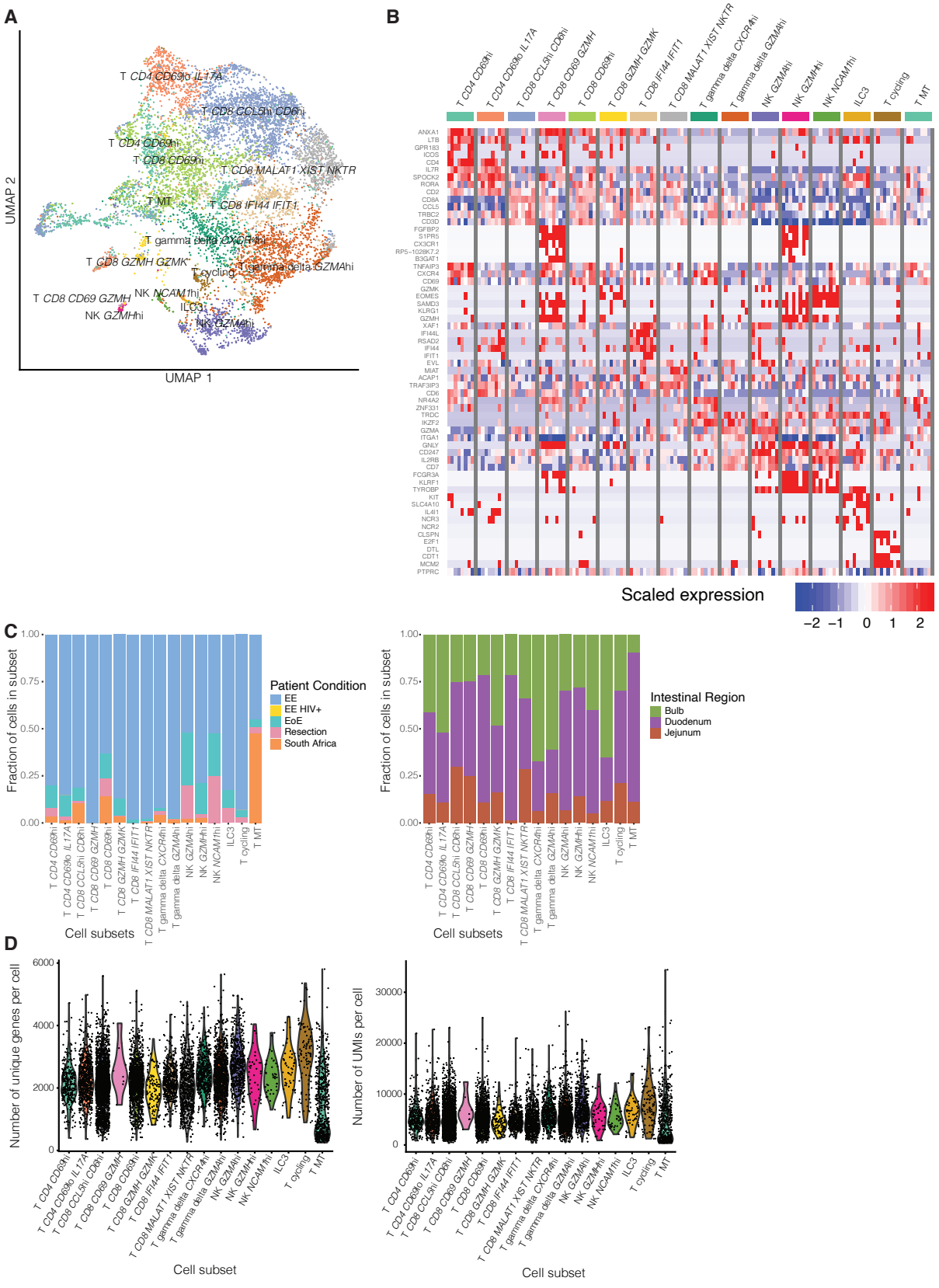

A
UMAP 2
UMAP 1
T CD4 CD69lo IL17A
T CD8 CCL5hi CD6hi
T CD4 CD69hi
T CD8 CD69hi
T CD8 MALAT1 XIST NKTR
T MT
T CD8 IFI44 IFIT1
T gamma delta CXCR4hi
T CD8 GZMH GZMK
T cycling
T gamma delta GZMAhi
T CD8 CD69 GZMH
NK NCAM1hi
ILC3
NK GZMHhi
NK GZMAhi
B
T CD4 CD69hi
T CD4 CD69lo IL17A
T CD8 CCL5hi CD6hi
T CD8 CD69 GZMH
T CD8 CD69hi
T CD8 GZMH GZMK
T CD8 IFI44 IFIT1
T CD8 MALAT1 XIST NKTR
T gamma delta CXCR4hi
T gamma delta GZMAhi
NK GZMHhi
NK NCAM1hi
ILC3
T cycling
T MT
ANXA1
LTB
GPR183
ICOS
CD4
IL7R
SPOCK2
RORA
CD2
CD8A
CCL5
TRBC2
CD3D
FGFBP2
S1PR5
CX3CR1
RP5−1028K7.2
B3GAT1
TNFAIP3
CXCR4
CD69
GZMK
EOMES
SAMD3
KLRG1
GZMH
XAF1
IFI44L
RSAD2
IFI44
IFIT1
EVL
MIAT
ACAP1
TRAF3IP3
CD6
NR4A2
ZNF331
TRDC
IKZF2
GZMA
ITGA1
GNLY
CD247
IL2RB
CD7
FCGR3A
KLRF1
TYROBP
KIT
SLC4A10
IL4I1
NCR3
NCR2
CLSPN
E2F1
DTL
CDT1
MCM2
PTPRC
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
South Africa
Cell subsets
Intestinal Region
Bulb
Duodenum
Jejunum
D
Number of unique genes per cell
6000
4000
2000
0
Cell subset
Number of UMIs per cell
30000
20000
10000
0

Fig. S2. Characterization of T and NK cell subsets.

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.

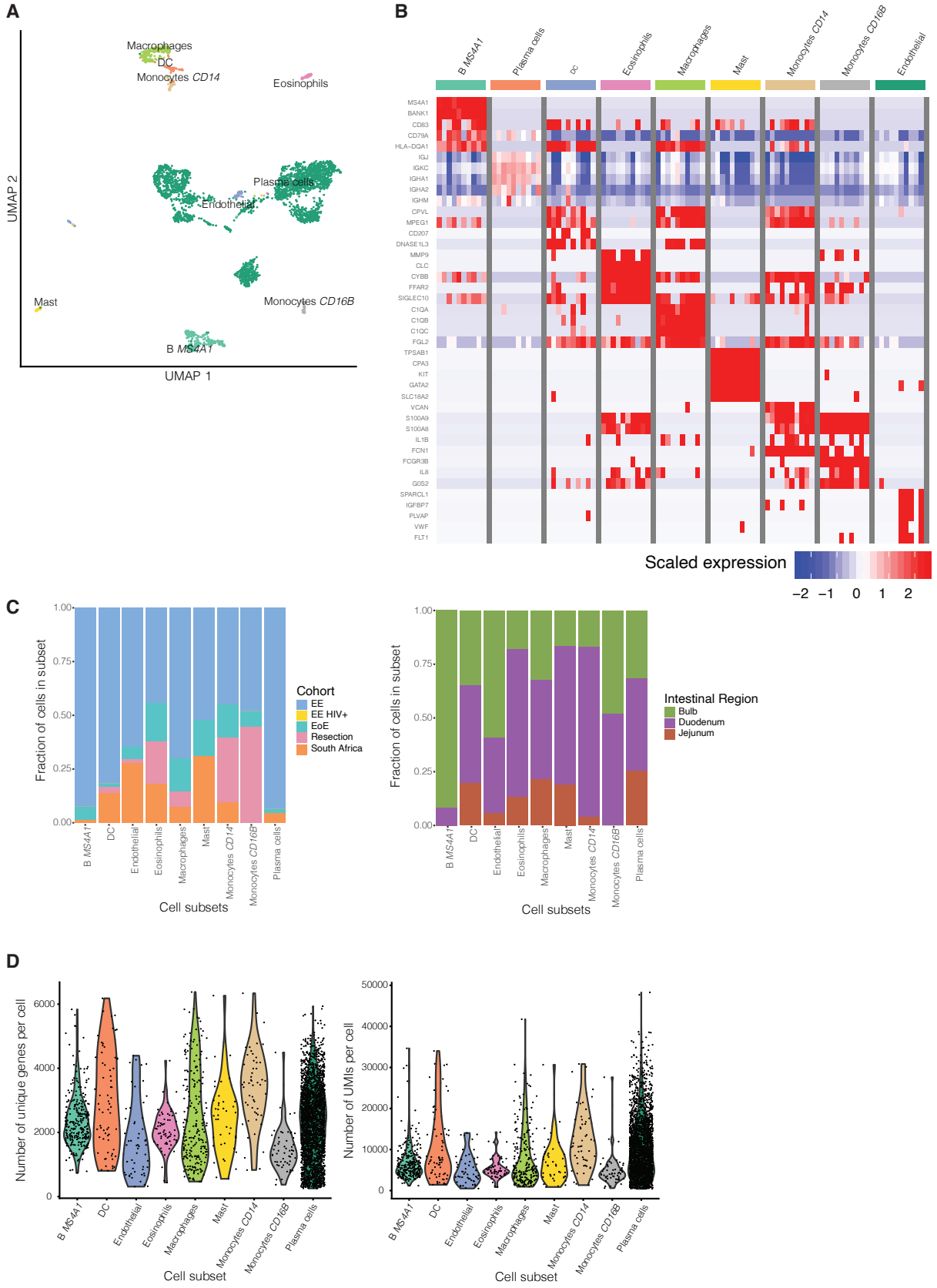

A
Macrophages
DC
Monocytes CD14
Eosinophils
Plasma cells
Endothelial
Mast
Monocytes CD16B
B MS4A1
UMAP 2
UMAP 1
B
B MS4A1
Plasma cells
DC
Eosinophils
Macrophages
Mast
Monocytes CD14
Monocytes CD16B
Endothelial
MS4A1
BANK1
CD83
CD79A
HLA-DQA1
IGJ
IGKC
IGHA1
IGHA2
IGHM
CPVL
MPEG1
CD207
DNASE1L3
MMP9
CLC
CYBB
FFAR2
SIGLEC10
C1QA
C1QB
C1QC
FGL2
TPSAB1
CPA3
KIT
GATA2
SLC18A2
VCAN
S100A9
S100A8
IL1B
FCN1
FCGR3B
IL8
G0S2
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
Cohort
EE
EE HIV+
EoE
Resection
South Africa
B MS4A1
DC
Endothelial
Eosinophils
Macrophages
Mast
Monocytes CD14
Monocytes CD16B
Plasma cells
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

Fig. S3. Characterization of B cell, myeloid, and stromal subsets.

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.

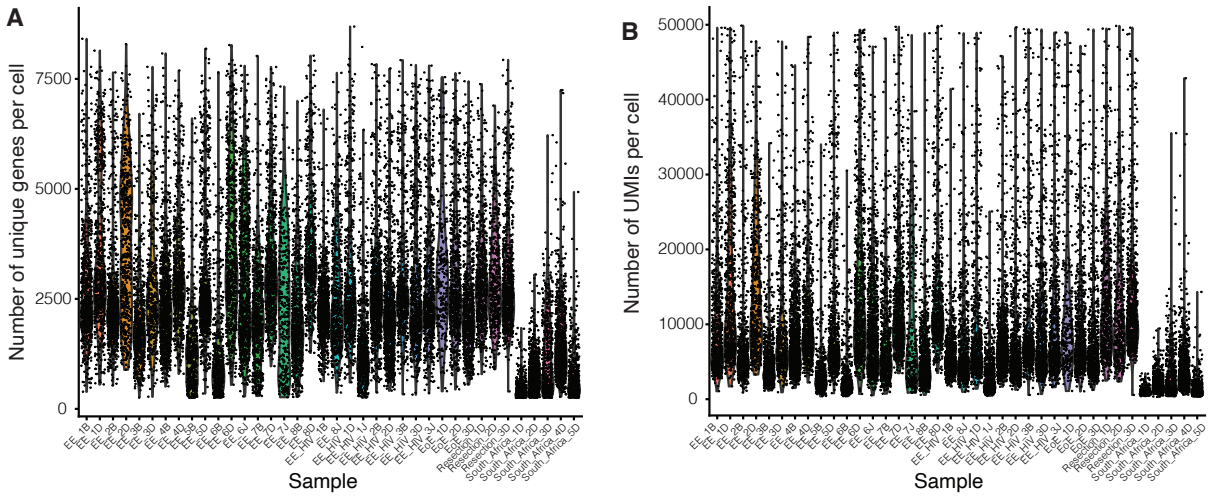

A
Number of unique genes per cell
7500
5000
2500
0
Sample
B
Number of UMIs per cell
50000
40000
30000
20000
10000
0
Sample

Fig. 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

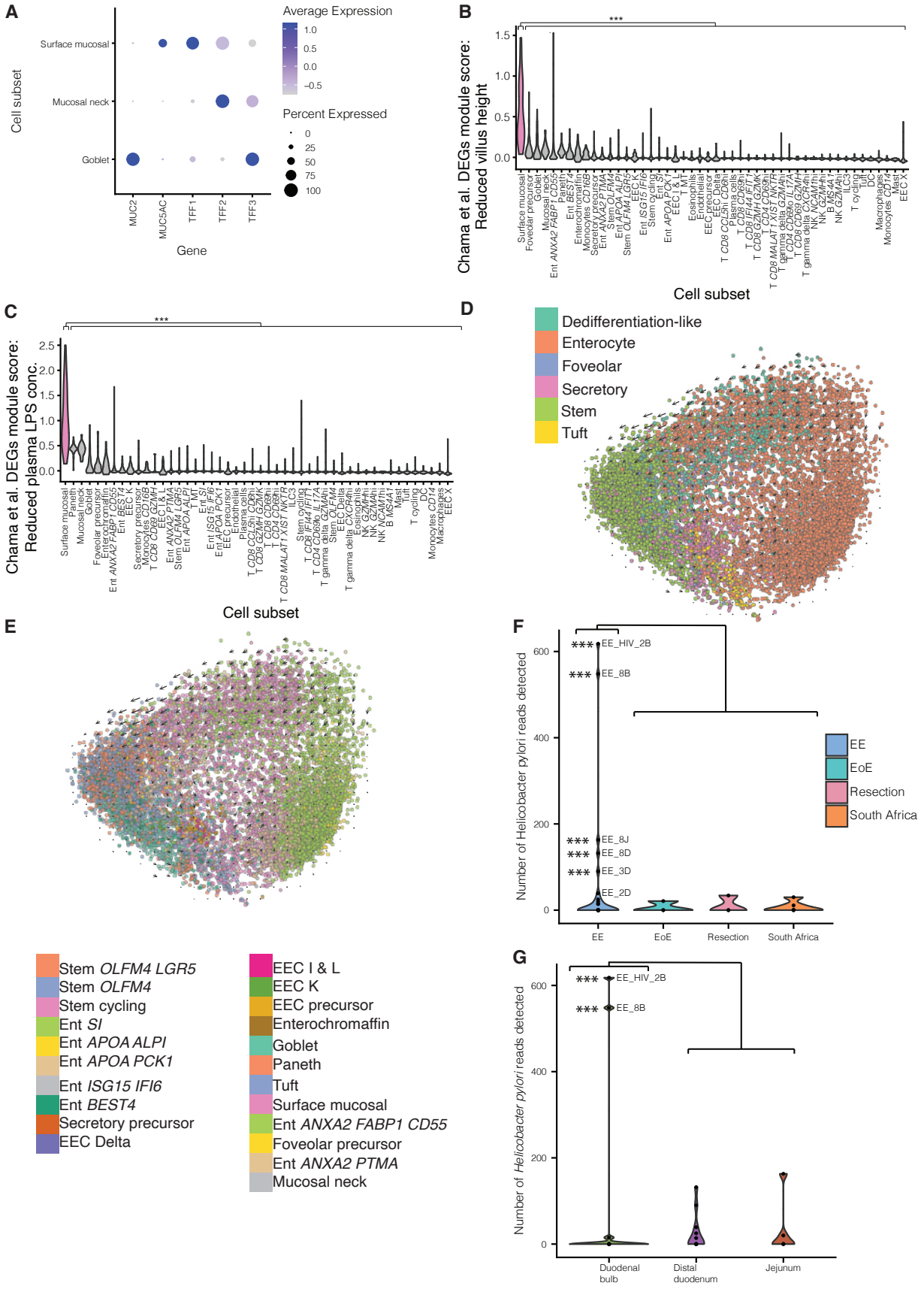

A
Average Expression
1.0
0.5
0.0
−0.5
Percent Expressed
0
25
50
75
100
Surface mucosal
Mucosal neck
Goblet
Cell subset
MUC2
MUC5AC
TFF1
TFF2
TFF3
Gene
B
Chama et al. DEGs module score:
Reduced villus height
***
Cell subset
C
Chama et al. DEGs module score:
Reduced plasma LPS conc.
***
Cell subset
D
Dedifferentiation-like
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
F
Number of Helicobacter pylori reads detected
EE_HIV_2B
EE_8B
EE_8J
EE_8D
EE_3D
EE_2D
EE
EoE
Resection
South Africa
G
Number of Helicobacter pylori reads detected
EE_HIV_2B
EE_8B
Duodenal bulb
Distal duodenum
Jejunum

Fig. 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

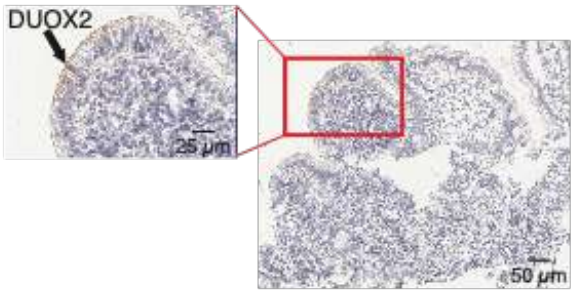
E, Velocyto results grouped by epithelial subset

F, Violin plots by participant cohort of the number of *H. pylori* reads detected by metagenomic alignment with Kraken2 in the sequencing reads from each sample (T-test: *,adj. $p < 0.05$; **, adj. $p < 0.01$; ***, adj. $p < 0.001$)..

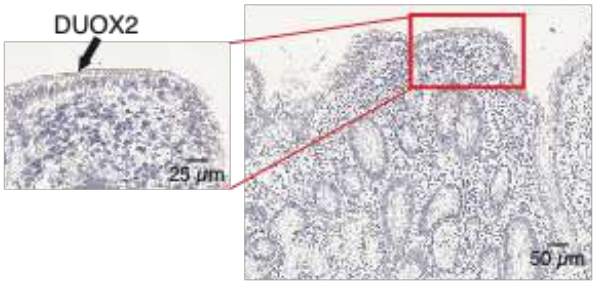
G, Violin plots by intestinal region of the number of *H. pylori* reads detected by metagenomic alignment with Kraken2 in the sequencing reads from each sample from participants in the Zambian EE cohort (T-test: *,adj. $p < 0.05$; **, adj. $p < 0.01$; ***, adj. $p < 0.001$).

A)

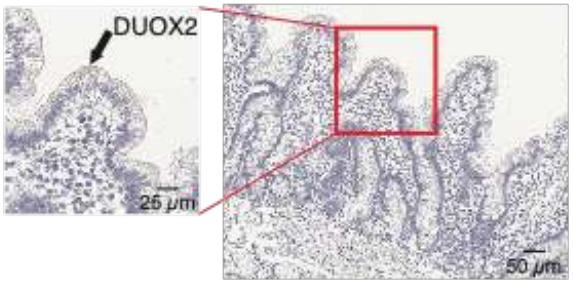
EE 7B: DUOX2+



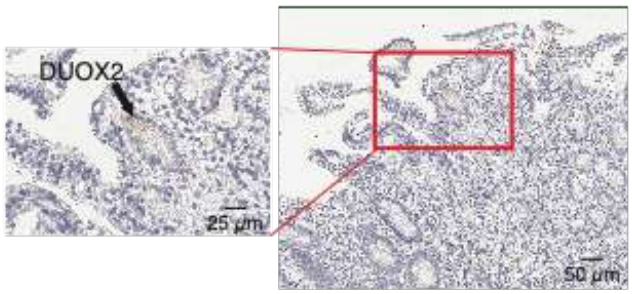
EE HIV 3B: DUOX2+



EE 6B: DUOX2+

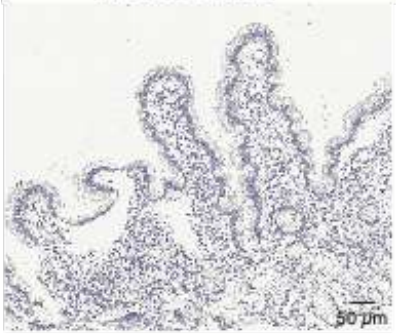


EE 5B: DUOX2+

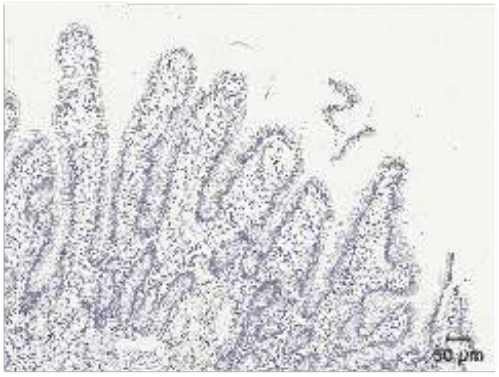


B)

EE 4B: DUOX2-



EE 2B: DUOX2-



EE 8B: DUOX2-

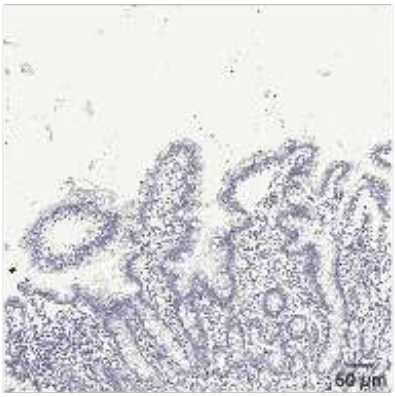


Fig. S6. Immunohistochemical staining for DUOX2 protein.

A, H&E images (purple, H&E) that stained positive for DUOX2 (brown, DUOX2) from duodenal bulb samples from participants in the Zambian EE cohort

B, H&E images (purple, H&E) that did not stain positive for DUOX2 (brown, DUOX2) from duodenal bulb samples from participants in the Zambian EE cohort

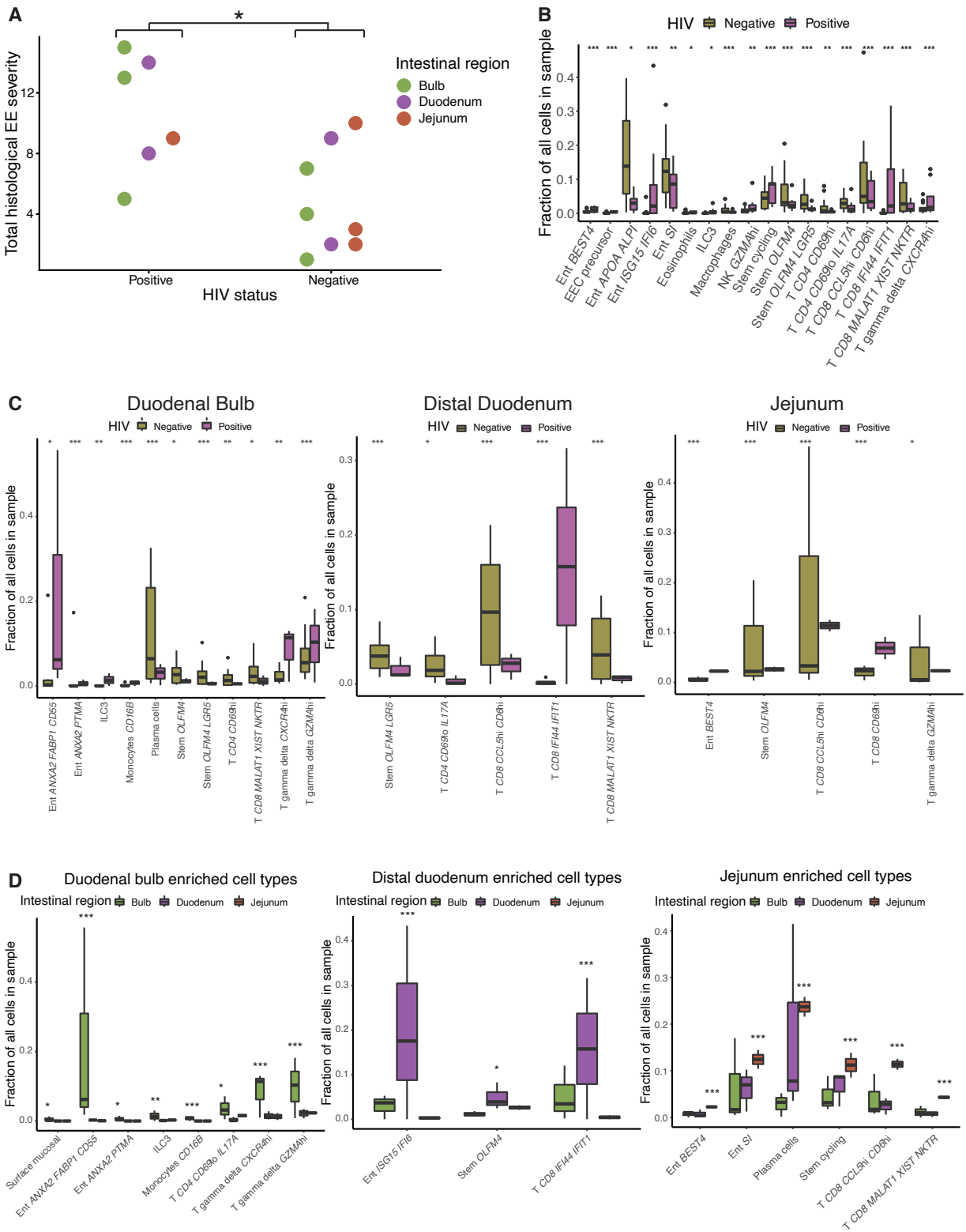

A
Total histological EE severity
HIV status
Positive
Negative
Intestinal region
Bulb
Duodenum
Jejunum
B
HIV
Negative
Positive
Fraction of all cells in sample
Ent BEST4
EEC precursor
Ent APOA ALPI
Ent ISG15 IFI6
Ent SI
Eosinophils
ILC3
Macrophages
NK GZMAhi
Stem cycling
Stem OLFM4
Stem OLFM4 LGR5
T CD4 CD69hi
T CD4 CD69lo IL17A
T CD8 CCL5hi CD8hi
T CD8 IFI44 IFIT1
T CD8 MALAT1 XIST NKTR
T gamma delta CXCR4hi
C
Duodenal Bulb
Distal Duodenum
Jejunum
D
Duodenal bulb enriched cell types
Distal duodenum enriched cell types
Jejunum enriched cell types
Surface mucosal

Fig. S7. 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 ($p < 0.05$; Wilcoxon test).

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. $p < 0.001$).

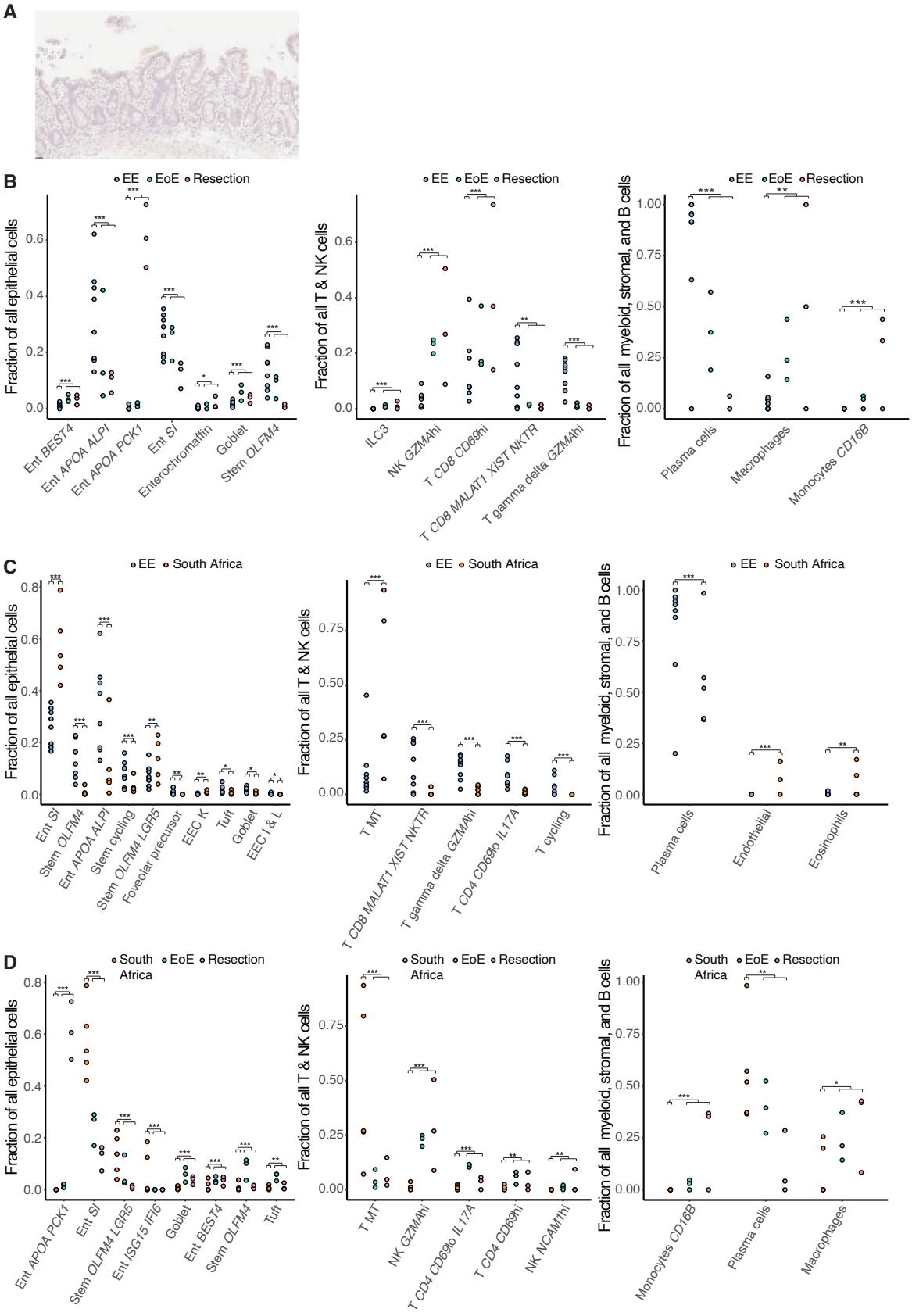

A
B
EE
EoE
Resection
Fraction of all epithelial cells
Ent BEST4
Ent APOA ALPI
Ent APOA PCK1
Ent SI
Enterochromaffin
Goblet
Stem OLFM4
Fraction of all T & NK cells
ILC3
NK GZMAhi
T CD8 CD69hi
T CD8 MALAT1 XIST NKTR
T gamma delta GZMAhi
Fraction of all myeloid, stromal, and B cells
Plasma cells
Macrophages
Monocytes CD16B
C
EE
South Africa
Ent SI
Stem OLFM4
Ent APOA ALPI
Stem cycling
Stem OLFM4 LGR5
Foveolar precursor
EEC K
Tuft
Goblet
EEC I & L
T MT
T CD8 MALAT1 XIST NKTR
T gamma delta GZMAhi
T CD4 CD69lo IL17A
T cycling
Plasma cells
Endothelial
Eosinophils
D
South Africa
EoE
Resection
Ent APOA PCK1
Ent SI
Stem OLFM4 LGR5
Ent ISG15 IFI6
Goblet
Ent BEST4
Stem OLFM4
Tuft
T MT
NK GZMAhi
T CD4 CD69lo IL17A
T CD4 CD69hi
NK NCAM1hi
Monocytes CD16B
Plasma cells
Macrophages

Fig. S8. Samples from South African participants display features of EE.

A, H&E staining from a biopsy taken from a participant at the same clinical site in South Africa where the participants in this study were profiled.

B, Cell subsets with significant shifts in relative abundances between the Zambian EE cohort and both U.S. cohorts (*,adj. $p < 0.05$; **, adj. $p < 0.01$; ***, adj. $p < 0.001$; Fischer's exact test).

C, Cell subsets with significant shifts in relative abundances between the Zambian EE cohort and the South African cohort (*,adj. $p < 0.05$; **, adj. $p < 0.01$; ***, adj. $p < 0.001$; Fischer's exact test).

D, Cell subsets with significant shifts in relative abundances between the South African cohort and both U.S. cohorts (*,adj. $p < 0.05$; **, adj. $p < 0.01$; ***, adj. $p < 0.001$; Fischer's exact test).

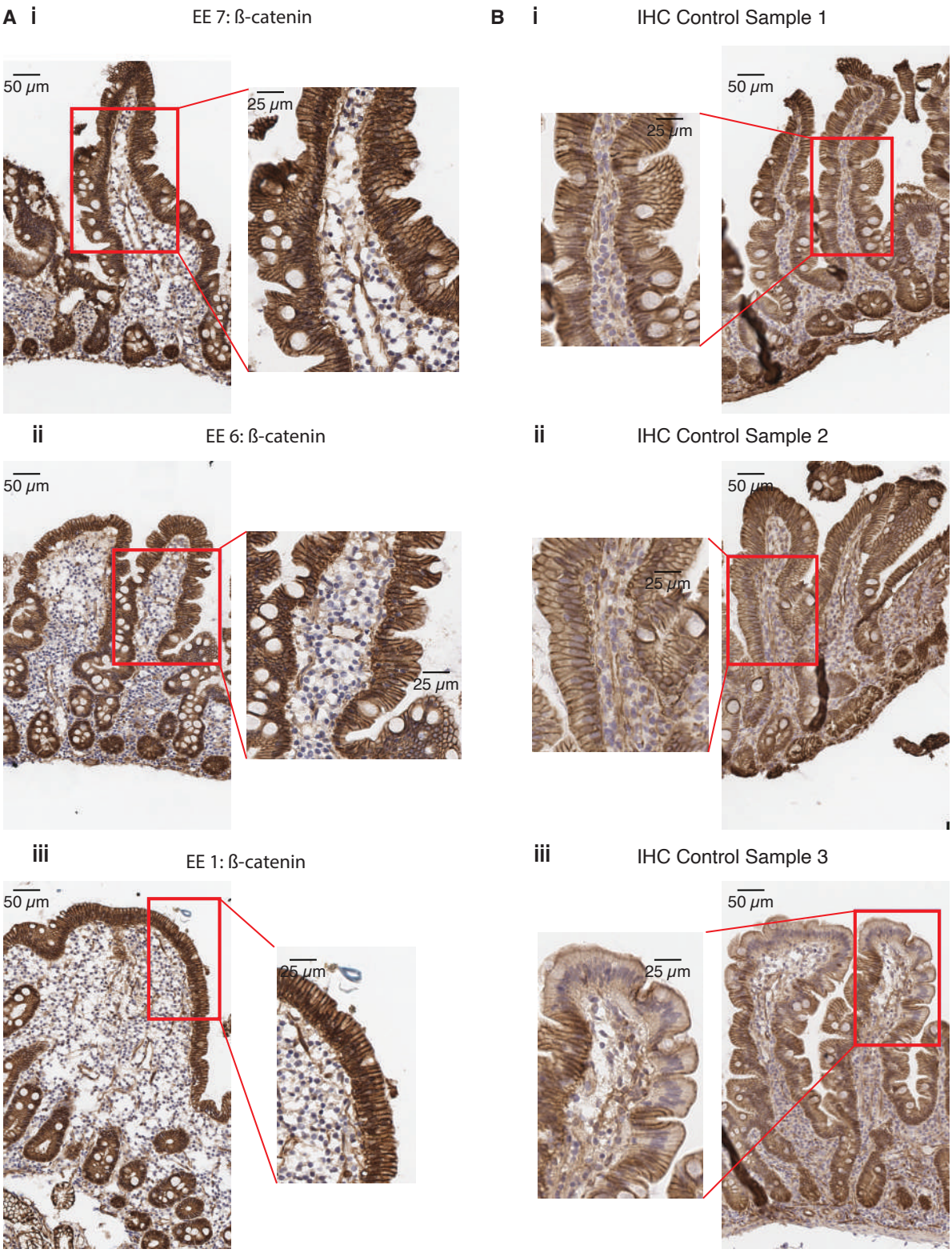

A i
EE 7: ß-catenin
50 μm
25 μm
ii
EE 6: ß-catenin
50 μm
25 μm
iii
EE 1: ß-catenin
50 μm
25 μm
B i
IHC Control Sample 1
25 μm
50 μm
ii
IHC Control Sample 2
25 μm
50 μm
iii
IHC Control Sample 3
25 μm
50 μm

Fig. S9. Immunohistochemical staining for ß-catenin.

A, H&E (purple) images stained for ß-catenin (brown) from distal duodenal biopsies from HIV-negative participants with EE

B, H&E (purple) images stained for ß-catenin (brown) from normal intestine from the U.S.

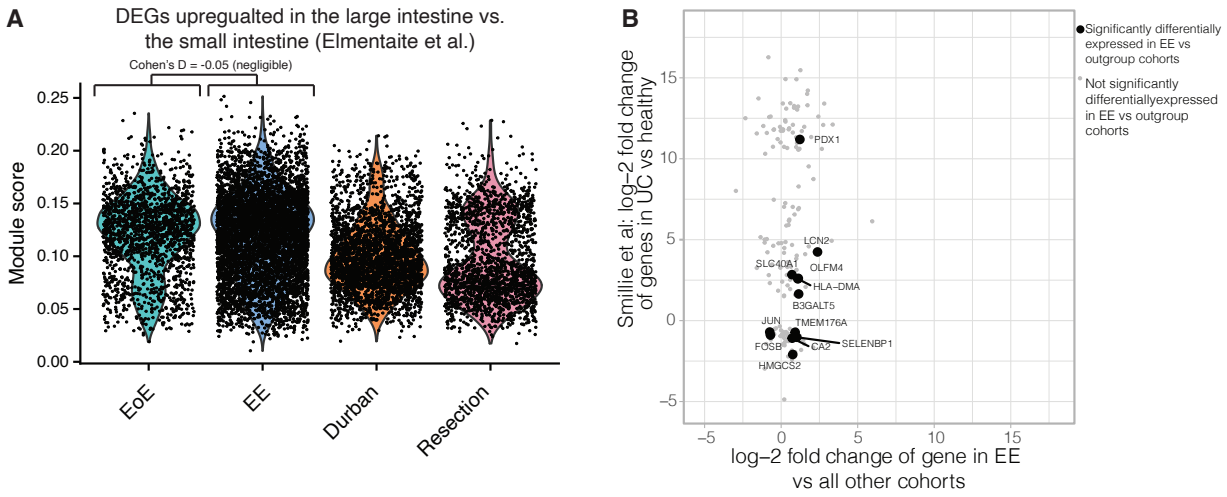
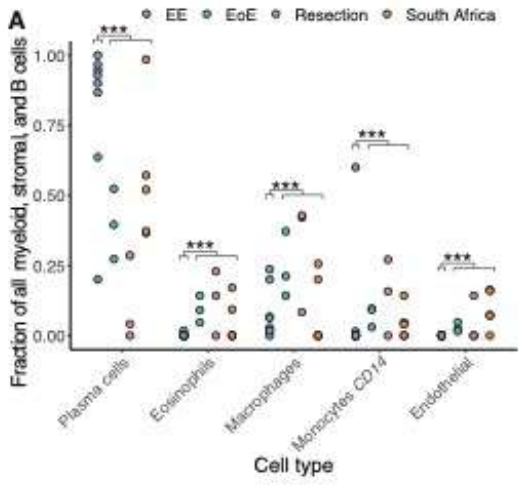
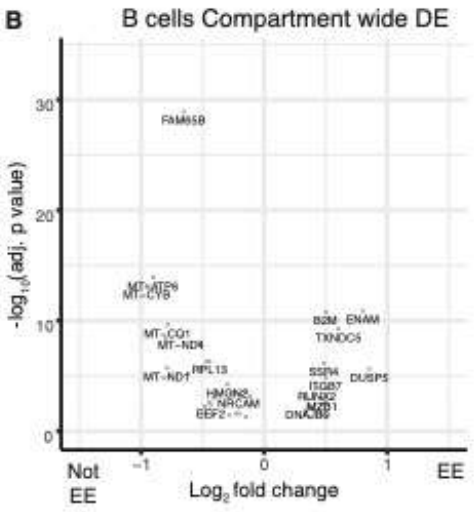

A
DEGs upregulated in the large intestine vs. the small intestine (Elmentaite et al.)
Cohen's D = -0.05 (negligible)
Module score
0.25
0.20
0.15
0.10
0.05
0.00
EoE
EE
Durban
Resection
B
Smillie et al: log−2 fold change of genes in UC vs healthy
log−2 fold change of gene in EE vs all other cohorts
15
10
5
0
−5
PDX1
LCN2
SLC40A1
OLFM4
HLA−DMA
B3GALT5
JUN
TMEM176A
FOSB
CA2
SELENBP1
HMGCS2
Significantly differentially expressed in EE vs outgroup cohorts
Not significantly differentiallyexpressed in EE vs outgroup cohorts

Fig. S10. Contextualizing EE epithelial cells with existing intestinal scRNA-seq signatures.
A, Violin plots of module scores for a set of genes enriched in the large intestine relative to the small intestine for all HIV-negative distal duodenal samples from each participant cohort.
B, Scatterplot of the genes differentially expressed in the EE epithelium relative to all three control cohorts and the genes differentially expressed in the large intestine with ulcerative colitis relative to the healthy large intestine.


A
EE
EoE
Resection
South Africa
Fraction of all myeloid, stromal, and B cells
1.00
0.75
0.50
0.25
0.00
***
Plasma cells
Eosinophils
Macrophages
Monocytes CD14
Endothelial
Cell type


B
B cells Compartment wide DE
-log10(adj. p value)
30
20
10
0
FAM65B
MT-ATP6
MT-CYB
MT-CO1
MT-ND4
MT-ND1
RPL13
B2M
ENAM
TXNDC5
SSR4
DUSP5
ISG20
RLN2
MZB1
DNAJB9
HMGN3
NRCAM
EEF2
Not EE
−1
0
1
EE
Log2 fold change

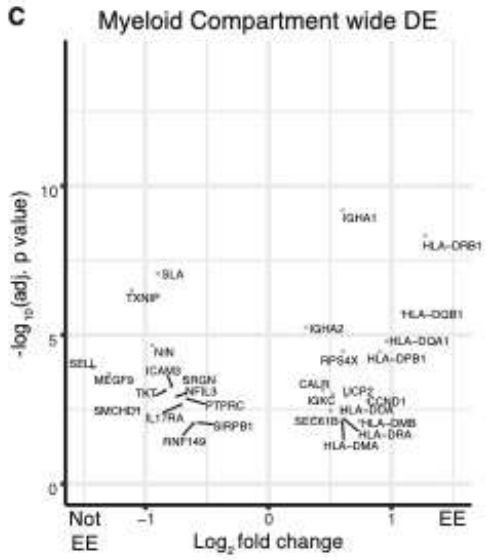

C
Myeloid Compartment wide DE
-log10(adj. p value)
10
5
0
IGHA1
HLA-DRB1
SLA
TXNIP
HLA-DQB1
IGHA2
HLA-DQA1
RPS4X
HLA-DPB1
SELL
NIN
MEGF9
ICAM3
SRGN
TKT
AFTL3
PTPRC
SMCHD1
IL17RA
SIRPB1
RNF149
CALR
KIAA1551
UCP2
CCND1
HLA-DOA
SEC61B
HLA-DMB
HLA-DRA
HLA-DMA
Not EE
−1
0
1
EE
Log2 fold change

Fig. S11. Analysis of B cells and myeloid cells between EE and controls.

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

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.

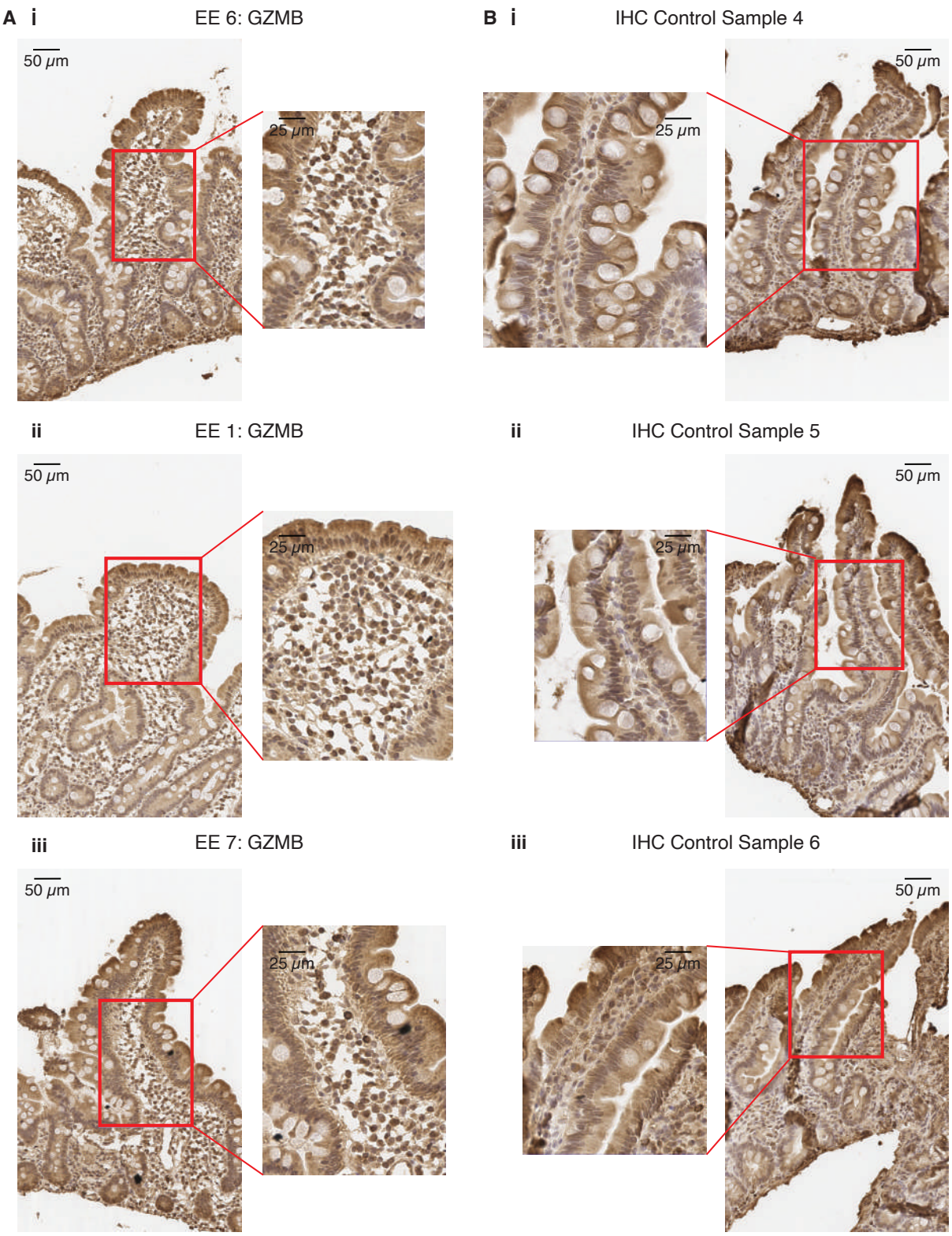

A i
EE 6: GZMB
50 μm
25 μm
ii
EE 1: GZMB
50 μm
25 μm
iii
EE 7: GZMB
50 μm
25 μm
B i
IHC Control Sample 4
25 μm
50 μm
ii
IHC Control Sample 5
25 μm
50 μm
iii
IHC Control Sample 6
25 μm
50 μm

Fig. S12. Immunohistochemical staining for GZMB protein.

A, H&E (purple) images stained for GZMB (brown) from distal duodenal biopsies from HIV-negative participants with EE

B, H&E (purple) images stained for GZMB (brown) from normal intestine from the U.S.

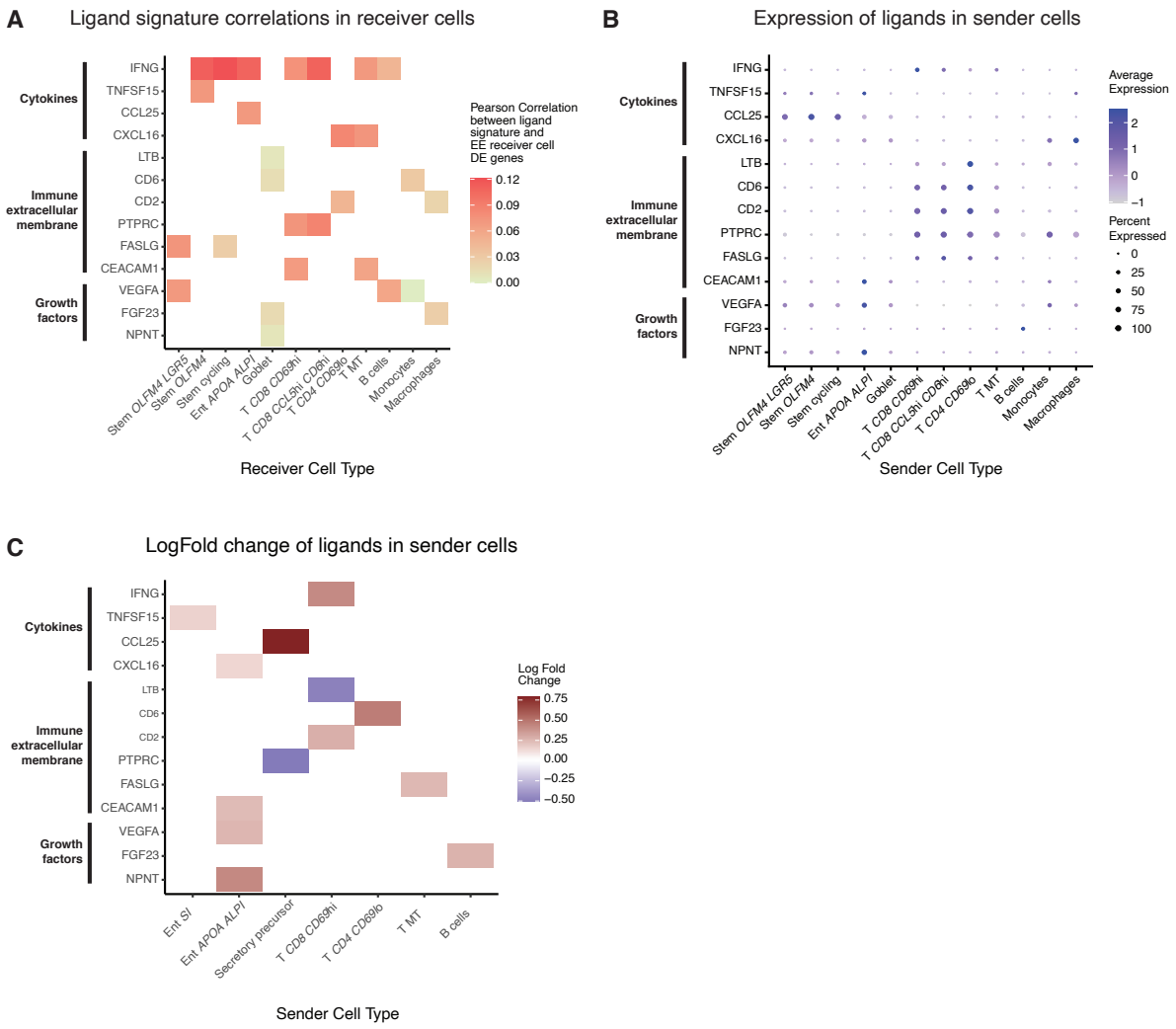

A
Ligand signature correlations in receiver cells
Cytokines
Immune extracellular membrane
Growth factors
IFNG
TNFSF15
CCL25
CXCL16
LTB
CD6
CD2
PTPRC
FASLG
CEACAM1
VEGFA
FGF23
NPNT
Stem OLFM4 LGR5
Stem OLFM4
Stem cycling
Ent APOA ALPI
Goblet
T CD8 CD69hi
T CD8 CCL5hi CD8hi
T CD4 CD69lo
T MT
B cells
Monocytes
Macrophages
Receiver Cell Type
Pearson Correlation between ligand signature and EE receiver cell DE genes
0.12
0.09
0.06
0.03
0.00
B
Expression of ligands in sender cells
Sender Cell Type
Average Expression
2
1
0
−1
Percent Expressed
0
25
50
75
100
C
LogFold change of ligands in sender cells
Ent SI
Ent APOA ALPI
Secretory precursor
Log Fold Change
0.75
0.50
0.25
0.00
−0.25
−0.50

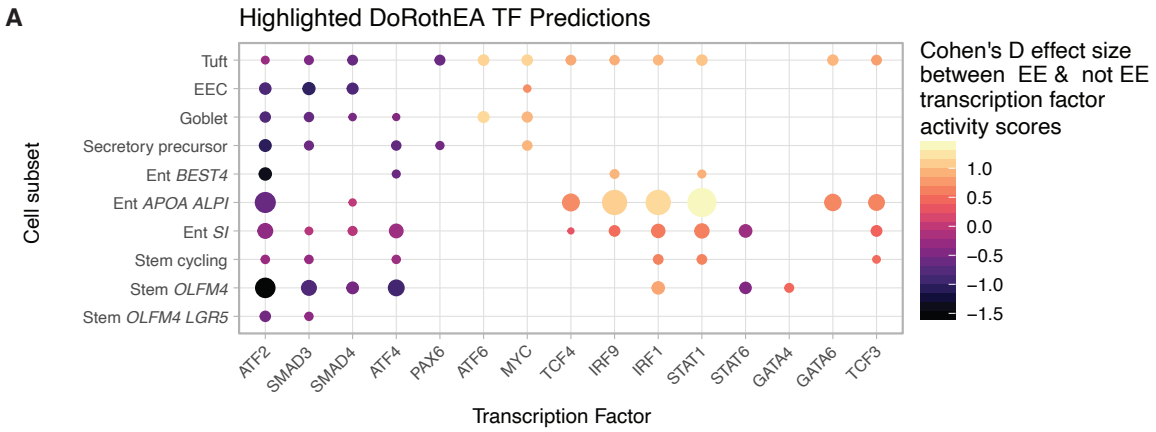
Fig. S13. NicheNet analysis of cell-cell signaling.

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

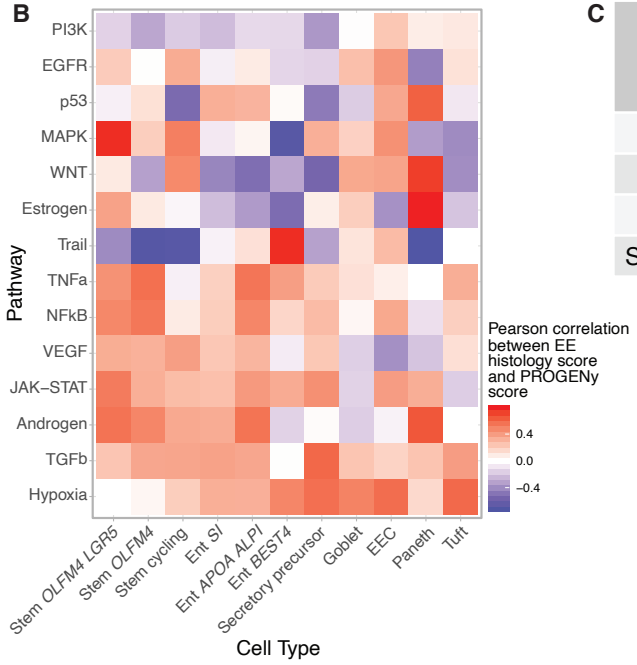
B, Expression levels of the genes for NicheNet ligands in all cell subsets.

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

A



B



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

Fig. S14. Epithelial signaling changes in the EE cohort relative to the U.S. cohorts.

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.

SUPPLEMENTARY TABLES

Tables S1-S9 have been uploaded as excel files in the Auxiliary Supplementary Materials section

Table S1. Clinical characteristics of donors of intestinal biopsy samples.

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

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 EE relative to control cohorts.

Table S7. KEGG, REACTOME, and PID pathway analysis.

Table S8. PROGENY pathway activation scores.

Table S9. Transcription factors with upstream activity predictions from DoRothEA.