Supplementary Information

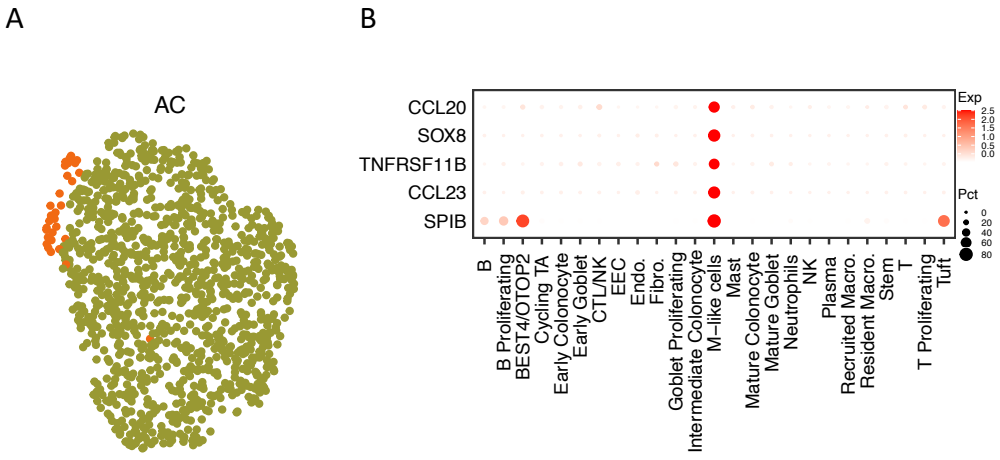
Identification and multimodal characterization of a specialized epithelial cell type associated with Crohn's Disease

Jia Li[1,2], Alan J. Simmons[3,4], Caroline V. Hawkins[5], Sophie Chiron[5], Marisol A. Ramirez-Solano[1,2], Naila Tasneem[3,4], Harsimran Kaur[3,6], Yanwen Xu[3,4], Frank Revetta[7], Paige N. Vega[3,4], Shunxing Bao[8], Can Cui[9], Regina N. Tyree[5], Larry W. Raber[5], Anna N. Conner[5], Jennifer M. Pilat[5], Justin Jacobse[5], Kara M. McNamara[5,10], Margaret M. Allaman[5], Gabriella A. Raffa[5], Alain P. Gobert[5,10,11], Mohammad Asim[5], Jeremy A. Goettel[5, 7,10,12], Yash A. Choksi[5,10,11,12], Dawn B. Beaulieu[5], Robin L. Dalal[5], Sara N. Horst[5], Baldeep S. Pabla[5], Yuankai Huo[8,9], Bennett A. Landman[8,9], Joseph T. Roland[3,13], Elizabeth A. Scoville[5,12], David A. Schwartz[5], M. Kay Washington[6,12], Yu Shyr[1,2], Keith T. Wilson[5, 7,10,11,12]*, Lori A. Coburn[5,10,11,12]*, Ken S. Lau[1,3,4,6,12,13]*, Qi Liu[1,2*]

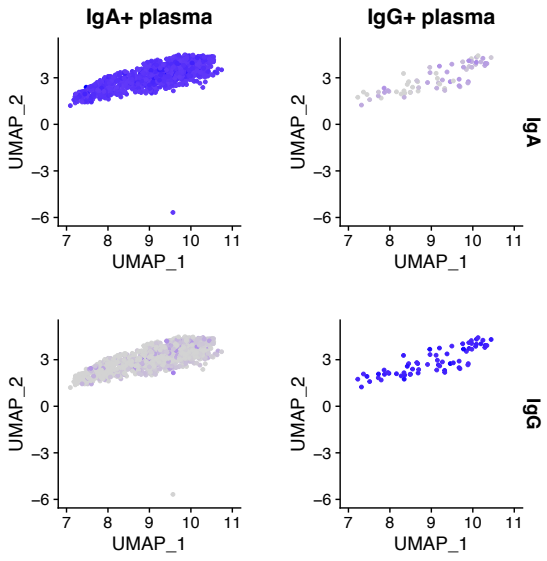
*Corresponding authors:
keith.wilson@vumc.org; lori.coburn@vumc.org; ken.s.lau@vanderbilt.edu; qi.liu@vumc.org

This file includes:

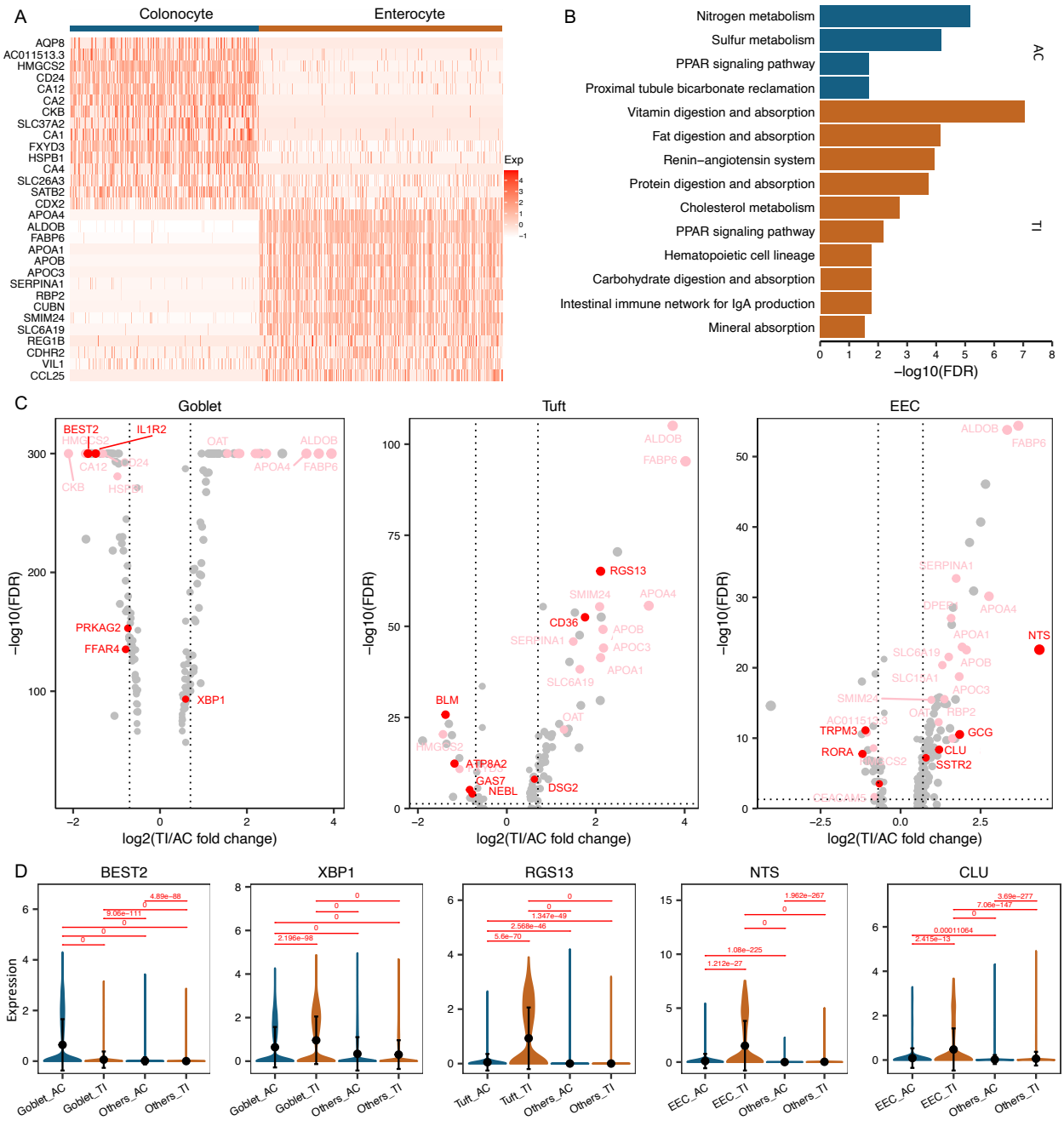
Supplementary Figures 1-26
Supplementary Note 1. Rare cell populations identified in the TI and AC
Supplementary Note 2. Distinct epithelial transcriptional programs between the TI and AC
Supplementary Note 3. Cell-type-specific differential expression between active CD, inactive CD and non-IBD controls
Supplementary Note 4. Ligand-receptor interactions between LND and immune cells



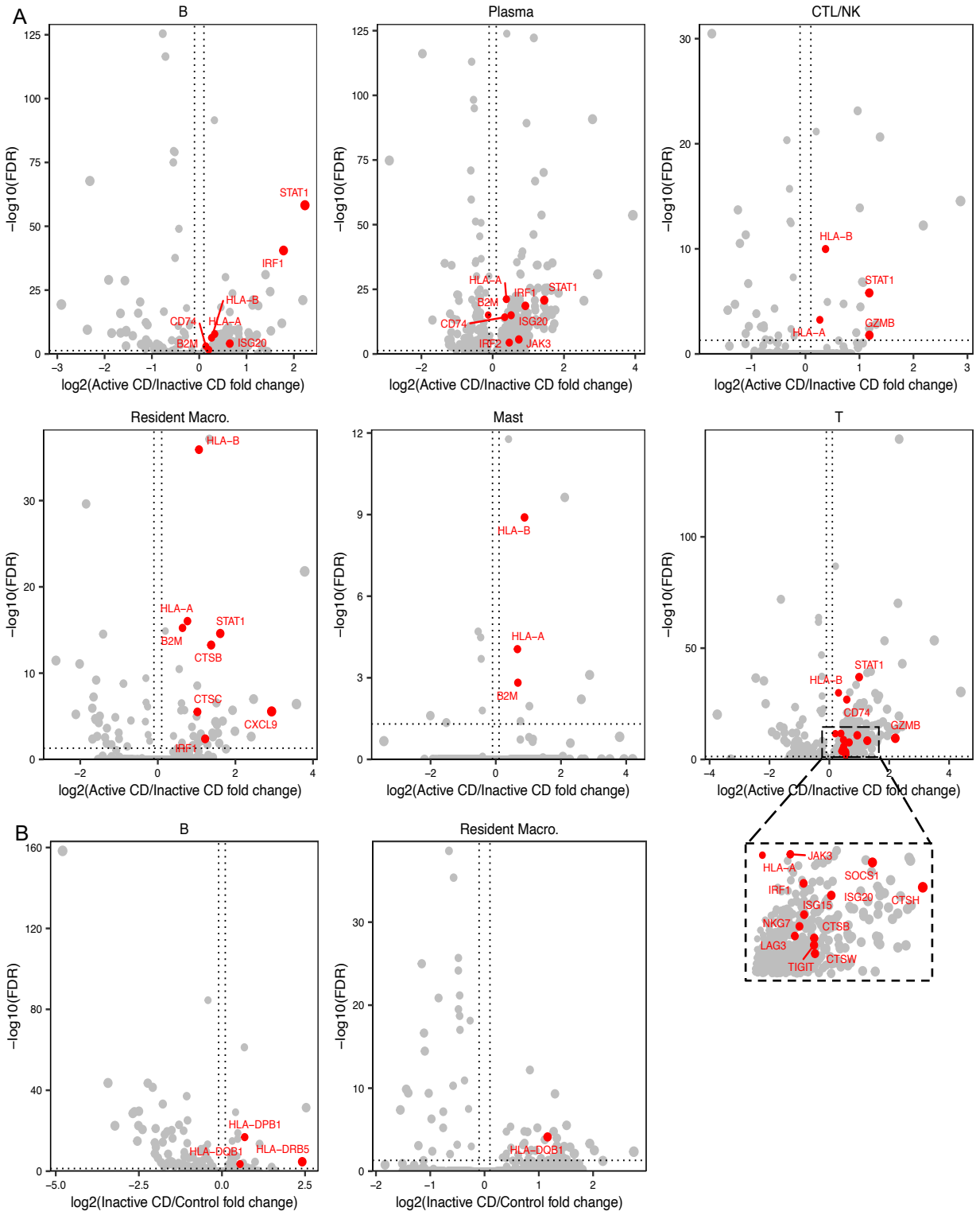
Supplementary Figure 1. **BEST4/OTOP2 and M-like cells.** A) UMAP of BEST4/OTOP2 (brown) and Microfold-like cells (red) in AC. B) Dotplot of markers in colonic Microfold-like cells.



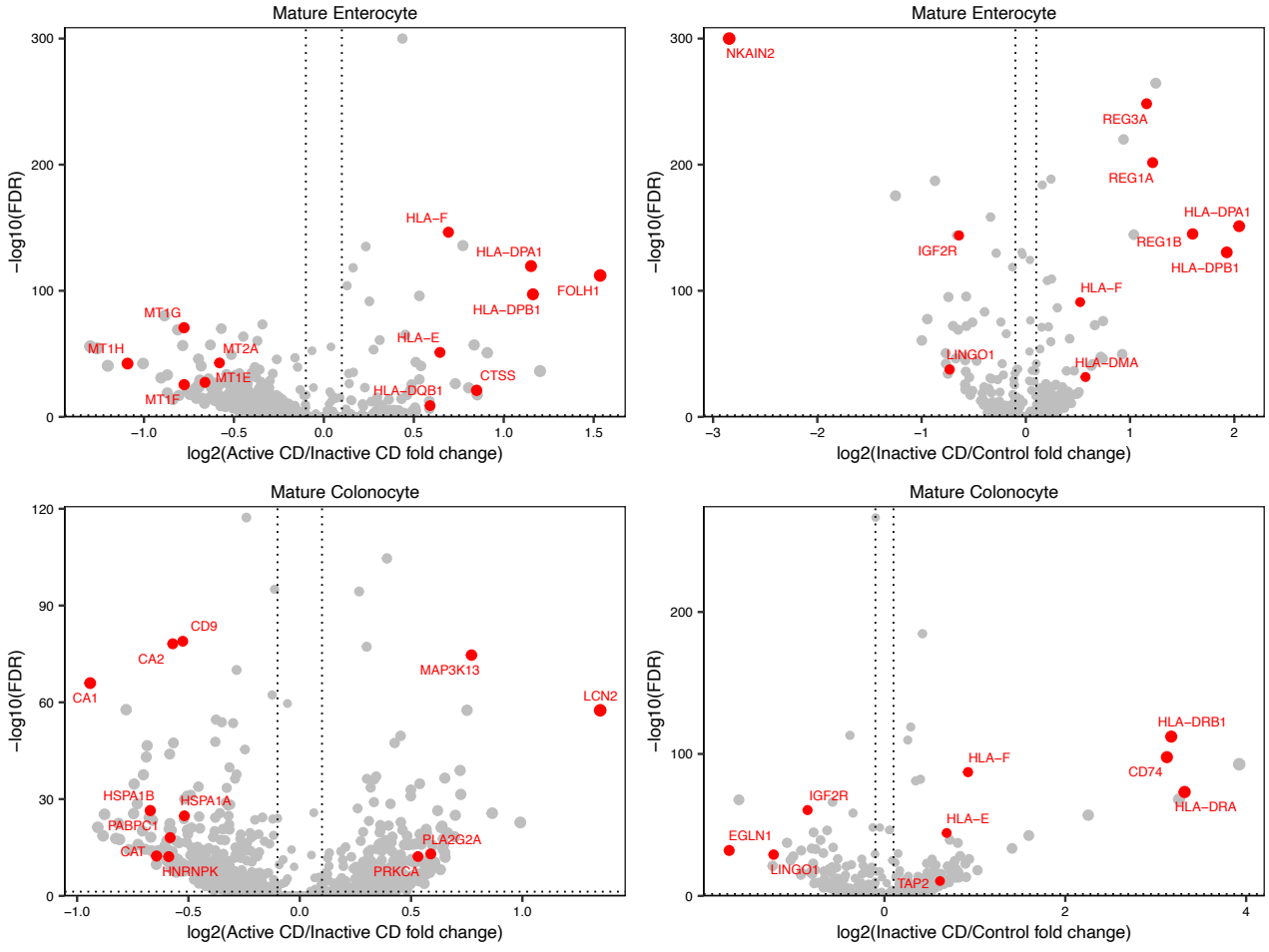
Supplementary Figure 2. **IgA+ and IgG+ plasma cells.** UMAP of expression of IgA (top) and IgG (bottom) heavy chain genes in the IgA+ (left) and IgG+ plasma cells (right).



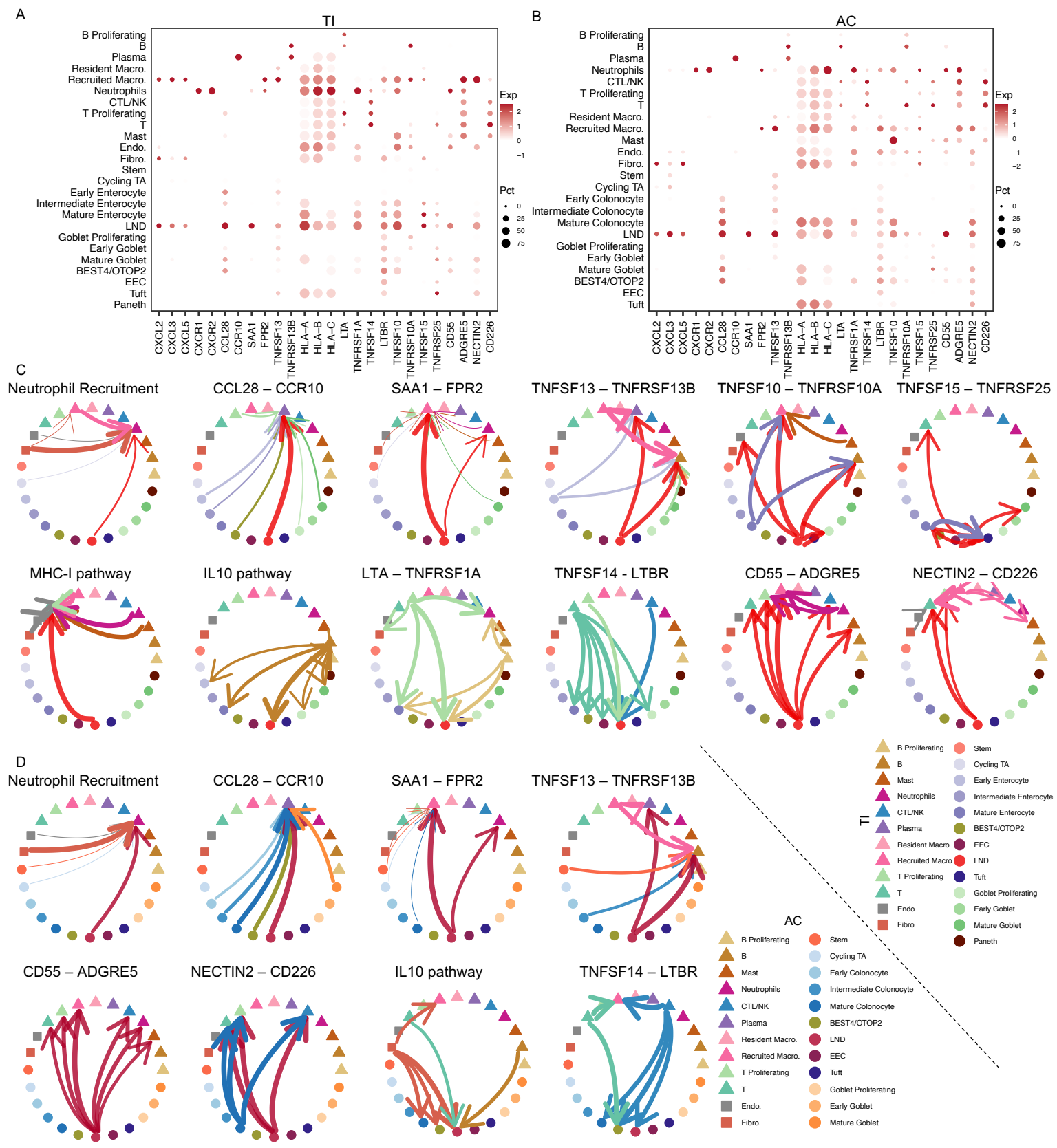
Supplementary Figure 3. Distinct epithelial transcriptional programs between the TI and AC. A) A heatmap of the top 15 differentially expressed genes between AC colonocytes and TI enterocytes. B) Functional enrichment results on differentially expressed genes between AC colonocytes (blue) and TI enterocytes (brown). C) Volcano plots of differential results between TI goblet and AC goblet (left), between TI Tuft and AC Tuft (middle), and between TI EEC and AC EEC (right) (pink denotes region-specific genes, while red denotes genes with both region and cell-type specific); D) Violon plots of gene expression with both region and cell-type specific. AC and goblet-specific gene, BEST2, TI and goblet-specific gene, XBP1, TI and tuft-specific gene, RGS13, TI and EEC-specific gene, NTS, and TI and EEC-specific gene, CLU. Two-sided Wilcoxon test with FDR adjustment. Data are represented as the mean ± SD.



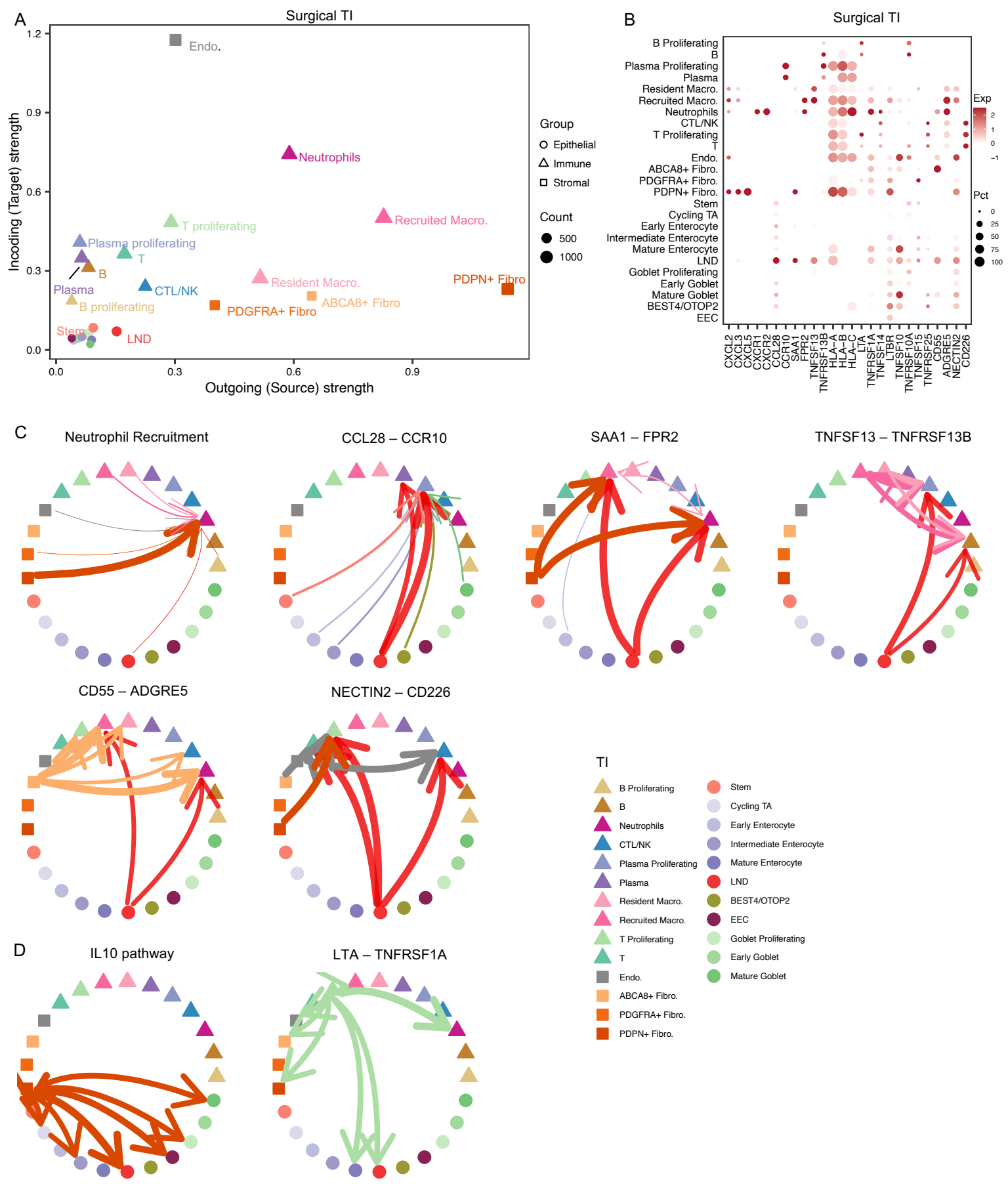
Supplementary Figure 4. **Immune cell-type-specific expression changes between active CD, inactive CD and non-IBD controls.** A) Volcano plots of differentially expressed genes between Active and Inactive CD. B) Volcano plots of differentially expressed genes between Inactive CD and controls in B and Resident Macrophages.



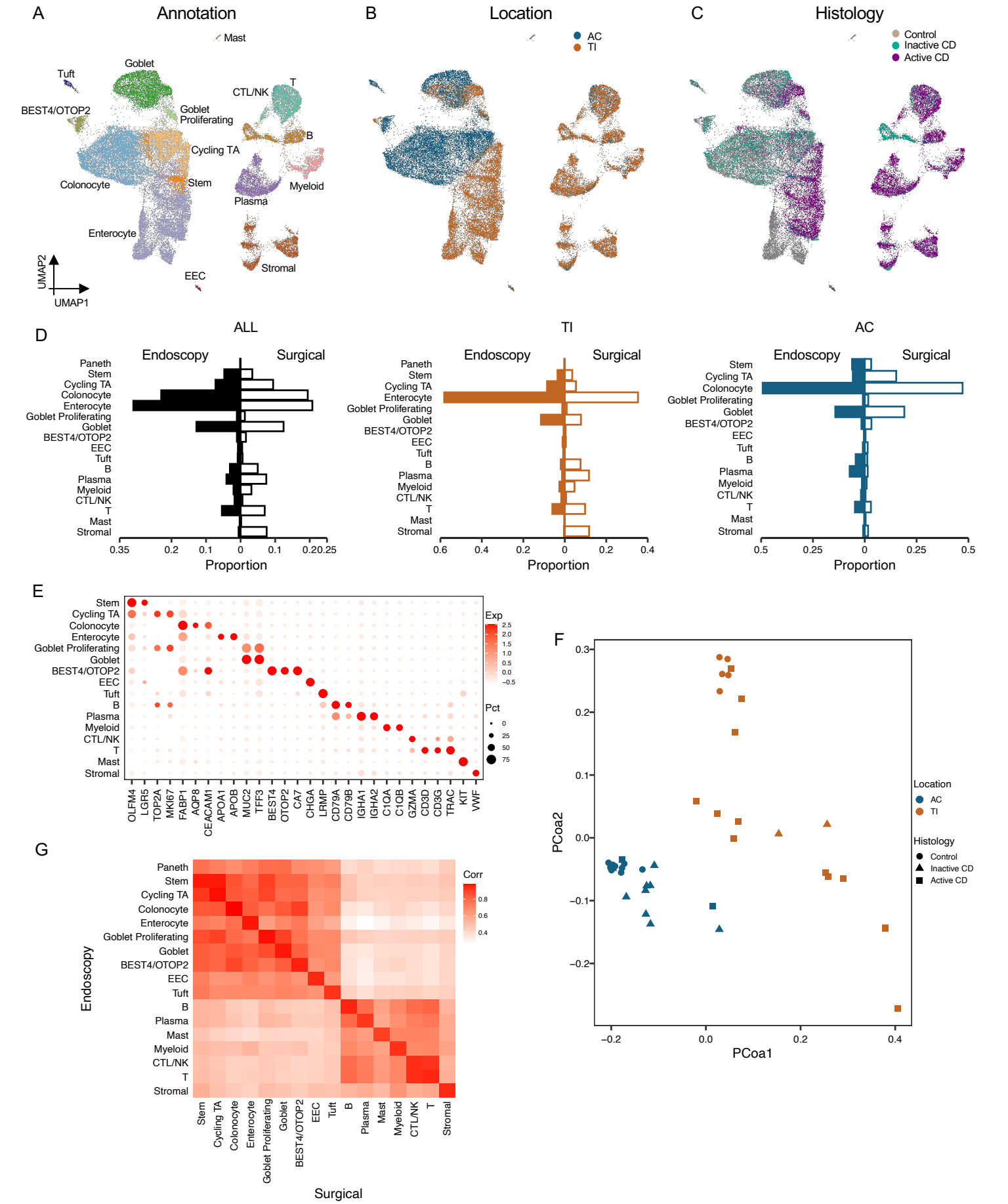
Supplementary Figure 5. **Mature enterocyte-specific expression changes between active CD, inactive CD and non-IBD controls**. Volcano plots of differentially expressed genes between active and inactive CD, inactive CD and controls in mature enterocytes in TI (up) and mature colonocytes in AC (down).



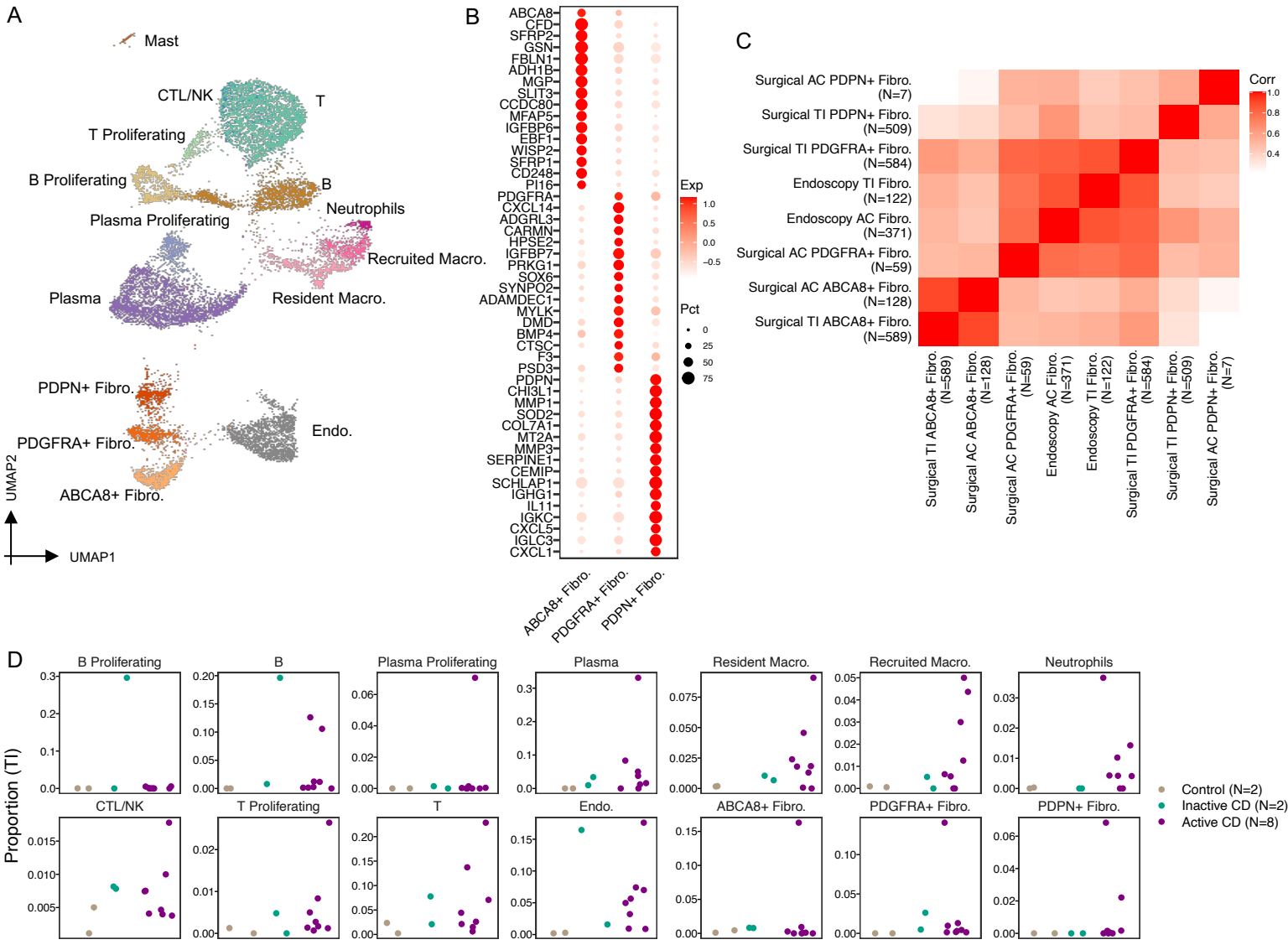
Supplementary Figure 6. Ligand-receptor interactions in endoscopic TI and AC samples. Dotplots of ligands and receptors expression in different cell types in TI (A) and AC (B). Ligand-receptor interactions between LND and immune cells in TI (C) and AC (D).



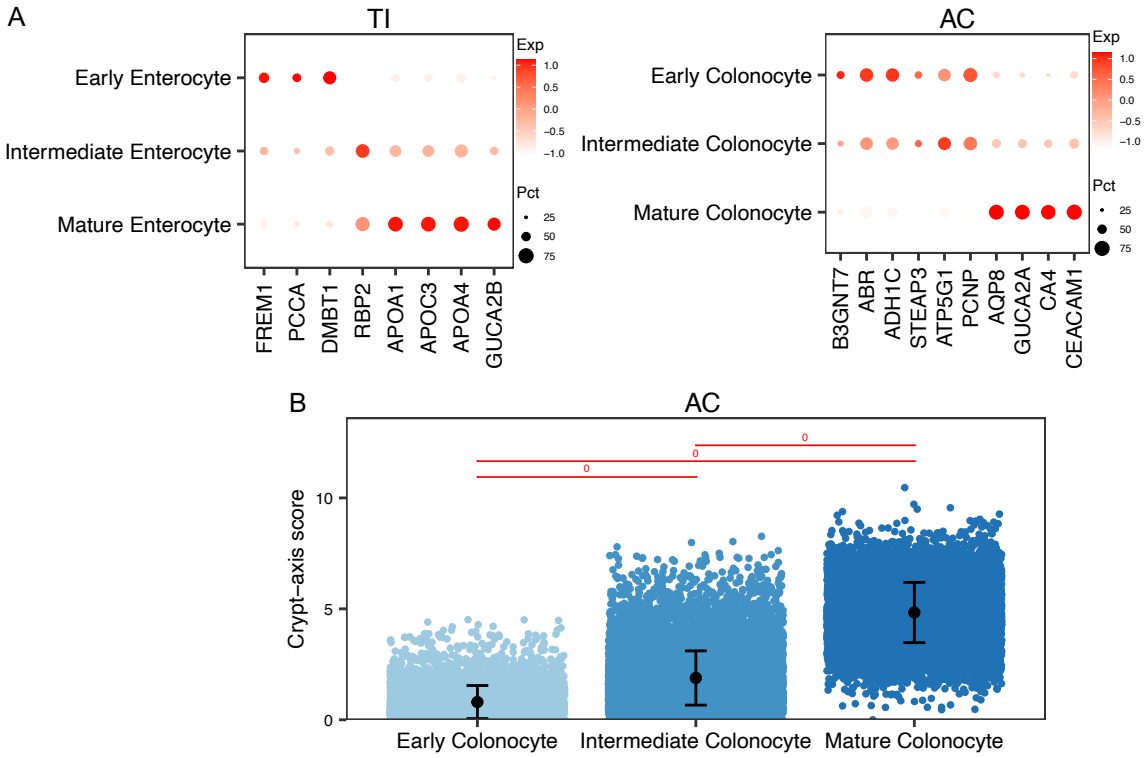
Supplementary Figure 7. Ligand-receptor interactions in surgical TI samples. A)A scatterplot of outgoing (x-axis) and incoming (y-axis) strength of each cell type in surgical TI samples. B) DotPlot of ligands and receptors expression in different cell types in surgical TI samples. C) Ligand-receptor interactions between LND and immune cells in surgical TI samples.



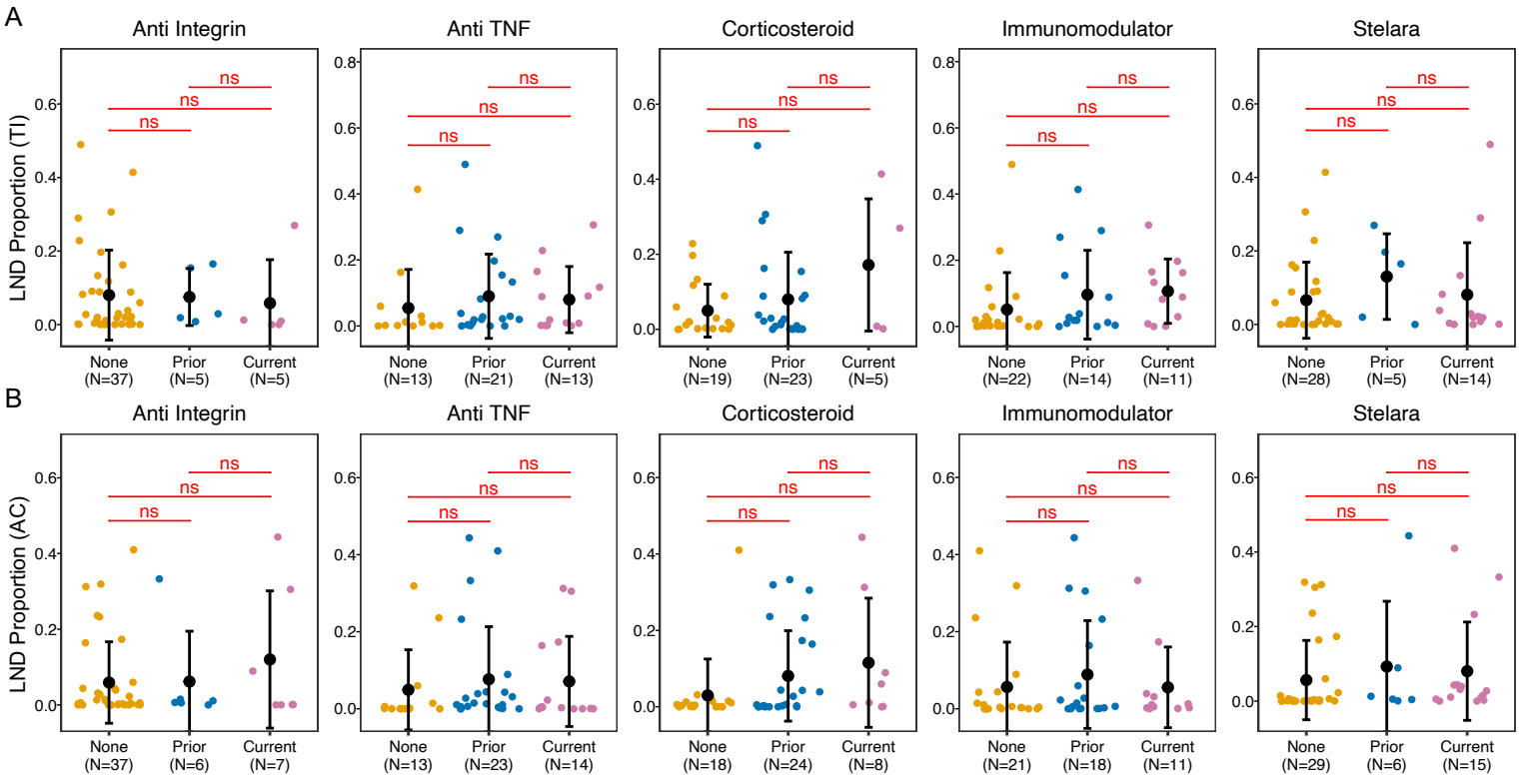
Supplementary Figure 8. Single-cell landscape in CD and non-IBD controls in surgical specimens. A) UMAP of of 47,266 cells from surgical specimens colored by cell type; B) UMAP of 47,266 cells from surgical specimens colored by tissue origin, TI(brown) and AC (blue). C) UMAP of 47,266 cells colored by disease status, non-IBD controls (brown), inactive (green) and active CD(red). D) Comparison of cell proportions between endoscopy (filled) and surgical (unfilled) samples. E) Dotplot of marker genes in each cell type. F) MDS plot of cell compositional difference across all surgical samples. G) Similarity of cell types between endoscopy and surgical samples.



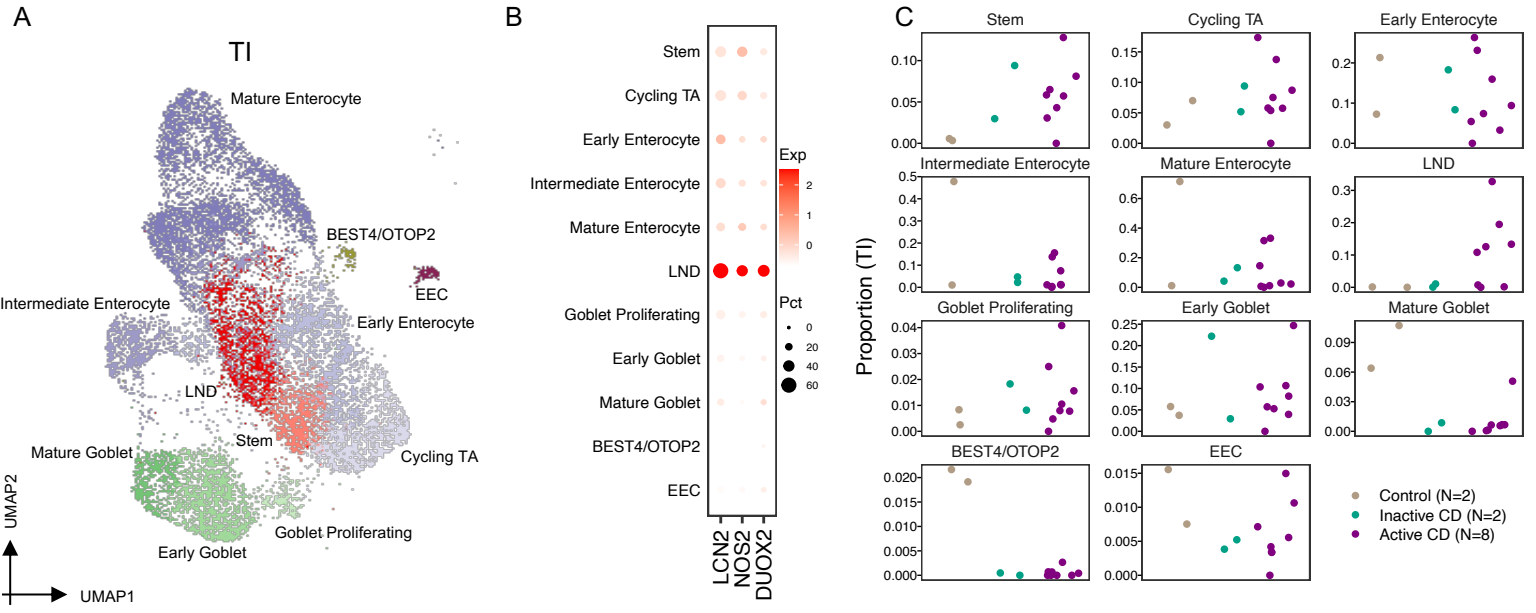
Supplementary Figure 9. Immune and stromal cells rewiring in CD patients in surgical samples. A) UMAP of immune and stromal cells in surgical samples. B) Dotplot of marker genes in three fibroblast cell types. C) Similarity of fibroblast cell types detected from surgical and endoscopic samples. D) Proportional changes of immune and stromal cells from controls to inactive and active CDs in surgical TI.



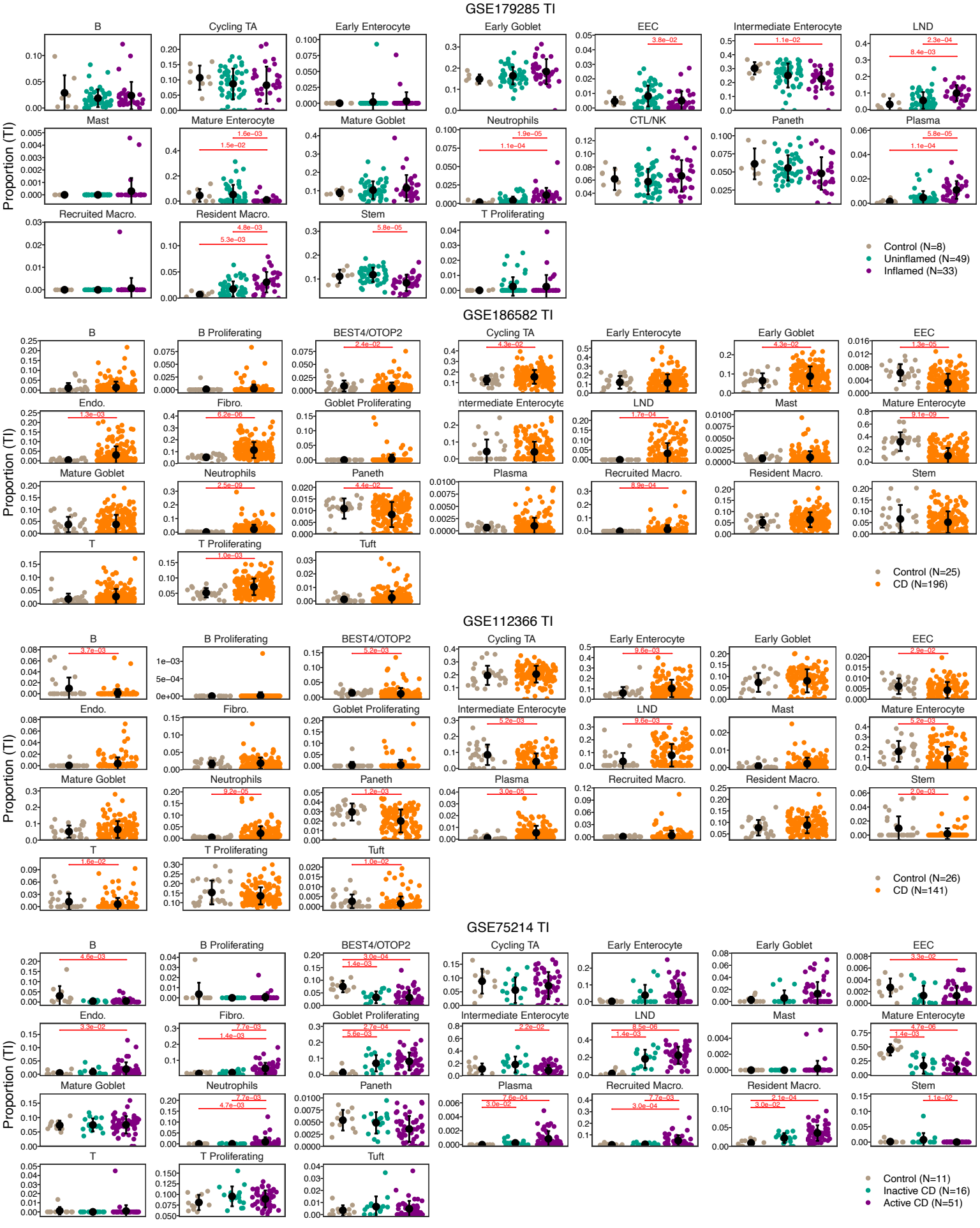
Supplementary Figure 10. Characterization of early, intermediate, and mature enterocytes. A) Dotplot of known markers in early, intermediate, and mature enterocyte (left) and colonocyte (right). B) Distribution of Crypt-axis scores in early, intermediate, and mature colonocyte. Two-sided Wilcoxon test with FDR adjustment. Data are represented as the mean ± SD.



Supplementary Figure 11. The association of LND proportions with medical exposure. LND proportional changes with different medical exposure in the TI (A) and AC(B). Two-sided Wilcoxon test with FDR adjustment. All data are represented as the mean ± SD.

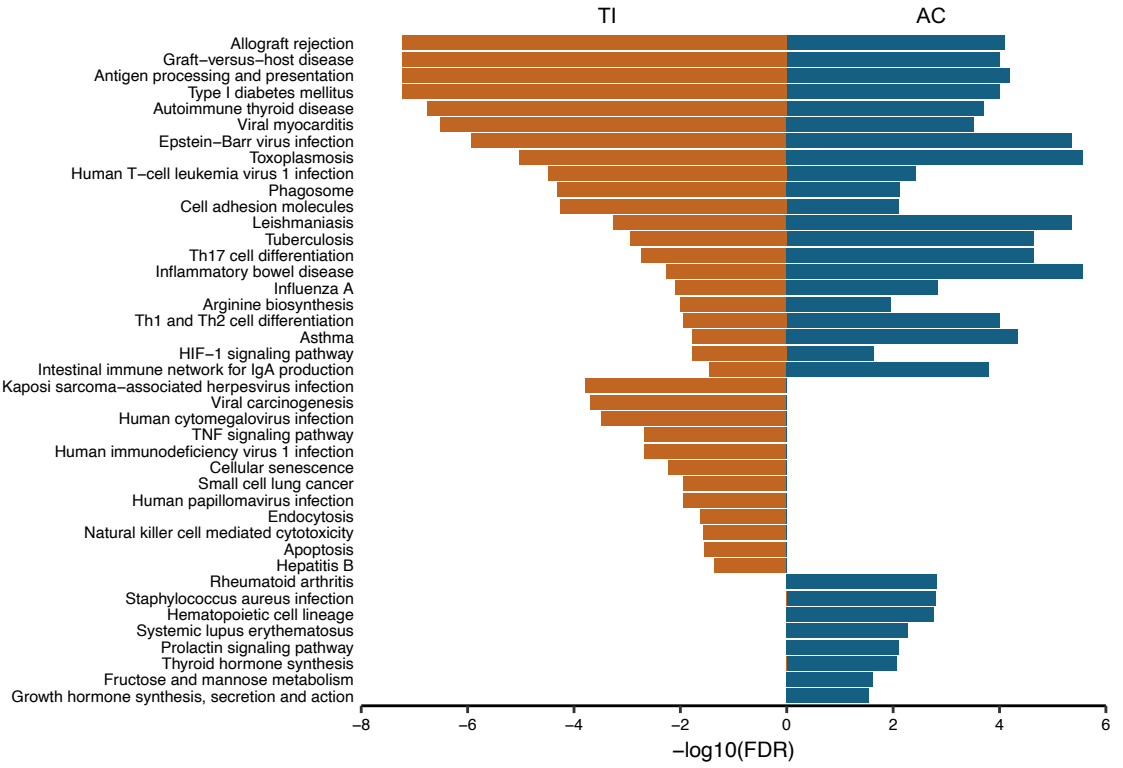


Supplementary Figure 12. Epithelial cell types in surgical TI samples. A) UMAP of epithelial cells from surgical TI samples colored by cell type. B) High expression of LCN2, NOS2 and DUOX2 in the LND cluster in the surgical TI. C) Cell proportional changes of cell clusters in surgical TI from controls, inactive and active CD.

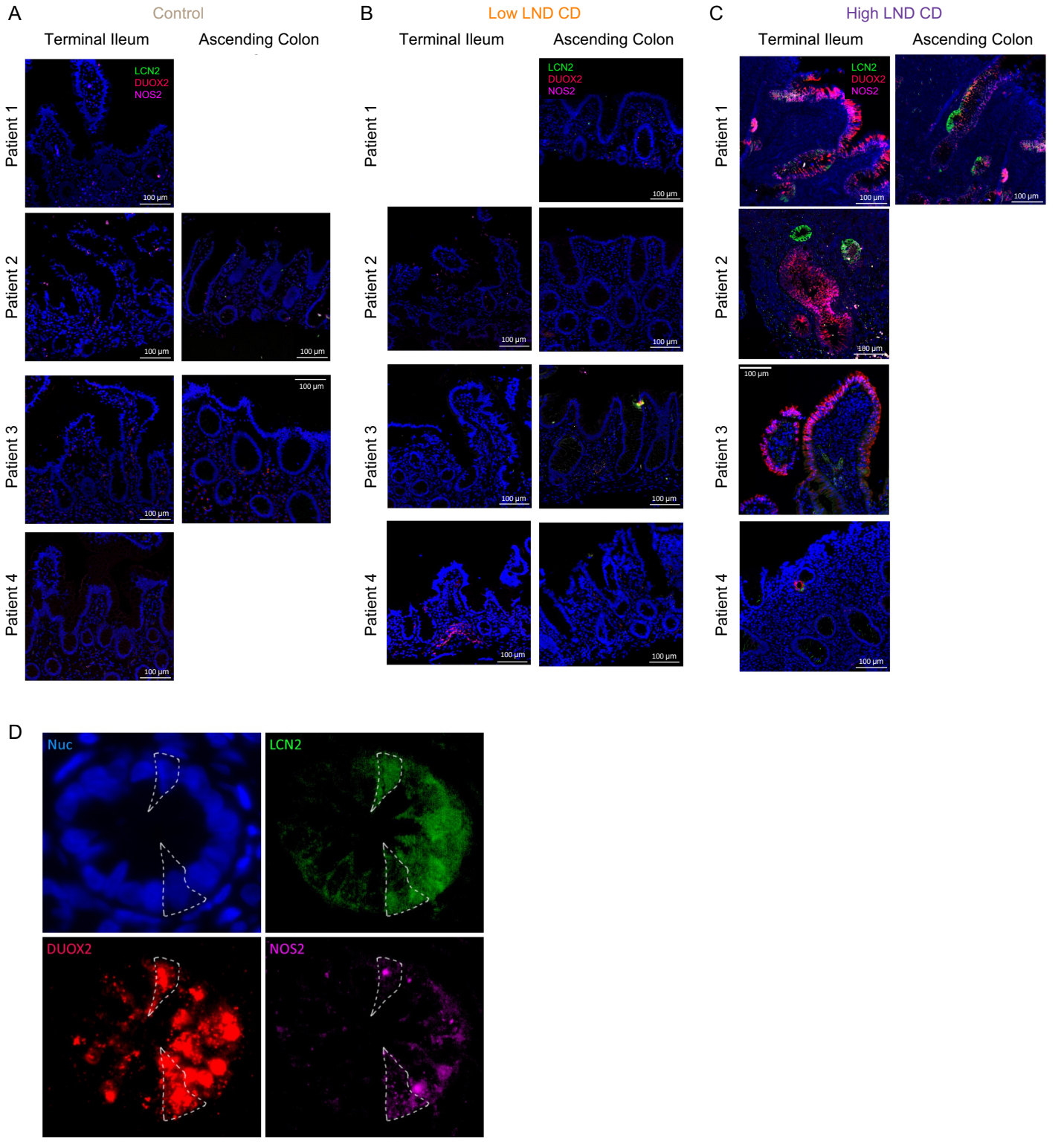


Supplementary Figure 13. Proportion changes of epithelial cells in four datasets from TI: GSE179285, GSE186582, GSE112366, and GSE75214. Two-sided Wilcoxon test with FDR adjustment. All data are represented as the mean ± SD.

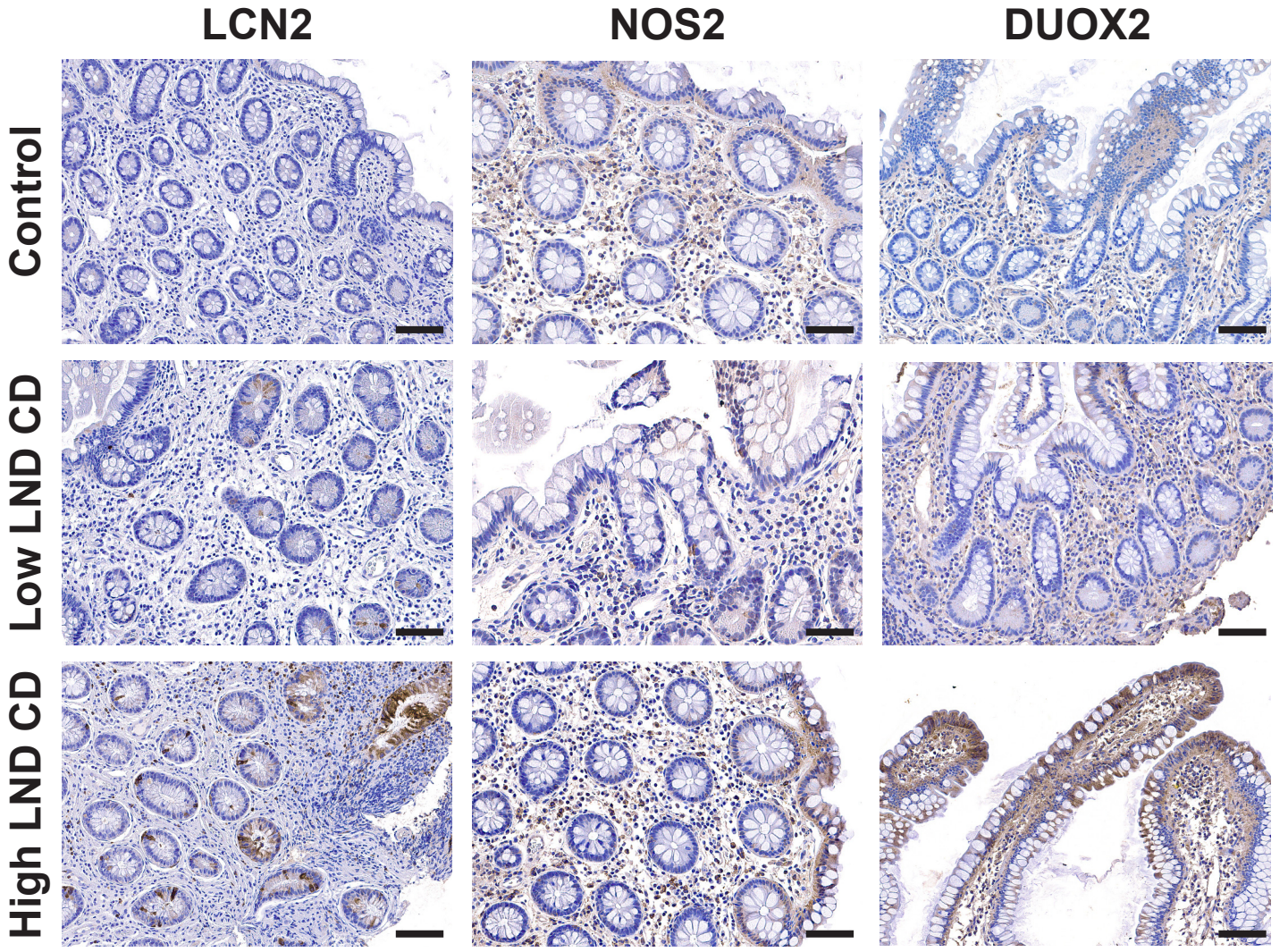
Supplementary Figure 14. Proportion changes of epithelial cells in four datasets from AC: GSE179285, GSE20881, GSE16879, and GSE75214. Two-sided Wilcoxon test with FDR adjustment. All data are represented as the mean ± SD.



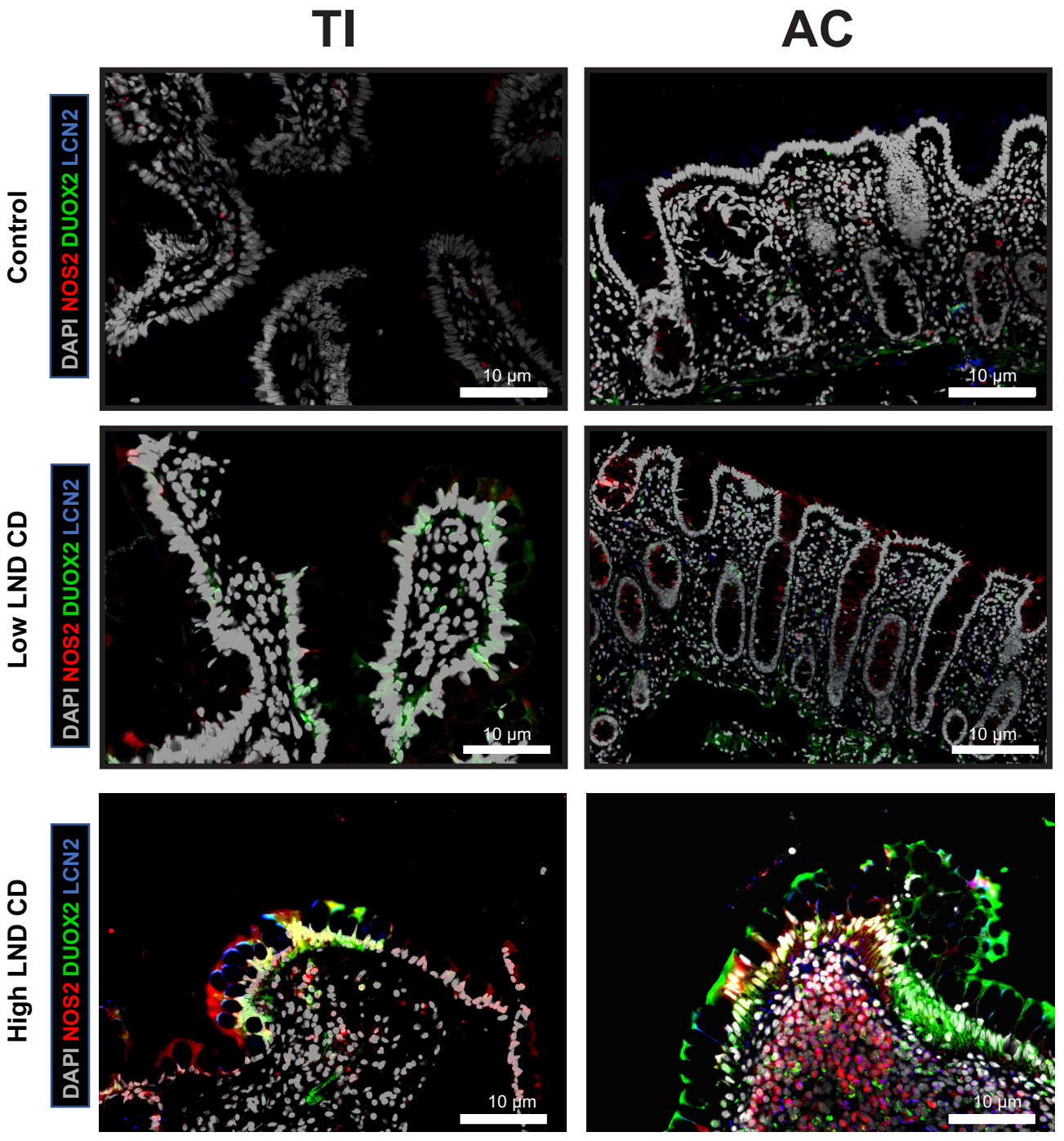
Supplementary Figure 15. Pathways enriched in the LND subpopulation.



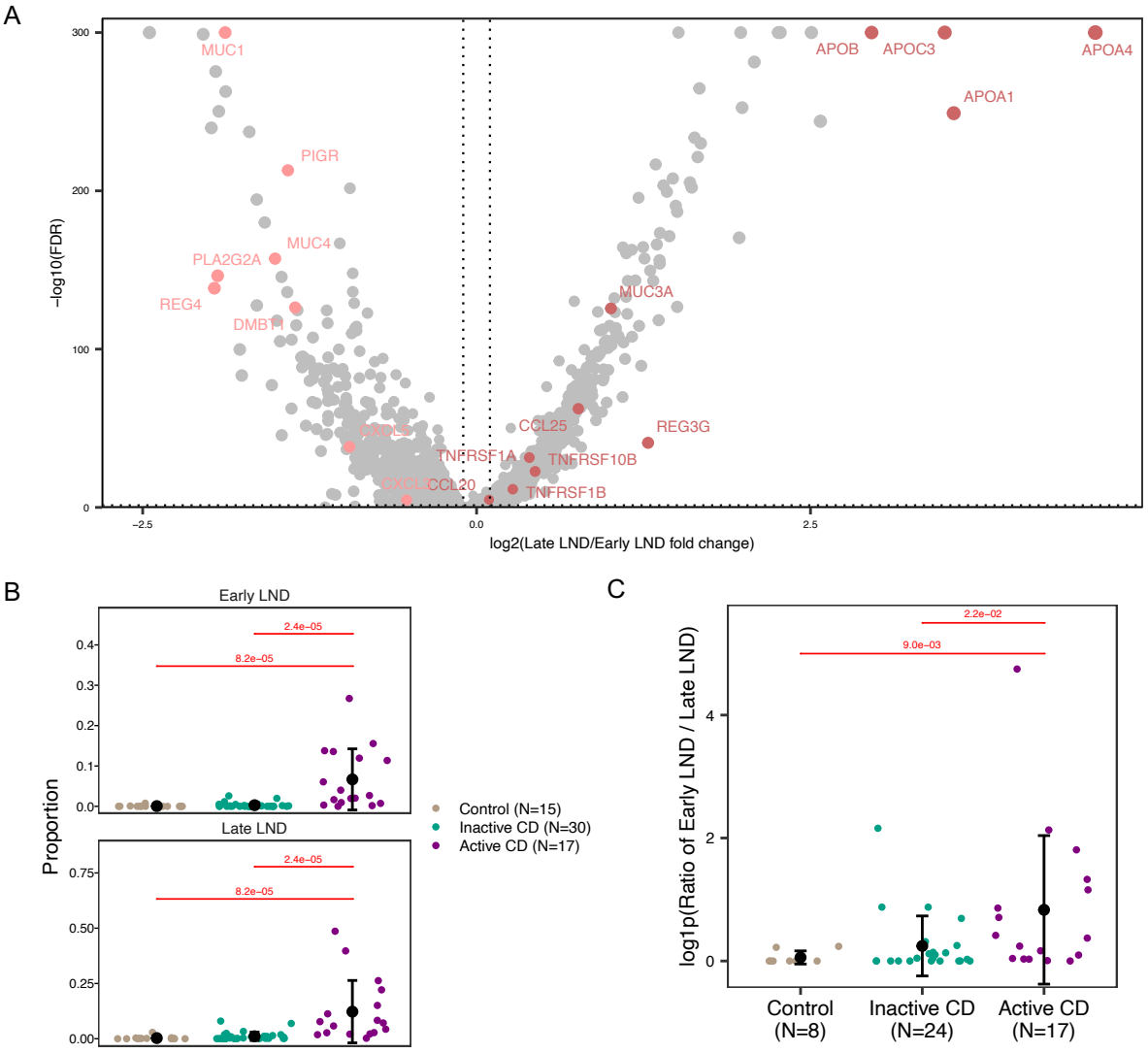
Supplementary Figure 16. RNA-FISH images of LCN2, NOS2 and DUOX2. RNA-FISH of LCN2, NOS2 and DUOX2 in controls (A), low LND CD (B) and high LND CD (C) in the TI and AC (Note that the TI/AC not always available due to morphology). D) Single cell segmentation.



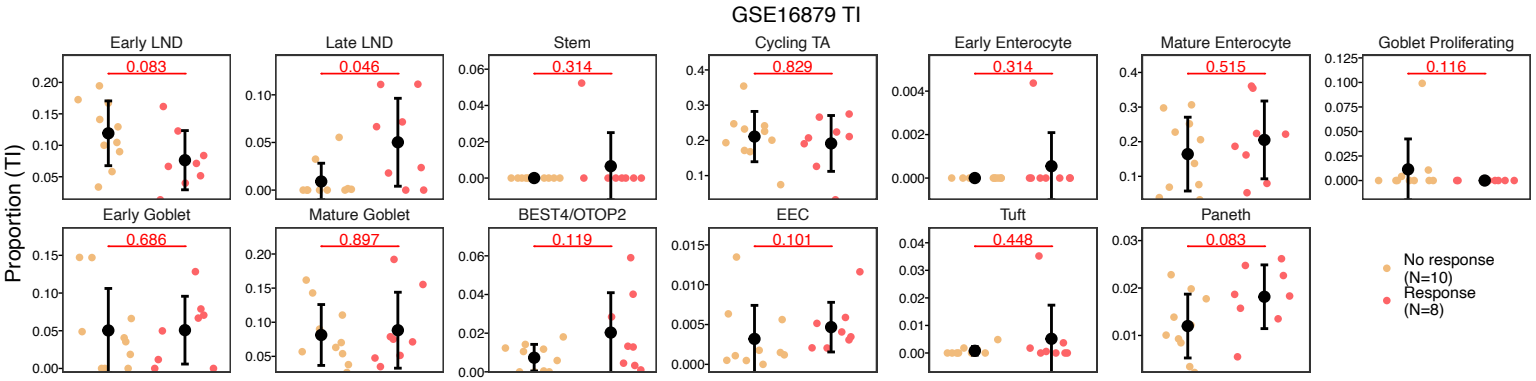
Supplementary Figure 17. Immunohistochemistry of LCN2, NOS2 and DUOX2 in controls, low LND CD, and high LND CD.



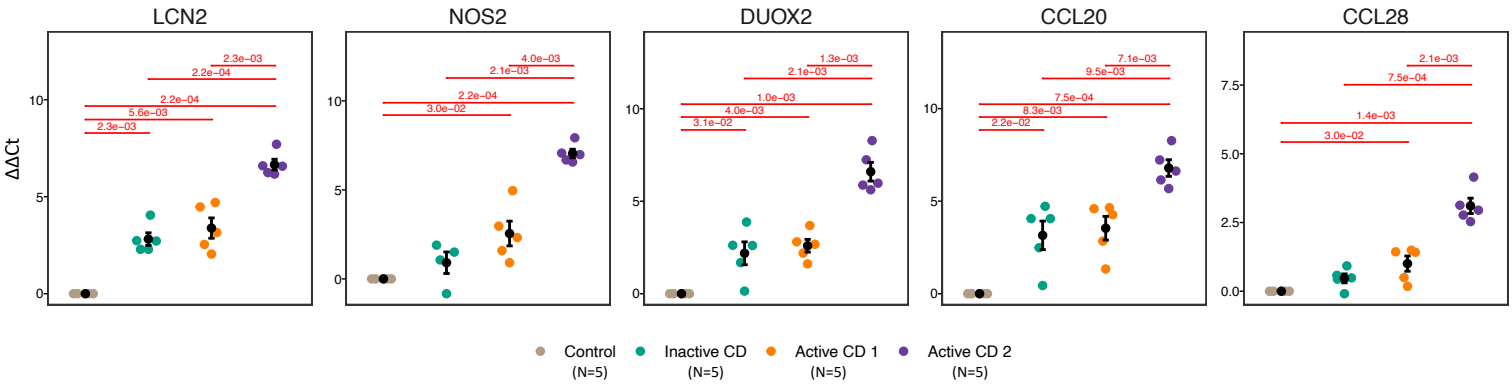
Supplementary Figure 18. Multiplex immunofluorescence of LCN2 (blue), NOS2 (red) and DUOX2 (green) in controls, low LND CD, and high LND CD.



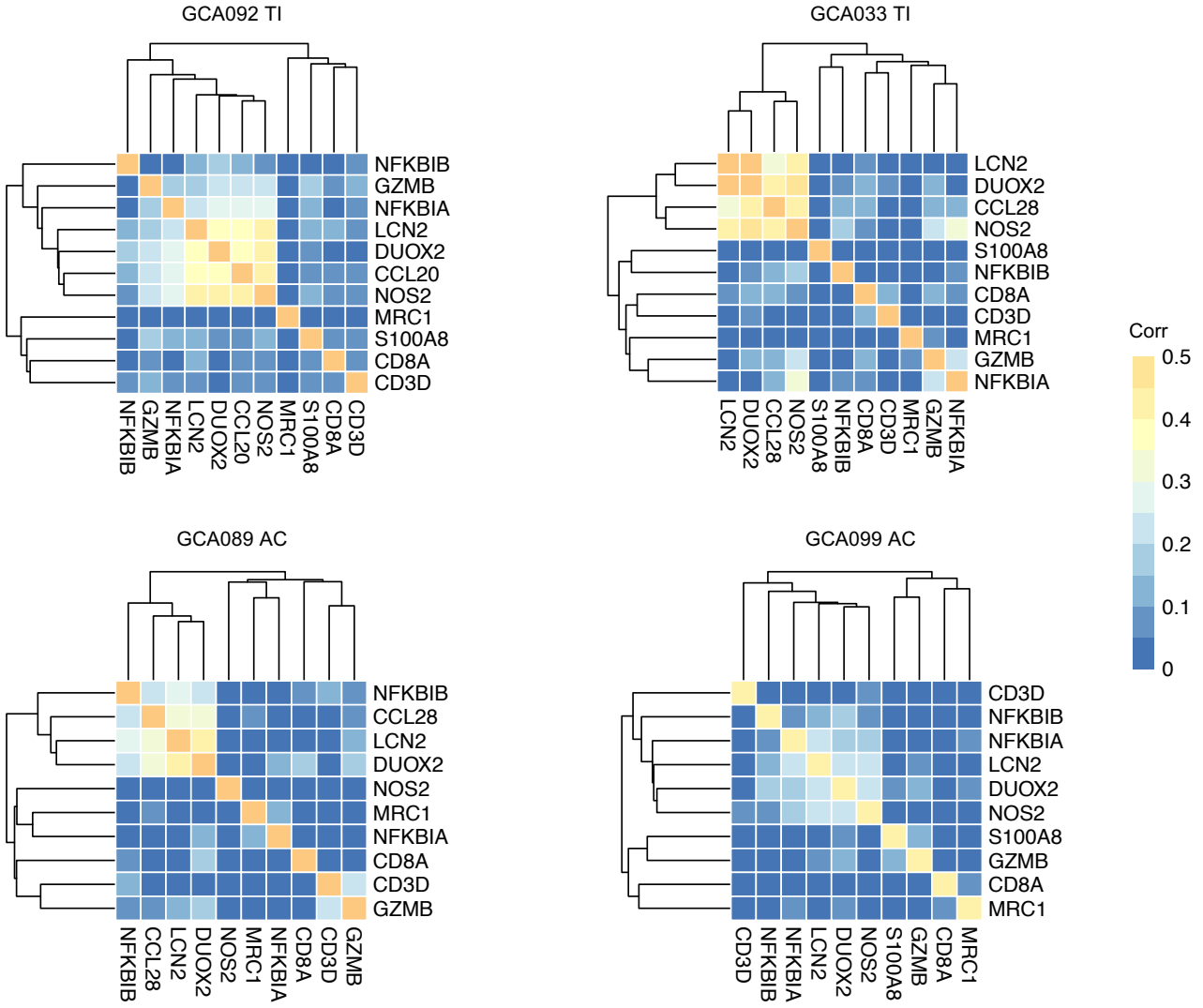
Supplementary Figure 19. Transcriptional and proportional difference between early and late LND. A) Volcano plot of differential expression between the early and late LND in TI. B) Cell proportions of early and late LND in TI in controls, inactive and active CD. Two-sided Wilcoxon test with FDR adjustment. C)The ratio of early to late LND in TI in controls, inactive and active CD. Two-sided Wilcoxon test with FDR adjustment. All data are represented as the mean ± SD.



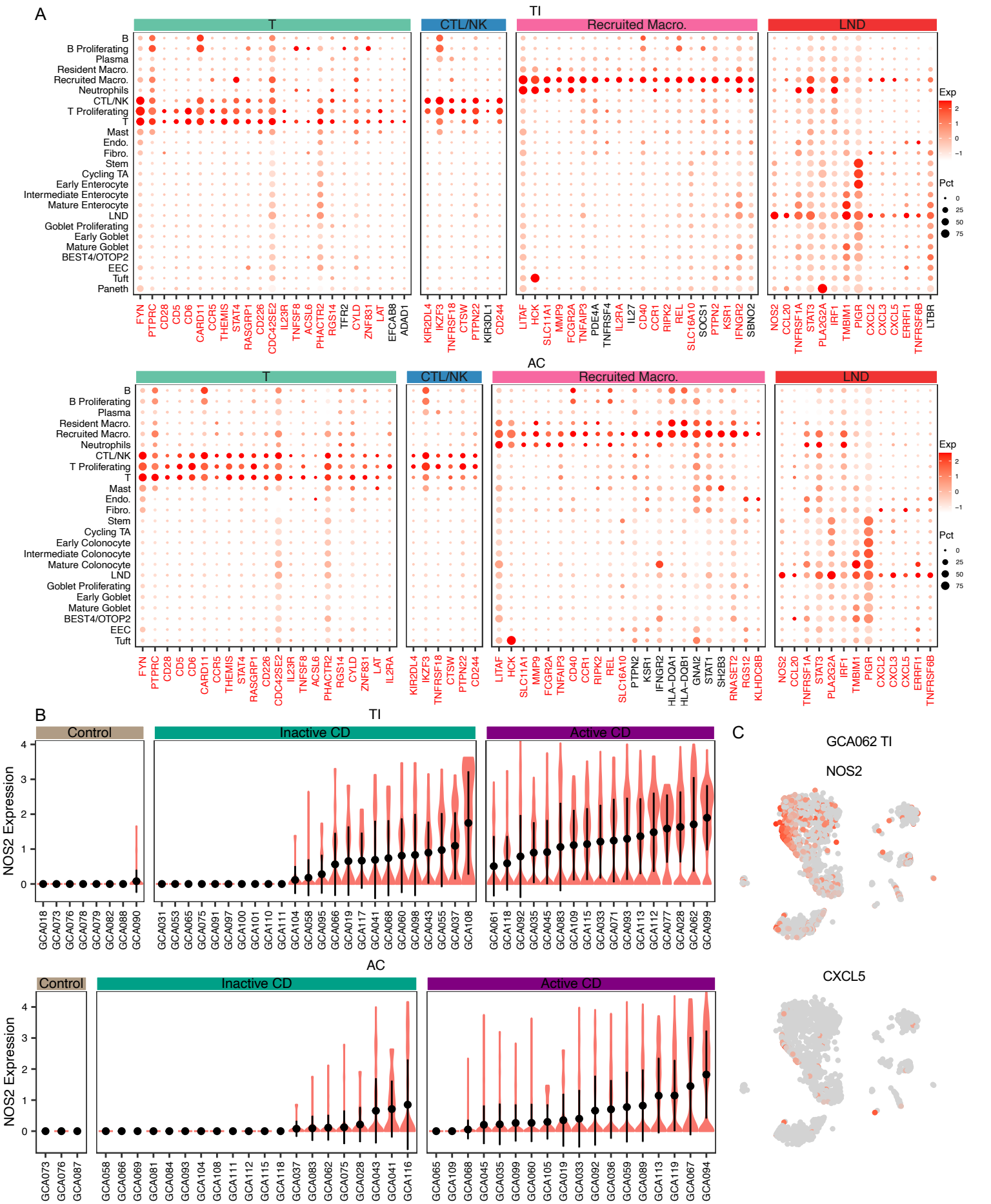
Supplementary Figure 20. Proportional differences of each epithelial cell type between anti-TNF non-responders and responders. Two-sided Wilcoxon test. All data are represented as the mean ± SD.



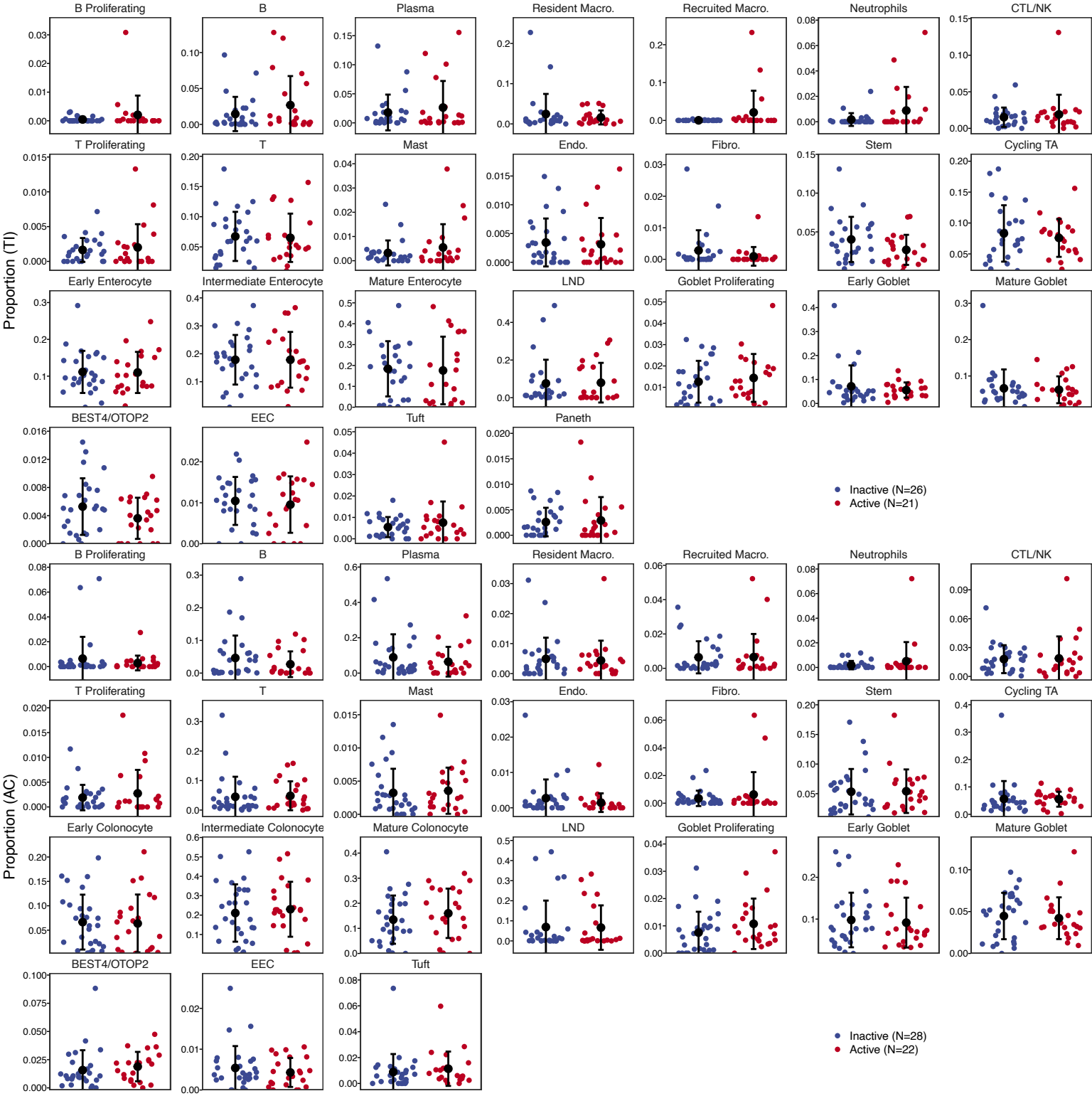
Supplementary Figure 21. Expression levels of LCN2, NOS2, DUOX2, CCL20 and CCL28 in the four organoid lines, one from control, one from inactive CD, two from active CD. Two-sided Wilcoxon test with FDR adjustment. All data are represented as the mean ± SD.



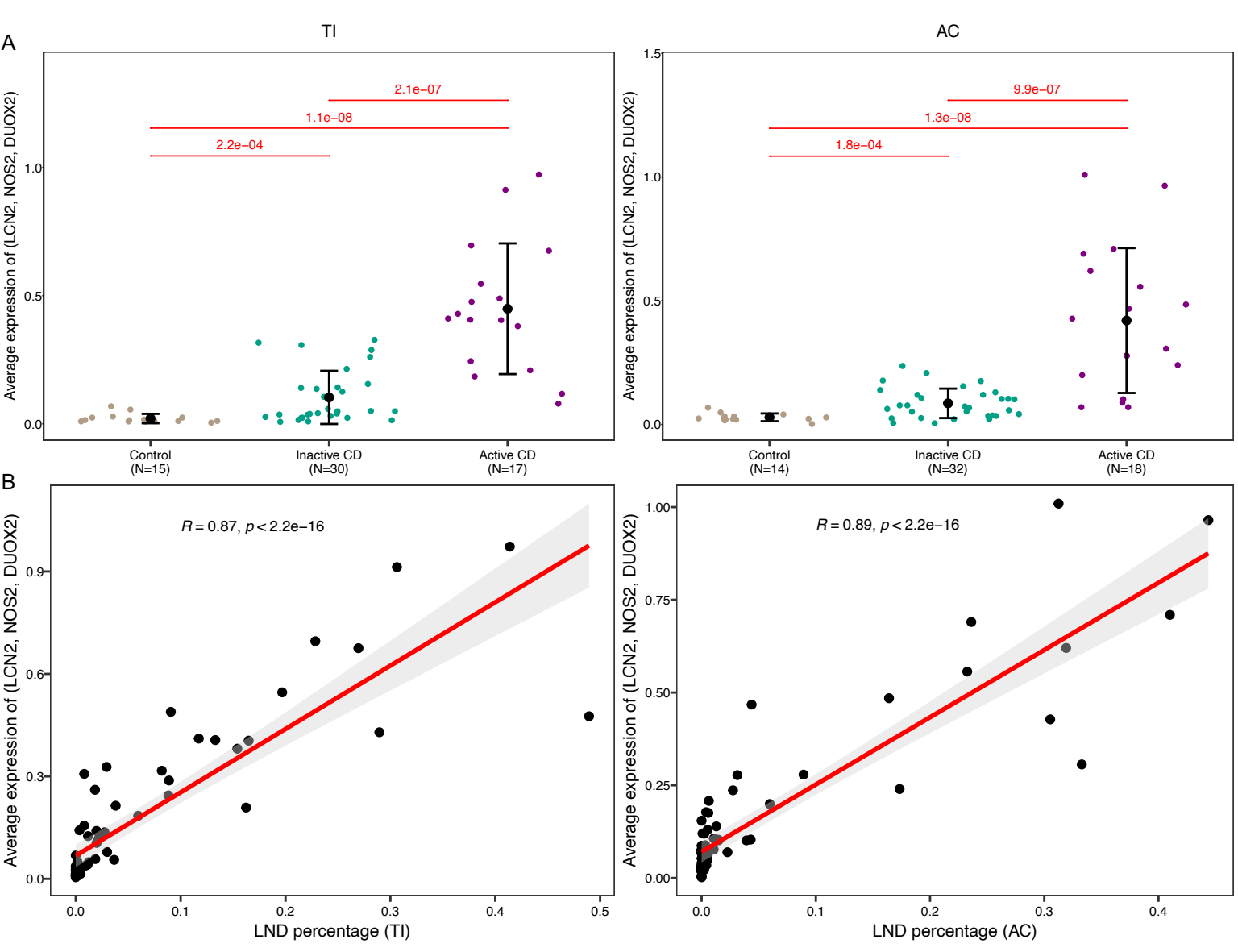
Supplementary Figure 22. **Expression correlation of LND markers (LCN2,DUOX2, NOS2, CCL20 or CCL28) and immune signatures (CD3D, CD8A, GZMB, MRC1, S100A8, NFKBIA, and NFKBIB) in the four 10x Visium datasets.**



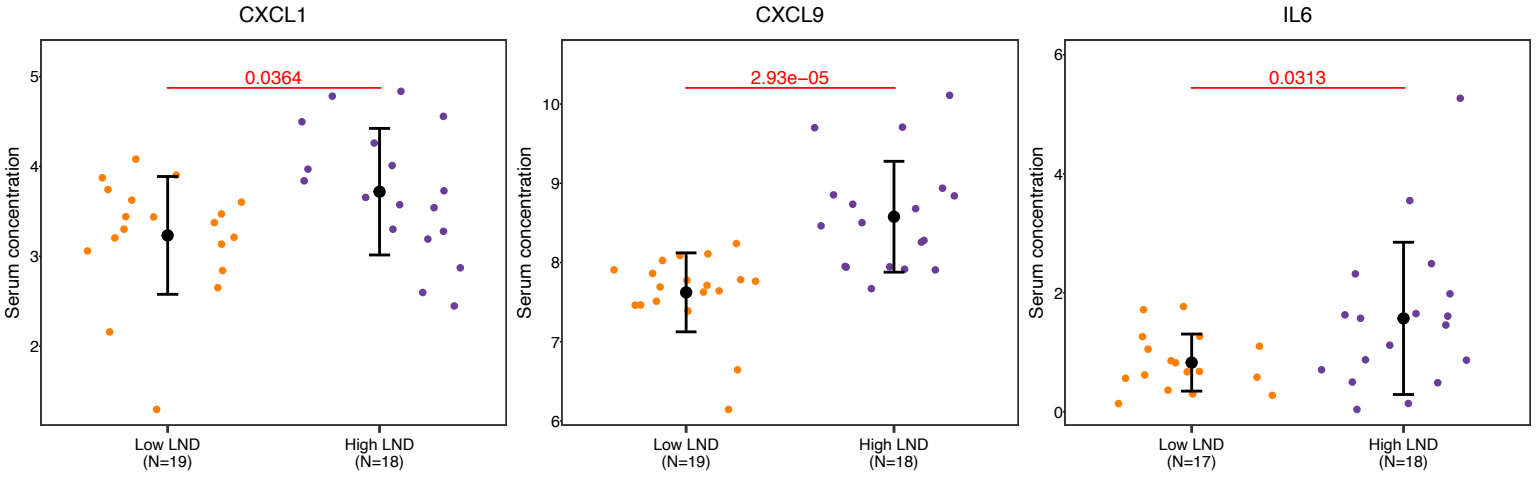
Supplementary Figure 23. Expression of IBD/CD-risk genes. A) Dotplots of expression of risk genes contributing to significant disease association of T cells, CTL/NK, recruited macrophages, and LND in TI(top) and AC (bottom). B)Expression heterogeneity of NOS2 in LND across all TI (top) and AC (bottom) samples. Data are represented as the mean ± SD. C) UMAP plot of NOS2 and CXCL5 expression in the TI from CD patient GCA062.



Supplementary Figure 24. Proportional changes of each cell type between inactive and active patients defined by CDAI scores in endoscopy TI (up) and AC (down). All data are represented as the mean ± SD.



Supplementary Figure 25. A) Average expression of LCN2, NOS2, and DUOX2 in non-IBD control, inactive, and active CD in the TI (left) and AC (right). Two-sided Wilcoxon test with FDR adjustment. Data are represented as the mean ± SD. B) Scatterplot of average expression of LCN2, NOS2, and DUOX2 (y-axis) and LND proportions (x-axis) in TI (left) and AC (right). Pearson correlation.



Supplementary Figure 26. Serum cytokine levels between low LND and high LND patients. Two-sided Wilcoxon test with FDR adjustment. Data are represented as the mean ± SD.

Supplementary Note 1. Rare cell populations identified in the TI and AC

We identified rare microfold-like cells by employing a divide-and-conquer strategy to regroup the BEST4/OTOP2 cell cluster. Using an innovative clustering approach based on an adaptive k-nearest neighbor[1], we isolated a small cluster in the AC characterized by high expression of *SPIB*, *SOX8*, *TNFRSF11B*, *CCL23*, and *CCL20*, recognized markers of colonic microfold-like cells[2,3] (Supplementary Fig. 1). Notably, most microfold-like cells came from one patient in the AC and this cell population were not identified in the TI. Furthermore, within the cluster of IgA+ plasma cells, we noted a minimal presence of IgG+ plasma cells, constituting only 1% (Supplementary Fig. 2), as determined by the criteria outlined in a previous study.

Supplementary Note 2. Distinct epithelial transcriptional programs between the TI and AC

The largest transcriptomic difference was derived from absorptive cells from TI or AC. The most distinguishing marker genes for TI enterocytes were *APOA1* and *APOB*, which encode apolipoproteins and play a key role in intestinal lipid absorption. In contrast, AC colonocytes were marked by *AQP8*, a major transcellular water transporter. Differential transcriptional programs between TI enterocytes and AC colonocytes were identified (Supplementary Fig. 3A), which reflected tissue-specific functions. Genes highly expressed in the TI were enriched in programs of protein/fat/vitamin absorption and processing, while genes highly expressed in the AC were enriched in nitrogen/sulfur metabolism (Supplementary Fig. 3B), consistent with the nutrient digestion function of the small intestine and the interaction of the colon with microbial metabolites. Furthermore, signatures of region-specific development programs were also different between TI and AC. AC cells expressed hindgut-specific transcription factors such as *SATB2* and *CDX2*[4-6], as well as the colon-specific chloride anion exchanger *SLC26A3*, also known as DRA (Down-regulated in adenoma)[7] (Supplementary Fig. 3A). In contrast, TI cells exhibit increased expression of machinery that organizes the brush border, including *VIL1* (Villin) and *CDHR2*, also known as protocadherin-24[8] (Supplementary Fig. 3A).

Differences in goblet, EEC, and tuft cells between human TI and AC have not been fully investigated previously. We found that expression of region-specific genes identified in absorptive cells were also generally different in goblet, EEC, and tuft cells between TI and AC (Supplementary Fig. 3C). Interestingly, there are region-specific genes that are also cell type-specific (Supplementary Fig. 3C). For instance, *BEST2* was exclusively expressed in AC goblet cells, but not in TI goblet cells or other AC epithelial cell types (Supplementary Figs. 3C and 3D)[9]. *XBP1*, an ER stress response factor commonly expressed highly in secretory cells, was more highly expressed in TI goblet cells (Supplementary Figs. 3C and 3D). SNPs in *XBP1* have been found to confer increased risk for developing IBD[10], pointing towards an overload of protein production and ER stress that can potentially lead to secretory cell and barrier dysfunction, tipping the balance toward IBD. *RGS13* was exclusively expressed in TI tuft cells (Supplementary Figs. 3C and 3D). *NTS* and *CLU* were exclusively expressed in TI EEC cells (Supplementary Figs. 3C and 3D). EEC cells have been shown to act as facultative stem cells upon certain types of

damage[11]. Clusterin expression in TI EEC cells may implicate stem cell activation in the context of TI epithelial restitution as a response to damage, as clusterin was recently identified as a marker of revival stem cells[12]. Regarding immunological genes, AC goblet cells expressed more *IL1R2*, suggesting they were more sensitive to inflammatory stimuli (Supplementary Fig. 3C). TI enterocytes expressed more *CCL25*, which is a chemokine ligand for *CCR9* expressed on T cells responsible for inflammatory T cell recruitment[13] (Supplementary Fig. 3A).

Supplementary Note 3. Cell-type-specific differential expression between active CD, inactive CD and non-IBD controls

Transcriptional upregulation of pro-inflammatory genes was observed in active CD compared to inactive CD within each immune cell type. Inflammatory signaling programs characteristic of the type 1 immune response appear to be upregulated in almost all immune cells; these include components of the JAK/STAT (*JAK3*, *STAT1*, *SOCS1*) pathway and a large variety of interferon regulatory factors (IRFs) and interferon stimulated genes (ISGs). Consequently, antigen-presentation genes, such as *CD74*, *B2M*, *HLA-A*, *HLA-B*, and others, were upregulated in both lymphoid and myeloid cell types (Supplementary Fig. 4A). T cells were characterized by increases in both cytotoxic (*GZMB*, *NKG7*) and exhaustion phenotypes (*LAG3*, *TIGIT*) (Supplementary Fig. 4A). Resident macrophages in active CD specifically increased expression of various cathepsins (such as *CTSB* and *CTSC*) that are important in microbial defense, as well as *CXCL9* that is important in further recruitment of *CXCR3*+ T cells (Supplementary Fig. 4A).

Transcriptional upregulation of proinflammatory genes was also observed in a cell type-specific manner in inactive CD compared to non-IBD controls in immune cells, such as overexpression of MHCII genes, *HLA-DRB5* and *HLA-DPB1* in B cells, and *HLA-DQB1* in both B cells and resident macrophages (Supplementary Fig. 4B).

Additionally, cell-specific transcriptional changes with disease activity were observed in mature enterocytes/colonocytes and other epithelial cell types (Supplementary Data 4). For example, in mature absorptive epithelial cells, antigen presentation genes (*HLA-DPB1*, *HLA-DPA1*,*HLA-F*, *HLA-E*, and *CTSS*) were upregulated in active compared to inactive CD in the TI, while MAPK signaling genes (*MAP3K13*, *PRKCA*, and *PLA2G2A*) were increased in active compared to inactive CD in the AC. Metallothionein (*MTIE*, *MTIF*, *MTIG*, *MTIH*, and *MT2A*) were downregulated in active compared to inactive CD in the TI, while focal adhesion genes (*CD9*, *HSPA1A*, and *HSPA1B*) were decreased in active compared to inactive CD in the AC (Supplementary Fig. 5). Moreover, antigen presentation genes were also upregulated in inactive CD compared to non-IBD controls in both TI and AC (Supplementary Fig. 5).

Supplementary Note 4. Ligand-receptor interactions between LND and immune cells

Compared to other epithelial cells, LND cells showed much stronger cytokine-receptor interactions especially as signaling sources, suggesting their specialized role in regulating mucosal immunity. *CXCL2*, *CXCL3*, and *CXCL5,* which encode known neutrophil-attracting chemokines acting through CXCR1/CXCR2 receptors, were expressed significantly higher in LND cells (Supplementary Figs. 6A and 6B). The LND-neutrophil interaction showed a similar interaction strength as the recruited macrophage-neutrophil interaction, implicating similar importance of LND in neutrophil recruitment and inflammatory response orchestration (Supplementary Figs. 6C and 6D). LND was also the primary cell type for recruiting plasma cells through CCL28-CCR10 signaling (Supplementary Figs. 6C and 6D), where *CCL28* was significantly expressed in LND cells and *CCR10* in plasma cells (Supplementary Figs. 6A and 6B). The CCL28-CCR10 interaction is critical for regulating intestinal IgA production under homeostasis or infection[14]. SAA1 is an acute phase protein in response to inflammation and tissue injury, which has strong chemotactic activity for neutrophils and macrophages mediated through FPR2[15-17]. *SAA1* was significantly upregulated in LND cells (Supplementary Figs. 6A and 6B), leading to strong LND-neutrophil and LND-recruited macrophage interactions (Supplementary Figs. 6C and 6D). We also detected LND-B/Plasma cells interactions through TNFSF13-TNFRSF13B (Supplementary Figs. 6C and 6D), LND-T cells by NECTIN2-CD226, and LND-immune cells by CD55-ADERG5 (Supplementary Figs. 6C and 6D). Furthermore, we identified interactions between LND and T cells via TNFSF15-TNFRSF25, LND-recruited macrophages via TNFSF10-TFNRSF10B and strong MHC-I signaling, but only in TI not in AC (Supplementary Fig. 6C).

LND cells not only actively participated in immune responses as signaling senders, but also responded to signals from immune cells as signaling receivers, although their activities as signal receivers were similar to other absorptive cells. A few examples of LND cells acting as signal receivers included B cell-LND by IL10 signaling and T/NK-LND by TNFSF14-LTBR in both TI and AC (Supplementary Figs. 6C and 6D), and proliferative B/T cell-LND by LTA-TNFRSF1A only in TI (Supplementary Fig. 6C).

Similar LND-immune interactions were observed for surgical TI samples (Supplementary Fig. 7A). LND cells acted as both source and target for immune interactions. Consistent with the results found in endoscopic samples, LND cells were the primary cell type for recruiting plasma cells through CCL28-CCR10 signaling. Strong LND-neutrophil and LND-recruited macrophage interactions via SAA1-FPR2, LND-B/Plasma cells interactions through TNFSF13-TNFRSF13B, LND-T cells by NECTIN2-CD226, and LND-immune cells by CD55-ADERG5 were also detected (Supplementary Fig. 7C). Furthermore, LND cells acted as signal receivers in the IL10 signaling and LTA-TNFRSF1A interaction (Supplementary Fig. 7D). Interestingly, the PDPN+ fibroblast population served as a strong signaling sender and was a main source for releasing *CXCL2*/*CXCL3*/*CXCL5* in neutrophil recruitment (Supplementary Figs. 7B and 7C).

References

1 Li, J., Shyr, Y. & Liu, Q. Single-cell and Spatial Transcriptomics Clustering with an Optimized Adaptive K-Nearest Neighbor Graph. *bioRxiv*, doi:10.1101/2023.10.13.562261 (2023).

2 Serigado, J. M. *et al*. Ulcerative Colitis: Novel Epithelial Insights Provided by Single Cell RNA Sequencing. *Front Med (Lausanne)* **9**, 868508, doi:10.3389/fmed.2022.868508 (2022).

3 Smillie, C. S. *et al*. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730 e722, doi:10.1016/j.cell.2019.06.029 (2019).

4 Gu, W. *et al*. SATB2 preserves colon stem cell identity and mediates ileum-colon conversion via enhancer remodeling. *Cell Stem Cell* **29**, 101-115 e110, doi:10.1016/j.stem.2021.09.004 (2022).

5 Munera, J. O. *et al*. Differentiation of Human Pluripotent Stem Cells into Colonic Organoids via Transient Activation of BMP Signaling. *Cell Stem Cell* **21**, 51-64 e56, doi:10.1016/j.stem.2017.05.020 (2017).

6 Stringer, E. J. *et al*. Cdx2 determines the fate of postnatal intestinal endoderm. *Development* **139**, 465-474, doi:10.1242/dev.070722 (2012).

7 Chatterjee, I. *et al*. CDX2 upregulates SLC26A3 gene expression in intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol* **313**, G256-G264, doi:10.1152/ajpgi.00108.2017 (2017).

8 Crawley, S. W. *et al*. Intestinal brush border assembly driven by protocadherin-based intermicrovillar adhesion. *Cell* **157**, 433-446, doi:10.1016/j.cell.2014.01.067 (2014).

9 Ito, G. *et al*. Lineage-specific expression of bestrophin-2 and bestrophin-4 in human intestinal epithelial cells. *PLoS One* **8**, e79693, doi:10.1371/journal.pone.0079693 (2013).

10 Kaser, A. *et al*. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* **134**, 743-756, doi:10.1016/j.cell.2008.07.021 (2008).

11 Vega, P. N., Lau, K. S. & Goldenring, J. R. Not All Insults Are Created Equal for Awakening Dormant Stem Cell Abilities. *Cell Mol Gastroenterol Hepatol* **7**, 619-621, doi:10.1016/j.jcmgh.2018.12.003 (2019).

12 Ayyaz, A. *et al*. Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature* **569**, 121-125, doi:10.1038/s41586-019-1154-y (2019).

13 Wendt, E. & Keshav, S. CCR9 antagonism: potential in the treatment of Inflammatory Bowel Disease. *Clin Exp Gastroenterol* **8**, 119-130, doi:10.2147/CEG.S48305 (2015).

14 Hu, S., Yang, K., Yang, J., Li, M. & Xiong, N. Critical roles of chemokine receptor CCR10 in regulating memory IgA responses in intestines. *Proc Natl Acad Sci U S A* **108**, E1035-1044, doi:10.1073/pnas.1100156108 (2011).

15 Badolato, R. *et al*. Serum amyloid A is a chemoattractant: induction of migration, adhesion, and tissue infiltration of monocytes and polymorphonuclear leukocytes. *J Exp Med* **180**, 203-209, doi:10.1084/jem.180.1.203 (1994).

16 Liang, T. S., Wang, J. M., Murphy, P. M. & Gao, J. L. Serum amyloid A is a chemotactic agonist at FPR2, a low-affinity N-formylpeptide receptor on mouse neutrophils. *Biochem Biophys Res Commun* **270**, 331-335, doi:10.1006/bbrc.2000.2416 (2000).

17 Dufton, N. *et al*. Anti-inflammatory role of the murine formyl-peptide receptor 2: ligand-specific effects on leukocyte responses and experimental inflammation. *J Immunol* **184**, 2611-2619, doi:10.4049/jimmunol.0903526 (2010).