

Divergent T Cell Phenotypes Define Pediatric Crohn's Disease and Ulcerative Colitis

Leonard Nettey^{1,2,3,4,5*}, Hengqi Betty Zheng^{6*}, Joseph C. Devlin^{7*}, Julie E. Horowitz^{7*}, Sara Rosenbaum⁸, Nathalie Fiaschi⁷, Wei Keat Lim⁷, Kyle Kimler^{2,3,4,5,8,9}, Christina Adler⁷, Annabelle Lee⁷, Min Ni⁷, Yi Wei⁷, Peter J. Ehmann⁷, Shane E. McCarthy⁷, Manuel A.R. Ferreira⁷, William Galbavy⁷, Eli Stahl⁷, Sokol Haxhinasto⁷, Sandra Coetzee⁷, Yoko Yabe⁷, Michael Dobosz⁷, Paula Keskula⁹, Jillian Zavistaski⁹, Alexandre Albanese⁹, Zoë Steier^{2,3,4,5}, Joshua de Sousa Casal^{1,4,8}, Veronika Niederlova^{3,4,8,10}, Lusine Ambartsumyan⁶, Ghassan Wahbeh⁶, David L. Suskind⁶, Vanessa Mitsialis^{8,11}, Sara Hamon⁷, Stephen M. Carpenter¹², Jennifer D. Hamilton⁷, Alex K. Shalek^{1,2,3,4,5,13,14**}, Scott B. Snapper^{8,11,13**}, Matthew A. Sleeman^{7**}, George D. Kalliolias^{7**}, Andrea T. Hooper^{7**}, Jose Ordovas-Montanes^{1,3,4,8,15**#}, Leslie S. Kean^{9,16**}

1. Program in Immunology, Harvard Medical School, Boston, MA 02115, USA
2. Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02142, 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. Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
6. Division of Gastroenterology and Hepatology, Seattle Children's Hospital and University of Washington, Seattle, WA 98105, USA
7. Regeneron Pharmaceuticals Inc., Tarrytown, NY 10591, USA
8. Division of Gastroenterology, Hepatology, and Nutrition, Boston Children's Hospital, Boston, MA 02115, USA
9. Division of Pediatric Hematology/Oncology, Boston Children's Hospital, Boston, MA 02115, USA
10. Laboratory of Adaptive Immunity, Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic
11. Brigham and Women's Hospital Department of Medicine, Division of Gastroenterology, Hepatology, and Endoscopy, Boston, MA 02115, USA
12. Division of Infectious Diseases and HIV Medicine, Department of Medicine, University Hospitals, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA
13. Harvard Medical School, Boston, MA 02115, USA
14. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
15. Harvard Stem Cell Institute, Cambridge, MA 02138, USA
16. Dana Farber Cancer Institute, Boston, MA 02215, USA

*these authors contributed equally to the work

**these senior authors contributed equally to the work

#lead contact

Correspondence to Leslie S. Kean (Leslie_Kean@dfci.harvard.edu), Jose Ordovas-Montanes (jose.ordovas-montanes@childrens.harvard.edu), and Alex K. Shalek (shalek@mit.edu)

ABSTRACT

Pediatric Inflammatory Bowel Disease (IBD) remains challenging to treat and difficult to prognosticate. Although multiple immune cell types coordinate pathology in both Crohn's disease (CD) and ulcerative colitis (UC), specifying which cell types and cell states portend better or worse response to major IBD treatment strategies, including anti-TNF therapies (the only FDA-approved therapy for pediatric IBD), remains challenging. Here, we present the results of the PREDICT study, which enrolled 79 treatment-naïve pediatric patients at the time of diagnostic endoscopy, enabling a comprehensive transcriptomic, histologic, and serologic analysis of 40 CD patients and 16 UC patients, as well as 23 patients with functional gastrointestinal disorders (FGID) who served as pediatric non-inflamed controls. Leveraging these data, we performed a comprehensive analysis of colonic immunology in each of these clinical conditions. Our results indicate that, within the complex landscape of immune pathology in pediatric CD and UC, there is a coordinated shift across the T_H1 -to- T_H17 immune activation continuum among T cells that is pertinent to anti-TNF response. For CD, this shift defines partial response to anti-TNF treatment. For UC, the landscape is more complex, with both T_H17 and T_{FH} biology defining disease, and pre-treatment T_H17 biology contributing to anti-TNF treatment resistance. Related to this, polyreactive TCR phenotypes within UC T_{FH} cells are correlated with both germinal center activity and the frequency of IgG1 plasma cells, yet opposed to the T_H17 signatures associated with anti-TNF nonresponse. The elucidation of these distinct mechanisms of T cell-dependent disease pathology, treatment response, and TCR polyreactivity suggests a model in which sustained T_H17 signaling in CD and baseline T_H17 signaling in UC are associated with disease pathogenesis, forming a basis for a generalized understanding of IBD pathogenesis and underscoring the need for endotype-specific approaches to IBD therapy.

INTRODUCTION

Inflammatory bowel disease (IBD), for which Crohn’s disease (CD) and ulcerative colitis (UC) are the most common subtypes, encompasses a group of debilitating inflammatory gastrointestinal conditions with increasing incidence¹. Although pediatric and adult disease have many clinical and biologic commonalities, pediatric IBD can be particularly devastating, given that this disease is often treatment refractory, more predisposed to rapid progression², and requires effective treatment for many decades. Recent efforts have catalogued immunologic associations with treatment resistance in biologic-, but not treatment-naïve, adults; however, no study has yet defined the comparative landscape of these diseases at diagnosis (naïve to all treatments), nor provided a multiomic map of these diseases in pediatric patients³. Understanding the cellular and molecular immunopathologic networks causing pediatric disease, elucidating the mechanisms of treatment failure in these young patients, and identifying novel avenues for therapeutic intervention⁴ remain major unmet clinical needs.

Acknowledging the central role that T cells play in both CD and UC, multiple therapies targeting T cell trafficking, IL-12/IL-23, and lymphocyte activation are in use. However, anti-TNF monoclonal antibodies remain first-line biologic treatment for both CD and UC^{5,6}, and the only FDA-approved biologic therapies for pediatric IBD. The success of anti-TNF agents is highly variable, and there remains no diagnostic paradigm for prioritizing other agents due to limited understanding of TNF-resistant cell states in CD or UC. Such functional knowledge is urgently required, given that differing mechanisms of IBD pathology may be better treated by biologic therapies targeted towards disease endotypes.

To address these knowledge gaps, we created a prospective natural history study, the Precision Diagnostics in Inflammatory Bowel Disease, Cellular Therapy, and Transplantation study (“PREDICT”, NCT03369353), which enrolled and evaluated a cohort of treatment-naïve pediatric patients (≥ 6 years old) with GI distress requiring endoscopy ($n = 79$), and who were diagnosed

with one of four conditions: non-inflammatory ‘Functional GI Disease’ (FGID), ileal-only CD (iCD), colon-involving CD (cCD), or UC. Key to the design of this study was the enrollment of patients prior to diagnostic endoscopy and any treatment, such that the distinctive immunobiology of these conditions could be determined, unaffected by concomitant immunosuppressive treatments. PREDICT combines detailed clinical evaluation and follow-up with multiomic biologic evaluation including 5'-scRNA-seq/scTCR-seq, multiplexed immunohistochemistry (IHC), and human proteome array-based autoantibody detection. This study enabled us to uncover distinct cellular disease mechanisms that drive pathology and treatment resistance within pediatric CD and UC, providing an evidence-based foundation for future evaluations of molecular disease-targeted therapies in these patients.

RESULTS

PREDICT: Clinical Study Enrollment, Clinical Outcomes, and Follow-Up

PREDICT IBD study participants were enrolled from November 2017 to January 2022, during which time 23 FGID, 7 iCD, 33 cCD (1 with perianal-only disease, and 2 without successfully processed diagnostic scRNA-seq biopsies), and 16 UC patients were enrolled (**Fig. 1a**). In addition to the diagnostic biopsies, 4 iCD, 10 cCD, and 2 UC patients contributed follow-up biopsies that were processed for downstream analysis. At diagnosis, enrolled patients underwent clinical and laboratory evaluation, and both upper and lower endoscopy to evaluate disease and obtain biopsies. Following endoscopy, standard clinical and histologic features were used by the treating physicians to classify participants as FGID, iCD, cCD, or UC, and all treatment decisions were made by the treating physicians⁷. After their initial diagnosis, patients with IBD were followed clinically for up to 2 years. Patients with FGID were followed clinically only as needed. Of the FGID, iCD, cCD, and UC patients analyzed, the median age at diagnosis was 15.7, 17.6, 11.4, and 12.8, respectively ($p = 0.013$, ANOVA, **Supplemental Table 1**). Males represented 56.5, 71.4,

54.5, and 56.3% of patients with FGID, iCD, cCD, and UC, respectively ($p = 0.88$, chi-squared test, **Supplemental Table 1**).

The majority (70%) of total IBD patients were treated with anti-TNF, with both infliximab and adalimumab used, based on treating physician decision-making (**Fig. 1a, Supplemental Table 1**). Among patients with successfully processed diagnostic biopsies, 5/7 iCD, 25/31 cCD, and 8/16 UC patients received anti-TNF. For these patients, we defined anti-TNF response status at 2 years post-diagnostic endoscopy. Anti-TNF treatment response categorization [Full Response (FR), Partial Response (PR), Nonresponse (NR)] was adjudicated by two pediatric gastroenterologists (HBZ and SR) based on the following criteria: FR was defined as clinical symptom control and biochemical response (normalization of C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR)) on maintenance anti-TNF therapy with dose adjustments only due to insufficient blood drug levels. PR was defined as improvement in clinical symptoms or CRP/ESR without full clinical symptom control or CRP/ESR normalization, with documented escalation of anti-TNF therapy or additional therapeutic agents added. NR was defined as lack of clinical and biochemical response to anti-TNF, using the criteria defined above for FR and PR. In PREDICT, 4/5 iCD, 15/25 cCD, and 0/8 UC patients achieved FR. 1/5 iCD, 10/25 cCD, and 4/8 UC patients achieved PR. 0/5 iCD, 0/25 cCD, and 4/8 UC patients were designated as NR (**Fig. 1a, Supplemental Table 1**). In addition, for each disease there were patients for whom the treating physician chose not to treat with anti-TNF agents (2/7 iCD, 6/31 cCD, and 8/16 UC). These patients are designated as “not on anti-TNF” (“NOA”), and were treated with a variety of other immunomodulators and antimetabolic agents, based on their treating physician’s choice.

scRNA-seq and IHC analysis define parenchymal, mesenchymal, and immunological features in FGID, CD and UC

We performed cellular profiling on fresh (non-frozen) diagnostic and follow-up colon biopsies through linked 5' scRNA-seq and scTCR-seq (**Fig. 1a**). During sample processing, biopsies were

separated into lamina propria and epithelial fractions, and samples from each fraction were separately prepared for scRNA-seq^{8,9}. The use of 5' chemistry enabled us to simultaneously interrogate the gene expression profiles and TCR identities of the profiled T cells using scTCR-seq. Following quality control and cellular annotation, 650,501 high-quality gene expression transcriptomes from diagnostic and follow-up biopsies were further analyzed (**Supplemental Table 1**).

To annotate cells, we first identified major “cell types” through low-resolution clustering (see Methods)¹⁰. These cell types were analyzed agnostic to GI layer of origin and included colonic epithelium; stromal and endothelial cells; myeloid cells; plasma cells; B cells; and T cells, NK cells & innate lymphoid cells (**Fig. 1b-g, Supplemental Fig. 1a-g, 2a-f, Supplemental Table 2**). Each cell type was separately analyzed to identify additional levels of detail, which we termed “cell subtypes” and “cell states” (**Supplemental Fig. 1h**). We manually annotated a total of 82 cell states within 31 cell subtypes across the 6 cell types (**Supplemental Table 3**)^{8,11,12}. Each final annotated cell state was represented by multiple samples, regardless of the number of cells comprising the cell state (**Supplemental Fig. 1a, Supplemental Fig. 2a-f**). In addition, IHC was performed on biopsies from 44 PREDICT participants (8 FGID, 6 iCD, 15 UC, 15 cCD). Two immunophenotyping panels were used, identifying B cells (CD20), T cells (CD3), myeloid cells (CD68), T_{reg} cells (FOXP3), CD8 T cells (CD8), plasma cells (BCMA), eosinophils (ECP), neutrophils (MPO), dendritic cells and monocytes (CD11c), NK cells (NCR1), and epithelial cells (panCK) (**Fig. 1h, i**).

Transcriptomic and Histologic Analysis Reveal Numerous Cellular Distinctions Between FGID and IBD

To statistically characterize major differences in the cell state frequencies between FGID controls and IBD at diagnostic endoscopy, we used Milo, a tool that generates differential abundance statistics for scRNA-seq data by assigning cells to partially overlapping neighborhoods on a k -

nearest neighbor graph¹³. Milo neighborhoods were created within each cell type to determine which cell states contributed most to the variation observed amongst FGID, cCD, and UC colon biopsies. When comparing IBD samples with colonic inflammation (both cCD and UC) to FGID controls, we found enrichment in multiple activated cell types, including inflamed epithelial cells, inflammatory fibroblasts, inflammatory macrophages, mononuclear phagocytes (MNPs), plasmacytoid dendritic cells (pDCs), neutrophils, and multiple activated and inflammatory T cell populations (**Fig. 2a, Supplemental Table 4**). These findings demonstrate a diverse mucosal inflammatory response in both cCD and UC. IHC analysis corroborated the overarching scRNA-seq findings, identifying increased densities of CD11c+ cells, MPO+ cells, total CD3+ cells, as well as CD3+CD8-FOXP3+ cells (T_{regs}) in IBD samples (**Fig. 2b, Supplemental Table 5**).

Transcriptomic and Histologic Distinguishers of cCD and UC: Evidence for Increased Epithelial Damage in UC

While pediatric cCD and UC can be challenging to distinguish endoscopically and histologically¹⁴, Milo differential abundance identified multiple cell states that could clearly distinguish these two IBD entities (**Fig. 2c**). Compared to UC, cCD biopsies were enriched in epithelial cell states suggestive of retained epithelial function, including non-inflamed absorptive colonocytes and goblet cells (**Fig. 2c**). In contrast, most epithelial cell states were depleted in diagnostic samples from UC patients, and the cells that remained in UC were skewed towards *LGR5*+ epithelial stem cells, *CCL24*-expressing goblet cells, and an epithelial population with expression of antimicrobial and T_H17-promoting genes (*PI3* and *SAA1*, respectively), previously demonstrated to be induced by proinflammatory cytokines, including TNF, IL-1 β , and IL-6 (**Fig. 2c, Supplemental Fig. 2a**)^{15–19}. These cell states are consistent with characteristic UC histopathology findings that reveal regenerating stem cells in the setting of inflamed epithelium lacking normally functioning absorptive and secretory cells²⁰.

Consistent with the epithelial damage in UC, the fibroblasts that support differentiating epithelium (annotated as *ABCA8*⁺ and *VSTM2A*⁺ fibroblasts) were generally absent from UC biopsies, but present in cCD biopsies (**Fig. 2c**). We also identified inflammatory fibroblast populations that were nearly exclusive to either cCD (“TFPI2Hi fibroblasts”) or UC (“PDPNHi fibroblasts”) (**Fig. 2c**). TFPI2Hi fibroblasts exhibited greater expression of chemokines *CCL13* and *CXCL10*, both of which are induced by IFN γ signaling (**Supplemental Fig. 2b**)^{21–23}. In contrast, PDPNHi fibroblasts were enriched for genes involved in extracellular matrix organization and vascular remodeling, as have been described in adult UC (**Supplemental Fig. 2b**)¹². In line with increased vascular remodeling, UC biopsies demonstrated an enrichment in the venule cell state and pericytes involved in capillary constriction and vascular remodeling (**Fig. 2c**)^{11,24}.

Transcriptomic and Histologic Distinguishers of cCD and UC: Evidence for Increased Inflammatory Myeloid Activity and Enhanced Germinal Center Activity in UC

While enrichments in inflammatory myeloid populations were often shared between both cCD and UC biopsies compared to FGID (with notable elevations in MNPs, inflammatory macrophages, and pDCs, **Fig. 2a**), UC biopsies were more skewed towards these monocyte, macrophage, and dendritic cell inflammatory states compared to cCD (**Fig. 2c**). IHC analysis also revealed a marked increase in MPO⁺ cells in UC, indicating neutrophilia that was not found in cCD samples (**Fig. 2d**). Plasma cells were significantly shifted towards IgG₁ and IgG₃ in UC (**Fig. 2c**), as has been previously reported^{25–27}. We also observed that, concomitant with the increase in IgG, there was a coordinated decrease in both IgM and IgA plasma cells among UC biopsies, with the latter dominated by a reduction in IgA₂, rather than all IgA cells (**Fig. 2c**). Potentially connected to the shifts in the immunoglobulin balance, UC could be distinguished from cCD by the presence of proliferating B cells and germinal center-like B cells, which were both increased in UC versus cCD (**Fig. 2c**). UC samples were also enriched for PD1Hi T_{FH} cells, which are likely to be germinal center T_{FH} cells^{28,29} (**Fig. 2c**). The combination of increases in these cell state frequencies

suggested elevated germinal center activity in colonic mucosa from UC patients compared to both FGID and cCD, which likely contributed to the observed plasma cell shifts. These scRNA-seq findings were corroborated by IHC analysis which demonstrated increased CD20+ (B cells) and BCMA+ (plasma cells) cell density in UC relative to cCD (**Fig. 2d**).

Transcriptomic and Histologic Distinguishers of cCD and UC: Evidence for skewing towards T_H1 immunity for CD, and T_H17 immunity for UC at the time of diagnosis

T cells play a major role in IBD, coordinating interactions with immune, stromal, and parenchymal cells that contribute to inflammation, with known instructive roles for T_H1 and T_H17 cells in CD and UC^{3,8,9,30–32}. We found that both T_H1 and T_H17 cell states were enriched in cCD and UC relative to FGID (**Fig. 2a**). Moreover, testing for differences between cCD and UC highlighted a notable shift towards T_H1 cells in cCD and towards T_H17 cells in UC (**Fig. 2c**). Milo-based differential abundance analysis was corroborated with per-biopsy comparisons of FGID, iCD, cCD, and UC cell state frequencies (**Fig. 2e-n, Supplemental Table 5**). Given the well-established connections between inflammatory T cell signaling and T cell proliferation³³, we explicitly interrogated proliferating T cell states in CD versus UC. Because proliferating T cell states did not comprise enough cellular neighborhoods for Milo analysis (see Methods), this analysis was limited to per-biopsy comparisons. Our findings were consistent with the observation described above, of more PD1Hi T_{FH} in UC versus CD (**Fig. 2c**), with a trend towards more proliferating T_{FH} cells also observed in UC (**Fig. 2o**). Because proliferating T cells often express transcriptomic hallmarks of both T_H1 and T_H17 cells^{34,35}, we defined a proliferating T cell state with features of both (including *CXCR3*, *IFNG*, *IL17A*, and *GNLY*, termed “Prlf. $T_H17.1$ ”). Interrogation of Prlf. $T_H17.1$ cells revealed that they were enriched in both cCD and UC compared to FGID, and of the two disease states, most enriched in patients with UC (**Fig. 2p, Supplemental Fig. 2f**).

T_H cells are associated with increased T cell activation and exhaustion in UC

Having identified increased T_H1 frequency in cCD and increased T_H17 frequency in UC, we investigated whether detailed analysis of these T cell populations could further elucidate distinct signatures in each IBD subtype. Given the plasticity of helper T cells and their divergent association with cCD and UC, we performed differential gene expression comparing UC and cCD within the “T helper” cell subtype, which includes T_H1, T_H17, and Prlf. T_H17.1 cells. This analysis identified genes enriched in UC versus cCD related to T cell activation, co-stimulation, or exhaustion, including *TOX*, *TNFRSF4* (OX40), *TNFRSF18* (GITR), *LAG3*, and *CD247* (CD3ζ) (**Supplemental Fig. 3a, Supplemental Table 6**). We further delineated these DE genes within the T_H1 and T_H17 cell states. We found that these genes, as well as *HAVCR2* (TIM-3), were increased in both UC T_H17 and UC T_H1 cells, confirming that their enrichment in the T_H cell subtype was not solely due to differences in cell state abundance (**Supplemental Fig. 3b, Supplemental Table 7**).

Next, we investigated the ambiguity of the Prlf. T_H17.1 cell state by analyzing the expression of T_H1- and T_H17-identifying genes within the Prlf. T_H17.1 cell state, as well as within T_H1, and T_H17 cell states. Even amongst cells annotated as T_H1, we found increased expression of T_H1-associated genes (*IFNG*, *TNF*, *ANXA1*, *CXCR3*, *GZMK*) in cCD compared to UC (**Supplemental Fig. 3c**). Importantly, *TNF*, *ANXA1*, *CXCR3*, and *GZMK* were also increased in cCD Prlf. T_H17.1 cells, suggesting that T_H1 cells may be the primary cell state of origin in cCD Prlf. T_H17.1 cells (**Supplemental Fig. 3c**). Conversely, genes *IL17A*, *IL17F*, *IL26*, and *GNLY* were increased in UC Prlf. T_H17.1 cells relative to cCD, suggesting an association with the T_H17 cell state (**Supplemental Fig. 3d**), and distinguishing the origin of these cells in the two IBD endotypes. Notably, we observed increased *IL17A* expression in UC T_H1 cells. In the context of increased T cell activation and exhaustion in UC, these data suggest that UC T_H1 cells may actually be ex-T_H17 cells, known to take on a T_H1-like phenotype under prolonged inflammatory conditions³⁶.

Clonal overlap within T_H cell states highlights enrichment for $CD8A^+$ T_H17 cells in UC

We took advantage of the 5' scRNA-seq chemistry to simultaneously track linked TCR sequences, to determine the extent to which T_H1 or T_H17 clones shared TCRs with proliferating $T_H17.1$ cells in cCD or UC (**Supplemental Fig. 3e-g, Supplemental Table 8**). We found that Prlf. $T_H17.1$ clones came from both T_H1 and T_H17 cell states in cCD, with a trend towards increased representation of T_H1 clones (**Supplemental Fig. 3h**). In contrast, nearly all clonal overlap in UC was between Prlf. $T_H17.1$ cells and T_H17 cells, reinforcing the likelihood that actively proliferating T_H cells in UC are primarily of T_H17 origin. We also noted $CD8A$ expression in the $CD4^+$ T_H17 cell state (**Supplemental Fig. 3i**). Given our previous investigation demonstrating expansion of $CD8^+$ T_H17 cells in adult UC⁸, we sought to determine their activity in pediatric UC. We found that dual $CD4^+CD8A^+$ T_H17 cells were significantly enriched for clonal overlap with proliferating $T_H17.1$ cells, suggesting active clonal expansion of this cell substate in UC (**Supplemental Fig. 3j**).

T_H1 , T_H17 , and $PD1^+ T_{FH}$ cells exhibit distinct cell state correlation networks

To explore the extent to which other immune pathways interacted with T_H1 , T_H17 , and $PD1^+ T_{FH}$ populations, we performed a cell state abundance correlation analysis between these three $CD4$ subpopulations and each of the other annotated cell states (**Fig. 3a, Supplemental Table 9**). Our analysis revealed several cell states across all cell types that were significantly associated with each of these $CD4$ subpopulations. T_H1 cells were associated with NK cells, inflammatory macrophages, and *TFPI2*-high fibroblasts, consistent with a type 1 immune response^{37,38}. Amongst T_H17 cells, significant correlations were observed for a variety of cell states, with the strongest correlations between MNPs and venules, as well as between T_H17 cells and the *PI3*⁺ inflamed epithelium or *PDPN*⁺ remodeling fibroblasts. Importantly, while both T_H17 and $PD1^+ T_{FH}$ cells were enriched in UC patients, their correlation networks were largely distinct, with $PD1^+ T_{FH}$ being primarily associated with IgG₁ plasma cells, GC-like and proliferating B cells, and plasmablasts, consistent with their role in T cell-mediated B cell immune activation.

The T_H1 correlation network is enriched in cCD and associated with $STAT1$ -high cell states

We calculated the per-biopsy geometric mean of the cell states involved in each correlation network to determine the extent to which each network aligned with each disease state. We found that cCD was distinct from FGID, UC, and iCD in its enrichment for T_H1 -associated cell states (**Fig. 3b**). Because IFN γ is a predominant T_H1 cytokine, we next determined whether IFN γ may be associated with the cell states in the T_H1 network. We first examined *STAT1* expression in all cell states, given that *STAT1* both mediates and is induced by IFN γ signaling^{39,40}. We found that 4 of the 6 top *STAT1*-expressing cell states (Inflm. Mac, T_H1 , MNP, TFPI2Hi Fibro.) were present in the T_H1 correlation network (**Supplemental Fig. 4a**). We then compared *IFNG* expression of *IFNG*-expressing cell subtypes (including T_H , NK, CTL, and InvariantT; **Supplemental Fig. 4b-c**) to *STAT1* expression in T_H1 -associated cell states within the same biopsy, finding significant correlations between T_H *IFNG* expression and *STAT1* expression in inflammatory macrophages, T_H1 cells, T_{reg} s, and XCL+ NK cells (**Fig. 3c, Supplemental Table 10**). The specificity of this correlation was underscored by the fact that *IFNG* from other *IFNG*-expressing cell subtypes was not associated with *STAT1* in any T_H1 -associated cell states. (**Supplemental Fig. 4d**). These correlations were reproduced when we restricted the dataset from all enrolled patients to focus only on cCD and UC biopsies, suggesting that the T_H IFN γ association with *STAT1* expression is not solely due to the overall status of tissue inflammation (**Supplemental Fig. 4e**).

The T_H17 correlation network is most substantially enriched in UC, and is associated with epithelial antimicrobial and chemokine response

Although cCD was significantly enriched for T_H17 -associated states compared to both FGID and iCD, this network was most enriched in UC patients (**Fig. 3d**). To further interrogate T_H17 signaling, we next focused on IL-17A, the key T_H17 cytokine. IL-17A is known to have multiple effects, including enhanced recruitment of peripheral leukocytes and increased antimicrobial

defense^{36,41–43}. We therefore created gene set scores for IL-17-induced antimicrobial function, chemokine production, cytokine production, and tissue remodeling using genes downstream of IL-17 signaling in the KEGG IL-17 signaling pathway (**Supplemental Fig. 5a**)^{44–46}. These gene sets were active across multiple cell types (**Supplemental Fig. 5b, Supplemental Table 11**). Given IL-17 effects on barrier function and the presence of T_H17 cells within the epithelial fraction of our biopsies (**Fig. 3e, Supplemental Fig. 1a**), we focused our analyses on IL-17 effects in colonic epithelium. The antimicrobial and chemokine gene sets were active in colonic epithelium (**Fig. 3f**). We found strong associations between *IL17A* in T_H cells and activity of IL-17 gene sets in colonic epithelium, suggesting either a direct influence of *IL17A* on this epithelium, or a common process contributing to expression of these genes in distinct cell types (**Fig. 3g**). Strong correlations remained when we restricted analysis to cCD and UC biopsies only, but the correlations were not observed between *IL17A* and active IL-17 gene sets in other cell types (**Supplemental Fig. 5c**). The antimicrobial and chemokine gene sets, and associated antimicrobial and chemokine genes, were active in similar colonic epithelial cells, were significantly enriched in cCD compared to FGID, and were further enriched in UC (**Fig. 3h, Supplemental Fig. 5d-g**). Because many of the chemokines in the IL-17-induced chemokine gene set are known neutrophil chemoattractants, we contextualized the scRNA-seq results with multiplexed IHC and found profound neutrophilia in UC epithelium (**Supplemental Fig. 5i**). A similar increase in potentially inflammatory conventional CD4 T cells (CD3+CD8-FOXP3-), was not observed histologically in UC epithelium relative to cCD (**Supplemental Fig. 5j**).

The PD1Hi T_{FH} correlation network is enriched in UC and is associated with generation of humoral immunity

In addition to their enrichment in T_H17 networks, UC patients also demonstrated significant enrichment in the PD1Hi T_{FH} network (**Fig. 3i**). This group of differentially abundant cell states was highly suggestive of increased T cell-dependent B cell activation and tertiary lymphoid

structure activity. The frequencies of GC-like B cells and proliferating B cells were highly correlated ($R = 0.8$) (**Fig. 3a, Supplemental Table 9**). We considered that these cell states may be part of tertiary lymphoid structures (TLS). We thus combined both cell states into a “TLS-like B Cell” annotation and observed this grouping of cells to be increased in UC biopsies relative to both FGID and cCD (**Fig. 3j**). We then examined expression of the germinal center-organizing cytokine *CXCL13*. Its expression was increased in UC T_{FH} cells and was greatly enriched in UC follicular dendritic cells (FDCs) (**Fig. 3k**). Notably, *CXCL13* expression among FDCs correlated with TLS-like B cell frequency, but only in UC (**Fig 3l**). Histologically, we identified lymphoid aggregates of B and T cells, which contained a center of CD20+ cells surrounded by a ring of CD3+ cells (morphologically resembling a germinal center), in biopsies across all disease groups (**Fig. 3m-n**). Importantly, amongst biopsies with any identified TLS, we observed an increase in the number of TLS per biopsy in UC, as well as increased CD20+ cell density within these structures in UC relative to cCD (**Fig. 3o-p**).

Predictors of Anti-TNF Response in IBD: cCD

Differential gene expression at diagnosis does not distinguish anti-TNF treatment or response groups in cCD

Because anti-TNF agents are the first line (and only FDA-approved) biologic therapy in pediatric IBD, we next investigated whether we could determine differential expression signatures associated with response to these agents from data in diagnostic biopsies (obtained prior to any treatment) from IBD patients. We first performed per-cell type differential expression (DE) analysis between cCD patients who would be treated with anti-TNF and those not treated with anti-TNF in the 2-year follow-up period (the ‘NOA’ patients). This analysis identified a paucity of DE differences, with the only DE gene identified being *TNFAIP6*, for which the enrichment was largely driven by a single outlier sample (**Supplemental Fig. 6a-b, Supplemental Table 6**). We next performed a similar DE analysis to compare cCD FR to PR patients at the time of diagnostic

biopsy. This analysis identified only the plasma cell gene *PHGDH* (**Supplemental Fig. 6c-d, Supplemental Table 6**) differentially expressed between the two groups. This enzyme, critical for serine synthesis, was enriched in cCD PR vs. FR⁴⁷. These results underscore the challenge in creating a predictive signature at diagnosis for anti-TNF response in cCD^{48,49}.

Post-anti-TNF treatment cCD biopsies are defined by a decrease in IFNG signaling

Unlike the lack of differences in diagnostic cCD biopsies, an immunologic signature did emerge when post-treatment cCD biopsies were compared between FR and PR patients. To determine the cell types most likely to reflect an immunologic impact from anti-TNF therapies, we first scored each cell type for pre-treatment TNF pathway activity (**Fig. 4a**). Both myeloid cells and TNKILCs demonstrated elevated TNF pathway scores and were thus chosen for more detailed analyses. We found that TNF pathway scores were decreased for both myeloid cells and TNKILCs in post-treatment cCD relative to pre-treatment for both response groups, consistent with a biologic impact of anti-TNF on all treated patients (regardless of anti-TNF response) (**Fig. 4b-c**)^{50,51}. To more deeply interrogate the molecular impact of anti-TNF on myeloid and TNKILCs, we conducted a DE analysis in these cell types, comparing pre- and post-treatment cCD, without regard to treatment response (**Fig. 4d**). Several immune genes were found to be upregulated in pre-treatment cCD myeloid cells and TNKILCs relative to post-treatment, including *STAT1*, cytolytic genes (*GNLY*, *PRF1*), calprotectin (*S100A8*, *S100A9*), and IFN γ -induced genes (*CXCL9*, *CXCL10*, *GBP1*, *GBP2*, *GBP5*)⁵². Moreover, “Interferon gamma signaling” was the top Reactome term enriched in pre-treatment cCD differential expression, consistent with IFN γ -induced inflammatory states suggested in the T_H1 network (**Fig. 3c, 4e, Supplemental Table 6**). Based on these findings, we compared pre- and post-treatment cell state frequencies of T_H1-associated cell states among cCD biopsies, finding decreases in each cell state (**Fig. 3a, 4f**). Additionally, the two other top *STAT1*-expressing cell states, MNPs and T_H17 cells, were also decreased in post-treatment cCD, when all treatment response groups were analyzed together (**Fig. 4f**). Overall,

these findings suggest successful abrogation of IFN γ signaling through anti-TNF therapy in cCD, regardless of clinical treatment response.

Post-treatment cCD partial responders retain T_H17 signaling

With overall differences in pre- and post-treatment cCD delineated, we investigated whether post-treatment biopsies would illuminate differences in cCD FR and cCD PR. This analysis revealed decreases in the T_H1-associated inflammatory macrophage cell state in both FR and PR (**Fig. 4g**), and that T_H1 cells themselves were significantly decreased in FR with a trend towards reduction in PR (**Fig. 4h**). In contrast, the T_H17-associated MNP cell state (shown in **Fig. 3a**), as well as T_H17 cells themselves, were decreased only in cCD FR, while being maintained in PR (**Fig. 4i-j**). To more fully evaluate canonical T_H1 and T_H17 cytokine gene expression, we analyzed the expression of key T_H1 cytokine genes (*IFNG*, *TNF*) and T_H17 cytokine genes (*IL17A*, *IL17F*, *IL21*, *IL22*, *IL26*) within the T_H cell subtype. Consistent with the findings above, a gene signature created from these cytokines was *significantly reduced* in the T_H cells from cCD FR patients post-treatment relative to pre-treatment, but *significantly increased* in cCD PR patients post-treatment T_H cells relative to pre-treatment (**Fig. 4k**)⁵⁰. Single gene analysis of pre- and post-treatment T_H cells revealed trends toward reduction in *TNF*, *IFNG*, *IL22*, and *IL26* in cCD FR (**Fig. 4l**). In contrast, cCD PR T_H cells demonstrated unchanged *TNF* and *IFNG*, along with trends towards increases in *IL17A*, *IL17F*, *IL21*, and *IL22* after anti-TNF treatment (**Fig. 4l**). Together, these data point to persistent T_H17 immunopathology associated with anti-TNF PR in cCD.

Predictors of Anti-TNF Response in IBD: UC

UC diagnostic biopsies from anti-TNF-treated patients are enriched for markers of inflammation and epithelial cell cycling

In the PREDICT cohort, the number of UC patients with repeat biopsies (2/16) was too few to rigorously define the post-treatment impact of anti-TNF therapies on immune cells and

interactions. However, unlike for CD, UC diagnostic biopsies were discriminative of anti-TNF response. To discover this, we first performed an analysis comparing diagnostic biopsies from UC patients later treated with anti-TNF to NOA patients, utilizing the per-cell type DE analysis paradigm described above. Unlike in cCD patients, where informative gene expression changes were not observed, in UC, several notable genes were transcriptomically distinct in patients for whom the treating physician initiated anti-TNF therapy versus the NOA patients. Among colonic epithelial cells in pre-treatment biopsies, genes enriched in UC NOA patients were notable for *CA1*, *CA2*, and *SLC26A3*, representing normal colonic absorptive function (**Supplemental Fig. 7a, Supplemental Table 12**). Genes enriched in the colonic epithelium from UC patients who would eventually be treated with anti-TNF therapies included cell cycle-associated genes, such as *RRM2*, *ZWINT*, and *CENPM* (**Supplemental Fig. 7a, Supplemental Table 12**). Reactome pathways enriched for the DE results included GTPase cell signaling pathways for UC NOA patients, and DNA synthesis and replication pathways for UC anti-TNF patients (**Supplemental Fig. 7b-c**). Additional notable DE genes in UC patients treated with anti-TNF therapies included the neutrophil chemoattractant *CXCL6* in stromal cells, and the inflammatory cytokine-induced ferritin transport genes *FTL* and *FTH1* in myeloid cells^{53,54} (**Supplemental Fig. 7a, Supplemental Table 6**).

At the time of diagnostic biopsy, UC patients who would become anti-TNF NR were enriched for extracellular remodeling and T cell differentiation genes

We next evaluated diagnostic biopsies from UC patients who were eventually treated with anti-TNF for per-cell type DE that could differentiate PR versus NR. Of note, PR was the best response achieved in the PREDICT UC cohort (although other studies have documented UC patients achieving FR to anti-TNF)⁵⁵⁻⁵⁷. In patients that partially responded to anti-TNF versus NR patients, pre-treatment UC colonic epithelium and TNKILCs were enriched for genes in cellular metabolism and maintenance pathways (**Fig. 5a-b**). *MIF*, which promotes *TNF* induction and whose release

is induced by microbial proteins, was enriched in UC PR versus NR myeloid cells, B cells, and TNKILCs (**Fig. 5a**)^{58–60}. Additional notable genes enriched in UC PR versus NR included *LYZ* (an antimicrobial enzyme) in myeloid cells and interferon-stimulated genes *IFI6*, *IFITM1*, and *ISG15* in plasma cells (**Fig. 5a**)⁶¹.

In contrast, UC NR versus PR colonic epithelium demonstrated enrichment in pathways governing extracellular degradation and remodeling (**Fig. 5a, c**). Notably, *FCER2* (encoding CD23) and *JAK3* (induced upon CD40 ligation or bacterial stimulation, necessary for lymphocyte proliferation and differentiation, and for which there is an FDA-approved therapeutic in adult UC) were enriched in UC NR B cells^{62–65}. Many of the genes enriched in UC NR TNKILC are involved in differentiation, activation, and migration of CD4 T cells, including *RORA* (T_H17 cytokine production), *CCR6* (T_H17 and T_{reg} migration), *FOXP3* (T_{reg} differentiation), *PRDM1* (encoding BLIMP-1, which mediates chronic activation/exhaustion), *IL12RB1* (subunit of IL-12 and IL-23 receptors), and *JAK3* (**Fig. 5a**)^{66–71}. We also examined per-biopsy expression of several TNKILC DE genes in the TNKILC cell type, which recapitulated the DE findings described above (**Fig. 5d**). Of note, in addition to therapeutics directed at *JAK3*, *IL12RB1* is involved in pathways targeted by anti-IL-12p40/anti-IL23p19 therapies, underscoring the potential availability of alternatives to anti-TNF in UC patients bearing anti-TNF nonresponse markers^{64,72}.

At the time of diagnostic biopsy, UC patients who would become anti-TNF NR were enriched for T_H IL17A

We further evaluated the role of T_H17 cytokine signaling in defining UC treatment response. We found that, compared to pre-treatment UC PR, UC NR T_H cells were enriched in T_H17 cytokines, with all UC NR biopsies having greater T_H *IL17A* and *IL21* than UC PR biopsies (**Fig. 5e**). Thus, *IL17A* was identified as a marker of anti-TNF therapy failure in both cCD and UC.

A distinct T peripheral helper (T_{PH}) subset was not detected in pediatric IBD patients

In addition to T_H17 cells, adult UC has previously been associated with T_{PH} cells²⁶. Thus, we explored whether T_{PH} cells were present in the pediatric IBD patients profiled in PREDICT. We found that neither the PD1Lo T_{FH} nor the PD1Hi T_{FH} population clearly expressed T_{PH} markers (**Supplemental Fig. 8a-b, Supplemental Table 7**)⁷³. Additionally, no clear subpopulation of cells expressed a combination of canonical T_{PH} markers⁷³, assessed through UMAP visualization and through gene expression across a pseudotime diffusion vector spanning CD4 T cells (**Supplemental Fig. 8c-e, Supplemental Table 7**). This may represent a phenotypic difference in adult and pediatric UC.

UC T_H17 cells exhibit T_{FH} features and are inversely correlated with a UC-specific shift towards IgG₁ plasma cells

While no evidence for an association between T_{PH} and UC patients was observed, we did identify a subset of T_{FH} cells that were altered in PREDICT's UC patients: We found that although total T_{FH} cells were present in similar relative frequencies in all PREDICT disease groups, UC could be distinguished by the fact that T_{FH} cells in these patients were more shifted towards the PD1Hi T_{FH} cell state (**Fig. 6a-b**). Moreover, amongst T_{FH} genes involved in B cell help⁷³, we found *CXCL13* to be increased in UC PD1Hi T_{FH} relative to cCD (**Fig. 6c, Supplemental Fig. 8a**). During cell state annotation, we also observed that B cell help genes were present in T_H17 cells (**Supplemental Fig. 2f**), with *MAF*, *PDCD1*, *IL21*, and *CXCL13* being increased in UC T_H17 cells compared to cCD T_H17 cells (**Fig. 6d**).

Given these data, we considered a model wherein UC T_H17 cells, but not cCD T_H17 cells, exhibit T_{FH}-like plasticity. To test this model, we assessed which set of paired CDR3αβ sequences were detectable between non-proliferating T_{FH} cells and other T_H cells within a biopsy (**Fig. 6e-f**). Our analysis demonstrated that, in 6/14 UC diagnostic biopsies, there was evidence for T_{FH}-T_H clonal overlap, while only 3/30 cCD diagnostic biopsies demonstrated T_{FH}-T_H clonal overlap (**Fig. 6g-h**,

Supplemental Table 8). Clonal sharing was primarily between T_H17 and PD1Hi T_{FH} , with only one clonal pair involving T_H1 or PD1Lo T_{FH} cells (**Supplemental Table 8**). These data indicate that, although a small absolute number of shared clonotypes was identified, UC biopsies were enriched for identifiable T_{FH} - T_H clonal overlap (**Fig. 6g-h**).

To determine whether the coordination of these phenotypic shifts would be distinct in cCD versus UC, we compared the relative frequencies of T_H17 cells among all T_H cells, and PD1Hi T_{FH} cells among all T_{FH} cells. Notably, we found opposite patterns in the two IBD endotypes, with T_H17 and PD1Hi T_{FH} cells being positively correlated in cCD, but inversely correlated in UC (**Fig. 6i**, **Supplemental Table 10**). Considering the implications of differences in T cell phenotypes on B cell help and the development of antibody-secreting cells, we explored whether PD1Hi T_{FH} cells or T_H17 cells were associated with the shift towards IgG₁ plasma cells observed in UC (**Fig. 2i**). We found that in both cCD and UC, PD1Hi T_{FH} cell proportion and *CXCL13* expression were positively correlated with IgG₁ plasma cells (**Fig. 6j-k**). However, in T_H17 cells, we found a *strong inverse correlation* of T_H17 cell state frequency or *IL17A* expression and IgG₁ plasma cells in UC, despite both T_H17 and IgG₁ plasma cells both being increased in UC compared to CD (**Fig. 2i**, **2n**, **6j-k**). These data suggest an antagonism between T_H17 -associated and T_{FH} -associated signaling in UC.

UC T_{FH} TCRs exhibit polyreactivity signatures that are directly correlated with UC-specific plasma cell shifts and inversely associated with T_H17 -derived *IL17A*

Because of the prevalence of T-B interactions observed in UC, and known clinical associations with autoreactivity among IBD patients⁷⁴, we hypothesized that UC TCRs may bear features that distinguish them from cCD TCRs, and that TCR phenotypes could play a role in the plasticity observed among UC T cells.

Previous studies have suggested that increased CDR3 β hydrophobicity may predispose TCRs to polyreactivity or autoreactivity through non-specific interactions^{75–77}. Pointing towards poly- and/or

autoreactivity in UC T cells, we observed a relative increase in usage of large, aliphatic, and hydrophobic amino acid residues, with a corresponding decrease in polar amino acids, in the polymorphic middle region of CD4 T cell CDR3 β sequences in patients with UC (**Fig. 7a-b, Supplemental Table 8**). To relate these changes to functional differences in TCRs, we applied the previously-defined middle-TCR intrinsic regulatory potential (mTiRP) score⁷⁸. mTiRP is a TCR β -based metric created from the association of CD4 TCRs with the T_{reg} phenotype, given that T_{regs} derived from the thymus are trained to recognize self and commensal antigens, and are thus more likely to be self-reactive^{78,79}.

Our analysis demonstrated that UC biopsies exhibited increased mTiRP scores within CD4 T cells overall, and within T_H cells, T_{reg} cells, and T_{FH} cells (**Fig. 7c-d**). Moreover, mTiRP scores were further increased in UC PD1Hi T_{FH} cells relative to PD1Lo T_{FH} cells, and in clonally expanded UC PD1Hi T_{FH} cells relative to singlets, suggesting that more active T_{FH} populations in UC bear increased features of polyreactivity (**Fig. 7e-f**). PD1Hi T_{FH} mTiRP scores were significantly associated with TLS-like B cell ($R^2 = 0.53$, $p = 0.011$) and IgG₁ plasma cell ($R^2 = 0.65$, $p = 0.003$) frequencies in UC, but not in FGID or cCD (**Fig. 7g-h**), potentially suggesting this TCR phenotype helps drive germinal center reactions in UC.

Considering the overlapping and opposing T_H17 and PD1Hi T_{FH} phenotypes observed in UC, we examined whether PD1Hi T_{FH} mTiRP would have any relationship to T_H17 gene expression. Among canonical T_H17 cytokines, we identified a strong inverse association between UC PD1Hi T_{FH} mTiRP score and UC T_H17 *IL17A* expression ($R^2 = 0.87$, $p < 0.001$) (**Fig. 7i**). We note that UC NR pre-treatment (and post-treatment) biopsies bore the lowest average PD1Hi T_{FH} mTiRP score among UC participants, further suggesting orthogonal mechanisms of pathology (**Fig. 7j**).

Taken together, these data demonstrate that UC T_{FH} cells display a distinct phenotype from T_{FH} cells in both inflammatory (cCD) and non-inflammatory (FGID) study participants. These phenotypes are associated with polyreactivity and correlate strongly with UC-specific increases

in the development of humoral immune activation. However, they are inversely correlated with the T_H17 features associated with anti-TNF nonresponse, suggesting non-overlapping mechanisms of immune pathology in UC.

Autoantibody detection through human proteome array analysis reveals high intra-group variability in pediatric IBD patients

In previous studies, in addition to germinal center activation, adult patients with UC have demonstrated enhanced autoantibody production along with other signs of humoral immune pathogenesis^{26,27}. Because of this, we performed a proteome-wide screen of plasma from PREDICT participants to determine the burden of autoantibodies on a per-patient basis. To accomplish this, we used the HuProt microarray (CDI Labs)^{80–83}, which assesses immunoglobulins for binding to full-length recombinant human proteins produced in yeast, with coverage against approximately 80% of the human proteome^{80–83}. For comparison, we included samples (N=5) from patients diagnosed with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), a rare condition caused by mutations in *AIRE* that is characterized by a broad range of autoantibodies, which can be considered positive controls⁸⁴. As shown in **Supplemental Fig. 9a** and **Supplemental Table 13**, compared to APECED controls, the most striking finding in PREDICT patients was the high intra-group variability that was evident when comparing the median number of autoantigen hits per patient in each group: patients were present in all PREDICT cohorts with low and high median numbers of autoantigen hits. Overall, the median (interquartile range [IQR]) number of IgG hits was 185.5 (135.75-284.75) for FGID, 244 (124.5-361.5) for iCD, 301.5 (167-391) for cCD, and 215 (127-322) for UC, with each cohort containing patients demonstrating similar numbers of hits compared to the APECED controls (379 [364-437]). Similar variability was observed among IgA autoantigens (**Supplemental Table 13**). While many factors may contribute to the high patient-to-patient variability observed in PREDICT, some of this variability was likely explained by differences in antigen presentation driven by

specific HLA:TCR interactions. An example of this, wherein 2 UC patients expressing the HLA-DRB1*15:01 allele developed a distinct pattern of autoantigens, is described below.

Patients with the HLA-DRB1*15:01 UC risk allele illustrate the T_{FH} : T_H 17 inverse relationship and a distinct autoantibody signature

Given the association of TCR characteristics and UC phenotypes, together with the pronounced inter-patient variability in autoantibody burden we discovered with HuProt analysis, we investigated whether HLA antigen presentation may influence the T_{FH} / T_H 17 dichotomy, as well as the presence of patient-specific auto-antibody patterns, that we observed in UC. To accomplish this, we analyzed genomic data from PREDICT UC participants to map their HLA alleles, an analysis which detected several previously-identified HLA-DRB1 alleles associated with UC risk (**Supplemental Fig. 9b, Supplemental Tables 1, 12**)⁸⁵. HLA-DRB1*15:01, classically associated with multiple sclerosis, but also identified as a risk allele in UC, was present in two UC participants (participant IDs 001-036 and 001-038, **Supplemental Fig. 9b**)^{85,86}. Evaluation of TCR phenotypes revealed that biopsies from these two individuals had the highest PD1Hi T_{FH} mTiRP scores (**Supplemental Fig. 9c**). We also found substantially decreased frequency of T_H 17 cells and decreased expression of T_H 17 *IL17A* in these patients, relative to UC participants without the HLA-DRB1*15:01 allele (**Supplemental Fig. 9d-e**), consistent with the dichotomous relationship between T_H 17- and T_{FH} biology we previously documented in UC (**Fig. 6i-k, 7i**). As predicted, PD1Hi T_{FH} frequency and *CXCL13* expression in both T_H 17 cells and PD1Hi T_{FH} cells were increased in these individuals (**Supplemental Fig. 9d, f**). IgG₁ plasma cells were also increased in these individuals, while IgA₂ plasma cells were decreased (**Supplemental Fig. 9g**). Interestingly, the “PI3+ Epi.” cell state, an inflamed epithelial cell state likely reflecting microbial interaction, was also increased in the individuals with the HLA-DRB1*15:01 allele (**Supplemental Fig. 9h**).

Further analysis of the HuProt microarray data demonstrated that these participants (IDs 001-036 and 001-038) were jointly reactive to multiple antigens, some of which were uniquely present in this pair. These include autoantibodies directed against ENO1, an autoantigen previously identified in multiple autoimmune disorders⁸⁷⁻⁹¹, as well as two novel targets, YWHAQ⁹², MED18⁹³ (**Supplemental Fig. 9i-j**). Together, these data are consistent with a model wherein the combination of HLA type, antigen presentation, and T cell biology contribute to the unique autoantibody profiles observed in UC patients and may further explain the intra-group autoantibody heterogeneity observed across PREDICT.

Three axes of immunopathology in pediatric IBD

The combined scRNA-Seq, scTCR-Seq, IHC, and autoantibody analysis have led to a concerted model of pediatric IBD that encompasses three axes of immunopathology, with prominent features associated with T_H1 , T_H17 , and T_{FH} cells (**Supplemental Fig. 10a, Supplemental Table 14**) differentially represented across IBD endotypes. While cCD was primarily enriched in T_H1 features with a moderate increase in T_H17 features compared to FGID, UC was strongly enriched in T_H17 and T_{FH} features compared to both patient groups (**Supplemental Fig. 10b**). Although we observed no baseline differences in cCD patients who were eventually treated with anti-TNF that would predict their anti-TNF response status, analysis of post-treatment biopsies identified persistent T_H17 immunopathology as a signature of less-responsive cCD patients (**Supplemental Fig. 10c-d**). T_H17 cells were further implicated in anti-TNF response when we examined pre-treatment UC biopsies, which demonstrated an increase in T_H17 features among NR compared to PR patients (there were no FR to anti-TNF in the PREDICT UC cohort) (**Supplemental Fig. 10e**). Paradoxically, although both T_{FH} and T_H17 features were enriched in UC, we found orthogonal skewing of these factors: patients with greater enrichment for T_{FH} features had decreased association with T_H17 features (**Supplemental Fig. 10f**)⁸⁷⁻⁹¹. Taken together, these

data provide evidence for multiple disease endotypes within classically defined IBD that can be distinguished by their guiding immunopathologic signatures.

DISCUSSION

Here we present a multiomic examination of pediatric IBD from patients enrolled in the PREDICT study, identifying the distinguishing cellular and molecular features of cCD and UC, and their response to treatment with anti-TNF therapies. Our in-depth profiling provides a comprehensive cellular and molecular landscape of parenchymal, mesenchymal, and immune cells in the pediatric IBD colon prior to biologic or non-biologic treatment, with critical contextualization provided by non-colon-involved CD, and by pediatric FGID patients, who have GI symptoms, but lack IBD-associated inflammation. Most cellular profiling of adult or pediatric IBD has focused on either CD or UC, without direct comparison between the two disease states^{8,9,30,32,94–97}. Although these studies have been crucial for CD and UC characterization, the lack of direct comparison within the same study prevents the elucidation of critical nuances both between and within the two diseases. The PREDICT data underscore the importance of these direct comparisons, given that, while there were multiple immune states that were dysregulated in both CD and UC compared to FGID, close examination revealed distinct patterns of pathology between these two subtypes, revealing cell states and immune pathways specifically enriched in CD or UC. Previously published studies that have directly compared single-cell transcriptomic profiles of CD and UC have been of much smaller scale (often comparing 6 or fewer patients in one or both groups), lack contextualization with non-inflamed contemporaneously-enrolled controls, and lack a fully treatment-naïve group^{31,98–100}, which would limit the disease-specific inferences that could be drawn.

This work is most complementary to a recent study of transcriptomic differences in biologic-naïve (but pre-treated) adults with CD and UC, which utilized 3' scRNA-Seq to describe 109 cell states with similar phenotypes associated with either CD (including expansion of T_H1 cells and robust

interferon stimulation signatures), or UC (including IgG plasma cells, *CXCL13*⁺ T_{FH}/T_{PH} cells, and T_H17 cells)³. The similarity of these overarching features suggests the pathogenic mechanisms governing the generation of these cell states are preserved between pediatric and adult populations. Our study has additionally enabled an in-depth analysis of T cell plasticity (providing evidence for ex-T_H17 cells comprising the UC T_H1 cell state, as well as clonal proliferation of CD8⁺ T_H17 cells in UC), TCR phenotypes (demonstrating increased polyreactivity in UC T_{FH} cells), and T_{FH} vs. T_H17 dynamics (identifying an inverse correlation between T_{FH} features and T_H17 features in UC), highlighting the central role T cells play in delineating IBD subtypes and treatment response paradigms.

PREDICT demonstrated a profound reduction of T_H1 cells and T_H1//*IFNG*-associated cell states upon anti-TNF treatment, consistent with the hypothesis that T cell abrogation is a primary mechanism of anti-TNF efficacy^{101–103}. *In vitro* attempts to validate this hypothesis have further suggested that anti-TNF's ability to prevent TNFR2 pro-survival signaling, likely from myeloid cells, is a primary mechanism of drug efficacy^{104,105}. Our results are consistent with a model whereby anti-TNF limits T_H1 differentiation, proliferation, or survival, reducing pathologic IFN γ in the inflamed tissue and contributing to a shift away from inflammatory T cell, NK cell, and macrophage populations. Whereas anti-TNF successfully abrogates T_H1-mediated inflammation, our results suggest that it is less efficacious at correcting T_H17-associated pathology¹⁰⁵. In PREDICT, *IL17A* gene expression was associated with lack of response to anti-TNF, either prior to treatment (in UC), or due to a shift in T cell phenotype following treatment (in cCD). Previous studies have identified a transition to increased T_H17 features, including increased *IL17A* expression, in treatment-refractory adult CD^{105,106}.

Given these results, it is, perhaps, counterintuitive that selected case studies have reported new onset IBD following anti-IL-17A treatment of other rheumatological diseases, and that the only IL-17A cytokine blockade trial in IBD was halted due to adverse events in CD patients receiving the

anti-IL17A treatment^{107–112}. Moreover, due to IL-17's other roles in host defense and tissue remodeling, it is likely that this cytokine contributes to both persistence of inflammation and resolution of injury^{36,113}. Rather than targeting IL-17 directly, our data suggest that preventing T_H17 differentiation, for example via JAK inhibition or IL-12/IL-23 blockade, may be more effective in treatment-naïve IBD patients who exhibit greater T_H17 features, inclusive of *IL17A* expression. This is further reinforced by the emerging use of JAK and IL-12/IL-23 inhibition in IBD treatment, including in those who have failed anti-TNF therapy^{72,114–118}. Additionally, neutrophils have also been previously linked to anti-TNF nonresponse¹². In PREDICT, IHC data demonstrated that neutrophil density was significantly increased in UC. Preventing neutrophil recruitment or activation, for example through IL-1 inhibition, may also be a therapeutic strategy for these patients¹².

Our analysis provides a framework for combining gene expression and TCR analysis to identify distinct, cell state-influenced TCR phenotypes in autoimmune diseases. Although UC CD4 T cells overall demonstrated an increased disposition towards polyreactive TCR features (described primarily through mTiRP⁷⁸), this signature was particularly enriched in UC PD1Hi T_{FH} cells. This could represent increased reactivity to self-antigens, or potentially to microbial antigens, such as those of the commensal microbiome^{78,79}. Additionally, the association of these TCR phenotypes with TLS-like B cells and IgG₁ plasma cell frequencies may suggest that lymphoid aggregate formation in UC is due to UC-specific antigenic targets in the tissue. Previous studies have suggested that antibodies in UC primarily target commensal antigens, although some targeting of pathogenic bacteria has been observed as well^{27,119,120}.

Our evaluation of polyreactive TCR features in T_{FH} highlighted the opposing T_H17 and T_{FH} biology within UC, which was consistent with increased germinal center activity in some UC patients. To further evaluate this, we performed a proteome-wide autoantibody screen, to determine if the increased T cell polyreactivity in UC was mirrored by increased antibody-mediated autoreactivity.

Our results demonstrated a complex autoantibody landscape in all PRECICT patients, with substantial inter-patient variability being the most prominent feature of the data, and with examples of patients with high autoantibodies in FGID, CD and UC. These data are consistent with multiple factors likely influencing the largely patient-specific autoantibody profile that was measured in PREDICT, including HLA genotype, variations in dietary and microbial exposures, and robustness of humoral immune response. The influence of HLA genotype on the autoantibody profile was highlighted by an evaluation of two UC patients who both expressed the HLA-DRB1*15:01 allele, which has been associated with T cell autoreactivity and cross-reactivity in multiple autoimmune diseases^{86,121–124}. These two UC patients were both characterized by T_{FH} features rather than T_H17 features, and both developed autoantibodies to several antigens not present in any other PREDICT patient, including to ENO1 (a known autoantigen in autoimmune diseases including Hashimoto's encephalopathy, asthma, juvenile idiopathic arthritis, and Behçet's disease)^{87–91}. These patients highlight the (often unmeasured) influence that HLA genetics can have on the propensity for autoantibody production, underscoring the complex integration of multiple factors in IBD patients that ultimately produce each patient's unique clinical presentation.

Intriguingly, an experimental murine model of periodontal disease, which is dominated by T_H17 immunopathology, demonstrated that T_H17-to-T_{FH} plasticity was critical for generating IgG, restricting growth of oral commensals, and decreasing pathologic tissue damage¹²⁵. Given the similarities to the findings in the PREDICT study, we hypothesize that polyreactive TCR phenotypes may predispose T cells to T_H17-to-T_{FH} plasticity that could limit T_H17-mediated colonic inflammation and prevent nonresponse to anti-TNF therapy. While further work is required to address the relationship between T_H17 and T_{FH} cell biology, it is interesting to note that, in contrast to recent adult UC cohorts which revealed a dominant T_{PH} cell signature together with plasma cell infiltration in UC, pediatric UC harbors a co-existence of T_H17 and T_{FH} cells of related clonotypes.

Thus, this careful study of pediatric participants may be helping to reveal an earlier “transition state” in this disease prior to the formation of more organized lymphoid aggregates.

In summary, our results provide novel insights into the overlapping and distinct biology of pediatric cCD and UC, identifying three prominent axes of immunopathology (T_H1 , T_H17 , and T_{FH}) whose interactions and relative magnitude define these two IBD subtypes. They delineate which immune pathways portend suboptimal response to anti-TNF therapy and nominate data-driven approaches to IBD treatment.

ACKNOWLEDGEMENTS

The authors would like to thank Rachid Hamid for leading study and sample operations at Regeneron for this study, Jason Mighty for operational support for the genomic analyses, Baijun Kou for pathology support, and Mehrdad Rajaei for contributions to development of immunohistochemistry process for this study.

J.O.M is a New York Stem Cell Foundation - Robertson Investigator. J.O.M was supported by the AbbVie-Harvard Medical School Alliance, the AGA Research Foundation's AGA-Takeda Pharmaceuticals Research Scholar Award in IBD - AGA2020-13-01, the Leona M. and Harry B. Helmsley Charitable Trust, The Pew Charitable Trusts Biomedical Scholars, The Broad Next Generation Award, The Chan Zuckerberg Initiative Pediatric Networks, The Mathers Foundation, The Kenneth Rainin Foundation, The Massachusetts Consortium for Pathogen Readiness, The New York Stem Cell Foundation, NIH R01 HL162642, NIH R01 DE031928, NIH P30 DK034854, the NIH CIDER Network, Project NextGen, and The Cell Discovery Network, a collaborative funded by The Manton Foundation and The Warren Alpert Foundation at Boston Children's Hospital. L.S.K is supported by NIH P01 1P01HL158504, R01 5R01HL095791, U19 U19AI051731, and by the Helmsley Charitable Trust. A.K.S. was supported, in part, by the Searle Scholars Program, the Beckman Young Investigator Program, a Sloan Fellowship in Chemistry, and the NIH (5U24AI118672, 2R01HL095791). L.N. was supported by the National Institute of General Medical Sciences (T32GM007753 and T32GM144273). The research reported is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. V.N. was supported by the International mobility of research, technical and administrative staff of research organizations (CZ.02.2.69/0.0/0.0/18_053/0016981). S.B.S. is supported by NIH grants P30 DK034854 and RC2 DK122532, and the Helmsley Charitable Trust.

All figures created in *BioRender*.

DECLARATION OF INTERESTS

J.O.M. reports compensation for consulting services with Tessel Biosciences, Radera Biotherapeutics, and Passkey Therapeutics. L.S.K. is on the scientific advisory board for HiFiBio and Mammoth Biosciences. She reports research funding from Kymab Limited, Magenta Therapeutics, BlueBird Bio, and Regeneron Pharmaceuticals. She reports consulting fees from Equillium, FortySeven Inc, Novartis Inc, EMD Serono, Gilead Sciences, Vertex Pharmaceuticals, and Takeda Pharmaceuticals. She reports grants and personal fees from Bristol Myers Squibb that are managed under an agreement with Harvard Medical School. A.K.S. reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, JnJ, Repertoire Immune Medicines, Hovione, Third Rock Ventures, Ochre Bio, Danaher, Parabilis Medicines, Passkey Therapeutics, Quotient Therapeutics, Bio-Rad Laboratories, Senda Biosciences, Relation Therapeutics, Empress Therapeutics, IntrECate Biotherapeutics, and Dahlia Biosciences unrelated to this work. He has received research support from Merck, Novartis, Leo Pharma, Janssen, the Gates Foundation, the Moore Foundation, the Pew-Stewart Trust, Foundation MIT, the Chan Zuckerberg Initiative, Novo Nordisk, IBM, Wellcome Leap, Break Through Cancer, Coca Cola, UK Research and Innovation, St. Jude Children's Research Hospital, and the NIH unrelated to this work. S.B.S. declares the following: Scientific advisory board participation for Pfizer, Lilly, IFM therapeutics, Merck, Pandion, and Takeda Inc., and grant support from Pfizer, Novartis, Amgen, Takeda Consulting for Hoffman La Roche and Amgen. L.A. is a consultant for Takeda Pharmaceuticals. G.W. reports Research funding from Abbvie, Jansen, Takeda, Allakos; SAB for Abbvie, Bristol Myers Squibb; DSMB for Abbvie. D.L.S. is the co-founder, CM, president of NiMBAL Health. A.K.S., L.S.K., J.O.M., H.B.Z., and K.K. are co-inventors on a provisional patent application relating to methods of stratifying and treating IBD. J.C.D., J.D.H., N.F., W.K.L., C.A., A.L., M.N., Y.W., J.E.H., P.J.E., S.E.M., M.A.R.F., W.G., E.S., S.H., S.C., Y.Y., M.D., S.H., M.A.S., G.D.K., and A.T.H. are employees and shareholders of Regeneron Pharmaceuticals. S.M.C. is a former employee and shareholder of Regeneron Pharmaceuticals.

METHODS

Data availability

All raw and processed gene expression data will be made available upon publication of the peer-reviewed version of this manuscript.

Participants and sample collection

Study recruitment and clinical parameters

Pediatric patients were enrolled in the Precision Diagnostics in Inflammatory Bowel Disease, Cellular Therapies, and Transplantation (PREDICT) trial ([clinicaltrials.gov #NCT03369353](https://clinicaltrials.gov/ct2/show/study/NCT03369353)), as described previously⁹. Inclusion criteria required participants to be 6 years or older, weigh 10 kg or greater, and to be undergoing evaluation for new diagnosis of inflammatory bowel disease (IBD) or functional gastrointestinal disorder (FGID). Enrollment took place in accordance with Fred Hutchinson Cancer Center Institutional Review Board (Protocol #9730, ethical approval given) and Boston Children's Hospital Institutional Review Board (Protocol IRB-P00030890, ethical approval given) approved protocols, with written informed consent and assent when applicable. Enrollment for the trial began on May 1, 2017. For this analysis, only enrollees who underwent diagnostic endoscopy prior to February 1, 2022 were included. Diagnoses of FGID, Crohn's disease (CD), and ulcerative colitis (UC) were made by the patient's clinician based on endoscopic, histopathologic, and clinical features. Patients were designated as ileal-only Crohn's disease (iCD) if inflammation was present in the terminal ileum and not in any section of the colon. Designation of colon-involved Crohn's disease (cCD) required inflammation anywhere in the large intestine or in perianal area. Patients were excluded from the analysis if diagnosed with indeterminate IBD (2 cases), inflammation due to infectious etiology (1 case), or a non-FGID disorder (1 case revised to neurogenic bowel dysfunction). Clinical variables were recorded from enrollment through 2 years post-diagnostic endoscopy. Clinical variables recorded included age, sex, race, disease location and phenotype (using Montreal Criteria), clinical disease severity

(calculated using Pediatric Crohn's Disease Activity Index (PCDAI) or Pediatric Ulcerative Colitis Activity Index (PUCAI) for CD and UC, respectively), anti-TNF use, and anti-TNF response category.

For patients administered anti-TNF (5 iCD patients, 27 cCD patients, 8 UC patients), therapy was initiated within 90 days of diagnostic endoscopy. IBD patients who were not administered anti-TNF within this window were denoted as not on anti-TNF (NOA). Anti-TNF response status was determined at 2 years after diagnostic endoscopy. Full response (FR) was defined as clinical symptom control and biochemical response with a PCDAI score less than 12.5 or PUCAI score less than 10 on maintenance anti-TNF therapy. Partial response (PR) was defined as a lack of complete clinical symptom control and biochemical response with documented dose escalation of anti-TNF therapy and a possible change to a new therapeutic. Nonresponse (NR) was defined as complete lack of clinical and biochemical response to anti-TNF therapy.

Colon and blood collection

Colon and whole blood biopsies were obtained during the diagnostic endoscopy visit. Whole blood was drawn into EDTA anticoagulant and frozen at -80° C until required for genomic DNA experiments. Per protocol, up to 2 pinch biopsies were taken from non-necrotic areas of the colon, with preference for areas with marked erythema or edema. If no inflammation was observed, biopsies were obtained from left colon by default. Pinch biopsies were immediately placed into RPMI 1640 (ThermoFisher, 21870-076) on ice before subsequent processing.

Experimental methods

Pinch biopsy processing

Pinch biopsy processing was performed using a modified version of a previously published protocol^{9,126}. While intact, biopsy pinches were handled using a P1000 pipette applying gentle suction, and all centrifugation steps done in a temperature-controlled 4°C centrifuge. Biopsy

pinches were first rinsed in 25 mL PBS (ThermoFisher 14190-144) and allowed to settle. Each individual pinch was then transferred to 10 mL epithelial cell solution (ECS) (HBSS Ca/Mg-Free [ThermoFisher 14175-103], 100 U/mL penicillin [ThermoFisher 15140-122], 100 µg/mL streptomycin [ThermoFisher 15140-122], 10 mM HEPES [ThermoFisher 15630-080], and 2% FBS [ThermoFisher SH3007103HI]), freshly supplemented with 10 mM EDTA [ThermoFisher AM9261]. Separation of the epithelial layer (EPI) from the underlying lamina propria (LP) was performed for 15 minutes at 37°C with rotation at 700 RPM. The tube was then removed and placed on ice immediately for 10 minutes before shaking vigorously 20 times and undergoing 10 seconds on high vortex. Visual macroscopic inspection of the tube at this point yielded visible epithelial sheets.

The LP tissue fraction was carefully removed and placed into a large volume (45 mL) of ice-cold PBS to rinse before transferring to 10 mL of enzymatic digestion mix (Base: RPMI1640 [ThermoFisher 21870-076], 100 U/ml penicillin [ThermoFisher 15140-122], 100 µg/mL streptomycin [ThermoFisher 15140-122], 10 mM HEPES [ThermoFisher 15630-080], 2% FBS [ThermoFisher SH3007103HI], and 50 µg/mL gentamicin [ThermoFisher 15750-060]), freshly supplemented immediately prior with 100 µg/mL of Liberase TM [Roche 5401127001] and 100 µg/mL of DNase I [Sigma D5025-150KU]), at 37°C with rotation at 700 rpm for 30 minutes. LP enzymatic dissociation was quenched by addition 80 µL of 0.5M EDTA and placed on ice for five minutes. Samples were typically fully dissociated at this step and after gentle trituration with a P1000 pipette filtered through a 40 µm cell strainer (VWR 21008-949) into a new 50 mL conical tube and rinsed with ice-cold PBS to 35 mL total volume. This tube was spun down at 500g for 10 minutes and resuspended in 1.2 mL ECS. The tube was then spun down at 800g for 2 minutes and resuspended in 500 µL of ACK lysis buffer (Gibco A10492-01) and placed on ice for 3 minutes. 500 µL 0.2% FBS (ThermoFisher SH3007103HI) in DPBS (ThermoFisher 14-190-144) was added to post-lysis solution before centrifugation at 800g for 2 minutes and resuspension in 700 µL ice-

cold PBS with 0.4% BSA (ThermoFisher AM2616). LP cells were filtered through a 40 µm strainer then rinsed with an additional 700 µL ice-cold PBS with 0.4% BSA. LP cells were centrifuged at 500g for 5 minutes, then resuspended in 60 µL PBS + 0.4% BSA.

During LP processing, ECS was added to EPI-containing tube to 35 reach mL before centrifugation at 600g for 10 minutes and resuspension in 1 mL ECS. Cells were transferred to a 1.5mL Eppendorf tube pre-coated with 0.4% BSA in PBS in order to minimize time spent centrifuging and obtain a more concentrated cell pellet. Cells were spun down at 800g for 3 minutes, resuspended with 1 mL ECS, and centrifuged again at 800g for 3 minutes. Pellets were resuspended in ACK lysis buffer [ThermoFisher A1049201] for 2 minutes on ice to remove red blood cells, even if no RBC contamination was visibly observed, in order to maintain consistency across samples. 500 µL ECS without FBS was added to EPI cells before centrifugation at 800g for 5 minutes. Cells were washed (resuspended in 1 mL ECS without FBS, centrifuged at 800g for 3 minutes) and resuspended in 1 mL 37° C TrypLE express enzyme [ThermoFisher 12604-013]. After adequate mixing, another 400 µL TrypLE solution was added before additional mixing. Cells were incubated in TrypLE for 4 minutes in a 37°C bath followed by gentle trituration with a P1000 pipette. Cells were incubated for another 3 minutes in a 37°C bath followed by gentle trituration, and an additional 3 minute incubation at 37° C. Cells were centrifuged at 800g for 3 minutes then resuspended in 700 µL PBS + 0.4% BSA. Resuspended cells were filtered through 40 µm strainer and rinsed with 700 µL PBS + 0.4% BSA before being transferred to a new BSA-coated tube. Cells were centrifuged at 500g for 5 minutes and resuspended in 50 µL PBS + 0.4% BSA.

Cells from both EPI and LP fractions were counted and prepared as a single-cell suspension for scRNA-seq.

Single-cell RNA sequencing, read alignment, and ambient mRNA correction

Single cells were loaded onto 5' library chips as per the manufacturer's protocol for Chromium Single Cell 5' Library and Gel Bead Kit (10X Genomics). Biopsies were processed with 5' v1 kits November 2020 and prior (Chromium Single Cell 5' Library and Gel Bead Kit, PN-1000006; Chromium Single Cell V(D)J Enrichment Kit, Human T Cell, PN-1000005), or 5' v2 kits (Chromium Next GEM Single Cell 5' v2 Library and Gel Bead Kit, PN-1000263; Chromium Single Cell V(D)J Enrichment Kit, Human T Cell, PN-1000252) since November 2020. The EPI and LP fractions were processed in separate channels of the 10X Chromium Single Cell Platform. An input of 20,000 single cells was added to each channel. Briefly, single cells were portioned into Gel Beads in Emulsion (GEMs) in the Chromium controller with cell lysis and barcoded reverse transcription of RNA, followed by cDNA amplification, enzymatic fragmentation and 5' adaptor and sample index attachment.

For single cell libraries prepared using the 5' v2 assay from 10x Genomics, paired-end sequencing was performed on Illumina NovaSeq 6000 for RNA-seq libraries (Read 1 26-bp for UMI and cell barcode, Read 2 80-bp for transcript read, with 10-bp i7 and 10-bp i5 reads) and for V(D)J libraries (Read 1 150-bp, 10-bp i7, 10-bp i5, Read 2 150-bp). For single cell libraries prepared using the 5' v1 assay from 10x Genomics, paired-end sequencing was performed on Illumina NovaSeq 6000 for RNA-seq libraries (Read 1 26-bp for UMI and cell barcode, Read 2 80-bp for transcript read, with 8-bp i7 and 0-bp i5 reads) and for V(D)J libraries (Read 1 150-bp, 8-bp i7, 0-bp i5, Read 2 150-bp). Quality-filtered base calls were converted to demultiplexed FASTQ files. For all RNA-seq libraries, Cell Ranger Single Cell Software Suite (10X Genomics, 7.2.0) was used to perform transcriptome alignment, filtering, and UMI counting. The human GRCh38-1.2.0 genome assembly was used for the alignment. Intronic reads were not included in alignment. For all V(D)J libraries, Cell Ranger Single Cell Software Suite (10X Genomics, 7.2.0) was used to perform de novo assembly of read pairs into contigs followed by alignment and annotations of contigs against

the germline segment V(D)J reference sequences from the GRCh38-alt-ensembl-2.0.0 genome assembly.

For all aligned RNA-seq libraries, CellBender “remove-background” was used to reduce contaminating signal due to ambient mRNA^{127,128}. Raw Cell Ranger-aligned counts matrix were used as input. Default remove-background settings were used for all samples, with exception of learning rate (set to 3×10^{-5} for all samples, instead of default 10×10^{-5})

scRNA-seq data processing

Cell x gene matrix outputs from CellBender were aggregated into a single object using Seurat (version 5.0)¹²⁹. Samples were converted to AnnData format using zellkonverter¹³⁰. The AnnData object was processed with scanpy log-normalization, variable gene selection (flavor = seurat_v3, batch = 10x version, n_top_genes = 3000), PCA dimensional reduction, and coarse leiden clustering (n_neighbors = 20, n_pcs = 20, leiden resolution = 0.1)¹⁰. Coarse clusters were grouped into one of 6 groups: Colonic Epithelium (ColEpi); Stromal and Endothelial Cells (Stroma); Myeloid Cells (Myeloid); Plasma Cells (Plasma); B Cells (BCell); and T Cells, NK Cells, and Innate Lymphoid Cells (TNKILC). Clusters that were low-quality based on mitochondrial gene expression, number of features detected, number of UMIs detected, or a combination of these metrics, were not included in any of the above groups and discarded from further processing. Sample-specific clusters were also not included.

After partitioning into coarse cell types, each cell type was log-normalized and underwent 3 rounds of scVI model training, clustering, differential expression, annotation, and low-quality cell removal. For each round, an scVI model was trained and the scVI latent space was used for dimensional reduction, and between-cluster differential expression performed within the scVI framework¹³¹. Settings for scVI: n_layers = 2, n_latent = 30, n_hidden = 128, dropout_rate = 0.1, gene_likelihood = “nb”, max_epochs = 250 (with early stopping), batch_key = “version_10x”.

AnnData objects were reduced to top n variable genes prior to scVI model training. Number of variable gene selected differed by cellgroup (number of genes used for scVI input in final round of annotation and QC: ColEpi = 2000; Stroma = 5000; Myeloid = 5000; Plasma = 750; BCell = 3000; TNKILC = 5000). For Plasma and BCell types, additional gene filtering was performed to exclude variable immunoglobulin genes (pattern matching for “^IGH[HKL][VJC]”). This step was taken to reduce the ability of variable immune receptor genes to dominate model fitting and latent space generation in plasma cells/plasmablasts (constant genes were retained). Additional genes matching the following patterns were removed from scVI training as they appeared to dominate clustering without clear biological function: “^CH17”, “^RP11”, “^IGLL”, “^AC[0-9]{5}”. Conversely, genes matching the following pattern were explicitly included as input (if not already an highly variable gene) for plasma cell scVI model training, to ensure constant genes would be included: “IGH[ADEGM][1-4]*\$”

For cluster annotation, annotation was based on marker genes identified through literature review. Annotations were defined at additional levels within each cell type (level 1): cell subtype (level 2) and cell state (level 3). If appropriate and supported by literature, clusters were recombined into overarching cell types to obtain “level 2 annotations.” Conversely, if clusters were observed to be heterogeneous through differential expression analysis and visualization, subclustering was attempted, along with differential expression and annotation to obtain level 3 annotations (“cell state”). Annotations from final round of scVI processing were used for this study.

scTCR-seq data processing

After Cell Ranger alignment to the V(D)J reference, the filtered contigs files from each sample were combined into a single data frame. In some cases, more than one alpha chain was detected per barcode. In very rare cases, more than one beta chain was detected per barcode. To simplify data analysis, a single alpha or beta chain was selected for downstream processing. First, barcodes with only a CDR3 listed were excluded (these chains lacked values for CDRs 1 and 2

as well as framework regions 1-4). The selection preference order was: (1) which chain had the most UMIs, (2) which chain had the most reads, (3) which chain is listed first (this condition was rarely reached in the dataset). Cell annotations from the RNA-seq processing step were merged with TCR data through paired barcode sequences.

Immunofluorescence Microscopy

A fully automated multiplex immunohistochemistry assay was performed on the Ventana Discovery ULTRA platform (Ventana Medical Systems, Tucson, AZ) as previously described^{9,132}. Optimal concentrations of each antibody were determined, and they were applied in the following sequence and detected with the indicated fluorophore.

Pan-Immune Panel 1:

1. Rabbit anti-BCMA (clone E6D7B, Cell Signaling, SC#88183) was detected with DISCOVERY Rhodamine 6G (Roche, Part number: 7988168001).
2. Rabbit anti-ribonuclease 3/ECP (clone EPR20357, Abcam, ab207429) was detected with DISCOVERY DCC (Roche, Part number: 7988192001).
3. Rabbit anti-myeloperoxidase (clone EPR20257, Abcam, ab208670) was detected with DISCOVERY Cy5 (Roche, Part number: 7551215001).
4. Rabbit anti-CD11c (clone EP1347Y, Abcam, ab52623) was detected with DISCOVERY FAM (Roche, Part number: 7988150001).
5. Rabbit anti-NCR1 (clone EPR23097-35, Abcam, ab233558) was detected with Cy7-Goat anti-Rabbit-IgG (H+L) (AAT Bioquest, Cat 16872).
6. Mouse anti-pan keratin (clones AE1/AE3/PCK26, Roche, Part number 5266840001) was detected with DISCOVERY RED 610 (Roche, Part number: 7988176001).

Pan-Immune Panel 2:

1. Rabbit anti-CD20 (clone SP32, Abcam, ab64088) was detected with DISCOVERY Rhodamine 6G (Roche, Part number: 7988168001).
2. Rabbit anti-CD3 (clone SP162, Abcam, ab135372) was detected with DISCOVERY DCC (Roche, Part number: 7988192001).
3. Rabbit anti-CD8 (clone SP239, Abcam, ab178089) was detected with DISCOVERY RED 610 (Roche, Part number: 7988176001).
4. Rabbit anti-CD68 (clone SP251, Abcam, ab192847) was detected with DISCOVERY Cy5 (Roche, Part number: 7551215001).
5. Rabbit anti-FOXP3 (clone SP97, Abcam, ab99963) was detected with DISCOVERY FAM (Roche, Part number: 7988150001).
6. Mouse anti-pan keratin (clones AE1/AE3/PCK26, Roche, Part number 5266840001) was detected with Cy7-Goat anti-Mouse-IgG (H+L) (AAT Bioquest, Cat 16856).

Following staining, the tissue was counter-stained and cover slipped with Invitrogen ProLong Gold Antifade Mountant with NucBlue. Whole slide imaging was performed on the Zeiss Axioscan which was equipped with a Colibri light source.

Whole exome sequencing

For analyses of common variants, we used Illumina GSA array genotyping data and performed imputation with the TOPMed imputation server and reference panel¹³³. Exome sequencing was performed at the Regeneron Genetics Center using a custom automated sample preparation approach. Samples were captured with IDT xGen v1 probes and sequenced using Illumina HiSeq 2500-v4 or Illumina NovaSeq instruments, with 75-bp paired-end reads and two index reads. The GRCh38 human genome reference sequence and Ensembl version 100 gene definitions were used for variant calling using DeepVariant and for annotation¹³⁴.

HLA imputation

We imputed HLA alleles using the Michigan Imputation Server¹³³, HLA reference panel multiethnic-hla-panel-4digit-v2@1.0.0¹³⁵. Prior to imputation, in order to maximize the number of informative variants for HLA imputation, the PREDICT IBD sample was merged with n=1,000 random individuals from a Regeneron Genetics Center multi-ethnic cohort genotyped on the same array version. Variants in the MHC region chr6:28-34Mb and present in the HLA reference panel were extracted from TOPMED imputed genotype probabilities and hard-called with certainty threshold 0.9. The resulting dataset was QC-filtered for 12,101 variants with less than 0.1% missing genotype calls and HWE Pval > 1e-10. (All samples had less than 1% missing genotype calls.)

Proteome Microarray

The HuProt proteome microarray (CDI Labs, USA) was used to assay plasma for autoreactivity against the human proteome^{80–83}. HuProt arrays and plasma samples were diluted with CDI proprietary CDISampleBuffer (CDI Labs, USA) at 1:1000 and incubated with CDIArrayBlock (CDI Labs, USA) at room temperature with gentle shaking for 1 hour. After the blocking period, each sample was probed onto a HuProt microarray at room temperature for 1 hour with gentle shaking. The arrays were then washed with TBS-Tween (1x TBS / 0.1% Tween-20) 3 times for 10 minutes each. Following washes, the arrays were probed with fluorescent anti-human IgA or IgG secondary for 1 hour in the dark with gentle shaking. 3 washes were performed with TBS-Tween for 10 minutes each, followed by 3 rinses with ddH₂O. The arrays were dried with compressed CO₂. GenePix 4000B scanner was used to scan arrays. Scanned images were aligned to the array layout. Raw fluorescence intensity values were extracted using GenePix software for further data processing.

Quantification and statistical analysis

Gene set enrichment analysis

Gene set enrichment analysis was performed using fgsea¹³⁶. The DESeq2 “stat” value ($\text{Log}_2\text{FC}/\text{SE}_{\text{Log}_2\text{FC}}$) was used as gene-level statistical input for fgsea. 1000 permutations were used for fgsea calculation of enrichment scores. Reactome gene sets were used as input (version 2023.2, obtained through MSigDB C2 curated gene set)^{137,138}. Gene sets were included if they had at least 15 genes and no more than 100 genes.

Cell frequency analysis

Per-sample mean cell frequencies were calculated using custom code. Cells were grouped first by their sample or biopsy of origin (relevant grouping listed in text and/or figure captions; “sample” will be used throughout methods), then by the annotation of interest (e.g. CD4 Naïve T cells). The number of cells per sample and annotation was tallied. Unless otherwise stated, the number of cells was divided by a relative grouping, e.g. a cell subtype (e.g. Naïve T cells), a cell type (e.g. T cells), or the entire independent sample (e.g. “Relative to Sample”). For per-biopsy group comparisons, biopsies were only included for analysis if the relative grouping had at least 10 cells. This was an empirical measure taken to prevent statistical conclusions driven by low overall cell detection. Statistical testing was performed with ggpubr’s “stat_compare_means()” function¹³⁹.

Quantitative feature analysis

Per-sample quantitative feature (e.g. gene expression, module score, mTiRP, etc.) values were calculated using custom code. Data were aggregated based on sample of origin, then by cellular annotation(s) of interest. Mean values were calculated per combination of sample and cell annotation. For gene expression, log-normalized values were used as input for mean calculation. Samples were required to meet a minimum number of cells within sample and cell annotation combination to be included for analysis. 10 cells were required for calculation of gene-based metrics, and 50 cells were required for TCR-based metrics.

Milo differential abundance analysis

Only baseline, non-iCD samples were used as input for analysis with MiloR¹³. Milo analysis was performed on a per cell type basis (ColEpi, Stroma, Myeloid, Plasma, BCell, TNKILC). For each cell type, the data object was downsampled to the number of cells in the smallest disease category between FGID, cCD, and UC. This step was performed to reduce the effect of differing total number of cells skewing differential abundance results. Downsampling was performed on a per disease category basis, irrespective of sample of origin within cell types. After downsampling, any remaining biopsies with fewer than 5 cells were removed before proceeding with analysis. Following biopsy removal step, the Milo object was created and kNN graph generated with the cell group scVI latent space used as the reduced dimension (all 30 scVI dimensions used as input). Because there were many samples in this analysis (~70), we set $k = 200$ to ensure the average neighborhood would have cells numbering greater than $5 \times$ number of samples (reducing the likelihood of a single sample comprising a neighborhood), per package maintainer recommendations¹³. We verified that each cell group fulfilled this heuristic, then calculated neighborhood distances and built the neighborhood graph, using Milo functions with default inputs and latent space inputs as described above. Following graph construction, we tested for differential abundance with the scVI latent space used as the reduced dimension and a design formula specifying 10x version as a batch effect: “~version_10x + disease_group”. Alpha value was set to 0.1 for identifying differentially abundant neighborhoods. Neighborhoods with annotation fraction < 0.7 were excluded from visualization. 12 cell states did not meet the annotation fraction threshold in any neighborhood and were not visualized: Lymph. Endo., IgG4 PC, Prlf. IgG PC, Prlf. IgM PC, IgD PC, IgE PC, Plasmablast, Prlf. Th17.1, Prlf. Tfh, Prlf. Treg, Vg9d2 TC, and NKT.

Cell state correlation analysis

Cell state correlation analysis was performed using per-cell type frequencies of annotated cell states. Cell states were excluded if they did not have at least 100 cells (2 cell states excluded:

IgD PC and IgE PC). Only baseline biopsies were included for correlation analysis. The previously described cell frequency calculation method was used to generate per cell type frequencies for each cell state in each biopsy. Pearson correlation was used to compare biopsy frequencies of T_{H1} , T_{H17} , and PD1hi T_{FH} cells against all cell states excluding IgD PC and IgE PC. Test statistics were generated using “cor.test” function from the R stats package (version 4.3.1). Cell state correlations with a p value < 0.001 were selected as part of T_{H1} , T_{H17} , or PD1Hi T_{FH} network. Cell states in network were visualized with ggraph (version 2.2.1). Only edges with a correlation coefficient > 0.35 were visualized. Graph layout algorithm set to “kk”, maximum iterations = 150. Node location manually adjusted when necessary for visualization of all cell states.

To provide a quantitative measure of the usage of cell states, we calculated the geometric mean of cell state frequencies, as has been performed previously³⁰. Briefly, we took the logarithm value of the associated cell state frequencies in each biopsy, calculated the mean of all cell state frequencies in logarithm space, then exponentiated the average value:

$$\exp \left(\prod_{i=1}^k \log (x_i) \right)$$

Where x is the cell state proportion, i is the index of the cell state in the correlation network, and k is the number of total cell states in the correlation network. Because calculation of the geometric mean requires a logarithmic transformation, we added a “pseudoproportion” of 1/1000 to any cell state which was not identified in a biopsy.

Pseudobulk differential gene expression analysis

For differential expression (DE) analysis, a single-cell object containing aggregated, annotated cells from each cell type was filtered to include only the biopsies of interest for that comparison. Raw counts from each cell type within each biopsy were aggregated into pseudobulk format using the Seurat “AggregateExpression()” function¹⁴⁰. Gene selection was performed by filtering for

genes with ≥ 250 counts in at least as many samples as were in the smaller of the two groups being compared (e.g. if comparing 15 cCD FR and 10 cCD PR samples, genes were included if there were ≥ 250 counts in at least 10 samples overall in the cell type of interest). For the combined myeloid and TNKILC pre- vs. post-treatment analysis, a 500 count minimum was used. All input genes are included in supplemental DE tables. DESeq2 was run with default settings, using UMI data from the pseudobulk object¹⁴¹.

Gene set scoring

Decoupler python package (version 1.5.0) was used for all gene set scoring⁵⁰. For TNF scoring, the PROGENy database was used to select genes involved in TNF pathway activity⁵¹. Top 500 genes were obtained initially. The genes were then filtered to only include those with p-value $< 1 \times 10^{-10}$, to increase confidence that the genes included in the network play a role in TNF signaling. This resulted in 225 input genes. These genes were used as input for decoupler's multivariate linear model (mlm), with weights provided by PROGENy database entries^{50,51}. Decoupler was performed on individual cell types (e.g. Myeloid, TNKILC) and mlm estimate values were used for data analysis. For IL-17 pathway scoring, genes that were listed as cell extrinsic and downstream of IL-17RA/IL-17RC signaling in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway "IL-17 SIGNALING PATHWAY" were investigated for gene expression in this dataset⁴⁶. Genes with robustly detectable expression were retained for further analysis and grouped into KEGG-provided gene groups: Chemokines (*CXCL1*, *CXCL2*, *CXCL5*, *CXCL8*, *CXCL10*, *CCL2*, *CCL7*, *CCL20*), Cytokines (*IL6*, *TNF*, *PTGS2*, *CSF2*, *CSF3*), Anti-microbial (*DEFA5*, *DEFA6*, *DEFB1*, *DEFB4A*, *DEFB4B*, *DEFB124*, *MUC5B*, *S100A7*, *S100A8*, *S100A9*, *LCN2*), Tissue remodeling (*MMP1*, *MMP3*, *MMP9*, *MMP13*). An unweighted network of these genes was created with the gene group as the source and the gene as the target. This network was used as input with each cell type for decoupler's weighted sum algorithm⁵⁰. However, equal weight was given to all genes in the network. Normalized wsum values were used for data analysis. The same analytical steps

as IL-17 gene set scoring were used for T_H1 and T_H17 cytokine (*IFNG*, *TNF*, *IL17A*, *IL17F*, *IL21*, *IL22*, *IL26*) scoring.

Regression analysis

Comparison of multiple metrics through regression analysis was performed using a combination of custom code and ggpmisc (version 0.5.5)¹⁴². Cell frequency and feature quantification was performed as described in “Cell frequency analysis” and “Quantitative feature analysis.” The regression analysis was performed using “stat_poly_eq” (for statistics) and “stat_poly_line” (for visualization) functions. The method was set to “lm” for all regressions. A 95% confidence interval used for shading.

T Cell Clonal Analysis

A T cell was defined as clonally expanded if its paired CDR3 α and CDR3 β amino acid sequences were observed in at least one other cell in the same biopsy. The number of cells in a clonotype (paired CDR3 α +CDR3 β) per biopsy was quantified and reported as the level of clonal expansion: 1 (singlet), 2, 3, 4, or 5+. Cells that did not have a CDR3 α or CDR3 β were not considered for clonal expansion analysis.

For T_H-T_{FH} cell clonal overlap analysis, the dataset was reduced to T_H1, T_H17, PD1Hi T_{FH}, and PD1Lo T_{FH} cell states. Additional filtering was performed to ensure paired CDR3 α +CDR3 β was present in each cell. Cells were grouped by paired CDR3 per biopsy and the number of cell subtypes within each biopsy+CDR3 pair was determined. Clonotypes spanning more than one cell subtype were considered to have clonal overlap. Visualization of clonal overlap was performed with ggraph (version 2.2.1).

Clonal overlap between Prlf. T_H17.1 and T_H1 or T_H17 was performed in an identical manner, except that cell states were used to define overlap rather than cell subtypes.

Fluorescence microscopy quantification

Quantitative image analysis was performed using HALO Indica Labs Hyperplex module (IndicaLabs, Albuquerque, NM). For each sample, images from the pan-immune panels and the pan-CK marker were fused to generate a single image. A classifier was first applied to detect the tissue in each section and an automatic annotation was generated as “whole section.” A second classifier was applied using panCK and DAPI channels to define the epithelium layer (panCK+) and the lamina propria (panCK-). Automated annotations were generated as “epithelium” and as “lamina propria”. Tertiary lymphoid structures (TLS) were annotated by a pathologist. Their number, frequency, size, and cell composition were measured. Immune cells were detected in each area of interest (whole section, epithelium, and lamina propria for pan-immune panel 1; whole section, epithelium, lamina propria, and tertiary lymphoid structures for pan-immune panel 2). T cells were defined as CD3+, B cells as CD20+, myeloid cells as CD68+. For each T cell subset, CD8 T cells were defined as CD3+CD8+FOXP3-, CD4 as CD3+CD8-, Tregs as CD3+CD8-FOXP3+ and CD4 conventional (non-Treg CD4) as CD3+CD8-FOXP3-. Plasma cells were defined as BCMA+. Eosinophils were defined as ECP+. Dendritic cells were defined as CD11c+. Neutrophils were defined as MPO+. NK cells were defined as NCR1+. Numbers of positive cells for each immune subset were counted and their density measured.

Pseudotime Analysis

Diffusion pseudotime was calculated using scanpy¹⁰. For pseudotime analysis, a subset of the TNKILC cell group was created from CD4 T cell states (NaiveCD4, MemoryCD4, Int. T_H, T_H1, T_H17, Prlf. T_H17.1, PD1Lo T_{FH}, PD1Hi T_{FH}, Prlf. T_{FH}, T_{reg}, LEF1+ T_{reg}, LAG3+ T_{reg}, Prlf. T_{reg}, CD4 CTL). Nearest neighbors were recalculated on this subset, with n_neighbors = 20 and all 30 scVI latent dimensions used. The starting position for the pseudotime analysis was selected as the cell with the most positive value in the UMAP2 dimension, which corresponded to a cell in the NaiveCD4 cell state. Diffusion pseudotime was calculated using scanpy’s “tl.dpt” function, with

n_branchings = 0 and n_dcs = 10. Gene expression of cells ordered by pseudotime was visualized using scanpy's "pl.paga_path" function, with "normalize_to_zero_one = True".

CDR3 Metrics

CDR3 length was quantified as the number of amino acids in the CDR3, as determined by Cell Ranger alignment and TCR classification. Previous studies have highlighted that the middle residues of the CDR3 β , determined by N and P nucleotides between regions belonging to the V gene and J gene, are most likely to directly participate in contacts between TCRs and antigenic peptides⁷⁵. Thus, we focused our subsequent biophysical analysis on the middle region of the CDR3 β . The CDR3 β middle region was calculated only for TCRs with CDR3 β lengths between 12-17 amino acids, in accordance with definition for T cell regulatory potential (TiRP) score⁷⁸. The Cell Ranger definition of CDR3 includes junctional amino acids at IMGT positions 104 and 118¹⁴³. The CDR3 β middle region was defined as all amino acids from the 5th amino acid from the beginning to the 6th from the end, corresponding to P108 and P112 in IMGT numbering¹⁴³. mTiRP score was calculated using publicly available code from Lagattuta et al. (<https://github.com/immunogenomics/TiRP>), with Cell Ranger-determined V gene and CDR3 β used as input⁷⁸. CDR3 β middle-region biophysical properties, including hydrophobicity ("grand average of hydrophobicity", or "gravy"), size, aliphatic propensity, and polarity, were calculated with "alakazam" R package (version 1.3.0). CDR3 β middle-region amino acids were used as input to alakazam aminoAcidProperties() function with nucleotides set to FALSE, all other settings default. Normalized z-scores were generated through transformation to standard score within the dataset: $z = \frac{X - \mu}{\sigma}$ where X is the vector of all metric values, μ is the metric's mean within the dataset, and σ is the metric's standard deviation within the dataset. Given the inherent heterogeneity in TCR features, we limited sample-comparison analysis to biopsies with at least 50 CDR3 β sequences detected in the cell grouping of interest.

Autoantigen analysis

Data from HuProt proteome microarray was processed for autoantigen reactivity among PREDICT participants^{80–83}. Briefly, median fluorescence intensity (MFI) of each replicate spot pair on the protein microarray was averaged. Replicates with values differing by two-fold or greater were discarded from further analysis. $\log_2(\text{MFI})$ was calculated to stabilize variance across cohorts. Microarray spots with readings 4-fold above secondary-only median were considered false-positives and removed. Spots that were not detected were also removed. Z-scores were calculated within each sample array for all spots passing QC. Autoantigen detection (binary variable) was defined as spots with readings 4-fold above secondary-only median for the relevant spot, and with a Z-score $> +3$ within that sample array. Arithmetic mean of values was used for visual representation when an antigen was present more than once on the proteome microarray, and thus had multiple independent readings. An antigen was considered a hit if the antigen met Z-score-based hit criteria in 1 or more positions on the microarray. ComplexHeatmap (version 2.24.0) was used to generate antigen heatmaps and any accompanying dendrogram using default settings^{144,145}.

T_H1, T_H17, and T_{FH} feature summary

Features that were associated with T_H1 (T_H1 cell state frequency, geometric mean of T_H1-associated cell state frequencies, T_H *IFNG* expression, myeloid and TNKILC cell type *STAT1* expression, and myeloid and TNKILC TNF pathway score), T_H17 (T_H17 cell state frequency, geometric mean of T_H17-associated cell state frequencies, T_H *IL17A* expression, ColEpi antimicrobial gene set score, ColEpi chemokine gene set score), and T_{FH} (PD1Hi T_{FH} cell state frequency, geometric mean of PD1Hi T_{FH}-associated cell state frequencies, T_{FH} *CXCL13* expression, IgG₁ plasma cell state frequency, and T_{FH} mTiRP score) were used as inputs to generate per-biopsy summaries of overall T_H1, T_H17, and T_{FH} tendencies. Per-biopsy values for each feature were generated as described for the relevant feature. Each feature was scaled from

0 to 1, corresponding to the minimum and maximum values of that feature within the dataset. The arithmetic mean of each feature within each feature set (T_{H1} , T_{H17} , or T_{FH}) was then averaged per biopsy. The average value of each scaled feature was used to represent biopsies in the per-biopsy radar plot. For group-based radar plots created with ggradar (version 0.2), the arithmetic mean of the scaled feature set among all biopsies within the group was used to represent the group value on the radar plot.

REFERENCES

1. Kaplan, G. G. & Windsor, J. W. The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 56–66 (2021).
2. Roda, G. *et al.* Crohn's disease. *Nat. Rev. Dis. Primer* **6**, 22 (2020).
3. Thomas, T. *et al.* A longitudinal single-cell atlas of anti-tumour necrosis factor treatment in inflammatory bowel disease. *Nat. Immunol.* **25**, 2152–2165 (2024).
4. Higuchi, L. & Bousvaros, A. Clinical presentation and diagnosis of IBD in Children .pdf. *UpToDate* (2022).
5. Zitomersky, N. & Bousvaros, A. Overview of the management of Crohn disease in children and adolescents Zitomersky Bousvaros .pdf. *UpToDate* (2022).
6. Bousvaros, A. & Kaplan, J. Management of mild to moderate ulcerative colitis in children and adolescents Bousvaros Kaplan UpToDate.pdf. *UpToDate* (2023).
7. Black, C. J., Drossman, D. A., Talley, N. J., Ruddy, J. & Ford, A. C. Functional gastrointestinal disorders: advances in understanding and management. *The Lancet* **396**, 1664–1674 (2020).
8. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).
9. Zheng, H. B. *et al.* Concerted changes in the pediatric single-cell intestinal ecosystem before and after anti-TNF blockade. *eLife* **12**, (2023).
10. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
11. Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space and time. *Nature* **597**, 250–255 (2021).
12. Friedrich, M. *et al.* IL-1-driven stromal–neutrophil interactions define a subset of patients with inflammatory bowel disease that does not respond to therapies. *Nat. Med.* **27**, 1970–1981 (2021).

13. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).
14. Feakins, R. M. Ulcerative colitis or Crohn's disease? Pitfalls and problems. *Histopathology* **64**, 317–335 (2014).
15. Moreau, T. *et al.* Multifaceted roles of human elafin and secretory leukocyte proteinase inhibitor (SLPI), two serine protease inhibitors of the chelonianin family. *Biochimie* **90**, 284–295 (2008).
16. Williams, S. E., Brown, T. I., Roghanian, A. & Sallenave, J.-M. SLPI and elafin: one glove, many fingers. *Clin. Sci.* **110**, 21–35 (2005).
17. Chen, R. *et al.* Serum amyloid protein A in inflammatory bowel disease: from bench to bedside. *Cell Death Discov.* **9**, 154 (2023).
18. 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 (1994).
19. Sano, T. *et al.* An IL-23R/IL-22 Circuit Regulates Epithelial Serum Amyloid A to Promote Local Effector Th17 Responses. *Cell* **163**, 381–393 (2015).
20. Zheng, L., Duan, S.-L., Wen, X.-L. & Dai, Y.-C. Molecular regulation after mucosal injury and regeneration in ulcerative colitis. *Front. Mol. Biosci.* **9**, (2022).
21. Nakayama, T. *et al.* Type 1 helper T cells generate CXCL9/10-producing T-bet⁺ effector B cells potentially involved in the pathogenesis of rheumatoid arthritis. *Cell. Immunol.* **360**, 104263 (2021).
22. Petering, H. *et al.* Cutting Edge: Detection of MCP-4 in Dermal Fibroblasts and Its Activation of the Respiratory Burst in Human Eosinophils¹. *J. Immunol.* **160**, 555–558 (1998).
23. Iwamoto, T. *et al.* A role of monocyte chemoattractant protein-4 (MCP-4)/CCL13 from chondrocytes in rheumatoid arthritis. *FEBS J.* **274**, 4904–4912 (2007).

24. Yamazaki, T. & Mukouyama, Y. Tissue Specific Origin, Development, and Pathological Perspectives of Pericytes. *Front. Cardiovasc. Med.* **5**, 78 (2018).
25. Uo, M. *et al.* Mucosal CXCR4+ IgG plasma cells contribute to the pathogenesis of human ulcerative colitis through FcγR-mediated CD14 macrophage activation. *Gut* **62**, 1734–1744 (2013).
26. Uzzan, M. *et al.* Ulcerative colitis is characterized by a plasmablast-skewed humoral response associated with disease activity. *Nat. Med.* **28**, 766–779 (2022).
27. Scheid, J. F. *et al.* Remodeling of colon plasma cell repertoire within ulcerative colitis patients. *J. Exp. Med.* **220**, e20220538 (2023).
28. Haynes, N. M. *et al.* Role of CXCR5 and CCR7 in Follicular Th Cell Positioning and Appearance of a Programmed Cell Death Gene-1^{High} Germinal Center-Associated Subpopulation1. *J. Immunol.* **179**, 5099–5108 (2007).
29. Victora, G. D. & Nussenzweig, M. C. Germinal Centers. *Annu. Rev. Immunol.* **40**, 413–442 (2022).
30. Martin, J. C. *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493-1508.e20 (2019).
31. Huang, B. *et al.* Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways. *Cell* **179**, 1160-1176.e24 (2019).
32. Boland, B. S. *et al.* Heterogeneity and clonal relationships of adaptive immune cells in ulcerative colitis revealed by single-cell analyses. *Sci. Immunol.* **5**, eabb4432 (2020).
33. Shi, Y., Zhang, H. & Miao, C. Metabolic reprogram and T cell differentiation in inflammation: current evidence and future perspectives. *Cell Death Discov.* **11**, 1–12 (2025).
34. Misra, D. P. & Agarwal, V. Th17.1 lymphocytes: emerging players in the orchestra of immune-mediated inflammatory diseases. *Clin. Rheumatol.* **41**, 2297–2308 (2022).

35. Broos, C. E. *et al.* Increased T-helper 17.1 cells in sarcoidosis mediastinal lymph nodes. *Eur. Respir. J.* **51**, (2018).
36. Mills, K. H. G. IL-17 and IL-17-producing cells in protection versus pathology. *Nat. Rev. Immunol.* **23**, 38–54 (2023).
37. Kobayashi, T. *et al.* Ulcerative colitis. *Nat. Rev. Dis. Primer* **6**, 74 (2020).
38. Camoglio, L., Te Velde, A. A., Tigges, A. J., Das, P. K. & Van Deventer, S. J. Altered expression of interferon-gamma and interleukin-4 in inflammatory bowel disease. *Inflamm. Bowel Dis.* **4**, 285–290 (1998).
39. Lehtonen, A., Matikainen, S. & Julkunen, I. Interferons up-regulate STAT1, STAT2, and IRF family transcription factor gene expression in human peripheral blood mononuclear cells and macrophages. *J. Immunol.* **159**, 794–803 (1997).
40. Ivashkiv, L. B. IFN γ : signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy. *Nat. Rev. Immunol.* **18**, 545–558 (2018).
41. Zenobia, C. & Hajishengallis, G. Basic biology and role of interleukin-17 in immunity and inflammation. *Periodontol. 2000* **69**, 142–159 (2015).
42. Robert, M., Miossec, P. & Hot, A. The Th17 Pathway in Vascular Inflammation: Culprit or Consort? *Front. Immunol.* **13**, 888763 (2022).
43. Konieczny, P. *et al.* Interleukin-17 governs hypoxic adaptation of injured epithelium. *Science* **377**, eabg9302 (2022).
44. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
45. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
46. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).

47. Snell, K. Enzymes of serine metabolism in normal and neoplastic rat tissues. *Biochim. Biophys. Acta* **843**, 276–281 (1985).
48. Wang, L.-F., Chen, P.-R., He, S.-K., Duan, S.-H. & Zhang, Y. Predictors and optimal management of tumor necrosis factor antagonist nonresponse in inflammatory bowel disease: A literature review. *World J. Gastroenterol.* **29**, 4481–4498 (2023).
49. Grad, S. *et al.* Predictors of Immunogenicity and Loss of Response to ANTI-TNF α Therapy in Crohn Disease—A Systematic Review. *Am. J. Ther.* **32**, e262 (2025).
50. Badia-i-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinforma. Adv.* **2**, vbac016 (2022).
51. Schubert, M. *et al.* Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**, 20 (2018).
52. Mao, H. *et al.* Protective and anti-inflammatory role of REG1A in inflammatory bowel disease induced by JAK/STAT3 signaling axis. *Int. Immunopharmacol.* **92**, 107304 (2021).
53. Wuyts, A. *et al.* Characterization of Synthetic Human Granulocyte Chemotactic Protein 2: Usage of Chemokine Receptors CXCR1 and CXCR2 and in Vivo Inflammatory Properties. *Biochemistry* **36**, 2716–2723 (1997).
54. Moreira, A. C., Mesquita, G. & Gomes, M. S. Ferritin: An Inflammatory Player Keeping Iron at the Core of Pathogen-Host Interactions. *Microorganisms* **8**, 589 (2020).
55. Martín-García, P., Alonso-Arroyo, A. & Catalá-López, F. Tumour necrosis factor (TNF) antagonist therapy for paediatric inflammatory bowel disease: A systematic review. *Med. Clínica Engl. Ed.* **160**, 501–516 (2023).
56. Hyams, J. *et al.* Induction and Maintenance Therapy With Infliximab for Children With Moderate to Severe Ulcerative Colitis. *Clin. Gastroenterol. Hepatol.* **10**, 391-399.e1 (2012).
57. Croft, N. M. *et al.* Efficacy and safety of adalimumab in paediatric patients with moderate-to-severe ulcerative colitis (ENVISION I): a randomised, controlled, phase 3 study. *Lancet Gastroenterol. Hepatol.* **6**, 616–627 (2021).

58. Kang, I. & Bucala, R. The immunobiology of MIF: function, genetics and prospects for precision medicine. *Nat. Rev. Rheumatol.* **15**, 427–437 (2019).
59. Calandra, T. & Roger, T. Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat. Rev. Immunol.* **3**, 791–800 (2003).
60. Roger, T., Chanson, A.-L., Knaup-Reymond, M. & Calandra, T. Macrophage migration inhibitory factor promotes innate immune responses by suppressing glucocorticoid-induced expression of mitogen-activated protein kinase phosphatase-1. *Eur. J. Immunol.* **35**, 3405–3413 (2005).
61. Ferraboschi, P., Ciceri, S. & Grisenti, P. Applications of Lysozyme, an Innate Immune Defense Factor, as an Alternative Antibiotic. *Antibiotics* **10**, 1534 (2021).
62. Hanissian, S. H. & Geha, R. S. Jak3 Is Associated with CD40 and Is Critical for CD40 Induction of Gene Expression in B Cells. *Immunity* **6**, 379–387 (1997).
63. Tortolani, P. J. *et al.* Regulation of JAK3 expression and activation in human B cells and B cell malignancies. *J. Immunol. Baltim. Md 1950* **155**, 5220–5226 (1995).
64. Sandborn, W. J. *et al.* Tofacitinib as Induction and Maintenance Therapy for Ulcerative Colitis. *N. Engl. J. Med.* **376**, 1723–1736 (2017).
65. Winthrop, K. L. The emerging safety profile of JAK inhibitors in rheumatic disease. *Nat. Rev. Rheumatol.* **13**, 234–243 (2017).
66. Hall, J. A. *et al.* Transcription factor ROR α enforces stability of the Th17 cell effector program by binding to a Rorc cis-regulatory element. *Immunity* **55**, 2027–2043.e9 (2022).
67. Wang, C., Kang, S. G., Lee, J., Sun, Z. & Kim, C. H. The roles of CCR6 in migration of Th17 cells and regulation of effector T-cell balance in the gut. *Mucosal Immunol.* **2**, 173–183 (2009).
68. Hsieh, C.-S., Lee, H.-M. & Lio, C.-W. J. Selection of regulatory T cells in the thymus. *Nat. Rev. Immunol.* **12**, 157–167 (2012).

69. Guo, H. *et al.* PRDM1 Drives Human Primary T Cell Hyporesponsiveness by Altering the T Cell Transcriptome and Epigenome. *Front. Immunol.* **13**, (2022).
70. Moschen, A. R., Tilg, H. & Raine, T. IL-12, IL-23 and IL-17 in IBD: immunobiology and therapeutic targeting. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 185–196 (2019).
71. Hu, X., Li, J., Fu, M., Zhao, X. & Wang, W. The JAK/STAT signaling pathway: from bench to clinic. *Signal Transduct. Target. Ther.* **6**, 402 (2021).
72. Sands, B. E. *et al.* Ustekinumab as Induction and Maintenance Therapy for Ulcerative Colitis. *N. Engl. J. Med.* **381**, 1201–1214 (2019).
73. Wacleche, V. S., Wang, R. & Rao, D. A. Identification of T Peripheral Helper (Tph) Cells. in *T-Follicular Helper Cells: Methods and Protocols* (ed. Graca, L.) 59–76 (Springer US, New York, NY, 2022). doi:10.1007/978-1-0716-1736-6_6.
74. Mitsuyama, K. Antibody markers in the diagnosis of inflammatory bowel disease. *World J. Gastroenterol.* **22**, 1304 (2016).
75. Stadinski, B. D. *et al.* Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat. Immunol.* **17**, 946–955 (2016).
76. Košmrlj, A., Jha, A. K., Huseby, E. S., Kardar, M. & Chakraborty, A. K. How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl. Acad. Sci.* **105**, 16671–16676 (2008).
77. Daley, S. R. *et al.* Cysteine and hydrophobic residues in CDR3 serve as distinct T-cell self-reactivity indices. *J. Allergy Clin. Immunol.* **144**, 333–336 (2019).
78. Lagattuta, K. A. *et al.* Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nat. Immunol.* **23**, 446–457 (2022).
79. Zegarra-Ruiz, D. F. *et al.* Thymic development of gut-microbiota-specific T cells. *Nature* **594**, 413–417 (2021).
80. Jeong, J. S. *et al.* Rapid Identification of Monospecific Monoclonal Antibodies Using a Human Proteome Microarray*. *Mol. Cell. Proteomics* **11**, O111.016253 (2012).

81. Hu, S. *et al.* Profiling the Human Protein-DNA Interactome Reveals ERK2 as a Transcriptional Repressor of Interferon Signaling. *Cell* **139**, 610–622 (2009).
82. Venkataraman, A. *et al.* A toolbox of immunoprecipitation-grade monoclonal antibodies to human transcription factors. *Nat. Methods* **15**, 330–338 (2018).
83. Zhu, H. *et al.* Global Analysis of Protein Activities Using Proteome Chips. *Science* **293**, 2101–2105 (2001).
84. Ferré, E. M. N., Schmitt, M. M. & Lionakis, M. S. Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy. *Front. Pediatr.* **9**, (2021).
85. Cleynen, I. *et al.* Inherited determinants of Crohn’s disease and ulcerative colitis phenotypes: a genetic association study. *The Lancet* **387**, 156–167 (2016).
86. Baranzini, S. E. & Oksenberg, J. R. The Genetics of Multiple Sclerosis: From 0 to 200 in 50 Years. *Trends Genet.* **33**, 960–970 (2017).
87. Lee, K. H. *et al.* Human alpha-enolase from endothelial cells as a target antigen of anti-endothelial cell antibody in Behçet’s disease. *Arthritis Rheum.* **48**, 2025–2035 (2003).
88. Yazici, Y. *et al.* Behçet syndrome. *Nat. Rev. Dis. Primer* **7**, 1–14 (2021).
89. Yoneda, M. *et al.* High prevalence of serum autoantibodies against the amino terminal of α -enolase in Hashimoto’s encephalopathy. *J. Neuroimmunol.* **185**, 195–200 (2007).
90. Moore, T. L., Gillian, B. E., Crespo-Pagnussat, S., Feller, L. & Chauhan, A. K. Measurement and evaluation of isotypes of anti-citrullinated fibrinogen and anti-citrullinated alpha-enolase antibodies in juvenile idiopathic arthritis. *Clin. Exp. Rheumatol.* **32**, 740–746 (2014).
91. Nahm, D.-H. *et al.* Identification of α -enolase as an autoantigen associated with severe asthma. *J. Allergy Clin. Immunol.* **118**, 376–381 (2006).
92. Gao, L., Liu, T. & Li, X. Identification and analysis of the endoplasmic reticulum stress hub genes in sepsis-associated ARDS. *Sci. Rep.* **15**, 31007 (2025).

93. Xuan, Y. & Wang, Y. Long non-coding RNA SNHG3 promotes progression of gastric cancer by regulating neighboring MED18 gene methylation. *Cell Death Dis.* **10**, 694 (2019).
94. Jaeger, N. *et al.* Single-cell analyses of Crohn's disease tissues reveal intestinal intraepithelial T cells heterogeneity and altered subset distributions. *Nat. Commun.* **12**, 1921 (2021).
95. Kinchen, J. *et al.* Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. *Cell* **175**, 372-386.e17 (2018).
96. Corridoni, D. *et al.* Single-cell atlas of colonic CD8+ T cells in ulcerative colitis. *Nat. Med.* **26**, 1480–1490 (2020).
97. Kong, L. *et al.* The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon. *Immunity* **56**, 444-458.e5 (2023).
98. Mitsialis, V. *et al.* Single-Cell Analyses of Colon and Blood Reveal Distinct Immune Cell Signatures of Ulcerative Colitis and Crohn's Disease. *Gastroenterology* **159**, 591-608.e10 (2020).
99. Huang, L. *et al.* Multiomics analyses reveal a critical role of selenium in controlling T cell differentiation in Crohn's disease. *Immunity* **54**, 1728-1744.e7 (2021).
100. Yokoi, T. *et al.* Identification of a unique subset of tissue-resident memory CD4 + T cells in Crohn's disease. *Proc. Natl. Acad. Sci.* **120**, e2204269120 (2023).
101. Ten Hove, T. Infliximab treatment induces apoptosis of lamina propria T lymphocytes in Crohn's disease. *Gut* **50**, 206–211 (2002).
102. Van Den Brande, J. M. H. *et al.* Prediction of antitumour necrosis factor clinical efficacy by real-time visualisation of apoptosis in patients with Crohn's disease. *Gut* **56**, 509–517 (2007).
103. Levin, A. D., Wildenberg, M. E. & Van Den Brink, G. R. Mechanism of Action of Anti-TNF Therapy in Inflammatory Bowel Disease. *J. Crohns Colitis* **10**, 989–997 (2016).

104. Atreya, R. *et al.* Antibodies Against Tumor Necrosis Factor (TNF) Induce T-Cell Apoptosis in Patients With Inflammatory Bowel Diseases via TNF Receptor 2 and Intestinal CD14+ Macrophages. *Gastroenterology* **141**, 2026–2038 (2011).
105. Schmitt, H. *et al.* Expansion of IL-23 receptor bearing TNFR2+ T cells is associated with molecular resistance to anti-TNF therapy in Crohn's disease. *Gut* **68**, 814–828 (2019).
106. Zorzi, F. *et al.* Distinct Profiles of Effector Cytokines Mark the Different Phases of Crohn's Disease. *PLoS ONE* **8**, e54562 (2013).
107. Deng, Z., Wang, S., Wu, C. & Wang, C. IL-17 inhibitor-associated inflammatory bowel disease: A study based on literature and database analysis. *Front. Pharmacol.* **14**, (2023).
108. Fobelo Lozano, M. J., Serrano Giménez, R. & Castro Fernández, M. Emergence of Inflammatory Bowel Disease During Treatment With Secukinumab. *J. Crohns Colitis* **12**, 1131–1133 (2018).
109. Darch, K. M., Holland, T. L. & Spelman, L. J. Secukinumab-Induced Inflammatory Bowel Disease in a Patient Treated for Chronic Plaque Psoriasis and Psoriatic Arthritis: A Case Report and Review of the Role of Novel Biologic Agents Targeting the p19 Subunit of IL-23. *Case Rep. Med.* **2020**, 9404505 (2020).
110. Onac, I. A. *et al.* Secukinumab as a potential trigger of inflammatory bowel disease in ankylosing spondylitis or psoriatic arthritis patients. *Rheumatol. Oxf. Engl.* **60**, 5233–5238 (2021).
111. Schreiber, S. *et al.* Incidence rates of inflammatory bowel disease in patients with psoriasis, psoriatic arthritis and ankylosing spondylitis treated with secukinumab: a retrospective analysis of pooled data from 21 clinical trials. *Ann. Rheum. Dis.* **78**, 473–479 (2019).
112. Hueber, W. *et al.* Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut* **61**, 1693–1700 (2012).

113. Lee, J. S. *et al.* Interleukin-23-Independent IL-17 Production Regulates Intestinal Epithelial Permeability. *Immunity* **43**, 727–738 (2015).
114. Feagan, B. G. *et al.* Induction therapy with the selective interleukin-23 inhibitor risankizumab in patients with moderate-to-severe Crohn's disease: a randomised, double-blind, placebo-controlled phase 2 study. *The Lancet* **389**, 1699–1709 (2017).
115. D'Haens, G. *et al.* Risankizumab as induction therapy for Crohn's disease: results from the phase 3 ADVANCE and MOTIVATE induction trials. *The Lancet* **399**, 2015–2030 (2022).
116. Visvanathan, S. *et al.* Selective IL-23 Inhibition by Risankizumab Modulates the Molecular Profile in the Colon and Ileum of Patients With Active Crohn's Disease: Results From a Randomised Phase II Biopsy Sub-study. *J. Crohns Colitis* **12**, 1170–1179 (2018).
117. Dayan, J. R. *et al.* Real World Experience With Ustekinumab in Children and Young Adults at a Tertiary Care Pediatric Inflammatory Bowel Disease Center. *J. Pediatr. Gastroenterol. Nutr.* **69**, 61–67 (2019).
118. Feagan, B. G. *et al.* Ustekinumab as Induction and Maintenance Therapy for Crohn's Disease. *N. Engl. J. Med.* **375**, 1946–1960 (2016).
119. Castro-Dopico, T. *et al.* Anti-commensal IgG Drives Intestinal Inflammation and Type 17 Immunity in Ulcerative Colitis. *Immunity* **50**, 1099-1114.e10 (2019).
120. Gronke, K. *et al.* Human Th17- and IgG3-Associated Autoimmunity Induced by a Translocating Gut Pathobiont. <http://biorxiv.org/lookup/doi/10.1101/2023.06.29.546430> (2023) doi:10.1101/2023.06.29.546430.
121. Wang, J. *et al.* HLA-DR15 Molecules Jointly Shape an Autoreactive T Cell Repertoire in Multiple Sclerosis. *Cell* **183**, 1264-1281.e20 (2020).
122. Hendrikse, J. *et al.* Paediatric autoimmune uveitis is associated with intraocular antibodies against Epstein–Barr virus Nuclear Antigen 1 (EBNA-1). *eBioMedicine* **115**, (2025).
123. Oka, S. *et al.* Association of HLA-DRB1 genotype with younger age onset and elder age onset rheumatoid arthritis in Japanese populations. *Medicine (Baltimore)* **98**, e18218 (2019).

124. Renaudineau, Y. *et al.* Across ancestries, HLA-B*08:01~DRB1*03:01 (DR3) and HLA-DQA*01:02 (DR2) increase the risk to develop juvenile-onset systemic lupus erythematosus through low complement C4 levels. *J. Transl. Autoimmun.* **10**, 100268 (2025).
125. McClure, F. A. *et al.* Th17-to-Tfh plasticity during periodontitis limits disease pathology. *J. Exp. Med.* **221**, e20232015 (2024).
126. Persson, E. K. *et al.* IRF4 Transcription-Factor-Dependent CD103+CD11b+ Dendritic Cells Drive Mucosal T Helper 17 Cell Differentiation. *Immunity* **38**, 958–969 (2013).
127. CellBender removes technical artifacts from single-cell RNA sequencing data. *Nat. Methods* **20**, 1285–1286 (2023).
128. Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).
129. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
130. Zappia, L. & Lun, A. theislab/zellkonverter. Theis Lab (2024).
131. Gayoso, A. *et al.* A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 (2022).
132. Marron, T. U. *et al.* Neoadjuvant cemiplimab for resectable hepatocellular carcinoma: a single-arm, open-label, phase 2 trial. *Lancet Gastroenterol. Hepatol.* **7**, 219–229 (2022).
133. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
134. Poplin, R. *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018).
135. Luo, Y. *et al.* A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nat. Genet.* **53**, 1504–1516 (2021).

136. Korotkevich, G. *et al.* *Fast Gene Set Enrichment Analysis*.
<http://biorxiv.org/lookup/doi/10.1101/060012> (2016) doi:10.1101/060012.
137. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
138. Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res.* **52**, D672–D678 (2024).
139. Kassambara, A. ggpubr: ‘ggplot2’ Based Publication Ready Plots. (2023).
140. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
141. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
142. Aphalo, P. J. *Learn R ...as You Learnt Your Mother Tongue*. (Leanpub, 2017).
143. Lefranc, M. P. Unique database numbering system for immunogenetic analysis. *Immunol. Today* **18**, 509 (1997).
144. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
145. Gu, Z. Complex heatmap visualization. *iMeta* **1**, e43 (2022).

LIST OF SUPPLEMENTAL TABLES

Supplemental Table 1 - Participant and Sample Details

Supplemental Table 2 - UMAP Coordinates

Supplemental Table 3 - Cell State Differential Expression and Markers

Supplemental Table 4 - Milo Differential Abundance

Supplemental Table 5 - IHC Density and scRNA-seq Cell State Frequency

Supplemental Table 6 - Pseudobulk Differential Expression and Gene Set Enrichment

Supplemental Table 7 - Log-Normalized Gene Expression

Supplemental Table 8 - TCR Sequences and Phenotypic Analysis

Supplemental Table 9 - Cell State Frequency Correlations

Supplemental Table 10 - Linear Regressions

Supplemental Table 11 - Gene Set Scoring

Supplemental Table 12 - HLA Alleles

Supplemental Table 13 - HuProt Proteome Microarray

Supplemental Table 14 - Summary of Key Data

MAIN FIGURES

Fig. 1 PREDICT Study Overview, Cell States, and Immunohistochemistry Panels	66
Fig. 2 Transcriptomic and histochemical profiling highlights cellular shifts in cCD and UC	68
Fig. 3 Crohn's disease is characterized by T _H 1 and T _H 17-associated cellular networks; ulcerative colitis is characterized by T _H 17- and PD1Hi T _{FH} -associated cellular networks	70
Fig. 4 Anti-TNF decreases T _H 1-associated cell states, but T _H 17-associated cytokine expression is persistent in post-treatment cCD partial responders	72
Fig. 5 <i>IL17A</i> expression is enriched in UC anti-TNF nonresponders prior to treatment.....	74
Fig. 6 T _H 17 and PD1Hi T _{FH} cells belong to shared clonotypes yet are inversely correlated in UC	75
Fig. 7 UC PD1Hi T _{FH} cells are enriched for polyreactive TCR features that are inversely correlated with <i>IL17A</i> gene expression	76

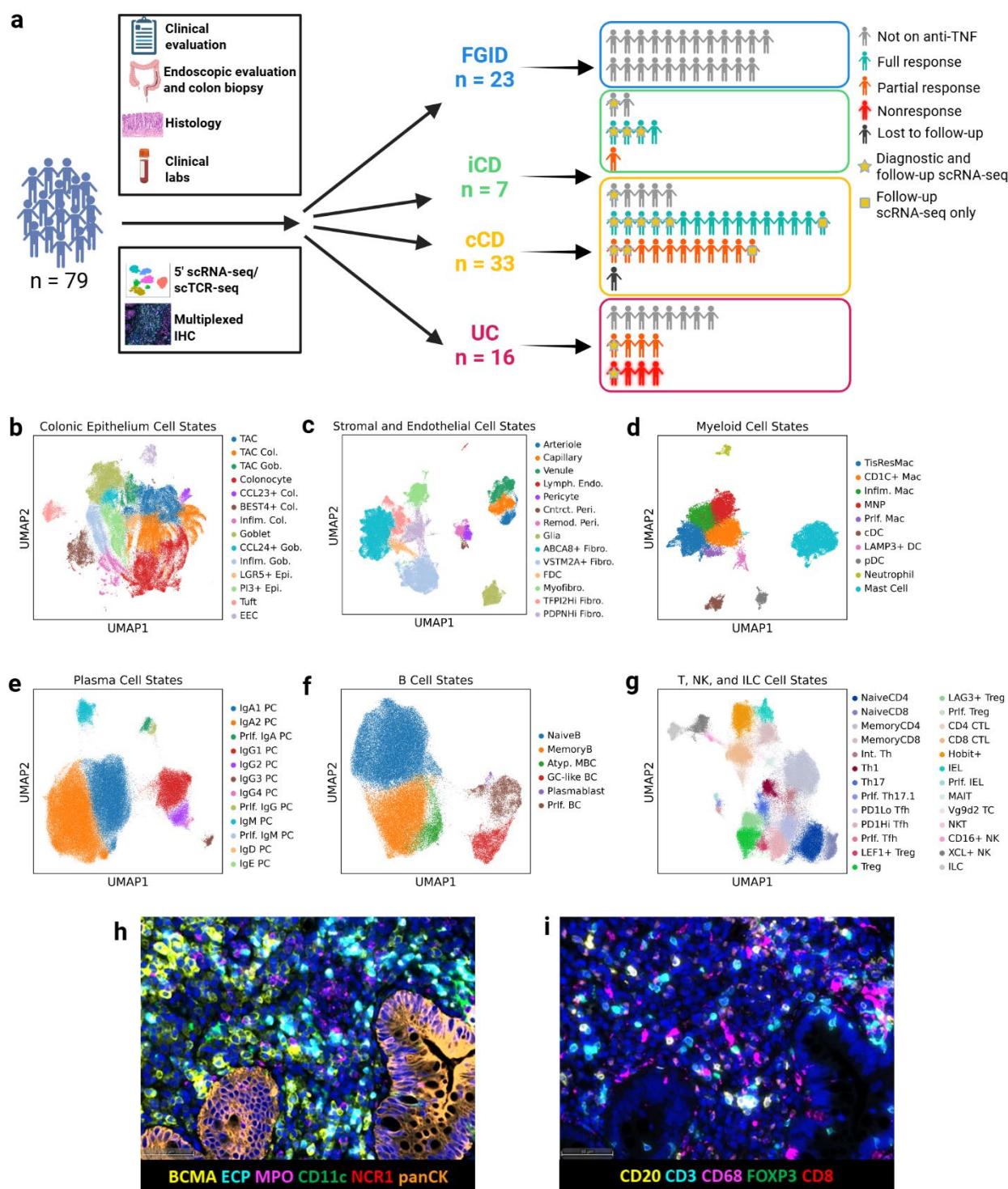


Fig. 1 | PREDICT Study Overview, Cell States, and Immunohistochemistry Panels

a, Outline of PREDICT data collection, specimen collection, and treatment response. During each endoscopy, colon biopsy and clinical metadata are obtained. Additional clinical metadata are obtained throughout the study's 2 year follow-up period. Additional colon biopsies obtained if clinically indicated. Primary assays performed with colon biopsies include 5' scRNA-seq and scTCR-seq of dissociated epithelium and lamina propria and *in situ* fluorescence microscopy. Visual representation of study participants with diagnostic colon biopsy successfully processed for scRNA-seq per disease category (FGID, n = 23; iCD, n = 7; cCD, n = 33; UC, n = 16; overall, n = 79). Participants with successfully processed diagnostic and follow-up biopsies are indicated by yellow star (FGID, n = 0; iCD, n = 4; cCD, n = 8; UC, n = 2). Participants with only follow-up biopsy successfully processed (i.e. without successfully processed diagnostic biopsy) indicated by yellow square (FGID, n = 0; iCD, n = 0; cCD, n = 2; UC, n = 0). Participant groups are

colored by disease category. Blue, FGID. Green, iCD. Yellow, cCD. Red, UC. Participant images are colored by anti-TNF response category. Dark gray, lost to follow-up. Light gray, NOA. Cyan, FR. Orange, PR. Red with glow, NR. **b, c, d, e, f, g**, UMAPs of cell state annotations for each cell type: **(b)** colonic epithelium, **(c)** stromal and endothelial cells, **(d)** myeloid cells, **(e)** plasma cells, **(f)** B cells, and **(g)** T cells, natural killer cells, and innate lymphoid cells (T, NK, and ILCs). **h, i**, Representative micrographs of immunohistochemical panels of colon biopsies from PREDICT study participants (images shown are from colon of participant 001-047). Two immunophenotyping panels were used. Immunophenotyping panel 1 contains markers for plasma cells (BCMA), eosinophils (ECP), neutrophils (MPO), dendritic cells and monocytes (CD11c), and NK cells (NCR1). Immunophenotyping panel 2 contains markers for B cells (CD20), T cells (CD3), myeloid cells (CD68), Treg cells (FOXP3), and CD8 T cells (CD8). panCK was used as marker for epithelial cells. Scale bar denotes 50 μ m. FGID, functional gastrointestinal disorder. iCD, ileal-only Crohn's disease. cCD, colon-involving Crohn's disease. UC, ulcerative colitis. TNF, tumor necrosis factor. NOA, not on anti-TNF. FR, full response. PR, partial response. NR, nonresponse. scRNA-seq, single-cell RNA sequencing. scTCR-seq, single-cell TCR sequencing. GEX, gene expression. TCR, T cell receptor. IHC, immunohistochemistry. UMAP, uniform manifold approximation and projection. CD, cluster of differentiation. BCMA, B cell maturation antigen. ECP, eosinophil cationic protein. MPO, myeloperoxidase. NCR1, natural cytotoxicity receptor 1. DAPI, 4',6-diamidino-2-phenylindole. panCK, pan-cytokeratin. FOXP3, forkhead box P3.

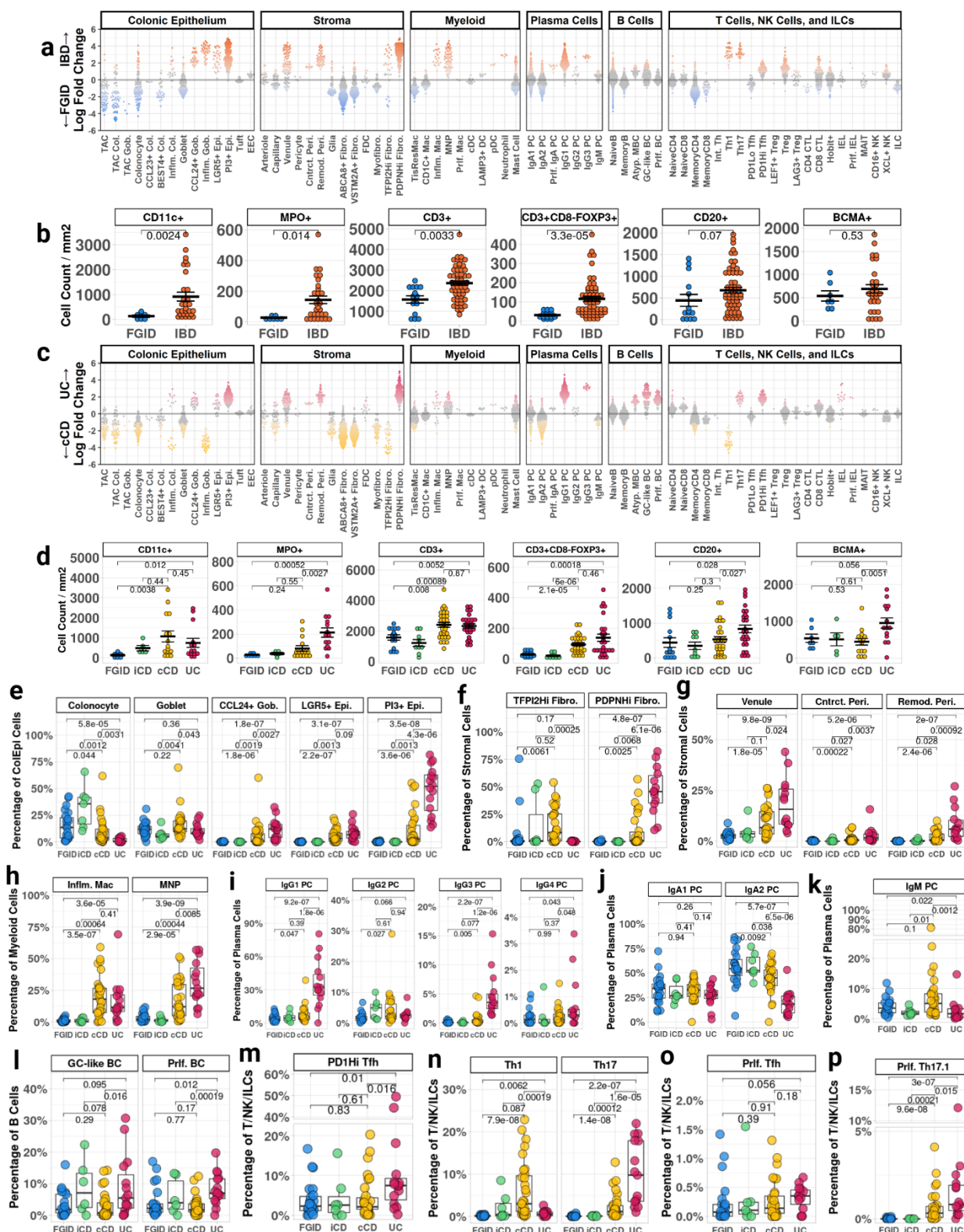


Fig. 2 | Transcriptomic and histochemical profiling highlights cellular shifts in cCD and UC

a, Milo-based differential abundance analysis performed within the six annotated cell types among all biopsies from diagnostic endoscopies, excluding biopsies with < 5 cells within the cell type after downsampling to equal cell numbers from FGID, cCD, and UC (FGID, n = 22; cCD, n = 31; UC, n = 16 for colonic epithelium. FGID, n = 23; cCD, n = 30; UC, n = 14 for stromal and endothelial cells. FGID, n = 23; cCD, n = 30; UC, n = 15 for myeloid cells and plasma cells. FGID, n = 21; cCD, n = 31; UC, n = 15 for B cells. FGID, n

= 23; cCD, n = 31; UC, n = 15 for T cells, NK cells, and innate lymphoid cells). Each dot represents a grouping of cells with similar transcriptomic profiles ("neighborhood"). Identified neighborhoods were assigned to cell states based on the cell state with $\geq 70\%$ representation within the neighborhood. Neighborhoods with no cell state occupying $\geq 70\%$ of neighborhood were removed from analysis. Cell states not shown did not meet criteria in any neighborhood. Comparisons made between FGID and IBD (cCD and UC). Individual Milo-derived neighborhoods are colored by disease if their spatial FDR < 0.1. Blue = FGID. Orange = IBD. **b**, Cell densities of select markers in immunophenotyping panels, comparing FGID to IBD (cCD and UC). Panel with markers BCMA, ECP, MPO, CD11c, and NCR1 was stained only once per study participant (FGID, n = 7; cCD, n = 15; UC, n = 14). Panel with markers CD20, CD3, CD68, FOXP3, and CD8 was stained 1-3 times per study participant (FGID, n = 8 participants, n = 14 samples; cCD, n = 15 participants, n = 30 samples; UC, n = 14 participants, n = 27 samples). **c**, Milo-based differential abundance analysis performed as in **(a)**, comparing cCD (yellow) and UC (red). **d**, Cell densities of select markers in immunophenotyping panels, performed as in **(c)**, comparing FGID, iCD, cCD, and UC. **e-p**, scRNA-seq cell state frequency (relative to cell type) comparisons for non-inflamed and inflamed epithelial cell states (**e**), TFPI2Hi and PDPNHi fibroblasts (**f**), venules, contractile pericytes, and remodeling pericytes (**g**), inflammatory macrophages and mononuclear phagocytes (**h**), IgG plasma cells (**i**), IgA plasma cells (**j**), IgM plasma cells (**k**), germinal center-like and proliferating B cells (**l**), PD1Hi T_{FH} cells (**m**), T_H1 and T_H17 cells (**n**), proliferating T_{FH} cells (**o**), and proliferating T_H17.1 cells (**p**). Only biopsies with at least 10 cells in the cell type were included for **e-p** cell state frequency analyses (FGID, n = 23; iCD, n = 7; cCD, n = 31; UC, n = 16 for colonic epithelium. FGID, n = 23; iCD, n = 7; cCD, n = 30; UC, n = 14 for stromal and endothelial cells. FGID, n = 23; iCD, n = 7; cCD, n = 30; UC, n = 15 for myeloid cells and plasma cells. FGID, n = 21; iCD, n = 6; cCD, n = 31; UC, n = 15 for B cells. FGID, n = 23; iCD, n = 7; cCD, n = 31; UC, n = 15 for T cells, NK cells, and innate lymphoid cells). Mann-Whitney U test, two-tailed, used for all pairwise comparisons. Error bars denote standard error of the mean.

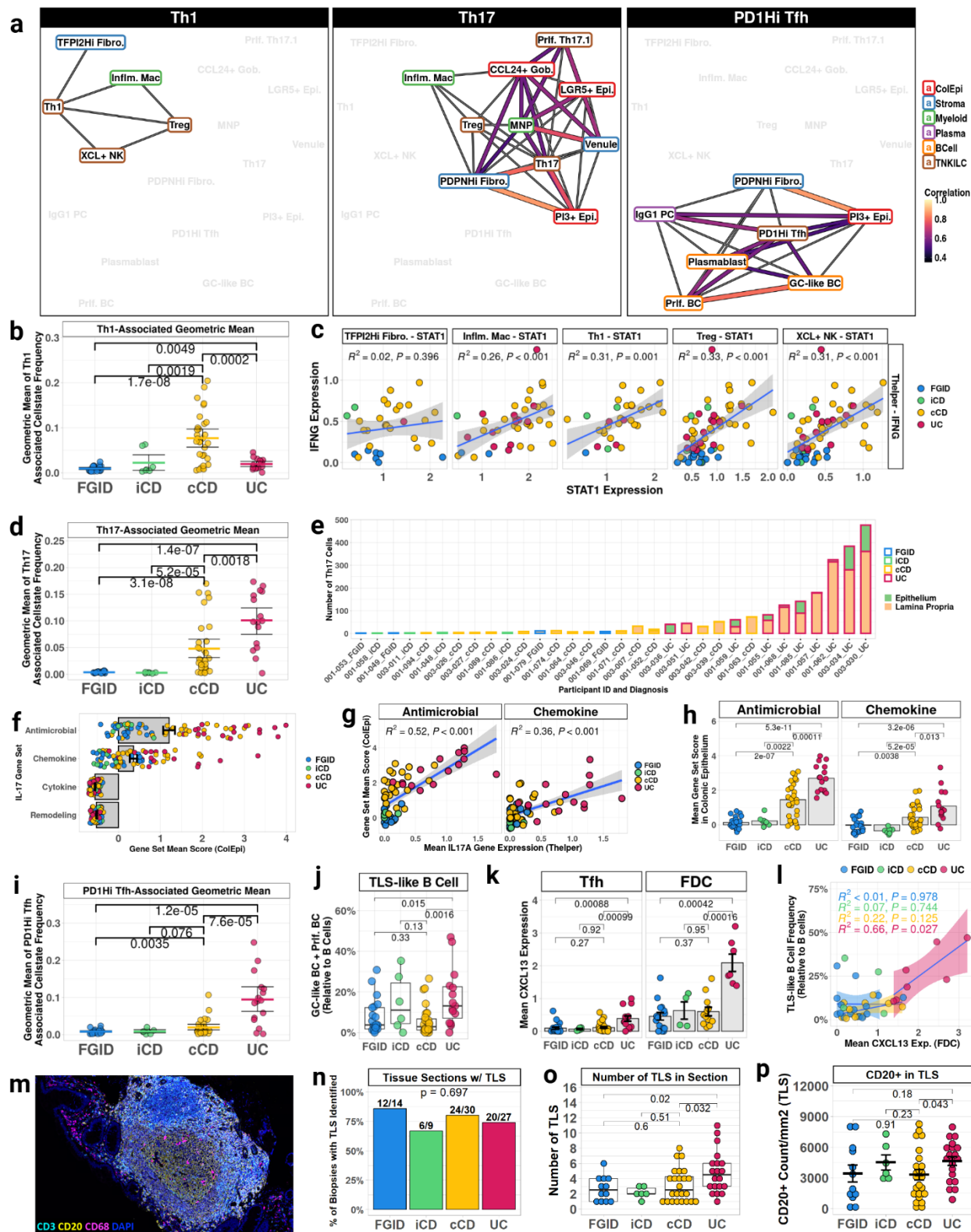


Fig. 3 | Crohn's disease is characterized by TH1 and TH17-associated cellular networks; ulcerative colitis is characterized by TH17- and PD1Hi THF-associated cellular networks

a, Cell state frequency Pearson correlation between TH1, TH17, and PD1Hi THF cell state and all other cell states with ≥ 100 cells (IgG and IgE plasma cells excluded). Cell state frequencies calculated as number of cells in the cell state relative to number of cells in cell type, per diagnostic biopsy (FGID, n = 23. iCD, n = 7. cCD, n = 31. UC, n = 16). Cell types labeled by color (colonic epithelium = red;

stroma = blue; myeloid cells = green; plasma cells = purple; B cells = orange, T cells, NK cells, and innate lymphoid cells = brown). Pearson correlation and statistical testing performed on per-biopsy cell state frequencies. Cell state frequencies associated with T_H1, T_H17, and PD1Hi T_{FF} frequency at $p < 1e-3$ are visualized as force-directed network plot. Edges are shown if Pearson correlation ≥ 0.35 . Edges gray if $0.35 < \text{Pearson correlation} < 0.5$, edges colored if Pearson correlation ≥ 0.5 . Width and color of edges correspond to strength of Pearson correlation ≥ 0.5 . **b**, Geometric mean analysis of cell state frequencies associated with T_H1 cell state, compared among disease categories. Cell state frequencies and biopsies used as input for correlation analysis also used as input for geometric mean analysis. **c**, Relationships between T_H cell subtype expression of *IFNG* and T_H1-associated cell states, modeled as simple linear regressions. **d**, Geometric mean analysis of cell state frequencies associated with T_H17 cell state, performed as in **(b)**. **e**, Number of T_H17 cells in EPI (green) and LP (beige) biopsy fractions, per biopsy with successfully processed LP and EPI fractions, ordered by total number of T_H17 cells. Disease group is denoted by bar outline color. Study participant ID and disease group are labeled on x-axis. **f**, IL-17 gene set scores in diagnostic biopsies with at least 10 colonic epithelial cells ($n = 77$). **g**, Relationship between per-biopsy mean expression of *IL17A* in T helper cells and IL-17 antimicrobial and chemokine gene set scores in colonic epithelial cells, modeled as a simple linear regression among diagnostic biopsies with at least 10 T helper cells and 10 colonic epithelial cells ($n = 66$). **h**, Mean antimicrobial and chemokine gene set score in diagnostic biopsies with at least 10 colonic epithelial cells (FGID, $n = 23$; iCD, $n = 7$; cCD, $n = 31$; UC, $n = 16$), compared across disease groups. **i**, Geometric mean analysis of cell state frequencies associated with PD1Hi T_{FF} cell state, performed as in **(b)**. **j**, Frequency of GC-like B cells and proliferating B cells combined into single group ("TLS-like B Cells"), relative to all B cells, per diagnostic biopsy with at least 10 B cells (FGID, $n = 21$; iCD, $n = 6$; cCD, $n = 31$; UC, $n = 15$). **k**, Mean T_{FF}- and FDC-derived *CXCL13* log-normalized expression among diagnostic biopsies with ≥ 10 T_{FF} cells (FGID, $n = 20$; iCD, $n = 5$; cCD, $n = 29$; UC, $n = 14$) or FDC cells (FGID, $n = 15$; iCD, $n = 4$; cCD, $n = 12$; UC, $n = 7$). **l**, Relationship between FDC expression of *CXCL13* and frequency of TLS-like B cells, modeled as a simple linear regression within each diagnosis category. Biopsies included for analysis contained ≥ 10 FDCs and ≥ 10 B Cells (FGID, $n = 15$; iCD, $n = 4$; cCD, $n = 12$; UC, $n = 7$). iCD confidence interval not displayed for clarity. **m**, Representative image from colon of IBD participant with lymphoid aggregate, stained for CD3 (cyan), CD20 (yellow), CD68 (magenta), DAPI (dark blue). **n**, Percentage of tissue sections with any TLS identified. Chi-squared test used for statistical analysis. **o**, Number of TLS annotated per tissue section with any TLS identified, comparing FGID ($n = 12$, 7 participants), iCD ($n = 6$, 5 participants), cCD ($n = 24$, 15 participants), UC ($n = 20$, 11 participants). **p**, Density of CD20+ cells in TLS per tissue section with any TLS identified. All error bars denote standard error of the mean. Mann-Whitney U test, two-tailed, used for all pairwise comparisons. Shaded areas of regressions show 95% confidence interval. TLS, tertiary lymphoid structure.

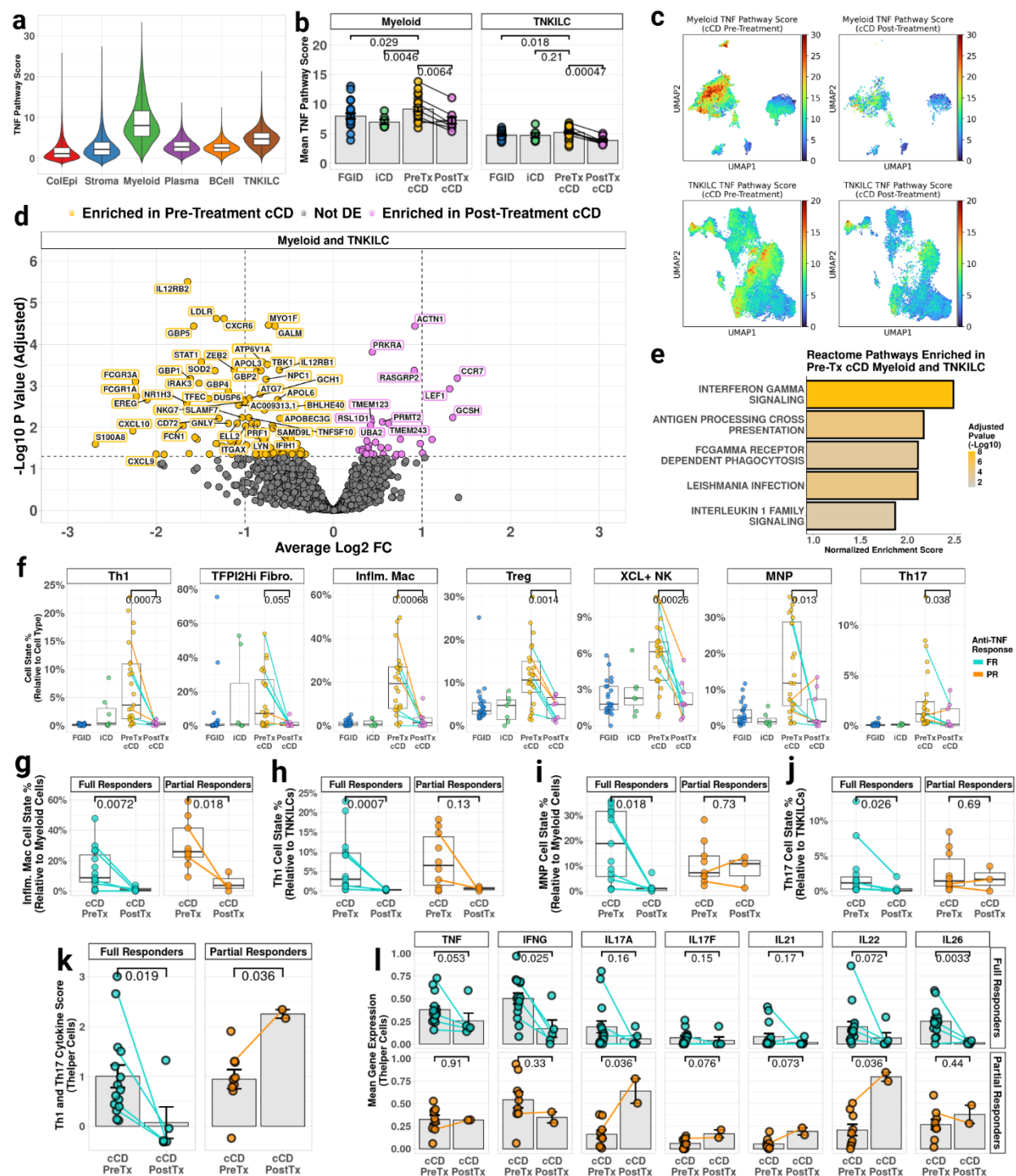


Fig. 4 | Anti-TNF decreases TH1-associated cell states, but TH17-associated cytokine expression is persistent in post-treatment cCD partial responders

a, Violin plot of TNF pathway scores in all cell groups. Cells were scored using decoupler multivariate linear model. PROGENY “TNFa” pathway genes and weights were used as input, filtered to genes with p-value < 1e-10 (225 genes). **b**, Pre- and post-treatment TNF scores in cCD, excluding cCD participants who did not receive anti-TNF. FGID and pre-treatment iCD included as reference. Biopsies used had ≥ 10 cells in each cell type (Myeloid: FGID, n = 22; iCD, n = 7; Pre-Tx cCD, n = 24; Post-Tx cCD, n = 9). TNKILC: FGID, n = 23; iCD, n = 7; Pre-Tx cCD, n = 25; Post-Tx cCD, n = 9). **c**, UMAP of TNF pathway score for myeloid (top) and TNKILC (bottom) cell groups among pre-treatment (left) and post-treatment (right) cCD biopsies. **d**, Differential expression analysis among myeloid cells and TNKILCs, comparing pre-treatment and post-treatment cCD (Pre-tx, n = 25; Post-tx, n = 9). Pseudobulk differential expression calculated with DESeq2. Genes labeled if adjusted p-value ≤ 0.01. **e**, Enriched terms from gene set enrichment analysis of differential expression between pre- and post-treatment cCD patients. Gene ranks computed using DESeq2 Wald test “stat” value

(log2FC/SE(log2FC)). Ranked genes were tested using Reactome 2023 v2 gene sets with $15 \leq \text{number of genes} \leq 100$ and sorted for normalized enrichment score. Terms enriched in pre-treatment cCD with the 5 lowest adjusted p-values are shown. **f**, Changes in cell frequencies among T_H1 -associated cell states, as well as MNP and T_H17 cell states (FGID, $n = 23$; iCD, $n = 7$; Pre-Tx cCD, $n = 24$ for stromal and myeloid cell states, $n = 25$ for TNKILC cell states; Post-cCD, $n = 9$). **g-j**, pre- and post-treatment cell frequency of inflammatory macrophages (**g**), T_H1 cells (**h**), mononuclear phagocytes (**i**), and T_H17 cells (**j**), relative to their respective cell type, among cCD full responder and partial responder biopsies with at least 10 cells in relevant cell type (cCD FR Pre, $n = 15$; cCD FR Post, $n = 6$; cCD PR Pre, $n = 9$ for Myeloid, $n = 10$ for TNKILC; cCD PR Post, $n = 3$). **k**, Mean module score of canonical T_H1 and T_H17 cytokine gene expression (module input genes: *IFNG*, *TNF*, *IL17A*, *IL17F*, *IL21*, *IL22*, *IL26*), grouped into timepoint and response categories for cCD biopsies with at least 10 T_H cells (cCD FR Pre-Tx, $n = 15$; cCD FR Post-Tx, $n = 5$; cCD PR Pre-Tx, $n = 9$; cCD PR Post-Tx, $n = 2$). **l**, Mean log-normalized expression of genes used as input for module score in (**k**) among T_H cell subtype, grouped into timepoint and response categories for cCD biopsies with at least 10 T_H cells. All error bars denote standard error of the mean. Mann-Whitney U test, two-tailed, used for pairwise comparisons. Lines connect pre- and post-treatment biopsies from the same study participant.

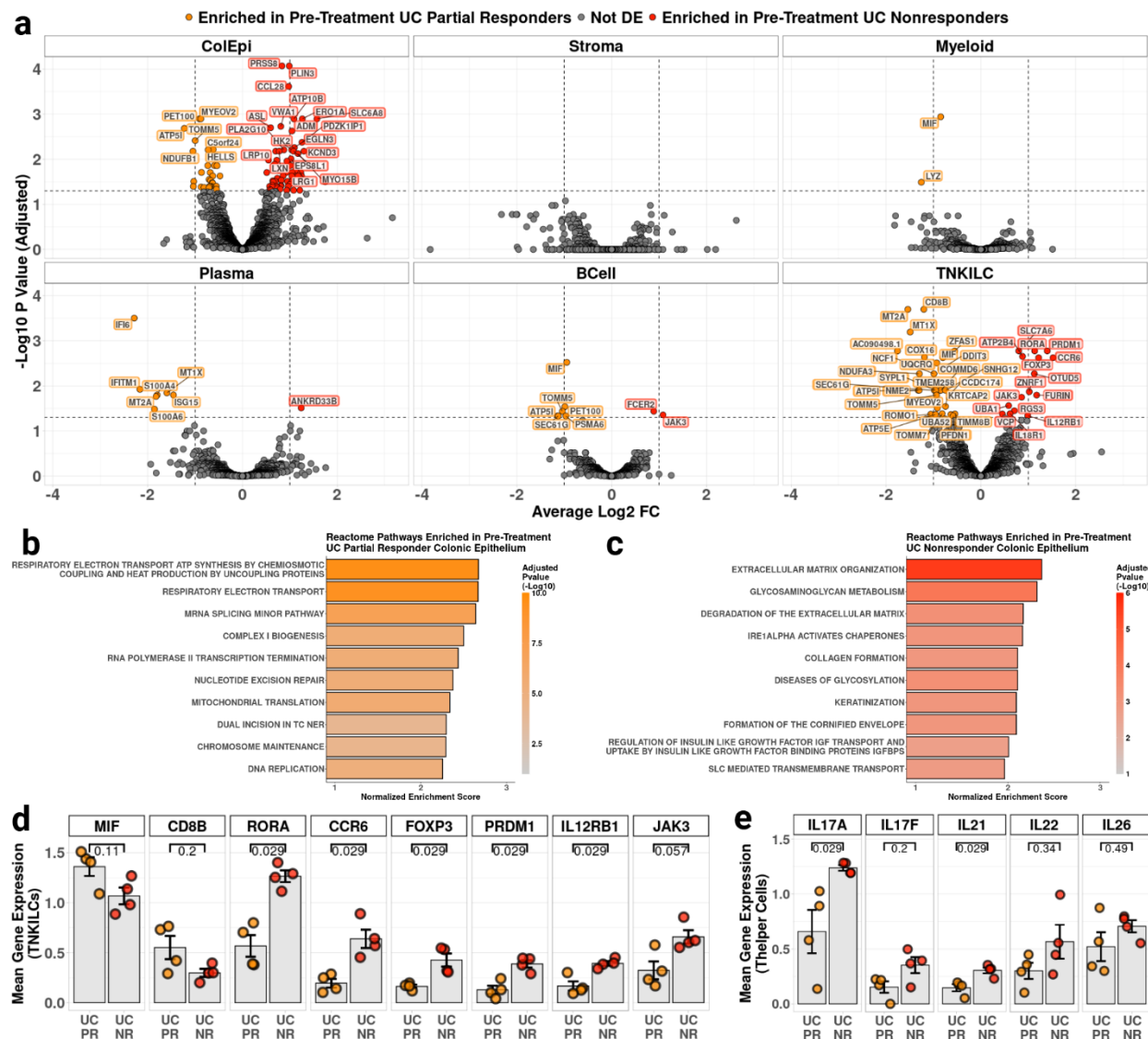


Fig. 5 | IL17A expression is enriched in UC anti-TNF nonresponders prior to treatment

a, Differential expression comparing differentially expressed genes in post-treatment UC full responders and partial responders within each annotated cell type. 4 PR and 4 NR post-treatment biopsies compared. Genes used as input determined through filtering by expression in a minimum number of samples. Genes shown if adjusted p value ≤ 0.05 , except for ColEpi (ColEpi DE genes shown if adjusted p value ≤ 0.01). **b**, **c**, Enriched terms from gene set enrichment analysis of differential expression between pre-treatment colonic epithelium from UC PR (**b**) and NR (**c**) participants. Top 10 terms for each treatment group shown. Gene ranks computed using DESeq2 Wald test "stat" value ($\log_2FC/SE(\log_2FC)$). Ranked genes tested using Reactome 2023 v2 gene sets with $15 \leq$ number of genes ≤ 100 and sorted for normalized enrichment score. **d**, Mean log-normalized expression of selected DE genes among TNKILCs, grouped into response categories for pre-treatment UC biopsies with at least 10 TNKILC cells (UC PR, $n = 4$; UC NR, $n = 4$). **e**, Mean log-normalized expression of selected T_H17 -associated cytokines among T_H cell subtype, grouped into response categories for pre-treatment UC biopsies with at least 10 T_H cells (UC PR, $n = 4$; UC NR, $n = 4$). All error bars denote standard error of the mean. Mann-Whitney U test, two-tailed, used for pairwise comparisons.

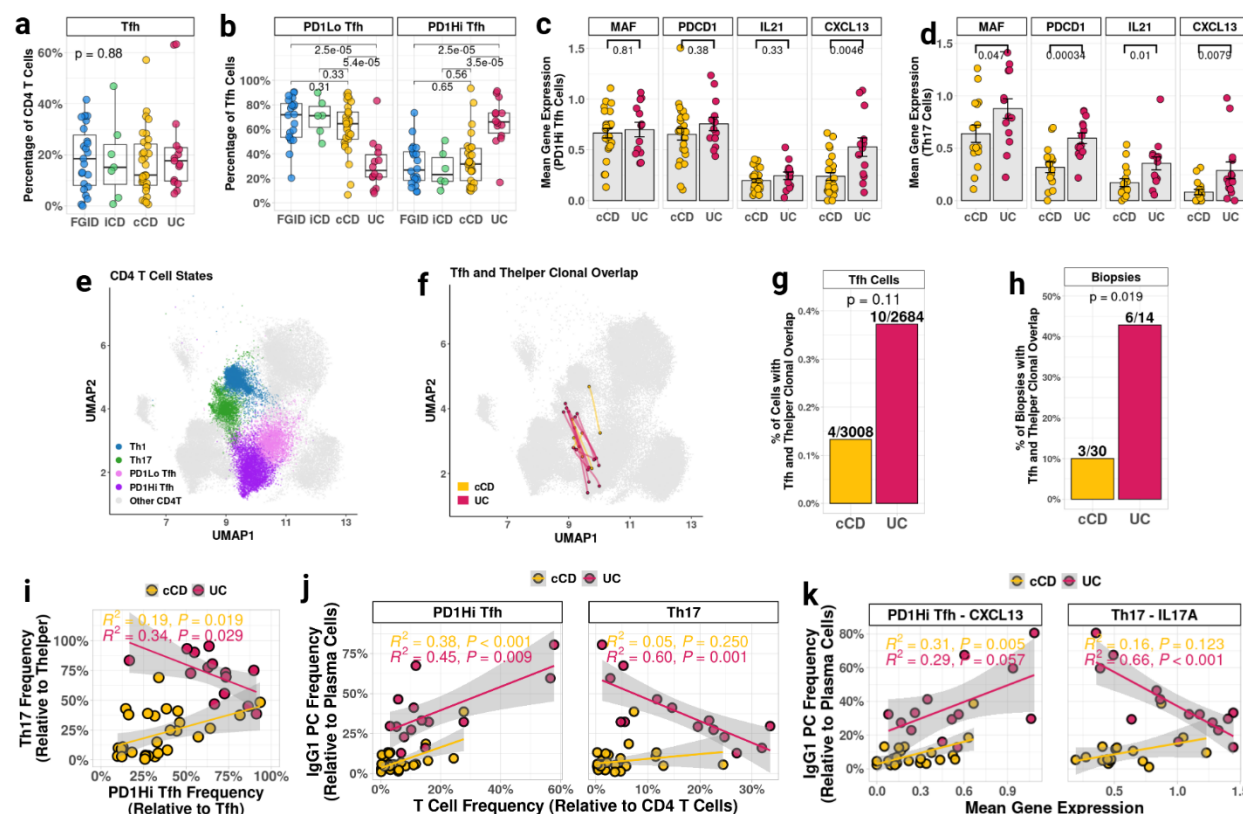


Fig. 6 | T_H17 and PD1Hi T_{FH} cells belong to shared clonotypes yet are inversely correlated in UC

a, Frequencies of T_{FH} cells relative to all CD4 T cells, per diagnostic biopsy with ≥ 10 CD4 T cells (FGID, $n = 23$; ICD, $n = 7$; cCD, $n = 31$; UC, $n = 15$). Kruskal-Wallis test used for statistical testing. **b**, Frequencies of T_{FH} cell states relative to all T_{FH} cells, per diagnostic biopsy with ≥ 10 T_{FH} cells (FGID, $n = 20$; ICD, $n = 6$; cCD, $n = 29$; UC, $n = 14$). **c**, **d**, Mean log-normalized per-biopsy gene expression of canonical T_{FH} B cell help genes in PD1Hi T_{FH} cells (**c**) or T_H17 cells (**d**), per cCD and UC diagnostic biopsy with ≥ 10 cells in cell state (PD1Hi T_{FH}: cCD, $n = 25$; UC, $n = 13$. T_H17: cCD, $n = 16$; UC, $n = 14$). **e**, UMAP of CD4 T cell states. T_H1, T_H17, PD1Lo T_{FH}, and PD1Hi T_{FH} cells highlighted. **f**, UMAP of shared clones between T_H and T_{FH} cells among T_H1, T_H17, PD1Lo T_{FH}, and PD1Hi T_{FH} cell states in cCD (yellow lines), or UC (red lines) diagnostic biopsies. Shared clones determined by paired CDR3 $\alpha\beta$ amino acid sequence. **g**, Percentage of cCD and UC cells in diagnostic biopsies with any identified clonal overlap between T_H and T_{FH} cells, as a fraction of total T_{FH} cells within disease group. Fisher's exact test was used to compare number of shared clones in cCD or UC cells. **h**, Percentage of cCD and UC diagnostic biopsies with any identified clonal overlap between T_H and T_{FH} cells, among biopsies with detected CDR3 $\alpha\beta$ in T_H1, T_H17, PD1Lo T_{FH}, and PD1Hi T_{FH} cell states. Fisher's exact test was used to compare presence or absence of any shared clones in cCD or UC biopsies. **i**, Linear regression between T_H17 frequency and PD1Hi T_{FH} frequency, relative to their respective cell subtypes (T_H and T_{FH}), within each disease group. Restricted to diagnostic cCD and UC biopsies with at least 10 cells in T_H and T_{FH} cell subtypes (cCD, $n = 28$; UC, $n = 14$). **j**, Frequency of IgG1 plasma cell relative to all plasma cells, compared to frequency of PD1Hi T_{FH} (left) or T_H17 (right), relative to CD4 T cells, per cCD and UC diagnostic biopsies with ≥ 10 plasma cells and ≥ 10 CD4 T cells (cCD, $n = 31$; UC, $n = 14$). **k**, Linear regression between IgG1 plasma cell frequency and expression of CXCL13 in PD1Hi T_{FH} cells and IL17A in T_H17 cells. Restricted to diagnostic cCD or UC biopsies with at least 10 cells in the compared cell states (IgG1 PC vs. PD1Hi T_{FH}: cCD, $n = 24$; UC, $n = 13$. IgG1 PC vs. T_H17: cCD, $n = 16$; UC, $n = 14$). All unspecified per-biopsy grouped comparisons were statistically tested with Mann-Whitney U, two-tailed. All error bars denote standard error of the mean. Shaded areas of regressions show 95% confidence interval.

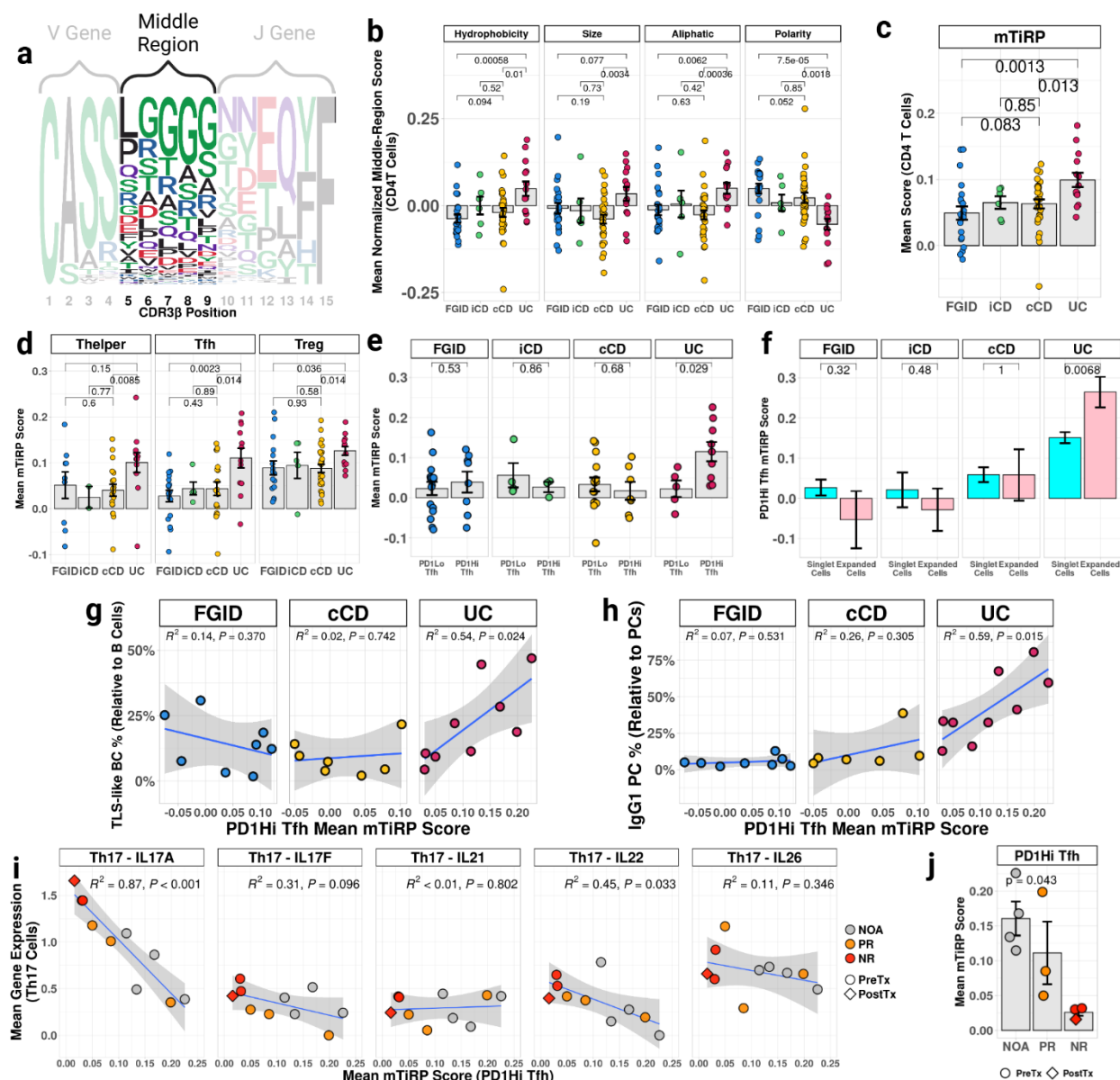
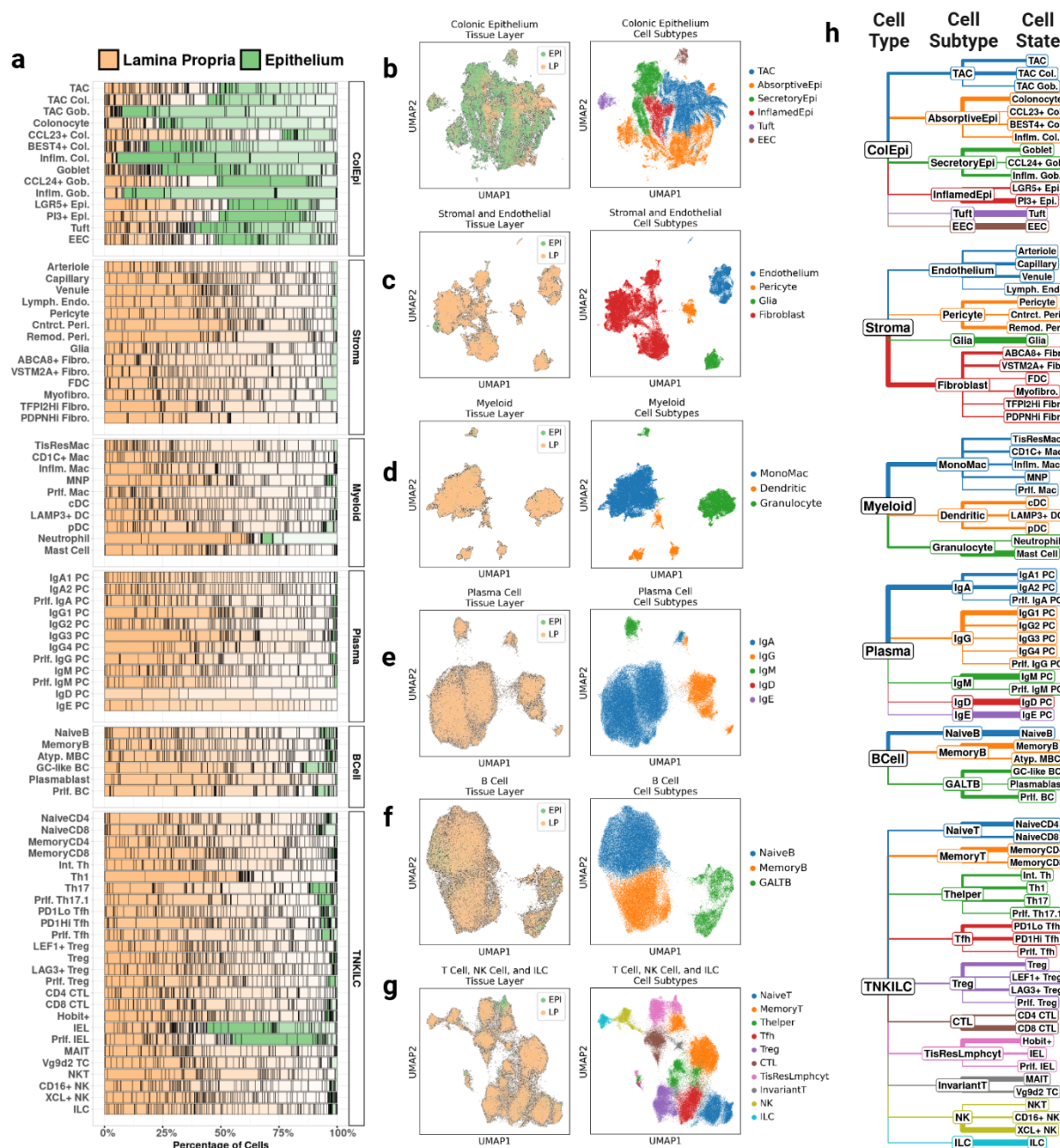


Fig. 7 | UC PD1Hi T_H cells are enriched for polyreactive TCR features that are inversely correlated with IL17A gene expression
a, Logo plot of CDR3β amino acid usage among baseline CD4 T cells, for length 15 CDR3β sequences. Brackets denote areas of V gene contribution, J gene contribution, and N or P nucleotide contribution (middle region). **b**, CDR3β middle-region metrics of biophysical properties among diagnostic biopsies with ≥ 50 CD4 T cells with CDR3β detected (FGID, n = 22; iCD, n = 6; cCD = 30; UC = 14). Quantification of CDR3β middle-region biophysical properties calculated using R package alakazam and Gaussian-normalized within PREDICT dataset. Mean score denoted by gray bar. **c**, Mean mTiRP score in CD4 T cells among diagnostic biopsies with ≥ 50 CD4 T cells with CDR3β detected. **d**, Mean mTiRP score calculated among CD4 T cell subtypes T_H, T_{FH}, and T_{reg} within FGID, cCD, and UC diagnostic biopsies. Biopsies included have ≥ 50 T_H cells (FGID, n = 13; iCD, n = 2; cCD, n = 24; UC, n = 14), T_{FH} cells (FGID, n = 19; iCD, n = 5; cCD, n = 27; UC, n = 13), or T_{reg} cells (FGID, n = 20; iCD, n = 5; cCD, n = 28; UC, n = 14) with CDR3β detected. **e**, Mean mTiRP score among PD1Lo and PD1Hi T_{FH} cells. FGID, cCD, and UC diagnostic biopsies with ≥ 50 cells in cell state with CDR3β detected included for analysis (FGID: PD1Lo T_{FH}, n = 16; PD1Hi T_{FH}, n = 8. iCD: PD1Lo T_{FH}, n = 4; PD1Hi T_{FH}, n = 3; cCD: PD1Lo T_{FH}, n = 15; PD1Hi T_{FH}, n = 7. UC: PD1Lo T_{FH}, n = 5; PD1Hi T_{FH}, n = 9). **f**, Comparison of singlet (CDR3αβ observed only once within a biopsy) and expanded (CDR3αβ observed more than once within a biopsy) PD1Hi T_{FH} mTiRP scores, within FGID, cCD, and UC cells (FGID, 945 singlets, 85 expanded; iCD, 272 singlets, 65 expanded; cCD, 1113 singlets, 127 expanded; UC, 2009 singlets, 250 expanded). T-test used for statistical comparisons. **g**, **h**, Frequency of TLS-like B cells (GC-like B cells and proliferating B cells) and IgG₁ plasma cells compared to mean PD1Hi T_{FH} mTiRP. FGID, cCD, and UC diagnostic biopsies included for linear regression analysis contained ≥ 50 PD1Hi T_{FH} cells with CDR3β detected and ≥ 10 TLS-like B cells or IgG₁ plasma cells (TLS-like B cell regression: FGID, n = 8; cCD, n = 9; UC, n = 10. IgG₁ regression: FGID, n = 8; cCD, n = 8; UC, n = 10). iCD excluded from analysis because only 3 biopsies met criteria. **i**, Regression between mTiRP in PD1Hi T_{FH} cells and canonical T_H17 gene expression in T_H17 cells among diagnostic or follow-up UC biopsies with ≥ 50 PD1Hi T_{FH} cells with CDR3β detected and ≥ 10 T_H17 cells (n = 10). **j**, Comparison of PD1Hi T_{FH} mTiRP score among diagnostic and follow-up UC not-on-anti-TNF (NOA), partial responder (PR) and

nonresponder (NR) biopsies with ≥ 50 CDR3 β detected (NOA: diagnostic biopsy, n = 4; follow-up biopsy, n = 0. PR: diagnostic biopsy, n = 3; follow-up biopsy, n = 0. NR: diagnostic biopsy, n = 2; follow-up biopsy, n = 1). Kruskal-Wallis test used to compare anti-TNF response groups. Two-tailed Mann-Whitney U test used for grouped comparisons unless otherwise specified. All error bars denote standard error. Shaded areas of regressions show 95% confidence interval.

SUPPLEMENTAL FIGURES

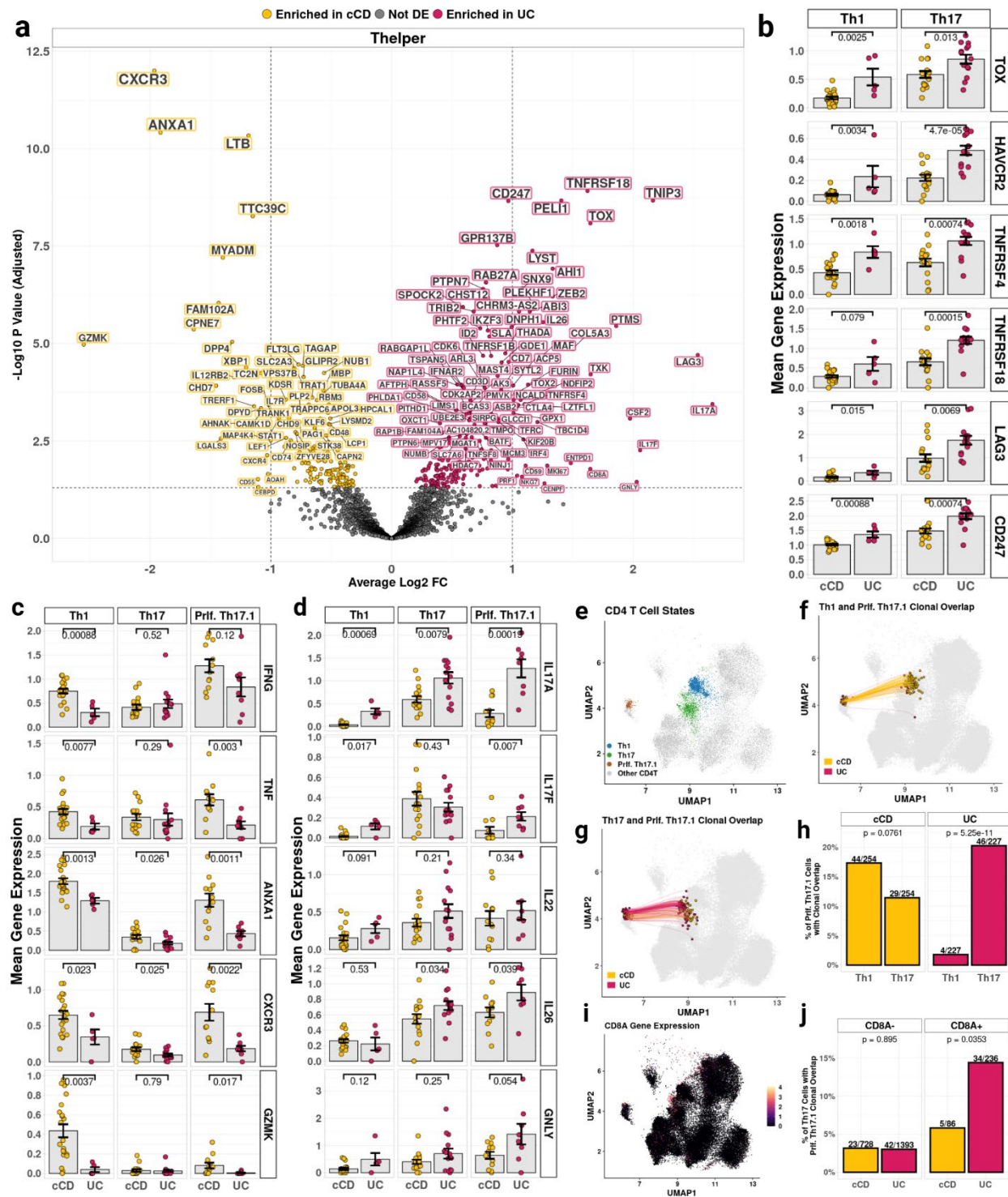
Supplemental Fig. 1 Comprehensive cellular annotation reveals cell state compartmentalization in epithelium or lamina propria.....	79
Supplemental Fig. 2 Cell state markers and biopsy diversity.....	80
Supplemental Fig. 3 Gene expression and clonal overlap in T _H 1, T _H 17, and Proliferating T _H 17.1 cells.....	81
Supplemental Fig. 4 T _H <i>IFNG</i> expression, but not other T/NK <i>IFNG</i> expression, is correlated with <i>STAT1</i> expression in T _H 1-associated cell states	83
Supplemental Fig. 5 <i>IL17A</i> expression is associated with IL-17-induced antimicrobial and leukocyte recruitment genes in colonic epithelial cells.....	84
Supplemental Fig. 6 Pre-treatment transcriptomic differential expression between cCD treatment groups and anti-TNF response groups	86
Supplemental Fig. 7 Pre-treatment transcriptomic differential expression between UC treatment groups	87
Supplemental Fig. 8 T _{FH} cells are identified without evidence of T _{PH} phenotype	88
Supplemental Fig. 9 HLA-DRB1*15:01+ UC participants are skewed towards PD1Hi T _{FH} -associated features and share features of autoreactivity.....	89
Supplemental Fig. 10 Summary of T _H 1-, T _H 17-, and PD1Hi T _{FH} -associated features among disease groups, anti-TNF response categories, and antigen reactivity	91



Supplemental Fig. 1 | Comprehensive cellular annotation reveals cell state compartmentalization in epithelium or lamina propria

a, Proportion of annotated cellular states identified within LP (beige) or EPI (green) layer from colonic biopsies. Individual samples are assigned to gradations of layer color, marked with vertical line between distinct samples. **b, c, d, e, f, g**, UMAPs highlighting cells attributed to EPI or LP samples (left), with corresponding cell subtype annotations for each cell type: colonic epithelium (**b**), stromal and endothelial cells (**c**), myeloid cells (**d**), plasma cells (**e**), B cells (**f**), and T, NK, and ILCs (**g**). (**h**) Visual tree of each cell state, relative to its corresponding cell subtype and cell type. Branch colors match the cell subtype annotation colors in **b-g**. Branch width corresponds to the proportion of cells in child node relative to parent (cell subtype/cell type; cell state/cell subtype). UMAP, uniform manifold approximation and projection. LP, lamina propria. EPI, epithelium. NK, natural killer. ILC, innate lymphoid cell.

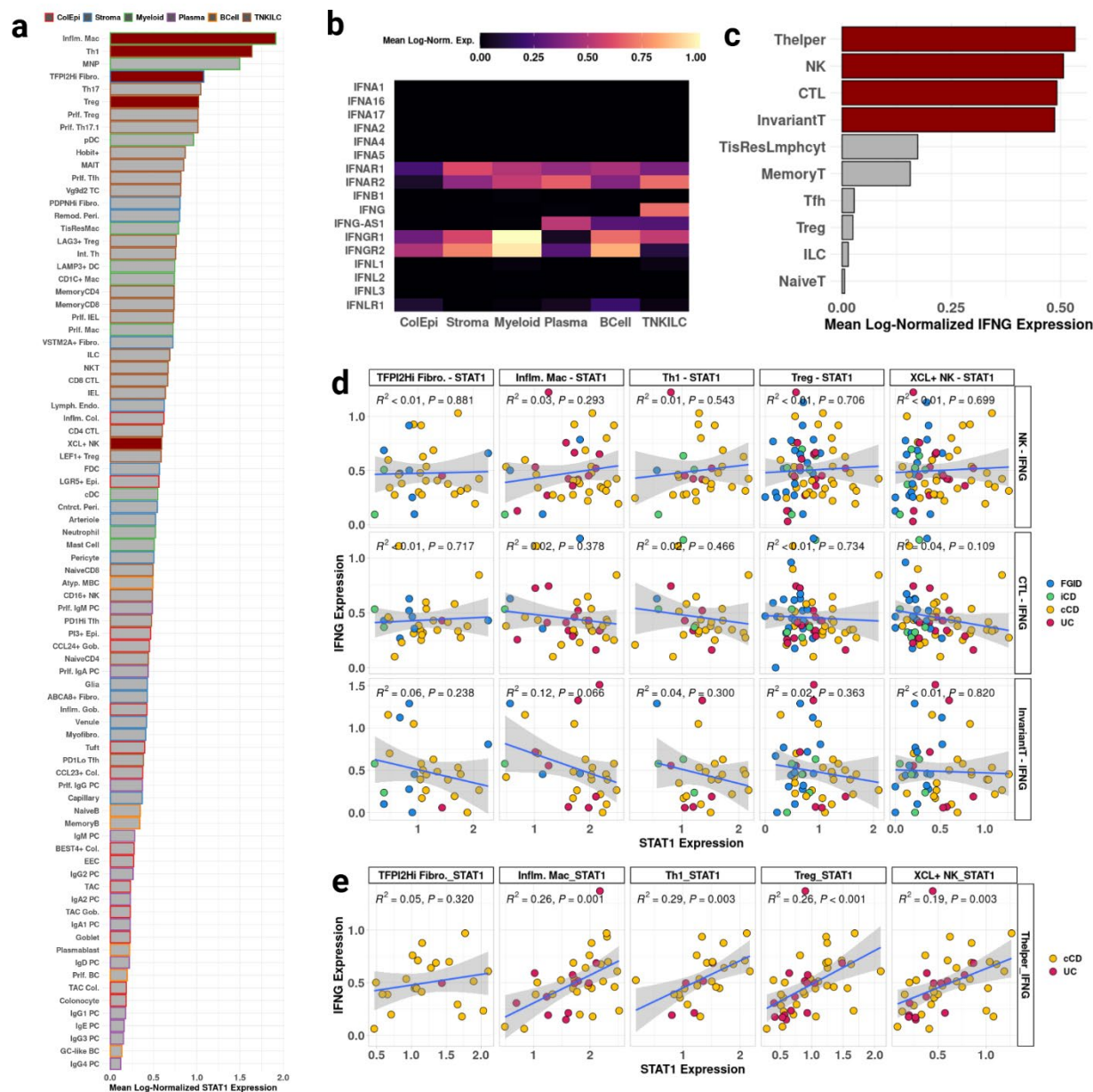
Supplemental Fig. 2 | Cell state markers and biopsy diversity



Supplemental Fig. 3 | Gene expression and clonal overlap in TH1, TH17, and Proliferating TH17.1 cells

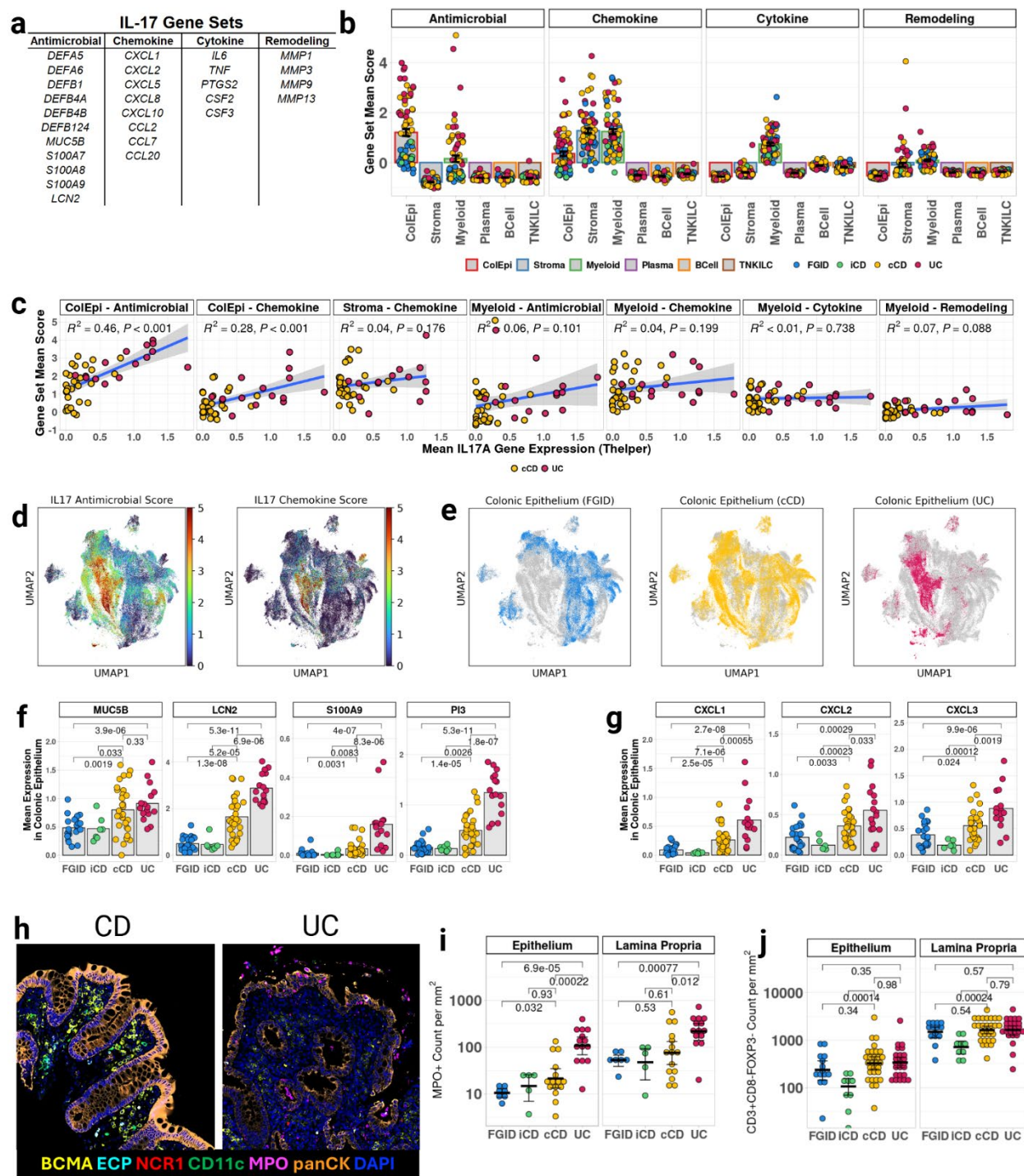
a, Differential expression analysis contrasting cCD and UC gene expression within pseudobulked TH cell subtype, among diagnostic biopsies with ≥ 10 TH cells (cCD, $n = 30$; UC, $n = 15$). **b**, Mean log-normalized expression of T cell activation, co-stimulatory and exhaustion genes within TH1 and TH17 cells among cCD and UC biopsies with ≥ 10 TH1 (cCD, $n = 21$; UC, $n = 5$) or TH17 (cCD, $n = 16$; UC, $n = 14$). **c**, **d**, Mean log-normalized expression of TH1-associated (**c**) and TH17-associated (**d**) genes within TH1 and TH17 cells among cCD and UC biopsies with ≥ 10 TH1 (cCD, $n = 21$; UC, $n = 5$) or TH17 (cCD, $n = 16$; UC, $n = 14$). **e**, UMAP of CD4 T cells among diagnostic biopsies with TH1 cells (blue), TH17 cells (green), and proliferating TH17.1 cells (brown) highlighted. **f**, **g** UMAPs of shared clones between proliferating TH17.1 cells and TH1 cells (**f**) or TH17 cells (**g**) within cCD (yellow lines) or UC (red lines) diagnostic biopsies. Clonotypes were determined by matched paired CDR3 α amino acid sequence. **h**, Percentage of cCD and UC proliferating

$T_H17.1$ cells with clonal overlap with T_H1 or T_H17 cells. All cells from diagnostic biopsies with at least 1 cell with paired CDR3 $\alpha\beta$ in Prlf. $T_H17.1$ cell state included (cCD, n = 24; UC, n = 14). i, UMAP of *CD8A* expression among CD4 T cells. j, Percentage of cCD and UC T_H17 cells with clonal overlap with the proliferating T_H17 cell state, split by T_H17 cells with or without *CD8A* expression. All cells from diagnostic biopsies with at least 1 cell with paired CDR3 $\alpha\beta$ in T_H17 cell state included (cCD, n = 30; UC, n = 14). Error bars denote standard error of the mean. CDR, complementarity determining region.



Supplemental Fig. 4 | T_H IFNG expression, but not other T/NK IFNG expression, is correlated with STAT1 expression in T_H1-associated cell states

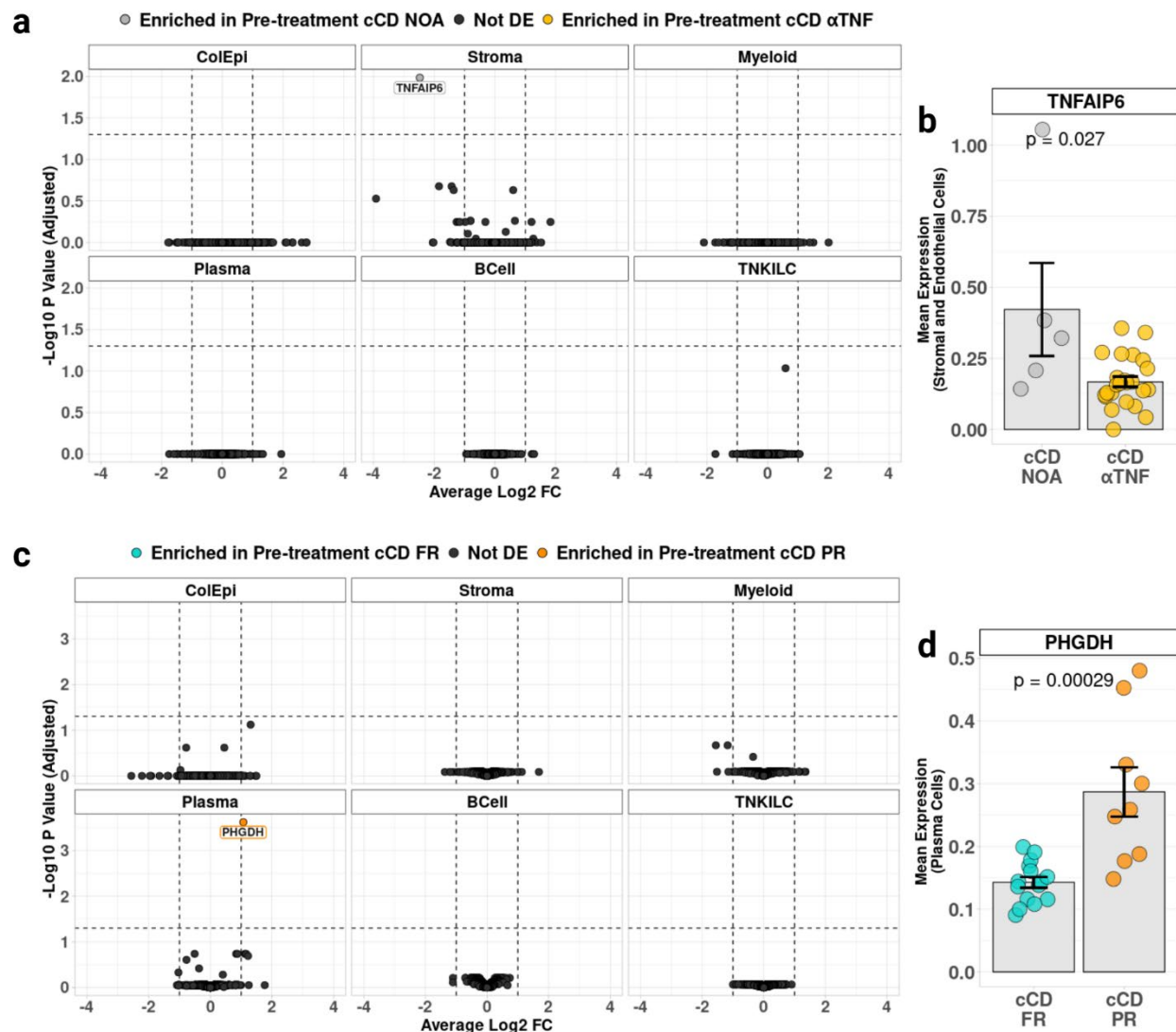
a, Expression of *STAT1* in all cell states. Bars outlined in color corresponding to respective cell group. Bars filled red if cell state was associated with T_H1 cell state in cell state frequency correlation analysis. **b**, Per-cell type mean log-normalized expression of type 1, type 2, and type 3 interferon cytokines and their receptors within pre-treatment biopsies. **c**, Mean log-normalized *IFNG* expression in T/NK/ILC cell subtypes pre-treatment biopsies. Bars filled red if cell subtype included in regression analyses in subfigures **d-e**. **d**, Relationships between NK, CTL, and InvariantT cell subtype expression of *IFNG* and T_H1-associated cell states, modeled as simple linear regressions. **e**, Relationships between T_H cell subtype expression of *IFNG* and T_H1-associated cell states in cCD and UC only, modeled as simple linear regressions. Shaded areas of regressions show 95% confidence interval.



Supplemental Fig. 5 | IL17A expression is associated with IL-17-induced antimicrobial and leukocyte recruitment genes in colonic epithelial cells

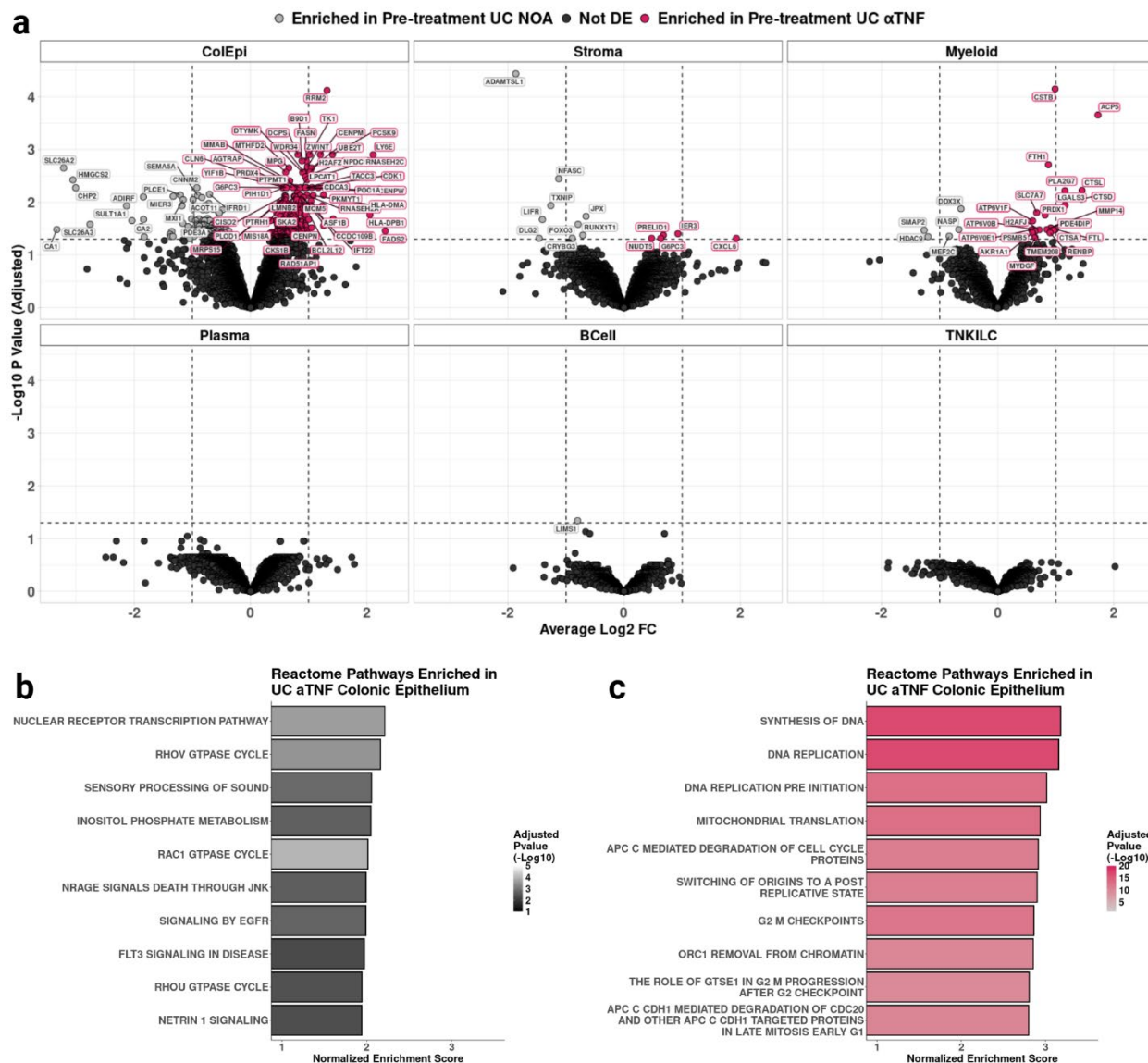
a, Table of genes associated with IL-17A signaling in KEGG “IL-17 Signaling Pathway.” **b**, IL-17 gene set scores in diagnostic biopsies with at least 10 cells in cell type (ColEpi, n = 77; Stroma, n = 74; Myeloid, n = 74; Plasma, n = 75; BCell, n = 72; TNKILC, n = 76) **c**, Relationship between mean expression of IL17A in T_H cells and IL-17 antimicrobial and chemokine gene set scores in cell types with positive mean expression of gene set, modeled as a simple linear regression among diagnostic cCD and UC biopsies with at least 10 T_H cells and 10 cells in the relevant cell type (cCD, n = 30; UC, n = 14 for stromal and endothelial cells, n = 15 for all other cell types). **d**, UMAP depicting scoring of colonic epithelium for IL-17 antimicrobial (left) and chemokine (right) gene sets in diagnostic biopsies. **e**, UMAPs highlighting FGID, cCD, and UC colonic epithelial cells in diagnostic biopsies. **f**, **g**, Mean log-normalized expression of antimicrobial (**f**) or chemokine (**g**) genes in colonic epithelium, among diagnostic biopsies with ≥ 10 colonic epithelial cells (FGID, n =

23; iCD, n = 7; cCD, n = 31; UC, n = 16). **h**, Representative images from colon of Crohn's disease and ulcerative colitis participants' epithelium and proximal lamina propria, stained for BCMA (yellow, plasma cells), ECP (cyan, eosinophils), MPO (magenta, neutrophils), CD11c (green, dendritic cells), NCR1 (red, NK cells), DAPI (dark blue, nucleated cells), and panCK (orange, epithelium). **i, j**, quantification of MPO+ (**i**) and CD3+CD8-FOXP3- (**j**) cells *in situ*, separated by epithelial and lamina propria layers. All fluorescence microscopy samples taken from diagnostic biopsies (MPO: FGID, n = 7; iCD, n = 5; cCD, n = 15; UC, n = 14. CD3+CD8-FOXP3-: FGID, n = 14; iCD, n = 9; cCD, n = 30; UC, n = 25). Two-tailed Mann-Whitney U test for all pairwise comparisons. Shaded areas of regressions show 95% confidence interval.



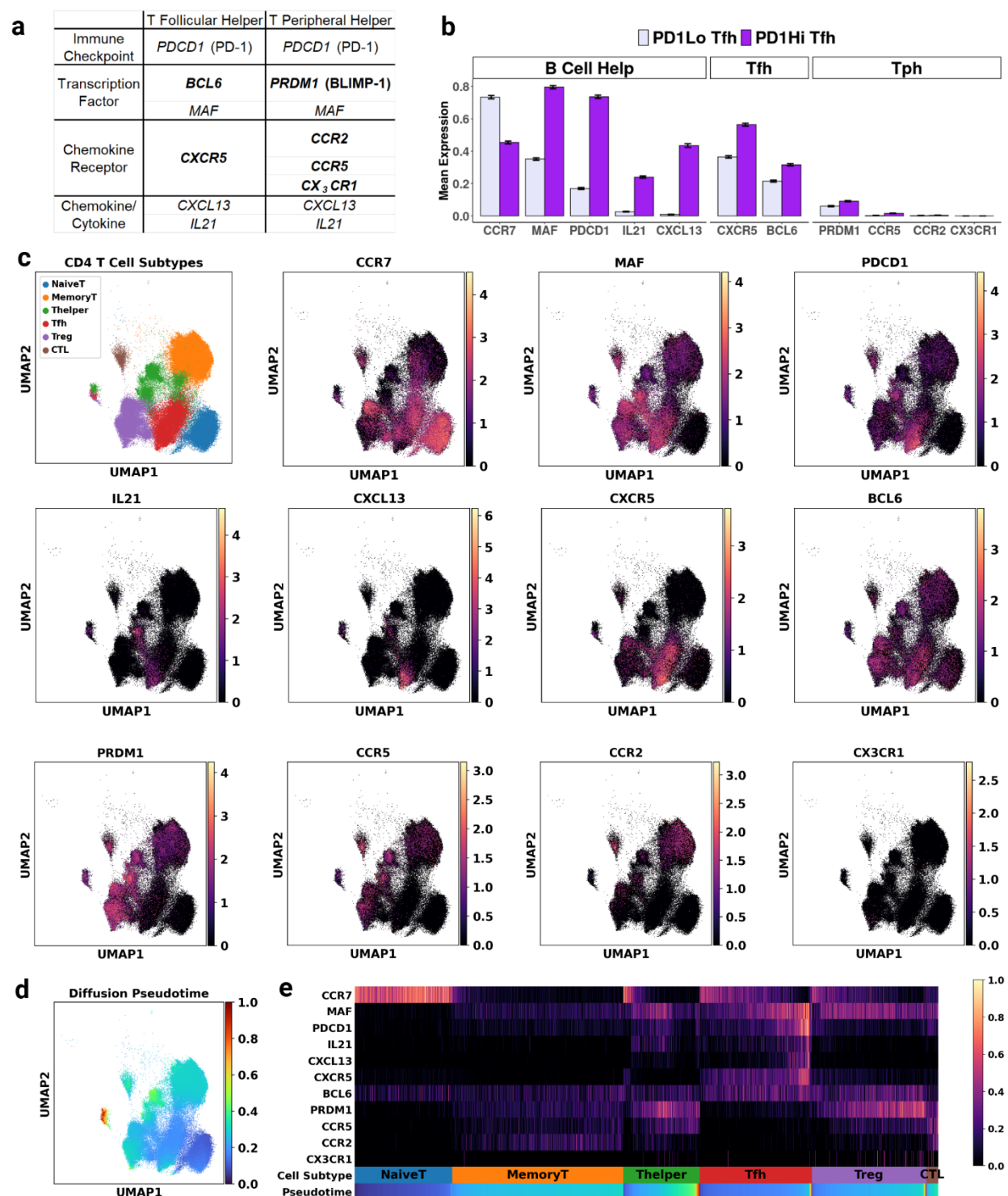
Supplemental Fig. 6 | Pre-treatment transcriptomic differential expression between cCD treatment groups and anti-TNF response groups

a, Volcano plot comparing differentially expressed genes in pre-treatment cCD NOA and anti-TNF recipients within each annotated cell type. Biopsies were included for analysis if they contained ≥ 10 cells within the cell type (NOA: $n = 5$ for all cell types. α TNF: stromal and endothelial cells, myeloid cells, plasma cells, $n = 24$; colonic epithelium, B cells, and T, NK, and ILC cells, $n = 25$). **b**, Mean log-normalized expression of *TNFAIP6* in stromal and endothelial cells, comparing cCD NOA and anti-TNF recipient diagnostic biopsies with ≥ 10 stromal and endothelial cells (NOA, $n = 5$; α TNF, $n = 24$). **c**, Volcano plot comparing differentially expressed genes in pre-treatment cCD FR and PR recipients within each annotated cell type. Biopsies were included for analysis if they contained ≥ 10 cells within the cell type (FR: $n = 15$ for all cell types. PR: stromal and endothelial cells, myeloid cells, plasma cells, $n = 9$; colonic epithelium, B cells, and T, NK, and ILC cells, $n = 10$). Two-tailed Mann-Whitney U test for all pairwise comparisons. **d**, Mean log-normalized expression of *PHGDH* in plasma cells, comparing cCD NOA and anti-TNF recipient diagnostic biopsies with ≥ 10 plasma cells (NOA, $n = 5$; α TNF, $n = 24$).



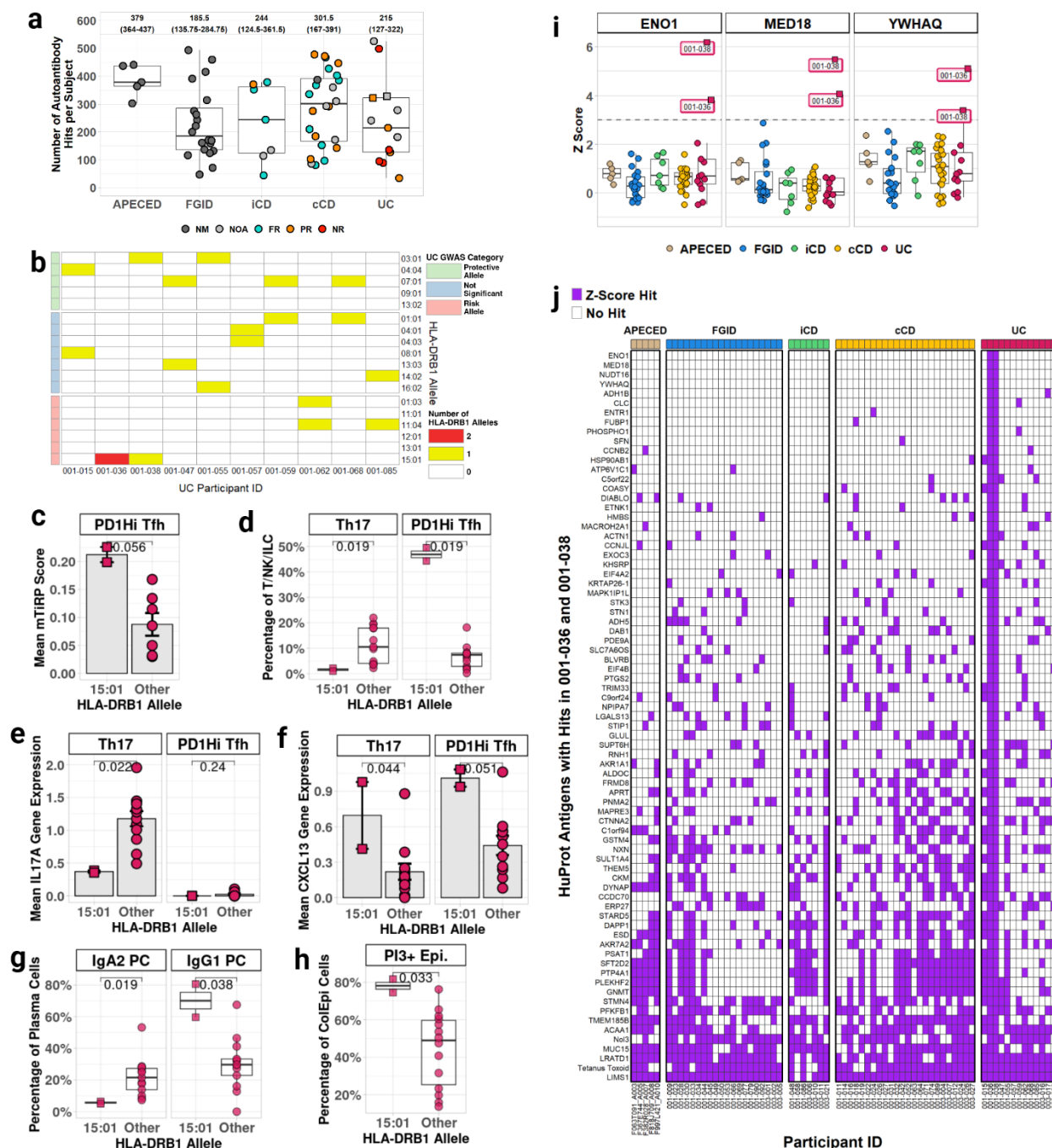
Supplemental Fig. 7 | Pre-treatment transcriptomic differential expression between UC treatment groups

a, Volcano plot comparing differentially expressed genes in UC NOA and anti-TNF recipients within each annotated cell type. Biopsies were included for analysis if they contained ≥ 10 cells within the cell type (NOA: colonic epithelium, $n = 8$; stromal cells, $n = 6$; all other cell types $n = 7$. αTNF: $n = 8$ for all cell types). **b**, **c**, Enriched terms from gene set enrichment analysis of differential expression between pre-treatment colonic epithelium from anti-TNF untreated (**b**) and treated (**c**) UC participants. Top 10 terms for each treatment group shown. Gene ranks computed using DESeq2 Wald test "stat" value ($\log_2FC/SE(\log_2FC)$). Ranked genes tested using Reactome 2023 v2 gene sets with $15 \leq \text{number of genes} \leq 100$ and sorted for normalized enrichment score.



Supplemental Fig. 8 | T_{FH} cells are identified without evidence of T_{PH} phenotype

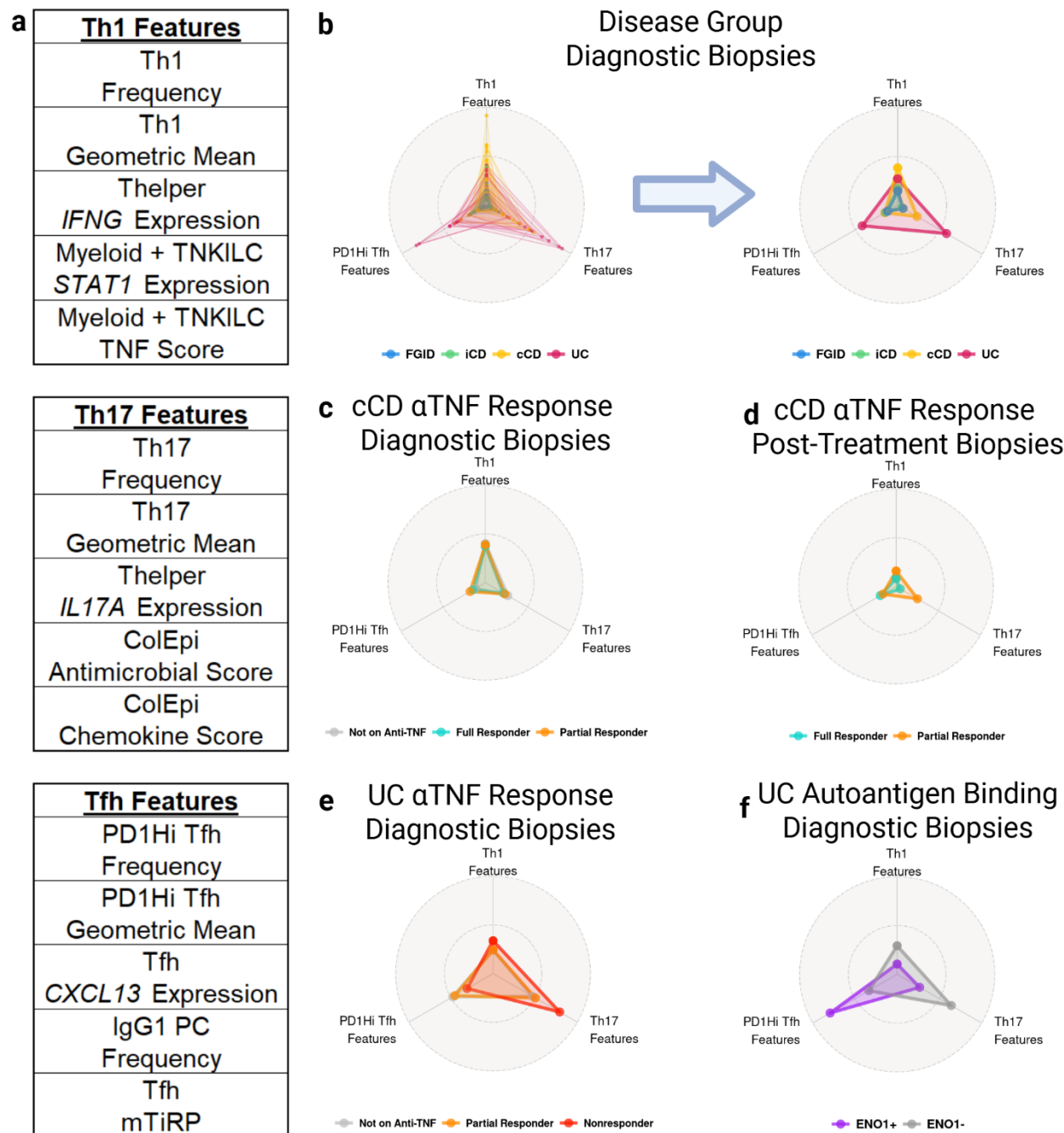
a, Table of genes canonically associated with T_{FH} and T_{PH} cells. Protein function listed in leftmost column. Bolded genes are differentiating between T_{FH} and T_{PH}. **b**, Mean log-normalized expression of genes involved in B cell help, T_{FH} identity, and T_{PH} identity amongst all PD1Lo T_{FH} cells and PD1Hi T_{FH} cells. **c**, UMAPs of CD4 T cells labeled by cell state, disease group. Log-normalized expression of selected T_{FH} and T_{PH} genes overlaid on UMAP. **d**, Diffusion pseudotime UMAP of CD4 T cells. **e**, Per-gene scaled expression (0-1) of T_{FH} and T_{PH} genes among CD4 T cells, ordered by cell subtypes and pseudotime value within cell subtypes. Cells from all disease groups, treatment groups, and timepoints included for all analyses.



Supplemental Fig. 9 | HLA-DRB1*15:01+ UC participants are skewed towards PD1Hi T_{FH}-associated features and share features of autoreactivity

a, Boxplots of number of autoantigen hits per disease group (FGID, ICD, cCD, UC, as well as 5 additional APECED samples). Median number of autoantigen hits is displayed above each group's boxplot, followed by the range of the 25th percentile and the 75th percentile number of autoantigen hits. Markers are colored by anti-TNF response status. APECED, n = 5; FGID, n = 20; ICD, n = 7; cCD, n = 24; UC, n = 13. **b**, Number of HLA-DRB1 alleles in UC participants with genotype data (n = 10). **c**, Mean mTIRP score in PD1Hi T_{FH} cell state among UC diagnostic biopsies with ≥ 50 PD1Hi T_{FH} cells (with HLA-DRB1*15:01 allele, n = 2; without HLA-DRB1*15:01 allele, n = 7). **d**, Frequency of T_H17 and PD1Hi T_{FH} cell states, relative to all T/NK/ILC, among UC diagnostic biopsies with ≥ 10 T/NK/ILCs (HLA-DRB1*15:01 allele, n = 2; without HLA-DRB1*15:01 allele, n = 13). **e**, **f**, Mean log expression of IL17A (**e**) and CXCL13 (**f**) in T_H17 and PD1Hi T_{FH} cell states, among UC diagnostic biopsies with ≥ 10 in the relevant cell state (T_H17: with HLA-DRB1*15:01 allele, n = 2; without HLA-DRB1*15:01 allele, n = 12. PD1Hi T_{FH}: with HLA-DRB1*15:01 allele, n = 2; without HLA-DRB1*15:01 allele, n = 11). **f**, **g**, Frequency of IgA2 PCs and IgG1 PCs (**g**), or PI3+ Epi. (**h**) cell states relative to their respective cell types, among UC diagnostic biopsies with ≥ 10 cells in cell type (Plasma cells: with HLA-DRB1*15:01 allele, n = 2; without HLA-DRB1*15:01 allele, n = 13).

= 13. Colonic epithelium: with HLA-DRB1*15:01 allele, n = 2; without HLA-DRB1*15:01 allele, n = 14). i, HuProt Z-score values for 3 antigens identified as hits only in both UC participants with HLA-DRB1*15:01 allele (001-036 and 001-038). A z-score cutoff of 3 is used to denote whether an antigen is a hit within a sample (APECED, n = 5; FGID, n = 20; iCD, n = 7; cCD, n = 24; UC, n = 13). j, Z-score hits from HuProt proteome microarray, filtered to 77 proteins with z-score hits among both UC participants with HLA-DRB1*15:01 allele (001-036 and 001-038). Antigens are ordered by number of distinct participants with hits. Two-tailed Mann-Whitney U test for all pairwise comparisons. In subfigures with participant markers, UC participants with HLA-DRB1*15:01 allele (001-036 and 001-038) denoted by square markers instead of circle markers.



Supplemental Fig. 10 | Summary of T_H1 -, T_H17 -, and PD1Hi T_{FH} -associated features among disease groups, anti-TNF response categories, and antigen reactivity

a, List of T_H1 , T_H17 , and T_{FH} features identified as distinguishing disease groups, treatment response groups, and/or autoantigen-binding groups throughout the PREDICT study. **b**, Radar plot demonstrating feature usage among diagnostic biopsies meeting all relevant cell minima and assays conducted (FGID, $n = 15$; iCD, $n = 4$; cCD, $n = 19$; UC, $n = 10$). Features were scaled between 0 and 1 (corresponding to per-feature minimum and maximum values, respectively), then averaged within each biopsy. Center of radar plot represents 0 scaled average value, outer edge represents 1 scaled average value. Radar plot shows individual diagnostic biopsies (left) or diagnostic biopsies grouped by disease (right). **c**, **d**, Radar plot comparing features among pre-treatment (**c**) and post-treatment (**d**) cCD biopsies, grouped by anti-TNF response status (Pre-treatment: NOA, $n = 4$; FR, $n = 10$; PR, $n = 5$. Post-treatment: FR, $n = 5$; PR, $n = 2$). **e**, Radar plot comparing features among pre-treatment UC biopsies, grouped by anti-TNF response status (NOA, $n = 3$; PR, $n = 4$; NR, $n = 3$). **f**, Radar plot comparing study features among pre-treatment UC biopsies, grouped by ENO1 reactivity (ENO1+, $n = 2$; ENO1-, $n = 8$).