Title

Single-cell transcriptomic analyses reveal distinct immune cell contributions to epithelial barrier dysfunction in checkpoint inhibitor colitis.

Journal

Nature Medicine, 30(5)

Authors

Thomas, Molly
Slowikowski, Kamil
Manakongtreecheep, Kasidet
et al.

Publication Date

2024-05-01

DOI

10.1038/s41591-024-02895-x

Peer reviewed

HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2024 December 27.

Published in final edited form as:

Nat Med. 2024 May ; 30(5): 1349–1362. doi:10.1038/s41591-024-02895-x.

Single-cell transcriptomic analyses reveal distinct immune cell contributions to epithelial barrier dysfunction in checkpoint inhibitor colitis

Molly Fisher Thomas[1,2,3,4,5,6,7,*,†], **Kamil Slowikowski**[1,2,3,4,*,†], **Kasidet Manakongtreecheep**[1,2,3], **Pritha Sen**[1,3,4,8], **Nandini Samanta**[1,2,3], **Jessica Tantivit**[1,2,3], **Mazen Nasrallah**[1,3,4,9], **Leyre Zubiri**[2,4,10], **Neal P. Smith**[1,2,3], **Alice Tirard**[1,2,3], **Swetha Ramesh**[1,2,3], **Benjamin Y. Arnold**[1,2,3], **Linda T. Nieman**[2,4], **Jonathan H. Chen**[2,3,4,11], **Thomas Eisenhaure**[3], **Karin Pelka**[2,3,4,12], **Yuhui Song**[2], **Katherine H. Xu**[2], **Vjola Jorgji**[2,11], **Christopher J. Pinto**[2,13], **Tatyana Sharova**[14], **Rachel Glasser**[5], **PuiYee Chan**[2,4,13], **Ryan J. Sullivan**[2,4,10], **Hamed Khalili**[3,4,5], **Dejan Juric**[2,3,4,10], **Genevieve M. Boland**[2,4,14], **Michael Dougan**[4,5], **Nir Hacohen**[2,3,4], **Bo Li**[1,3,4,15], **Kerry L. Reynolds**[2,4,10,16], **Alexandra-Chloé Villani**[1,2,3,4,16,†]

[1] -Center for Immunology and Inflammatory Diseases, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[2] –Krantz Family Center for Cancer Research, Massachusetts General Hospital, Boston, MA, USA

[3] -Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA

[4] -Harvard Medical School, Boston, MA, USA

[5] -Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

[6] -Current address: Division of Gastroenterology, Department of Medicine, Oregon Health and Sciences University, Portland, Oregon, USA

[7] –Department of Cell, Developmental, and Cancer Biology, Oregon Health and Sciences University, Portland, Oregon, USA

[8] -Division of Infectious Disease; Transplant, Oncology, and Immunocompromised Host Group, Department of Medicine, Brigham and Women's Hospital, and Dana-Farber Cancer Institute, Boston, MA, USA

[†]Correspondence to: thomamol@ohsu.edu; kslowikowski@mgh.harvard.edu; avillani@mgh.harvard.edu.
[*]These authors contributed equally: Molly Fisher Thomas, Kamil Slowikowski

Author Contributions Statement

M.F.T. and A.C.V. conceived of and led the study; M.F.T. and A.C.V. led experimental design; M.F.T. carried out experiments with assistance from K.M., J.T., P.S., M.N., A.T., and BY.A.; K.S. designed and performed computational analysis, with input and assistance from B.L., M.N., N.S., and S.R; M.F.T, L.T.N., J.H.C. designed and performed microscopy experiments with input and assistance from K.H.X., Y.S., and V.J.; T.E., K.P. provided input for sc/snRNA-Seq experiments, protocols and data interpretation; M.F.T., L.Z., C.J.P., T.S., R.G., P.Y.C., R.J.S., D.J., G.M.B., H. K., K.L.R, and M.D. provided clinical expertise, coordinated and performed sample acquisition and/or administrative coordination; M.F.T. and M.D. performed endoscopic exams; M.D. provided additional expertise on study design; N.H. contributed to biological expertise, study design, and advice; B.L. contributed to computational expertise and advice; A.C.V. managed and supervised the study; A.C.V. and K.L.R. raised funding for this work; M.F.T., K.S., A.C.V. wrote the manuscript, with input from all authors.

Code Availability

Source code for data analysis and the website is available on GitHub (https://github.com/villani-lab/ircolitis) and has been archived at Zenodo (https://zenodo.org/records/10650921).

[9] -Division of Rheumatology, North Shore Physicians Group, Department of Medicine, Mass General Brigham Healthcare Center, Lynn, MA, USA

[10]-Division of Hematology-Oncology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

[11] -Department of Pathology, Massachusetts General Hospital, Boston, MA, USA

[12]-Current address: Gladstone–UCSF Institute of Genomic Immunology, San Francisco, CA, USA

[13] -Clinical Research Center, Massachusetts General Hospital, Boston, MA, USA

[14] -Department of Surgery, Massachusetts General Hospital, Boston, MA, USA

[15] -Current address: Genentech, 1 DNA Way, South San Francisco, CA, 94080

[16]-These authors jointly supervised this work

Abstract

Immune checkpoint inhibitor (ICI) therapy has revolutionized oncology, but treatments are limited by immune-related adverse events, including checkpoint inhibitor colitis (irColitis). Little is understood about the pathogenic mechanisms driving irColitis, which does not readily occur in model organisms such as mice. To define molecular drivers of irColitis, we used single-cell multi-omics to profile ~300,000 cells from the colon mucosa and blood of 13 patients with cancer who developed irColitis (9 on anti-PD-1 or anti-CTLA-4 monotherapy and 4 on dual ICI therapy; most patients had skin or lung cancer), 8 controls on ICI therapy, and 8 healthy controls. Patients with irColitis showed expanded mucosal Tregs, *ITGAE*^{Hi} CD8 tissue-resident memory T cells expressing *CXCL13* and Th17 gene programs, and recirculating *ITGB2*^{Hi} CD8 T cells. Cytotoxic *GNLY*^{Hi} CD4 T cells, recirculating *ITGB2*^{Hi} CD8 T cells, and endothelial cells expressing hypoxia gene programs were further expanded in colitis associated with anti-PD-1/CTLA-4 compared to anti-PD-1 therapy. Luminal epithelial cells in patients with irColitis expressed PCSK9, PD-L1 and interferon-induced signatures associated with apoptosis, increased cell turnover, and malabsorption. Together, these data suggest roles for circulating T cells and epithelial-immune crosstalk critical to PD-1/CTLA-4-dependent tolerance and barrier function and identify potential therapeutic targets for irColitis.

Keywords

Immune-related adverse events; immune checkpoint inhibitors; CTLA-4; PD-1; single-cell RNA-seq; single-nuclei RNA-seq; colitis; tissue-resident memory cells; systems immunology

Introduction

Immune checkpoint inhibitors (ICIs) targeting PD-1 and CTLA-4 have revolutionized immuno-oncology. Unfortunately, treatments are limited by immune-related adverse events (IRAEs), which are inflammatory side effects ranging in severity from mild to potentially lethal. IRAEs occur in over 85% of patients on immunotherapy, leading to treatment discontinuation in 10–40% of patients[1]. IRAE colitis (irColitis) is the most common, severe IRAE and develops in up to 25% of patients[2]. irColitis presents with increased

intraepithelial lymphocytes, neutrophilic infiltrates, epithelial apoptosis, and/or isolated lymphocytic colitis[3,4]. Little is known about the mechanisms leading to irColitis, partly because mice treated with ICI-blocking antibodies do not develop colitis[5,6]. Correlational studies suggest roles for T regulatory (Treg) cells, CD21^{lo} B cells, TNF-α, IL-17A, MAIT cells, dysbiosis, and early T cell receptor (TCR) diversification[7–13]. Previous single-cell RNA sequencing (scRNA-seq) studies of colon mucosal immune cells have shown expanded tissue CD8 T cells[14,15], though because these studies did not examine epithelial or mesenchymal cells in tissue or circulating immune cells, it is unclear how these additional populations and crosstalk between immune and non-immune cells contribute to disease pathogenesis. irColitis is commonly treated with steroids as first line therapy and biologics targeting TNF-α or integrin $\alpha_4\beta_7$ for steroid-refractory disease[16,17]. Multiple studies have shown that treating IRAEs with high-dose steroids blunts ICI anti-tumor efficacy[18,19]. An improved understanding of IRAE pathogenesis is urgently needed to facilitate early diagnosis and inform new therapeutic targets that preserve anti-tumor responses.

Here we investigated five poorly understood aspects of irColitis: (1) the interactions between CD8 T cells and other immune and non-immune cells sustaining inflammation; (2) the contribution of circulating cells to disease; (3) the epithelial and mesenchymal defects defining colonic dysfunction; (4) the features of anti-PD-1 versus anti-PD-1/CTLA-4 irColitis; and (5) the identification of new, putative therapeutic targets. Our findings reveal the tremendous cellular transcriptional heterogeneity underlying irColitis and identify pathways for clinical diagnostics and therapeutic interventions.

Results

Study Design

To define cellular populations associated with irColitis, we generated scRNA-seq data from colon immune cells, epithelial and mesenchymal nuclei, and circulating peripheral blood mononuclear cells (PBMCs) from patients with irColitis, patients on ICI therapy without colitis, and healthy controls (Fig. 1A–C; Tables S1–S4; Extended Data Fig. 1; Supplementary Fig. 1). Most patients on ICI therapy had skin cancer and were treated with PD-1 blockade. Biospecimens were obtained within two weeks of symptom onset. Not all assays were performed for each patient due to limited sample availability (Table S2). We observed relatively few differences across cell lineages between the two control populations (Supplementary Fig. 1–2, Extended Data Fig. 2; Tables S5–7); these were thus aggregated for remaining analyses. The total number of differentially abundant clusters and differentially expressed genes between cases versus controls is summarized in Fig. 1D and highlights that disease-specific alterations are present in the colon mucosa but not in blood.

Tissue *ITGAE*^{Hi} and *ITGB2*^{Hi} CD8 T cells are expanded in irColitis

As reported[3], patients with irColitis had significantly increased colon mucosal immune cells, including T cells (Fig. 1C; Table S5). To characterize expanded populations, we subclustered lineages representing CD4 T cells, CD8 T cells/gamma delta T cells (GDT)/natural killer (NK) cells, mononuclear phagocytes (MPs), and B cells (Methods; Extended Data Fig. 1,Fig. Supplementary Figure 1; Tables S1–4). Compared to controls, patients with irColitis

had increased absolute numbers of all tissue CD8 T/GDT/NK subsets (Fig. S1A), which were dominated by expanded CD8 T cells with a T_{RM} phenotype defined by divergent *ITGAE* or *ITGB2* expression (Fig. 2A–D), as described[20]. ITGAE (*i.e.* CD103) and ITGB7 form a heterodimer mediating epithelial homing, while ITGB2 dimerizes with ITGAD/ITGAL(LFA-1)/ITGAM/ITGAX, which induce homing to ICAM/VCAM in inflamed tissue and modulate T cell differentiation and cytotoxicity[21]. *ITGAE*^{Hi} CD8 T cells represented a phenotypic gradient from naive/resting cells expressing *IL7R and TCF7* (clusters 1, 6), which were depleted in irColitis, to *GZMB*^{Hi} effectors (cluster 7), which were expanded (Fig. 2C; Table S6). CD8 T_{RM} had divergent expression of the transcriptional regulators *EOMES*, *ZNF683*, and *KZF2*, which was enriched in clusters 1 and 5 and is notably upregulated in exhausted T cells[22]. Indeed, CD8 T cell cluster 5 upregulated *KIR2DL4*, *HAVCR2*, and *TIGIT*, which comprise an exhausted T cell transcriptional program[22,23]. Cases showed expansion of both *ITGB2*^{Hi} CD8 T cell populations, including *GZMK*^{Hi} cluster 3 and *CX3CR1*^{Hi} cluster 11 cells, which expressed circulatory markers *S1PR1* and *SELL*, suggesting recent emigration from blood or recirculation (Fig. 2B–D). Additionally, *CX3CR1*^{Hi}, *GZMB*^{Hi} cluster 11 cells displayed a cytotoxic phenotype resembling described[24] blood-confined *CX3CR1*^{Hi} effector memory CD8 T cells (Fig. 2B–D).

Because combination therapy increases the risk of irColitis 5–10 fold compared to monotherapy[1], we investigated transcriptional differences in CD8 T cells from patients with irColitis on PD-1/CTLA-4 blockade versus anti-PD-1 (Fig. 2E; Supplementary Fig. 1M; Tables S6–7). Both *ITGB2*^{Hi} CD8 T populations (clusters 11, 3) were more enriched in patients receiving dual therapy (Fig. 2E; Table S6). Furthermore, CD8 T cells from dual therapy patients upregulated genes reflecting increased TCR signaling (*CD28*, *TNFRSF4*), homing to inflamed tissue (*SELL*, *ITGB2*), and homing to epithelium (*PXN*), and *LINC00861*, which positively correlates with PD-1 in cancer[25] (Fig. 2F–G; Supplementary Fig. 1M, 3; Table S7). In sum, irColitis is defined by the tissue expansion of cytotoxic *ITGAE*^{Hi} CD8 T_{RM} and likely recirculating *ITGB2*^{Hi} CD8 T cells, which are more strongly expanded in dual anti-PD-1/CTLA-4 than anti-PD-1 monotherapy and thus may increase irColitis risk.

Tissue CD8 T cell repertoire diversity increases in irColitis

ICI therapy remodels the CD8 TCR repertoire of tumor infiltrating CD8 T cells[26], yet the impact of irColitis on colon TCR diversity and TCR sharing with blood is unknown. Single-cell TCR analysis showed that the two tissue *ITGB2*^{Hi} CD8 T cell subtypes (clusters 3, 11) shared significantly fewer TCRs with other CD8 clusters (Fig. 2H–M; Extended Data Fig. 3A). In contrast, *ITGAE*-expressing CD8 T_{RM} cells variably shared TCRs across multiple clusters, suggesting a shared ontogeny across a phenotypic spectrum of *ITGAE*^{Hi} cells.

Blood CD8 TCR diversity correlates positively with ICI efficacy and IRAE development[12]. Our analysis demonstrated increased TCR diversity in tissue CD8 (but not CD4) T cells from cases versus controls (Fig. 2J; Extended Data Fig. 3). We did not observe tissue CD8

TCRs shared between patients. Increased tissue CD8 TCR diversity in cases suggests that T cell expansion is unlikely to be driven by a single dominant antigen.

Expanded *ITGB2*^{Hi} CD8 T cells in irColitis share clones with circulating cells

To determine if *ITGB2*^{Hi} CD8 T cell populations expressing circulatory markers were clonally related to blood T cells, we analyzed matched peripheral blood sampled at the time of tissue collection for each patient (Fig. 1A–B; Table S2). PBMCs were processed for scRNA-seq, subclustered (Methods; Fig. 2K; Extended Data Fig. 4; Supplementary Fig. 1,7; Tables S3–S4) and compared to blood cell types in a recently published dataset[27] (Extended Data Fig. 4F,K). Unlike in tissue, we found no differentially abundant blood immune populations or DE genes in cases versus controls (Extended Data Fig. 1D, 4, 5; Supplementary Fig. 1; Tables S6–S7).

TCR analysis revealed that an average of 2% of an individual's CD8 T cell clones were observed in both colon and blood (Extended Data Fig. 4D; Table S8), and this frequency did not vary between cases and controls. In contrast, *ITGB2*^{Hi} tissue CD8 T subsets had significantly increased blood and tissue TCR sharing compared to tissue *ITGAE*^{Hi} CD8 T cells (Fig. 2L). Cells in three blood CD8 T subsets (cycling cluster 12, *CX3CR1*^{Hi} *FGFBP2*^{Hi} cluster 5, and activated MHC-II^{Hi} *GZMK*^{Hi} cluster 2; Extended Data Fig. 4A–C) shared more CD8 TCRs with irColitis colon mucosa than controls (Fig. 2M). Blood cluster 5 *CX3CR1*^{Hi} CD8 T cells were transcriptionally similar to tissue cluster 11 *CX3CR1*^{Hi} CD8 T cells and described effector memory cells that re-express CD45RA (i.e., TEMRA)[27] (Extended Data Fig. 4F). Based on elevated *CX3CR1* and *GZMB*, we predict that blood CD8 cluster 5 and colon CD8 cluster 11 are intravascular T effector memory populations[24,28]. Blood CD8 cluster 2 cells were phenotypically similar to tissue CD8 cluster 3 cells expressing *GZMK* and *EOMES* and are phenotypically similar to described recirculating, non-cytotoxic effector memory CD8 T cells[24,27] (Extended Data Fig. 4F). We predict that circulating T cell populations have pathologic roles in irColitis as they are a predicted target of the integrin inhibitor vedolizumab, which prevents circulating immune cells from trafficking into inflamed gastrointestinal mucosa and is effective in treating irColitis[29].

ITGAE^{Hi} CD8 T_{RM} Cells upregulate *IL17A*, *IL26*, and *CXCL13* in irColitis

Differential gene expression (DGE) analysis across CD8 T cells showed that almost all the irColitis DE genes were in *ITGAE*^{Hi} CD8 T cells and not *ITGB2*^{Hi} CD8 T cells (Fig. 3A–B; Supplementary Fig. 4A; Table S7). Many DE effector genes were detected across CD8 T cell subsets, highlighting shared transcriptional programs (Fig. 3; Supplementary Fig. 4A; Table S7). Multispectral RNA microscopy confirmed that T cells co-expressing *CD3E* and *IFNG* were strongly enriched in irColitis tissue throughout the lamina propria (LP) (Fig. 3E; Supplementary Fig. 5).

Many TCR co-inhibitory/exhaustion genes, which are upregulated in activated T cells, were enriched in *ITGAE*^{Hi} *GZMB*^{Hi} CD8 T effector clusters (4, 5, 7) (Fig. 3C–D). Several *ITGAE*^{Hi} CD8 T populations upregulated the chemokine *CXCL13* – which promotes tertiary lymphoid structure (TLS) formation – and multiple Th17 genes (*IL17A*, *IL26*,

IL23R, BATF) (Fig. 3C–D; Table S7). *LDLR* was up in multiple CD8 T cell subsets (Supplementary Fig. 4) and notably polarizes T cells towards a Th17-like phenotype[30]. *IL17A*, *IL26*, and *CXCL13* expression was absent from *ITBG2*^{Hi} CD8 T cells (Fig. 3C; Table S7). In our data, 1.9% of CD8 T cells expressed *IL17A*, consistent with 1.5% we detected in another irColitis dataset[15] and an 1.7–2.8% in an IBD dataset[31] (Supplementary Fig. 4C). Given the possible protective role of IL-17A in IBD[32], we speculate that IL-17A expression may similarly play a role in limiting irColitis pathology, which will require further investigation in future work.

Together, our results suggest that *ITGAE*^{Hi} *GZMB*^{Hi} CD8 T_{RM} effectors undergo dramatic transcriptional regulation in irColitis towards a Th17 phenotype. In contrast, *ITGB2*^{Hi} CD8 T cells had few DE genes, suggesting tissue expansion driven by relocalization and/or bystander activation rather than TCR engagement.

Expanded *IL17A*^{Hi} *CXCL13*^{Hi} effector CD4 T cells and Tregs in irColitis

Subclustering of CD4 T cells demonstrated marked expansion of Th1/Th17 cells (cluster 7) and *FOXP3*^{Hi} Tregs including *SELL*^{Hi} follicular (cluster 5) and activated *TNFRSF4*^{Hi} *TNFRSF18*^{Hi} (cluster 9) Tregs (Fig. 4A–D; Extended Data Fig. 1; Supplementary Fig. 1; Tables S3–4, S6). Th1/17 cells (cluster 7) and *TNFRSF4*^{Hi} Tregs (cluster 9) upregulated genes involved in interferon stimulation, nucleic acid sensing, antigen presentation, cellular metabolism, transcription, apoptosis, and autophagy (Fig. 4E–G; Supplementary Fig. 4; Table S7). Th1/Th17 cells also upregulated Th17 genes (*IL17A*, *IL21*), *IFNG*, and *CXCL13*, which was similarly increased in T follicular helper (Tfh) and cycling cells (Fig. 4G–H; Table S7). No *CXCL13*^{Hi} CD4 T cells expressed *CX3CR1* (Table S4), so these were distinct from T peripheral helper CD4 cells reported in rheumatoid arthritis[33]. *TNFRSF4*^{Hi} Tregs upregulated Th1 genes (i.e., *TBX21*, *IL12RB1/2*, *CXCR3*) and TNF receptor genes (*TNFRSF4*, *TNFRSF18*) that abrogate Treg suppressive functions[34,35] (Fig. 4B,D,G–H; Table S7). Thus, Tregs enriched in irColitis may promote inflammation by dampening Treg immunosuppressive activity while expressing cytotoxic gene programs.

Lastly, compared to patients on anti-PD-1 therapy, patients on dual anti-PD-1/CTLA-4 inhibition had unique effector Th1/17 T cells (cluster 7) that upregulated cytotoxic genes *GNLY* and *GZMA* (Fig. 4I). These cells are akin to cytotoxic CD4 T cells in human cancer and viral infection[36] and are predicted to be pathologic in irColitis.

Validation of irColitis-specific signatures in an independent dataset

Two recent studies detected expanded CD8 T_{RM} in irColitis, however they did not report CD8 *CXCL13, IL17A*, or *IL26* expression nor expanded *ITGB2*^{Hi} *S1PR1*^{Hi} CD8 T cells. To assess this discrepancy with our study, we reanalyzed one dataset[15] using our computational methods. Reanalysis validated two expanded *ITGB2*^{Hi} CD8 T cell populations expressing *EOMES* and *CX3CR1*, respectively, and increased CD8 T *IL17A, IL26*, and *CXCL13* expression in patients with irColitis (Supplementary Fig. 6). While we found increase CD4 T *CXCL13* expression in both datasets, only our dataset showed CD4 T *IL17A* expression (Fig. 4; Supplementary Fig. 6; Table S7). Our ability to detect expanded *CXCL13*^{Hi} T cells,

IL17A^{Hi} *IL26*^{Hi} CD8 T cells, and *ITGB2*^{Hi} CD8 T cells in an independent irColitis dataset supports the reproducibility of these findings.

ISG^{HI} MPs expand in irColitis while B cell subsets remain unaltered

Subclustering of MPs showed three populations enriched in irColitis: *VCAN*^{Hi} inflammatory monocytes (cluster 5), *CD68*^{Hi}*C1QA*^{Hi}*CXCL10*^{Hi} MPs (cluster 6), and *CD68*^{Hi}*C1QA*^{Hi} MPs expressing cathepsin genes (cluster 2), which are upregulated by macrophages in IBD[37] (Extended Data Fig. 1, 5; Supplementary Fig. 1; Tables S3–4, S6). As in IBD[31], conventional *CLEC10A*^{Hi} dendritic cells (cDC2) were underrepresented in irColitis cases. Most DE MP genes in irColitis were ISGs (Extended Data Fig. 5; Table S7). RNA-ISH confirmed that the ISGs *CXCL10/11* are expressed predominantly by panCK^{-} cells in the irColitis LP (Fig. 3E–F), likely corresponding to *CXCL10*^{Hi} *CD68*^{Hi} cells (cluster 6). MPs also upregulated *CD274* (i.e., PD-L1), MHC-I genes, and genes involved in TCR co-stimulation (*CD80*) (Extended Data Fig. 5; Table S7), reflecting highly activated MPs with complex predicted effects on TCR signaling.

Contrary to MPs, tissue and blood B cells had few differentially abundant populations and DE genes in cases versus controls (Extended Data Fig. 1,6; Supplementary Fig. 1,7; Tables S3–4, S7). Unlike in IBD – where plasma B cells preferentially class switch to IgG[38] – we observed no difference in IgG-to-IgA plasma cell ratios in cases versus controls (Extended Data Fig. 6). Lastly, BCRs from irColitis tissue and blood showed no differences in diversity compared to controls (Extended Data Fig. 3C–D). Notably, one patient developed irColitis on anti-PD-1 therapy after B cell depletion (Extended Data Fig. 6; Tables S1–2, patient C14*). The presence of expanded *GZMB*^{Hi} *IFNG*^{Hi} CD8 T cells and the absence of colon mucosal B cells in this individual suggests that B cells are not necessary for irColitis development in all cases and that B cells are likely minor disease contributors.

Tissue immune alternations observed in ICI-treated controls

ICI-treated controls were compared to healthy individuals to understand how ICIs impact colon immunity in the absence of irColitis (Extended Data Fig. 2; Supplementary Fig. 1M). Compared to healthy controls, ICI-treated controls were depleted of *KIR*^{Hi} CD8 T cells (clusters 1, 5) defined by Treg transcriptional regulators (*ID3*, *IKZF2*), T cell exhaustion genes (*TIGIT*), and reminiscent of regulatory CD8 T cells described in autoimmunity[39] (Fig. 2A–D; Extended Data Fig. 2). While Treg abundance was not altered by ICI therapy, *ADAR* was repressed in *TNFRSF9*^{Hi} Tregs (Extended Data Fig. 2). ADAR promotes T cell self-tolerance and protects against colitis[40]. Among MP subsets, tissue *VCAN*^{Hi} inflammatory monocytes were expanded in ICI-controls while blood *CD14*^{Hi} monocytes were relatively depleted, raising the possibility that ICIs promote monocyte recruitment from blood to the colon mucosa, even in the absence of irColitis (Extended Data Fig. 2, S5; Supplementary Fig. 2). Furthermore, exposure to ICIs led to a decline in tissue cDC1 abundance and *HLA-DOB* expression (Fig S2). HLA-DO in DCs increases MHC antigen diversity and prevents autoimmunity[41], so its repression may promote autoimmunity. We suspect that these collective changes in response to ICI treatment may represent early stages of subclinical inflammation. Why these patients did not develop irColitis requires further study.

Inflammatory serum proteins do not track with irColitis status

39 secreted factors were analyzed in the serum of cases versus controls. These included factors with corresponding genes upregulated in irColitis by colon mucosal T cells (CXCL13, IFN-γ, IL-7, IL-10, IL-17α, IL22, CCL3/4) and MPs (GM-CSF, CXCL1/2/8/9/10, IL-1α/β, IL-1RA, IL15, IL27) (Tables S7, S9). None of these factors were differentially detected (Table S9), suggesting that peripheral blood does not readily mirror the tissue inflammatory milieu observed in patients with irColitis.

irColitis is associated with ISG-mediated defects epithelial function

Because epithelial damage is a histologic feature of irColitis[4], we next sought to define disease-associated colon epithelial and mesenchymal alterations. To maximize epithelial recovery, we performed single nuclei RNA-seq (snRNA-seq) from frozen mucosal specimens (Figs. 1, 5A–C; Extended Data Fig. 7; Supplementary Fig. 1; Tables S1–4). In irColitis, we observed a proportional decrease in *LGR5*^{Hi} stem cells (cluster 8) – which strongly upregulated ISGs (i.e., *STAT1*) – absorptive epithelial cells (clusters 5, 6), mature *CEACAM7*^{Hi} epithelial cells (cluster 20), and tuft cells (cluster 19) and expanded transit amplifying cells (cluster 3), suggesting increased epithelial turnover (Fig. 5B, D–G; Table S6).

The most upregulated genes across epithelial and mesenchymal populations from patients with irColitis patients were ISGs *(STAT1, CD274*/PD-L1) (Fig. 5D–H; Table S7), which were more strongly induced by dual PD-1/CTLA-4 blockade than monotherapy (Fig. 5I). Genes specifically upregulated in mature epithelial cells (clusters 2, 11) and goblet cells (cluster 10) were involved in T cell TCR signaling, trafficking, and metabolism (Fig. 5F–G; Extended Data Fig. 7; Table S7). Using a crypt axis score[42], we predicted that the cell types with the strongest ISG signature (clusters 2, 10, 11) resided at the top of epithelial crypts (Fig. 5A,C,F,G). Cluster 20 expressed top crypt markers (i.e., *SELENOP*) but was distinct from clusters 2 and 11 in expressing lower ISGs. Because cluster 20 was markedly depleted in irColitis, we suspect that interferons may induce the transition of cluster 20 cells into cluster 2 and 11 subsets.

irColitis-specific transcriptional programs impacting epithelial barrier function included upregulation of *CEACAM* adhesion genes and down-regulation of aquaporin genes and solute carrier genes associated with IBD (*SCL22A5*, *SLC22A23*)[43,44], which may impair colonic water absorption (Fig. 5F–G; Extended Data Fig. 7; Table S7). Interferons can downregulate *AQP* genes[45], which is consistent with high ISG and low *AQP* expression observed in top crypt epithelial clusters. In line with clinical histology[3,4], irColitis epithelial cells also displayed apoptotic gene signatures (*TNFRSF10A*, *CASP1*, *CASP8*).

To further assess the epithelial ISG crypt-axis gradient, we used protein immunofluorescence to detect PD-L1, which is strongly induced by interferons[46]. PD-L1 was readily detected at the top but not bottom of epithelial crypts in patients with irColitis (Fig, 5H; Supplementary Fig. 8–9) and in the LP. Our scRNA-seq demonstrated that *CD274* was strongly upregulated in ISG^{Hi} MPs (Table S7), suggesting that LP PD-L1 likely derives from CD68^{+} MPs, which in irColitis localize near the lumen (Fig, 5H;

Supplementary Fig. 8–9). Interferon signaling in top-crypt epithelial cells can induce both immune chemoattractants and factors that attenuate immune responses, such as *CD274*/PD-L1 and MHC genes. Together these data support that PD-1/CTLA-4 inhibition may lead to impaired interferon-induced immune tolerance programs and luminal epithelial barrier dysfunction.

Endothelial and fibroblast remodeling in irColitis

Most mesenchymal populations, including endothelial cells (cluster 13) and myofibroblasts (clusters 21,23) were more abundant in cases (Fig. 5A–B). Given epithelial upregulation of hypoxia and angiogenesis genes (*HIF1A*, *SERPINE1*, *VEGFA*) (Extended Data Fig. 7; Table S7), low oxygen tension may drive neovascularization in this context. Both *HIF1A* expression in epithelial cells and endothelial abundance were notably increased in patients on dual anti-PD-/CTLA-4 therapy versus monotherapy (Fig. 5I–J), though it is unclear if vasculogenesis contributes to or ameliorates irColitis.

Unlike in IBD[31], *CCL19*^{Hi} inflammatory fibroblasts (cluster 23) were not more abundant in irColitis (Fig.5B). However, top crypt fibroblast cluster 7 (*PDGFRA*, *BMP4/5/7*) upregulated genes implicated in fibroblast (*FGF7*) and epithelial (*BMP7*) differentiation, *OSMR*, which is upregulated by fibroblasts in IBD,[31] and *ROBO1*, which activates intestinal stem cell autophagy[47] (Fig. 5A,E; Extended Data Fig. 7; Tables S4,S7).

Putative roles for ICAM and CXCR3 ligands in irColitis

Given the central role of PD-1/PD-L1 in tolerance[48] and that most patients were on PD-1 inhibitors (Fig. 1B), we mapped interactions between *PDCD1*-expressing cells and cells expressing PD-1 ligands *CD274*(PD-L1) or *PDCD1LG2*(PD-L2), respectively. Results predicted that *PDCD1*-expressing T cells interact with *CD274*-expressing epithelial cells and *CD274/PDCD1LG2*-expressing MPs (macrophages and monocytes but not DCs) (Extended Data Fig. 5G, S8). Given elevated PD-L1 at the top of epithelial crypts in irColitis (Fig. 5H; Supplementary Fig. 8–9), PD-1 inhibition may preferentially disrupt PD-1/PD-L1 interactions at the top of epithelial crypts, leading to localized unmitigated cytotoxic T cell responses and type 2 interferon release.

We next analyzed 1,826 curated gene pairs[49] to predict putative cell-cell interactions based on ligand-receptor gene pair expression and identified 431 DE ligand-receptor pairs enriched in irColitis (Extended Data Fig. 8–9; Tables S10–11). We first examined *ITGAE*^{Hi} *GZMB*^{Hi} and *ITGB2*^{Hi} CD8 T cells (Fig. 2B), focusing on genes highly upregulated in MP, epithelial, and/or mesenchymal populations and T cell secreted factors (*CXCL13*, *IL26*, *IL17A*) (Fig. 3A, 4E; Extended Data Fig. 8C–D). *ITGB2–ICAM1/2/3* pairs were predicted to facilitate interactions between *ITGB2*^{Hi} CD8 T cells (clusters 3 and 11) and diverse mucosal cell types. In contrast, CXCR3 chemokine ligands/receptor pairs predicted limited interactions between MPs or top-crypt epithelial cells and *ITGAE*^{Hi} *GZMB*^{Hi} T_{RM} or *ITGB2*^{Hi} CD8 T cells (Extended Data Fig. 8). Similarly, Th17 cytokines *IL17A* and *IL26* were predicted to promote crosstalk between only *ITGAE*^{Hi} *GZMB*^{Hi} T_{RM} (but not *ITGB2*^{Hi}) CD8 T cells and the colon mucosa (Extended Data Fig. 8C–D). Interestingly, the per-patient abundance of exhausted, *IL17A*-expressing *HAVCR2*^{Hi} CD8 T cells (cluster 5) strongly correlated with

SPINK5^{Hi} *CCL20*^{Hi} epithelial cell (cluster 18) abundance only in cases (Fig S8F; Table S12). *SPINK5* is induced by IL-17 and may reinforce barrier function[50]. Because IL-17A blockade can worsen IBD[32], we speculate that IL-17A may curb cytotoxicity in irColitis.

We also identified 166 (of 1,826) significantly differential ligand-receptor pair interactions between all epithelial/mesenchymal cells and each tissue immune lineage in irColitis (Extended Data Fig. 9A–B). Several gene pairs were predicted to promote T cell endothelial homing (i.e., CX3CR1/CX3CL1). Furthermore, *ITGB2*^{Hi} *CX3CR1*^{Hi} CD8 T cell abundance was strongly correlated with *CX3CL1*^{Hi} endothelial cell abundance (Extended Data Fig. 8G; Table S12), suggesting possible interaction between these cell types. Lastly, both T and MP cells from irColitis cases upregulated cytokine/chemokine receptor genes that induce TLSs (*LTA/B, CCL19/20*).

Collectively, these results predict important roles for ICAM and CXCR3 ligands in the recruitment and retention of colon mucosal CD8 T cells in irColitis. Whether individual T cells are activated or suppressed in irColitis likely reflects the spatial distribution of TCR co-inhibitory/stimulatory ligands and homing ligands in the local tissue microenvironment.

Identification of putative drug targets to treat irColitis

We next defined cell-type expression patterns of ligand-receptor pairs disrupted by FDA-approved IBD medications and medications in clinical trials (Fig. 6A–B; Table S13). *TNF, ITGA4*, and *ITGB7* were particularly enriched in lymphocytes, while *ITGA4*/*ITGB7* ligands were additionally expressed in fibroblasts (e.g., *FN1*) and endothelial cells (e.g., *MADCAM1*), consistent with vedolizumab disrupting immune/mesenchymal interactions (Fig. 6B; Table S4). The primary targets of the JAK inhibitor tofacitinib, *JAK1/3*, were enriched in T cells, despite broad *JAK* expression (Fig. 6B). S1P receptors, which are modified by ozanimod treatment for IBD, showed striking enrichment in *ITGB2*^{Hi} CD8 T cells (cluster 3, 11), *SELL*^{Hi} *KLF2*^{Hi} CD4 T cells (cluster 1), follicular B cells (cluster 2), and endothelial cells (cluster 13) (Fig. 6B). Furthermore, endothelial cells expressed *CX3CL1* (Fig. 6B–C), which has impaired binding to CX3CR1-expressing immune cells in the presence of E6011[51]. Several drug targets now being tested in IBD trials (i.e. *HIF1A*, *NLRX1*) were particularly enriched in epithelial and mesenchymal cells (Fig. 6B–C).

We also identified putative therapeutic targets of FDA approved medications for non-colitis indications, including PCSK9 inhibitors, which decrease low density lipids[52] and may suppress T cell activation[53]. Compared to controls, patients with irColitis had higher *PCSK9* in colon epithelial cells and a strong correlation between epithelial *PCSK9* and CD8 T *LDLR* (Fig. 6D–F; Extended Data Fig. 9; Supplementary Fig. 10; Table S11). Given differential effects of PCSK9 inhibition on attenuating colitis[54] and promoting anti-tumor responses[53], future studies will have to determine if inhibiting versus promoting this ligand-receptor interaction could treat irColitis.

Distinct T cell signatures in irColitis tissue versus tumors

Given the strong upregulation of *CXCL13*, *IL17A*, and *IL26* by CD8 T_{RM} cells in irColitis, we queried a published cancer dataset[55] to determine the frequency of this signature in tumor-infiltrating CD8 T cells. While most tumor-infiltrating CD8 T cells expressed

CXCL13 (Fig. 6G), *IL17A* and *IL26* expression was limited and essentially absent from melanoma (Fig. 6G). Because IL-17 blockade may worsen IBD[56] and IL-26 can ameliorate murine colitis[57], we speculate that the IL-17/IL26 pathways may differentially impact tumor immune responses and irColitis.

We lastly examined Tregs given their importance in ICI anti-tumor responses and IBD. Both irColitis and tumor Treg populations expressed Th1 genes, including *CXCR3* (Fig. 6H). However, compared to Tregs in irColitis, tumor-derived Tregs had increased *TNFRSF9*, which is associated with poor cancer survival and its inhibition improves anti-tumor responses[58]. Whether Treg-depletion has comparable impacts on anti-tumor immune responses and irColitis will require future studies.

Discussion

The rigorous study of organ-specific IRAEs is essential for decoupling ICI efficacy and toxicity, given their intrinsic correlation[59]. Our study revealed that irColitis is defined by colon mucosal enrichment of recirculating and intravascular *ITGB2*^{Hi} CD8 T cells, *ITGAE*^{Hi} CD8 T_{RM} cells expressing *IL17A, IL26*, and *CXCL13*, CD4 T cells expressing *CXCL13* and *IL17A*, and a strong epithelial ISG signature (summarized in Extended Data Fig. 10). Since mice do not express IL26 and mouse T cells do not express appreciable CXCL13[60], this inflammatory milieu is likely unique to the human colon.

In melanoma, *CXCL13*-expressing T cells recognize tumor neoantigens and their frequency correlates with overall survival[61]. CXCL13, which promotes TLS formation and co-localizes with B cells in some but not all tumor microenvironments[62,63]. Further work is required to determine whether *CXCL13*-expressing T cells contribute to TLSs shaping irColitis immune responses. Our data showed that no B cell subset correlated with irColitis status and we observed colitis develop in a B cell-depleted patient, supporting that B cells are minor contributors to irColitis pathogenesis. Noteworthy, the co-expression of Th17 cytokines *IL17A* and *IL26* by *CXCL13*-expressing CD8 T cells from patients with irColitis distinguishes them from *CXCL13*-expressing T cells in most cancers including melanoma. This transcriptional program was enriched in *ITGAE*^{Hi} CD8 T_{RM} cells with exhaustion features (e.g., *HAVCR2*), similar to those in exhausted CD8 T cells described in melanoma and chronic viral infection[22,23]. Future studies will need to determine if this intestinal CD8 T cell subset arises through chronic TCR stimulation and is functionally exhausted or if it represents a more canonical memory T cell population.

As in irColitis, tissue CD8 T cells from patients with IBD also upregulate IL-17 and IL-26[37], yet the contribution of these Th17 cytokines to inflammatory gastrointestinal diseases is unclear. IL-26 attenuates a mouse model of colitis[31,57] and anti-IL17 can worsen pre-existing IBD and the risk of *de novo* IBD, raising the possibility that IL-17 is protective against human colitis[32,56]. IL-17 notably decreases intestinal permeability, promotes epithelial healing, and protects against epithelial infections[64–66]. We speculate that the intestinal microbial environment, which is correlated with ICI anti-tumor immune responses and irColitis[67,68], may be an important driver of the CD8 T cell Th17 gene

program. Future studies will need to assess how Th17 cytokines function distinctly in IRAEs versus tumors.

IrColitis was marked by the expansion of recirculating *ITGB2*^{Hi}*GZMK*^{Hi} and *ITGB2*^{Hi}*CX3CR1*^{Hi} intravascular effector memory populations. This is consistent with human intestinal transplant studies demonstrating that mucosal *GZMK*^{Hi} *KLRG1*^{Hi} and *ITGB2*^{Hi} CD8 T_{RM} cells are more frequently repopulated from blood than *ITGAE*^{Hi} CD8 T_{RM} populations, and IBD studies showing that *GZMK*^{Hi} *ITGB2*^{Hi} CD8 T cells have increased TCR sharing with circulating CD8 T cells[20,69]. *ITGB2*^{Hi} CD8 T cells were significantly more abundant in patients on dual anti-PD-1/CTLA-4 (i.e., more likely to develop colitis[1]) than those on anti-PD-1, suggesting a potentially pathogenic role in irColitis. Furthermore, *ITGB2*^{Hi}*CX3CR1*^{Hi} CD8 T effector memory populations exhibited gene signatures of cytotoxic anti-viral immune responses[70], and their prevalence in tumors correlates with anti-tumor immune responses[71]. Our study nominated several existing and novel IBD therapies (anti-α4β7, S1PR modulators, anti-CX3CL1) that are predicted to prevent circulating *ITGB2*^{Hi} CD8 T cells from homing to the inflamed colon mucosa and may be effective in irColitis while minimizing anti-tumor immune effects.

While two recent irColitis scRNA-seq studies also described the presence of cycling T cells populations, expanded Tregs, and a strong IFNG signature associated with irColitis,[14,15] neither study reported the presence of expanded *ITGB2*^{Hi} CD8 T cell populations from circulation or increased CD8 T cell expression of *IL26, IL17A*, or *CXCL13*. Nonetheless, our re-analysis of the CD4 and CD8 T cell dataset from the Luoma et al. irColitis cohort validated the dynamic transcriptional upregulation of *IL17A, IL26*, and *CXCL13* in *ITGAE*^{Hi} CD8 T_{RM} cells and increased tissue recruitment of *ITGB2*^{Hi} CD8 T cells in irColitis, further supporting the replicability of our findings. Our study provides complementary insights to these two reports[14,15] and further defines irColitis as a disease marked by altered global interactions between tissue and circulating immune cells with colon mucosal epithelial and mesenchymal cells.

Luminal epithelial cells strongly upregulated ISGs in irColitis. We suspect that many of these irColitis-associated epithelial transcriptional programs are the result of immune-mediated tissue damage and not the direct effect of ICI therapy, as many of these changes are observed in patients with IBD[72,31]. Our study found that several novel IBD therapies primarily acting on colon epithelial and mesenchymal cells may be potential targets to treat irColitis without abrogating anti-tumor immune responses. These include therapies promoting epithelial barrier function (e.g.,HIF-1α stabilizers) and novel, intra-luminal small molecules inhibiting the transcription factor NRLX1 and the E3 ubiquitin ligase Pellino[73–75]. These are attractive therapeutic options given the differences in epithelial cells in colon inflammation versus cancer and/or their predicted intra-luminal action that would have minimal impact on extraluminal tumors.

Our study had several limitations. Because scRNA-seq was performed using the 10X Genomics platform, granulocytes were not readily captured and so were not represented in our analysis of immune cells from tissue or blood. Furthermore, because most patients underwent flexible sigmoidoscopy, tissue biopsies were from the left colon. Future studies

will need to determine if there is regional heterogeneity in how irColitis presents across the length of the large bowel. Lastly, this study was underpowered to detect immune features of steroid-sensitive versus steroid-resistant irColitis.

The incidence of IRAEs will likely continue rising in the years to come as more patients become eligible for ICI therapy and new checkpoint therapies are introduced into clinical care. In defining detailed interaction networks between immune, mesenchymal, and colon epithelial cells, this study provides a comprehensive framework to dissect the pathologic underpinnings of IRAEs, identify tailored therapies that maintain anti-tumor immunity, and begin understanding which patients may be at risk of developing organ-specific IRAEs.

Methods

Patient selection, endoscopic tissue acquisition, blood and serum collection, PBMC isolation

Endoscopic colon biopsies and blood were obtained from patients with irColitis, patients on immunotherapy without colitis, patients with chronic diarrheal symptoms, and healthy controls at Massachusetts General Hospital (MGH). Control patients on ICI therapy had new onset gastrointestinal symptoms that prompted endoscopic evaluation, which revealed normal colon biopsies. Informed consent was obtained from all patients in accordance with protocols obtained from the Mass General Brigham and/or DANA-Farber/Harvard Cancer Center Institutional Review Boards (DFCI/HCC 11–181 and 13–416, Mass General Brigham 2015P001333). This consent included consent to publish indirect patient identifiers such as age, sex, and patient-identified race. Patients included both males and females spanning ages 30 – 81. Patient sex was obtained from electronic health records, which reflected a combination of self-reported and provider-reported sex. Patients were excluded if they were on immunosuppressive medications (other than chronic low-dose prednisone ≤ 10 mg daily for a non-colitis indication), antibiotics, had clinical evidence of an active infection including infectious colitis, biopsy-proven ischemic colitis, or pre-existing microscopic colitis, Crohn’s disease, or ulcerative colitis. A clinical pathologist confirmed all cases of irColitis. Clinical metadata for each patient can be found in Tables S1–2. Tissue biopsies were obtained during colonoscopy or flexible sigmoidoscopy with standard large endoscopic biopsy forceps and were taken from the descending colon, sigmoid colon, and rectum (unless otherwise indicated, Table S2). Because disease severity can differ between the ascending and descending colon, biopsies from all patients with irColitis were taken from the descending colon, sigmoid, and rectum and analyzed in aggregate[76,77]. Identically sized biopsy forceps were used to take biopsies from all patients, so all endoscopic biopsies were presumed to be the same size. For each patient, 6 tissue biopsies from each endoscopic region were placed in HypoThermosol medium (Stemcell Technologies 07935) and were transported on ice to the laboratory for further processing.

Blood was collected from patients undergoing colonoscopy no more than 60 minutes before their endoscopic procedure when tissue biopsies were collected. For selected patients with irColitis, follow-up blood samples were collected at an indicated time for routine clinical care. Blood and serum were collected in EDTA and heparin-coated serum vacutainer tubes (BD 366643, 668660), respectively, and processed within 3 hours of collection. PBMCs

were isolated from EDTA collection tubes using Ficoll-Paque density gradient centrifugation and frozen at −80°C in a 1:1 mixture of cells suspended in RPMI with 2% (v/v) human AB serum and CryoStor (BioLife Solutions 210102). Serum vacutainers were positioned upright for at least 30 minutes at room temperature, centrifuged at 2100 *g*, and supernatant was stored at −80° C.

Preparation of single cell suspensions from colon biopsies

At least 6 endoscopic forcep biopsies collected from multiple subregions of the colon were pooled together in equal proportion into phenol-free RPMI media (ThermoFisher 11835055) containing 1 mg/ml cold active protease from Bacillus licheniformis (MilliporeSigma P5380), 5 mM CaCl2, and 0.1 mg/ml DNAse I (MilliporeSigma DN25) for tissue digestion at 4°C as described[78]. Biopsies were cut into ≤ 1 mm pieces with standard laboratory tissue dissection scissors and incubated on a spinning rotor at 4°C for 40 minutes with tissue resuspension every 10 minutes using a P1000 pipette. Reaction was quenched by adding 10% (v/v) human AB serum (MilliporeSigma H4522). Digestion products were then passed through a 70 μM nylon mesh filter to give uniform cell suspensions, centrifuged at 350 g for 10 minutes, and resuspended in phenol-free RPMI. Cells were counted with a hemocytometer yielding at least 200,000 viable cells/biopsy from patients with irColitis and at least 50,000 viable cells/biopsy from patients without colitis. A portion of the cell suspension was then aliquoted to further perform $CD45^+$ cell selection.

Tissue biopsy flow cytometry, immune cell isolation, and preparation of tissue immune cells for 10x scRNA-sequencing profiling

Colon mucosal $CD45^+$ immune cells were isolated from single cell suspensions using either bead-based selection or sorting by flow cytometry (Table S14). For bead-based selection, single cell suspensions of enzymatically digested colon mucosal biopsies were depleted of dead cells using Annexin V conjugated beads according to manufacturer's protocol (Stemcell 17899). Red blood cells were then depleted from the suspension using antibodies directed against glycophorin A according to the manufacturer's protocol (Stemcell 18352). $CD45^+$ cells were then positively selected using CD45 antibodies conjugated to magnetic beads (Biolegend MojoSort CD45 nanobeads 480029). Viable cells were counted on a hemocytometer. At least 20,000 live immune cells were isolated from each patient for downstream analysis.

For selection of $CD45^+$ cells by flow cytometry, single cell suspensions of enzymatically digested biopsies were brought up in phenol-free RPMI with 2% (v/v) human AB serum and were incubated on ice for 30 minutes with the following antibodies: CD66b-FITC (1:100, Biolegend 305104), EpCAM-PE (1:100, Biolegend 324206), CD45-APC (1:150, Biolegend 304012), CD3 PerCP-Cy5.5 (1:150, Biolegend 300328), and CD235a PE-Cy7 (1:150, Biolegend 349112). Cells were washed once and resuspended in phenol-free RPMI with 2% (v/v) human AB serum containing DAPI (ThermoFisher 62248). Single-color controls were performed with these same antibodies using BD CompBeads (BD Biosciences 552843). Live (i.e., DAPI negative), singlet $CD66b^-$, $CD235a^-$, $EpCAM^-$, $CD45^+$ cells were sorted into phenol-free RPMI with 2% (v/v) human AB serum (see gating strategy in Supplementary Fig. 11). All sorting was performed on Sony SH800 or MA900 Cell Sorters.

At least 20,000 immune cells were sorted from each patient. Resorting of $CD45^+$ fractions showed >99% sorting purity of $CD45^+$ cells. Data were analyzed using FlowJo 10.7.2 (BD Life Sciences). Isolated $CD45^+$ cells were centrifuged and resuspended at a concentration of ~800 – 1200 cells/μl in RPMI with 2% (v/v) human AB serum in preparation for loading on the 10X Genomics Chromium instrument for downstream single-cell RNA sequencing.

Estimate of immune cells per colon endoscopic biopsy specimen

Multiple (6–16) endoscopic colon biopsies from one individual patient were pooled together and enzymatically digested as detailed above. The average total number of viable cells from each endoscopic tissue biopsy piece was estimated by counting the total number of viable, trypan blue-excluding cells following enzymatic digestion, and dividing by the total number of endoscopic colon biopsies. The frequency of immune cells was determined using the FACS panel detailed above. Immune cell frequency was estimated by determining the percentage of $CD45^+$ $EPCAM^-$ $CD235a^-$ cells among all $DAPI^-$ cells defined by an immune cell size gate. T cell frequency was estimated by determining the percentage of $CD45^+$ $CD3E^+$ $EPCAM^-$ $CD235a^-$ cells among all $DAPI^-$ cells defined by an immune cell size gate. The absolute number of $CD45^+$ immune cells and $CD45^+$ $CD3E^+$ T cells per endoscopic biopsy piece was then estimated by multiplying the immune cell frequency by the total number of viable cells per endoscopic biopsy specimen (Table S5).

PBMC $CD45^+$ and $CD8^+$ T cell enrichment, cell hashing, CITE-seq staining

PBMC samples that were cryopreserved and stored at −80°C from eight patients at the time (for a total of 23 samples; Table S15) were thawed at 37°C, transferred to individual 15 ml conical tubes with 10x volumes of RPMI with 10% (v/v) heat-inactivated FBS (Sigma), followed by centrifugation at 300 × *g* for 7 minutes. The cell pellet was resuspended in FACS buffer [1x PBS (ThermoFisher, 10010023), 2.5% (v/v) heat-inactivated FBS, 2mM EDTA] and transferred to a 96-well U-bottom plate (Corning Costar Assay Plate, 3788). Single cell suspensions were depleted of dead cells and red blood cells using a bead-based Annexin-V-conjugated bead kit (Stemcell 17899) and glycophorin A-based antibody kit (Stemcell 01738), respectively, with modifications made to the manufacturer protocol to decrease sample volume size to 150 μl. Resulting live cells were counted on the Bio-Rad TC20 automated Cell Counter with trypan blue, and 250,000 cells were resuspended in 100 μl of FACS buffer containing TruStain FcX blocker (Biolegend 422302) and MojoSort CD45 Nanobeads (Biolegend 480030). Individually titrated working solutions containing one of eight hashtag antibodies were added to each of the eight respective patient samples (Biolegend; Table S16) and incubated on ice for 30 minutes followed by three magnet-based washes in FACS buffer. Live cells were counted on the Bio-Rad counter with trypan blue, and 60,000 cells from each of the eight patient samples were combined together for a total of 500,0000 cells per pool. The same approach delineated above was employed to hash together $CD8^+$ T cells enriched from each patient, with the main difference that $CD8^+$ enrichment was achieved by incubating 1–2 million PBMCs with CD8 MojoSort nanobeads (Biolegend 480108) as opposed to incubating the cells with the MojoSort CD45 Nanobeads (Biolegend 480030). The remainder of the downstream protocol remained the same.

$CD45^+$ and $CD45^+$ $CD8^+$ cell suspensions respectively containing 500,000 cells from eight patients were subsequently filtered through a 40 μM strainer and centrifuged at 300 × *g* for 7 minutes, placed on a magnet, and resuspended in 25 μl of TotalSeq-C antibody cocktail containing 197 protein targets (Biolegend; Table S17). Cells were incubated on ice for 30 minutes and washed 4 times in 1.5 ml FACS buffer. The cell pellet was resuspended in RPMI with 10% (v/v) FBS, filtered again through a 40 μM strainer, and the live cells were counted and resuspended at a concentration of 1200 cells/μl. A total of 50,000 cells were loaded per 10X channel to generate downstream single-cell gene expression, TCR, BCR, and surface protein CITE-seq libraries (Table S15). CITE-seq protocol was executed following published methods[79].

Single nuclei isolation from colon tissue biopsies

Endoscopic colon biopsies were collected as indicated above. Individual biopsies intended for nuclei isolation were placed into cryo-vials, flash frozen on dry-ice, and stored at −80°C. Nuclei were extracted at a later date using dounce homogenization and lysis in a tween-based buffer as previously described[80]. In brief, flash frozen biopsies were thawed, rinsed in PBS, and dounce homogenized with a 2 ml Dounce Tissue Grinder (Sigma D8938) 20 times with pestle A and 20 times with pestle B. Tissue homogenization was performed in a lysis buffer containing 10% (w/v) tween, ST buffer (146 mM NaCl, 1 mM $CaCl_2$, 21mM $MgCl_2$, 10 mM Tris-HCl pH 9), and RNAse inhibitor (Takara 2313, diluted 1:1000). Homogenate was then passed through a 40 μM filter to remove debris, centrifuged at 500 × *g*, resuspended in ST buffer with RNAse inhibitor, and filtered through a 10 μM filter. Nuclei were counted with bright field microscopy and resuspended at a concentration of 800 – 1200 nuclei/μl in ST buffer in preparation for loading on the 10X Genomics Chromium instrument (Table S18).

Droplet-based scRNA-seq

Single immune cell or nuclei suspensions from colon tissue and hashed $CD45^+$ and $CD45^+CD8^+$ PBMC suspensions were generated as indicated above (Tables S14, S15, S18). An input of 12,000 single cells or nuclei from tissue was loaded to each 10X Genomics channel, aiming for a recovery goal of approximately 4,000 single cells or nuclei. An input of 50,000 hashed PBMC samples was loaded on a single channel in the 10X Chromium instrument for blood immune cell analysis, aiming for a recovery goal of 30,000 single cells. Each cell and nuclei suspensions were loaded along with reverse transcriptase reagents, 3’V2 or 5’V1 gel beads, and emulsification oil onto separate channels of a Single Cell A Chip, which was loaded into the 10X Genomics Chromium Controller instrument to generate emulsions. Emulsions were transferred to PCR strip tubes for immediate processing and reverse transcription. Library preparation was performed according to manufacturer’s recommendations.

All tissue-derived immune cell and nuclei gene expression libraries (Tables S14, S15, S18) were generated with the Chromium Single Cell 3’ (V2, 10X genomics PN-120237) or 5’ (V1, 10X Genomics PN-1000006) kits according to the manufacturer’s protocols. All hashed PBMC single cell libraries were generated with the Chromium Single Cell 5’ (V1, 10X Genomics PN-1000020) together with the 5’ Feature Barcode library kit

(10X Genomics PN- 1000080). For immune cells libraries generated with the Chromium Single Cell 5' V1 kit, PCR-amplified cDNA was used for TCR and BCR enrichment with the Chromium Single Cell V(D)J Enrichment kit (10 Genomics PN-1000005 and PN-1000016). cDNA and library quality were evaluated using an Agilent 2100 Bioanalyzer. All tissue single-cell and single-nuclei gene expression libraries were sequenced on an Illumina Nextseq 500/550 instrument using the high output v2.5 75 cycles kit with the following sequencing parameters: read 1 = 26; read 2 = 56; index 1 = 8; index 2 =0. TCR/BCR-enriched libraries were sequenced on an Illumina MiSeq instrument using the 300 cycles MiSeq kit v2 with the following sequencing parameters: read 1 = 26; read 2 = 91; index 1 = 8; index 2 =0. All hashed PBMC CITE-seq, feature barcode (i.e., ADT and HTO libraries) TCR and BCR libraries were sequenced on an Illumina Novaseq instrument using the S4 300 cycles flow with the following sequencing parameters: read 1 = 26; read 2 = 91; index 1 = 8; index 2 =0.

Luminex

Serum secreted protein measurements were performed using a custom Luminex ProcartaPlex Multiplex kit (ThermoFisher) to detect the following 39 analytes: CXCL13, CCL11, GM-CSF, CXCL1, IFN-α, IFN-γ, IL-1a, IL-1b, IL-1RA, IL-2, IL-4, IL-5, IL-6, IL-7, CXCL8, IL-9, IL-10, IL-12p70, IL-13, IL-15, IL-17A, IL-18, IL-21, IL-22, IL-23, IL-27, IL-31, IL-33, CXCL10, CCL2, CXCL9, CCL3, CCL4, CXCL2, CCL5, SDF-1a, CCL25, TNF-α, and TNF-β. Secreted proteins were detected per manufacturer's protocol. In brief, human serum stored at −80°C was thawed, centrifuged at 1000 × *g*, and all supernatants were diluted 1:2. All samples were aliquoted in technical duplicates to the same plate to avoid batch effects. Standard curves were prepared by serial dilutions of standards corresponding to secreted factor patterns supplied by the manufacturer. Plates were read with a FLEXMAP 3D instrument (Luminex) via xPONENT V3.1 software. Cloud-based Procartaplex Analysis Software V1.0 was used to calculate standard curves and mean fluorescence intensity (MFI) for each detected protein, and analyte concentrations (pg/mL) were averaged across duplicates. We tested for differences in irColitis cases versus controls for each analyte with a two-tailed t-test on log10 (concentration).

Multispectral RNA and Immunofluorescence Microscopy

5μm sections were cut from formalin-fixed paraffin-embedded endoscopically (FFPE) collected colon mucosal tissue blocks onto SuperFrost plus slides (Fisher Scientific 1255015). Slides were incubated at 65°C for 2 hours before use. Protein immunofluorescence staining for CD8A, CD68, FoxP3, PD-1, PD-L1, and PanCK was performed using the Motif PD-1/PD-L1 panel (Akoya Biosciences OP-000001) on a Leica Bond Rx instrument (Leica Biosystems) per manufacturer's protocol. Mixed RNAscope *in situ* hybridization and antibody antigen retrieval and staining was performed with RNAscope probes to *CXCL13*, *CXCL10/11*, *IFNG*, and *CD3E* and antibody to panCK on a Leica Bond Rx instrument using the RNAscope LS multiplex Fluorescent v2 Assay combined with a published immunofluorescence protocol (Advanced Cell Diagnostics (ACD) 322818-TN)[62]. The only two variations from the written protocol were (1) an open wash dispense after the peroxide step and (2) DAPI (Sigma D9542) was dispensed twice at the end of the protocol at a concentration of 1 μg/mL. RNAscope probes included

the following: CXCL13 (RNAscope LS 2.5 Probe-Hs-CXCL13, ACD 311328), CXCL10 (RNAscope LS 2.5 Probe-Hs-CXCL10-C2, ACD 311858-C2), CXCL11 (RNAscope LS 2.5 Probe-Hs-CXCL11-C2. ACD 312708-C2), IFNG (RNAscope LS 2.5 Probe-Hs-IFNG-C3, ACD 310508-C3), and CD3E (RNAscope LS 2.5 Probe-Hs-CD3E-C4, ACD 553978-C4). Immunofluorescence was performed using mouse anti-PanCK AE1/AE3 antibody (Agilent, M3515) and secondary Opal polymer HRP Ms + Rb (Akoya ARH1001EA). RNA probes and panCK antibody were detected with the following Opal fluorophores pairings: CXCL13 Opal 480 (Akoya FP1500001KT), CXCL10/11 Opal 520 (Akoya FP1487001KT), CD3E Opal 620 (Akoya FP1495001KT), IFNG Opal 690 (Akoya FP1497001KT), and panCK Opal 780 (Akoya FP1501001KT). Slides were rinsed with water (Fisher 23–751628) prior to coverslipping (Fisher 12–544C) with mountant (Life Technologies P36961). PCSK9 immunohistochemistry (IHC) staining was performed using a Leica Bond Rx automated stainer. Mouse monoclonal antibody PCSK9-2F1 (Thermo Fisher #MA5-32843) was labeled with DAB chromogen (Leica Bond Polymer Refine Detection DS9800) using citrate based pH 6 epitope retrieval condition and 0.1 μg/ml antibody concentration.

Immunofluorescence and RNAish slides were imaged using a Vectra Polaris microscope (Akoya Biosciences). Raw Vectra Polaris Images for each slide were unmixed with inForm software (Akoya Biosciences) using an algorithm built on a library of fluorescence spectra measured using single fluorophore-labeled control slides. Unmixed multi-layer image TIFFs from single fields of view were then stitched together into a single multilayer TIFF using Halo software (Indica Labs). For PCSK9 IHC staining, whole slide images were acquired at 40x (0.13 μM/pixel) resolution using a MoticEasyScan Infinity digital pathology scanner.

Image Quantification Analysis

We analyzed all RNAish, protein immunofluorescence (IF), and immunohistochemistry (IHC) images using the HALO® image analysis platform (Indica Labs). IF images were analyzed using HighPlex IF v 4.2.3 and FISH-IF modules v 2.1.5. PCSK9 IHC images were analyzed using FISH-IF module v3.1.3. For each patient biopsy specimen, we selected regions of each slide that were representative of the overall tissue architecture. We manually adjusted nuclear segmentation and fluorescence thresholds to reflect the visually identifiable cell separation and staining within an annotation area. For PCSK9 staining, we determined the average pixel coloration for positive cells for each slide within an annotation area. We used Halo software to determine which cells in an annotation were expressing each RNA probe and each protein marker.

For each annotation area, we normalized the total counts for each marker by the total area analyzed in square centimeters (cm^2). To model the relationship between case and expression of a certain marker, we fit a linear model to the normalized counts for each marker with the formula *log2(norm.count+1) ~ group*, where *group* is a categorical variable with two levels Case and Control. We then computed a false discovery rate[81] and reported results with FDR less than 5 percent as statistically significant.

Statistics & Reproducibility

No statistical method was used to predetermine sample size. For the blood samples, cases and controls were balanced across batches. For the tissue samples, we processed them as we collected them. Data distribution was assumed to be normal but this was not formally tested. All R software package names and versions used for data analysis can be found in Table S19.

Pre-processing of single cell and single nucleus RNA-seq data and quality control (QC) filtering steps

We used 10X Genomics Cell Ranger (version 3.1.0) to pre-process raw single-cell RNA-seq data, including demultiplexing FASTQ reads, aligning reads to the human reference genome (GRCh38, version 3.0.0 from 10X Genomics), and counting the unique molecular identifiers (UMIs) to produce a count matrix with one row for each gene and one column for each cell. We used the Terra platform (https://app.terra.bio/) to run Cell Ranger via the workflow script that is part of a collection called Cumulus[82]. We used R scripts to aggregate the data and filter out poor quality cells that had fewer than 500 genes detected (with at least 1 read) or more than 30% of reads from 13 mitochondrial genes (*MT-ND6, MT-CO2, MT-CYB, MT-ND2, MT-ND5, MT-CO1, MT-ND3, MT-ND4, MT-ND1, MT-ATP6, MT-CO3, MT-ND4L, MT-ATP8*). For the blood scRNA-seq data, we demultiplexed hashed data with R, using an expectation–maximization (EM) algorithm to filter out multiplets and assign single cells to the correct sample[83].

Normalizing count data

After excluding poor quality cells, we normalized the sequencing depth of each cell by dividing each cell's counts by the total counts in that cell, resulting in a matrix where the entries represent the proportion of a cell's reads allocated to each gene (i.e., values in the range [0,1]). To estimate a library size for each dataset, we summed the total counts in each cell, and then we took the median as the library size for the dataset. Next, we multiplied the proportions by the library size to get a 'counts per million' or CPM matrix that was normalized for sequencing depth. Finally, we transformed the normalized count matrix with log2(1 + CPM). We referred to this log-transformed quantity in the figures as log2CPM.

Selection of highly variable genes and principal component analysis

Briefly, we selected highly variable genes, centered and scaled each gene, computed principal components, and ran the molecular cross validation (MCV) algorithm to select the number of components for downstream analysis[84].

To identify highly variable genes, we first used the raw counts matrix to compute a mean with *Matrix::rowMeans()* and standard deviation with *proxyC::rowSds()* for each gene, excluding genes that are detected in fewer than 50 cells. Next, we fit a local 2nd degree polynomial regression with *stats::loess()* with the formula *log10(sd) ~ log10(mean)* to model the relationship between mean and variance for each gene's counts. We computed the residual variance for each gene, selected the top 80% of genes with the greatest residual variance, and centered and scaled each gene to have mean zero and unit variance. We used *RSpectra::svds()* to compute the truncated singular value decomposition (SVD) of the scaled

expression matrix. Then, we multiplied the left singular vectors by the singular values to obtain principal component scores for each cell.

To select the optimal number of principal components, we follow the MCV algorithm. For a raw count matrix, we randomly split the counts for each cell, resulting in two count matrices A and B whose sum is equal to the original matrix. With the centered and scaled matrix A', we compute principal components. Then, we reconstructed matrix A' from the top K PCA scores and loadings (UV^T) and estimated a reconstruction loss with the mean squared error. We also reconstructed matrix B' using the PCA scores computed from A' to estimate an MCV loss. We recorded the losses for each choice of top K principal components between 2 and 80. We repeated this procedure three times, resulting in three estimates of the MCV loss for each choice of the top K. Then, we chose the K that minimized the average MCV loss. These K principal components were used in downstream analyses.

Batch integration, cell clustering, and two-dimensional embedding

We used the Harmony algorithm to align PCA scores across batches of data[85] and then we created a nearest neighbor network of cells by connecting each cell to the 50 cells with the least Euclidean distance in the space of the harmonized PCA scores. We ran the Leiden algorithm[86] from the leidenalg Python package on the nearest neighbor network to identify cell clusters (10 iterations). We ran the UMAP algorithm[87] with uwot::umap() to embed the network in two dimensions (spread = 1, min_dist = 0.25).

Marker gene identification, cell annotation, and differential gene expression analysis

After generating single cell transcriptomes from colon immune cells, colon epithelial and mesenchymal cell nuclei, and circulating PBMCs, we used a two-step clustering strategy to define all cell subsets present in colon and blood. First, clustering was performed at low resolution to identify all main cell lineages, then clustering was repeated to define all cell subsets present in a given lineage (Methods). Immune cells were grouped into broad cell types including B cells (*MS4A1*, *JCHAIN*), T cells (*CD3D, CD8A, CD4*), and mononuclear phagocytes (MP) cells (*LYZ*) (Extended Data Fig. 1–2; Table S3–4). Colon mucosal single nuclei were broadly identified as deriving from epithelial cells (*EPCAM, CDH1*), epithelial stem cells (*LGR5, SMOC2*), or mesenchymal cells (*VIM, COL3A1*) (Extended Data Fig. 1; Table S3–4). All clusters captured cells (or nuclei) from most patients, despite patient differences in ICI exposure and underlying cancer status (Extended Data Fig. 1; Table S1–2).

We used complementary strategies to identify sets of marker genes uniquely expressed in each cell lineage and each cell subset. We used *presto::wilcoxauc()* to compute the area under receiver operator curve (AUROC or AUC) for the log2CPM value of each gene as a predictor of the cluster membership for each cluster[85]. We also created a pseudobulk expression matrix[88] for the cells in a cluster for each donor such that the pseudobulk matrix had one row for each gene and one column for each cluster from each patient. We normalized the pseudobulk counts to log2CPM as described above. Then we used *limma::lmFit()* to test for differential gene expression with the log2CPM pseudobulk matrix[89]. We also created another pseudobulk expression matrix by summing over all cells

in a lineage (tissue CD8, tissue CD4, tissue MP, tissue B, tissue non-immune). By creating these two different pseudobulk expression matrices, together with leveraging AUC statistics, we found gene expression differences (AUC ≥ 0.7; pseudobulk FDR < 0.05) at the level of cell lineage as well as at the level of each cell cluster. We then annotated cell subset by using marker genes (i.e., area under the curve (AUC) ≥0.7, pseudobulk FDR <0.05), which we also cross-referenced to published immune transcriptional signatures to make sure we contextualized our cell annotation labels with published literature[31,57,69,90] when relevant.

For the PBMC annotation, beyond the marker gene analysis, we also leveraged the CITE-seq protein profile data output to help annotate cellular subsets. Specifically, we projected the protein-single cell expression data of the 197 candidate proteins (Table S17) on the gene expression-based clustering solution generated from the same set of cells (Extended Data Fig. 1), similarly to what has been done previously[79]. This helped us relate our cell annotation of the immune cell subsets defined in an unbiased way with our scRNA-seq data to cell subsets that have been traditionally defined using surface protein analyzed by FACS.

For the comparison of cases and controls, the linear model fit to each gene is *Gene ~ 0 + sex + case* where *case* is a categorical variable with two levels *Case* and *Control* and *sex* is a binary indicator with levels *M* and *F*. To select the genes with significant differential expression between cases and controls, we require absolute fold-change > 1.5, false discovery rate < 5%, and more than 1% percent of cells expressing the gene.

For the comparison of dual therapy (anti-PD-1 and anti-CTLA-4) versus single therapy (anti-PD1), the linear model fit to each gene is *Gene ~ 0 + case_drug + sex* where *case_drug* is a categorical variable with four levels *ControlNone*, *ControlPD1*, *CasePD1*, *CasePD1CTLA4* and *sex* is a binary indicator with levels *M* and *F*. We then use limma::contrasts.fit() to test three contrasts *ControlPD1 - ControlNone*, *CasePD1 - ControlPD1*, and *CasePD1CTLA4 - CasePD1*. We reported the contrast *CasePD1CTLA4 - CasePD1* in the figures. We excluded spurious genes with low mean expression or detection in less than 2% of cells.

Cell type abundance analysis

We fit a logistic regression model with *lme4::glmer()* to determine whether the cells that belong to a given cluster were enriched or depleted from cases[91]. For each cluster, we defined a boolean vector Y with one value for each cell that is true if the cell belonged to the cluster and false otherwise. Then, we fit a null model $Y \sim 1 + sex + (1|donor)$ where $(1|donor)$ indicates a random effect that gives each donor a different intercept. Next, we fit an alternative model $Y \sim 1 + sex + case + (1|donor)$ where *case* is a boolean vector that is true if the cell belongs to a case donor. We used the likelihood ratio test with stats::anova() to compare the two models and record the *p*-values. We use the Benjamini & Hochberg (1995) method to compute false discovery rate (FDR) for the *p*-values and report results with FDR less than 5 percent as statistically significant. In the figures, we report the odds ratio (OR) and 95% confidence interval for each cluster. For the comparison of controls on anti-PD-1 versus healthy controls, we only used the cells that belonged to control donors. Here, we fit the null model $Y \sim 1 + (1|donor)$. The alternative model is $Y \sim 1 + drug + (1|donor)$ where *drug* is a binary indicator with levels *PD1* and *None*.

For the comparison of dual therapy (anti-PD-1 and anti-CTLA-4) versus single therapy (anti-PD-1), we used only the cells that belong to case donors on single or dual therapy. Here, we fit the null model $Y \sim 1 + sex + (1|donor)$ where *sex* is a binary indicator with two levels *M* and *F*. The alternative model is $Y \sim 1 + sex + drug + (1|donor)$ where *drug* is a binary indicator with levels *PD1* and *PD1CTLA4*.

All results are reported with boxplots showing patient cell type compositions, where each dot represents a patient. Composition of each patient is reported as the percent of cells from a patient in each cell cluster, and box plots show median and interquartile range. Error-bars reported for each analysis indicate logistic regression odds ratio (OR) for differential abundance of case cells for each cell cluster, and unadjusted likelihood ratio test *p*-values are shown.

Spearman correlation of cell type abundances

We computed the Spearman correlation to compare the abundances of pairs of cell clusters across donors. We computed three correlations using all donors, just cases, and just controls. This allows us to identify which correlations between cell types might be observed in all donors and which might be exclusively present in cases but not controls, or in controls but not in cases. For within-case and within-control analyses, we use a threshold of 20% FDR to select cell type pairs that might be differentially correlated. For analysis of all donors, we use a threshold of 5% FDR to select cell type pairs that might be correlated.

Differential expression analysis of sums of gene pairs

We obtained gene pairs from OmniPath[49] with OmnipathR::import_intercell_network() and then filtered to the pairs with the following criteria: consensus_score_intercell_source >= 3, curation_effort >= 1, n_resources >= 4. After filtering, we had 1,827 gene pairs that were likely to represent ligand-receptor pairs. Next, we looped through each pair of cell types (tissue CD8, tissue CD4, tissue MP, tissue B, tissue non-immune), looped through each gene pair, and summed the pseudobulk log2CPM values for the two genes. Finally, we used limma::lmFit() to fit the model $Gene \sim 1 + case$ where *Gene* is the sum of the log2CPM values of the gene pair and *case* is a binary indicator with levels *Case* and *Control*.

Spearman correlation of percent of cells with gene expression

We computed the percent of cells with non-zero expression of each gene for each cell lineage (Tissue CD8, Tissue CD4, Tissue MP, Tissue B, Tissue Non-Immune) in each donor. Then, we computed the Spearman correlation coefficient and *p*-value for each gene pair for each pair of cell types with *Hmisc::corr()*. This tested for correlation across all donors, so the result indicates whether the percent of cells in cell type 1 for gene 1 is correlated with the percent of cells in cell type 2 for gene 2. We used the same list of 1,827 gene pairs as in the previous section.

TCR and BCR diversity analysis

We defined a unique TCR sequence as the concatenation of four components: TCRα V-gene, TCRα CDR3 amino acid sequence, TCRβ V-gene, and TCRβ CDR3 amino acid sequence. Clones missing any of these features were not included in the analyses. Next, we used

alakazam::alphaDiversity() to compute the Hill numbers of order q (effective number of species)[92]. For BCR analysis, we defined a unique BCR clone as the heavy-chain CDR3 amino acid sequence and light chain CDR3 amino acid sequence. Clones that had a recovered heavy chain CDR3 amino acid sequence and either a lambda or kappa light chain CDR3 amino acid sequence were used for analysis.

Analysis of TCR sharing between tissue and blood

To estimate whether TCR clones observed in blood CD8 T cells were enriched or depleted from a particular tissue CD8 T cell cluster (or vice versa), we fit a logistic regression model with lme4::glmer()[93]. For each cluster, we defined a boolean vector Y with one value for each cell that is true if the cell belonged to the cluster and false otherwise. Then, we fit a null model $Y \sim 1 + (1|donor)$ where $(1|donor)$ indicates a random effect that gives each donor a different intercept. Next, we fit an alternative model $Y \sim 1 + tcr\_shared + (1|donor)$ where tcr_shared is a boolean vector that is true if the cell has a TCR that is observed in both blood and tissue cells. We used the likelihood ratio test with stats::anova() to compare the two models and recorded the p-values. We computed a false discovery rate (FDR)[81] for the p-values and reported results with FDR less than 5 percent as statistically significant. In the figures, we reported the odds ratio (OR) and 95% confidence interval for each cluster.

Analysis of TCR sharing between CD8 T cell clusters

We use a permutation strategy to quantify the amount of TCR clones observed in each pair of CD8 T cells from tissue. First, we discard cells without TCRs and focus only on those that have TCRs. For each pair of cell clusters A and B, we record the ratio of the number of unique TCR sequences that were observed in both clusters (i.e., the intersection $A \cap B$), divided by the number of sequences observed in either cluster (i.e., the union $A \cup B$). This is the observed statistic for each cluster pair. Next, within each donor, we shuffle the cell cluster labels for all the cells with TCRs. This shuffling disrupts any potential association between TCRs and cluster labels. With the shuffled data, we again record the ratio of TCRs (intersection divided by union). We repeat this 10,000 times and compute the empirical p-value for the ratio as the number of shuffled iterations that have a statistic equal to or more extreme than the observed statistic. We compute a relative enrichment or depletion of TCR sharing between clusters by dividing the observed statistic by the expected value (the mean of the shuffled statistics).

Crypt-axis score

The crypt-axis score was assigned for each cell and was based on the expression of a previously defined set of genes: *SELENOP, CEACAM7, PLAC8, CEACAM1, TSPAN1, CEACAM5, CEACAM6, IFI27, DHRS9, KRT20, RHOC, CD177, PKIB, HPGD, LYPD8*[42]. The log2CPM expression X_{gc} of gene g in cell c was divided by $Max(X_g)$ across all cells, resulting in a max-normalized value between 0 and 1 for each gene. We sum all of the max-normalized gene values to arrive at the crypt-axis score. The higher the score, the more likely a cell state is predicted to reside at the top of the crypt.

Re-analysis of data from Luoma et al 2020

We downloaded the data files provided by Luoma et al, *Cell*, 2020 (GSE144469) and then performed marker gene identification, cell annotation, and differential gene expression analysis, and cell type abundance analysis as described above.

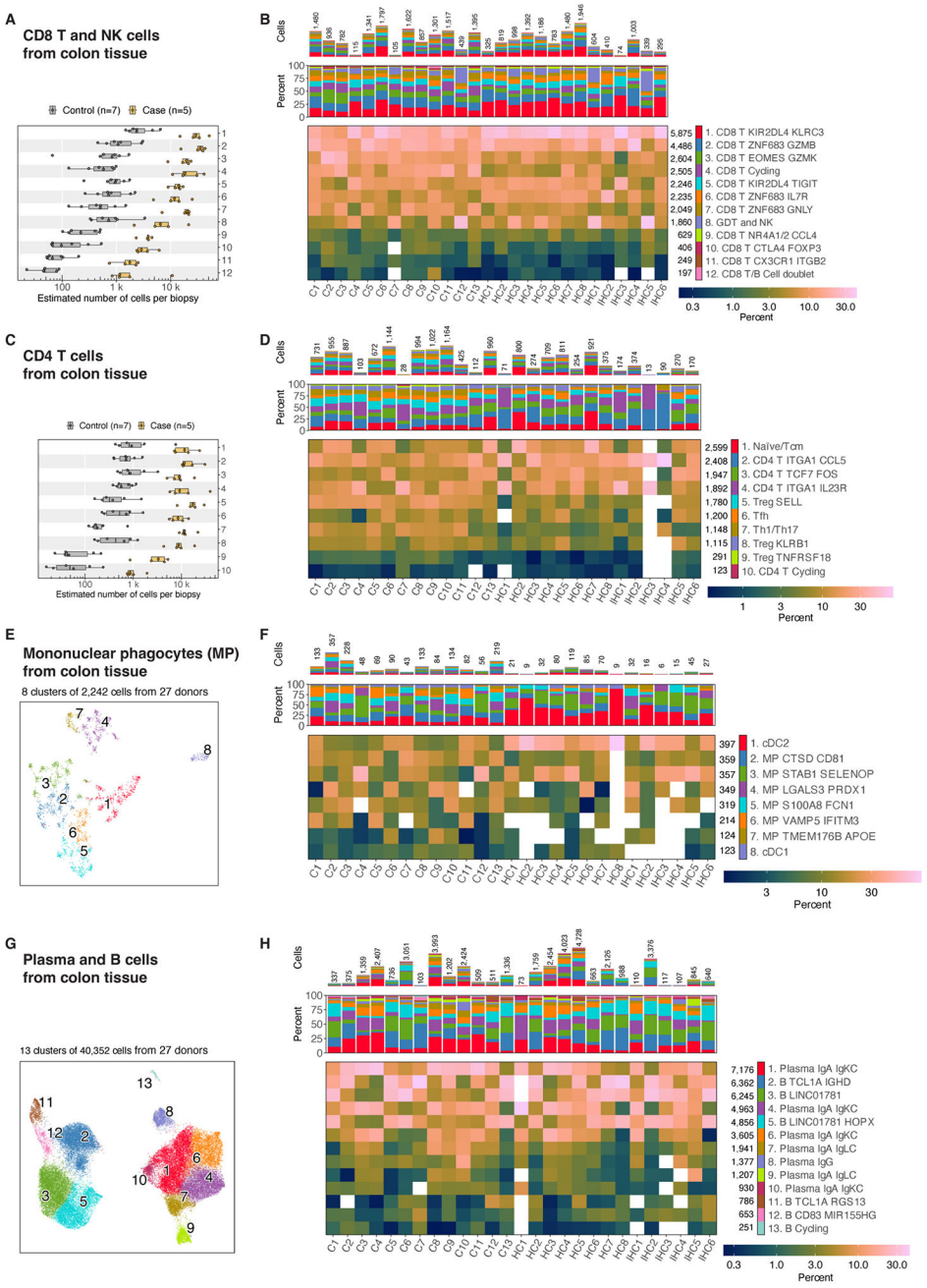
Comparison to T cell reference data from Giles et al 2022

We downloaded the data files provided by Giles et al, *Immunity*, 2022 (GSE179609) and then computed the average gene expression for each of the T cell subsets. Next, we computed the average gene expression for each of the clusters in our dataset. Finally, we computed the Spearman correlation between each pair of cell clusters and T cell subsets.

Comparison to cancer T cells from Zheng et al 2021

We downloaded the data files provided by Zheng et al 2021 (https://zenodo.org/record/5461803) and then computed, for each patient, the percent of cells that have expression for specific genes of interest (e.g. for CD8 T cells, *CXCL13, IL17A, IL26*).

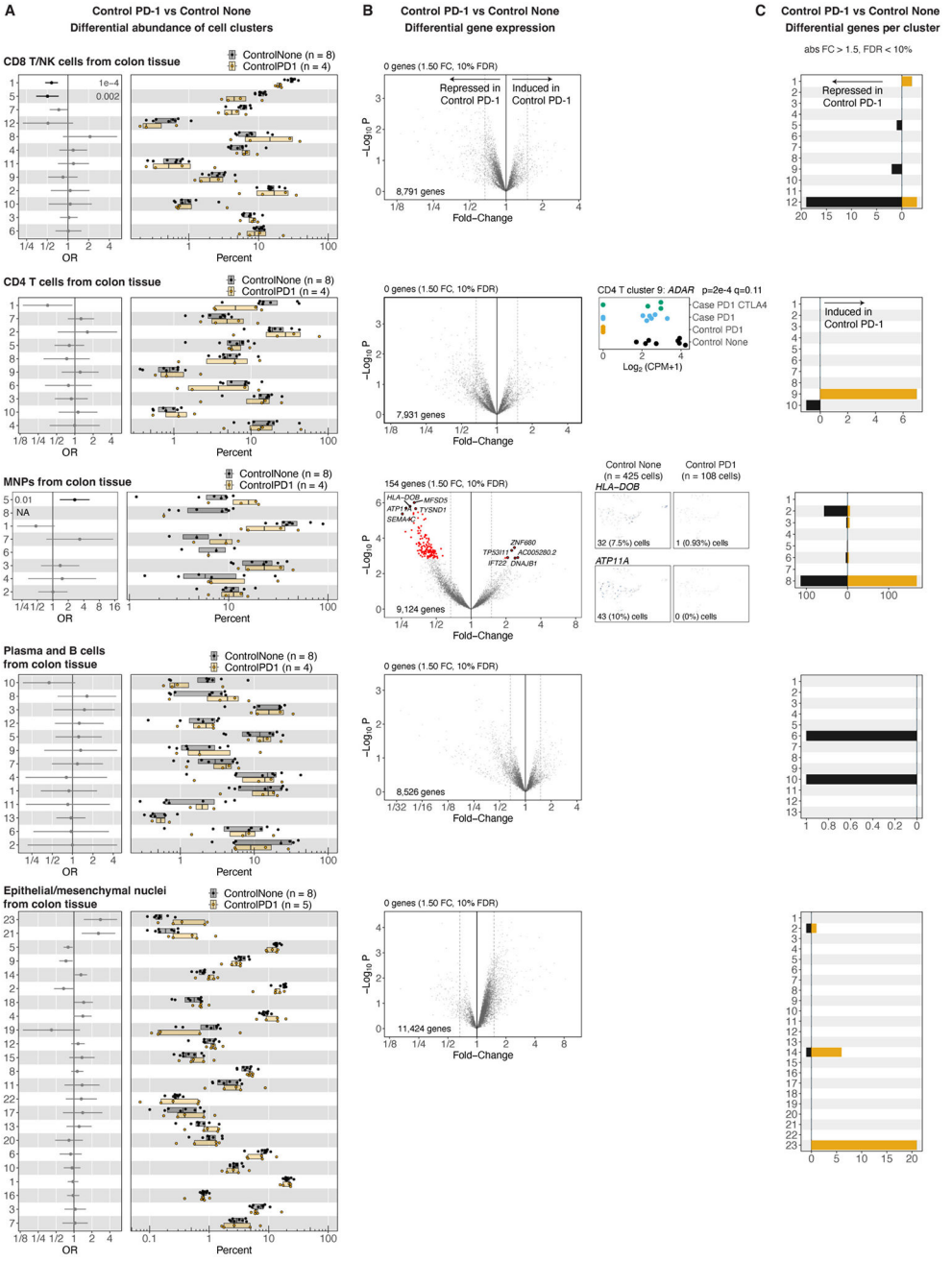
Extended Data



Extended Data Figure 1. Detailed cell type composition of each donor for immune cells from colon tissue.

(A) Estimated number of cells per biopsy in each CD8 T/GDT/NK cell cluster for cases (n=5) and controls (n=7). **(B)** Detailed composition of each patient across cell clusters for CD8 T/GDT/NK cells. **(C)** Estimated number of cells per biopsy in each CD4 T cell cluster for cases (n=5) and controls (n=7). **(D)** Detailed composition of each patient across cell clusters for CD4 T cells. Cell clusters depicted with different colors in UMAP embeddings of the **(E)** eight mononuclear phagocyte subsets and **(G)** 13 B cell subsets from colon tissue.

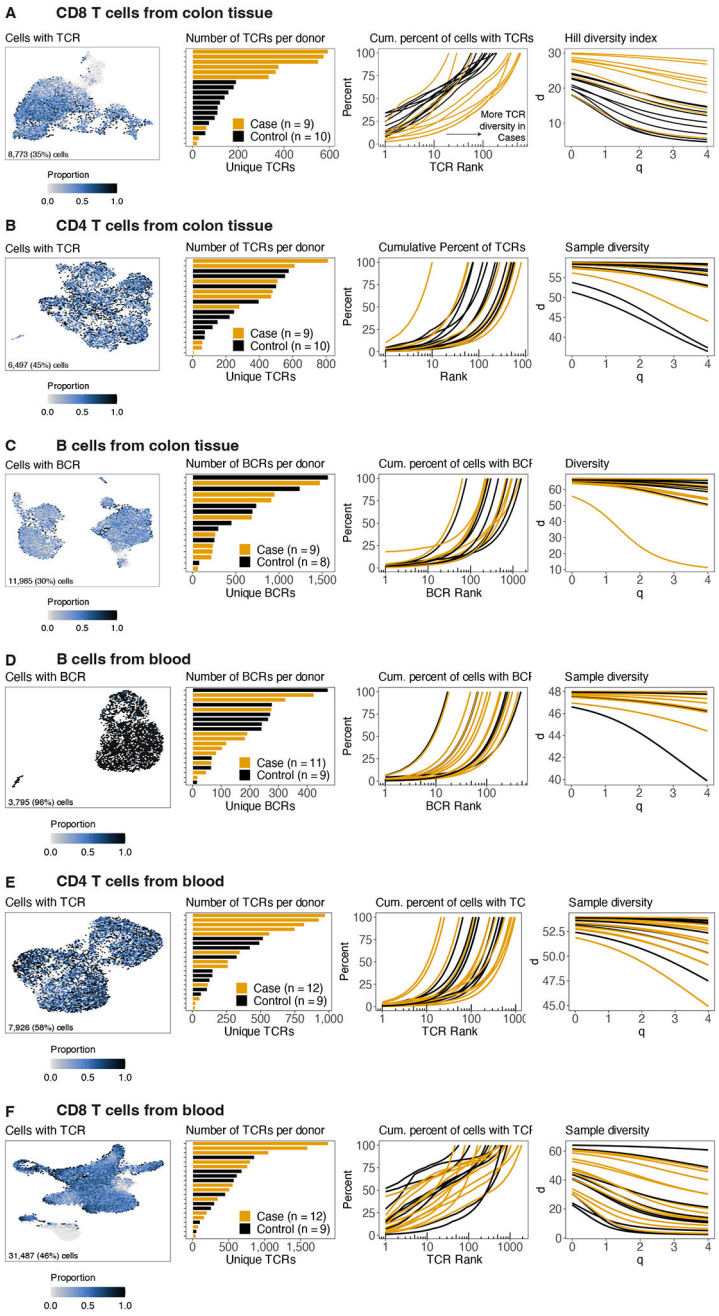
Detailed composition of each patient across cell clusters for **(F)** mononuclear phagocytes and **(H)** B cells from colon tissue. For panels B, D, F, and H, every column represents one individual donor referenced, labeled at the bottom of the heatmap. Boxplots show median and interquartile range. Heatmap color indicates percent of a patient's cells assigned to each cell cluster. NK natural killer. GDT gamma delta T cells. Related to Figures 2–4.



Extended Data Figure 2. Tissue cell type abundance associated with control patients on anti-PD-1 versus no therapy.

(A) Abundance analysis of tissue cell clusters in control patients on anti-PD-1 therapy (n=4) versus healthy controls not on ICI therapy (n=8) (i.e., those undergoing screening

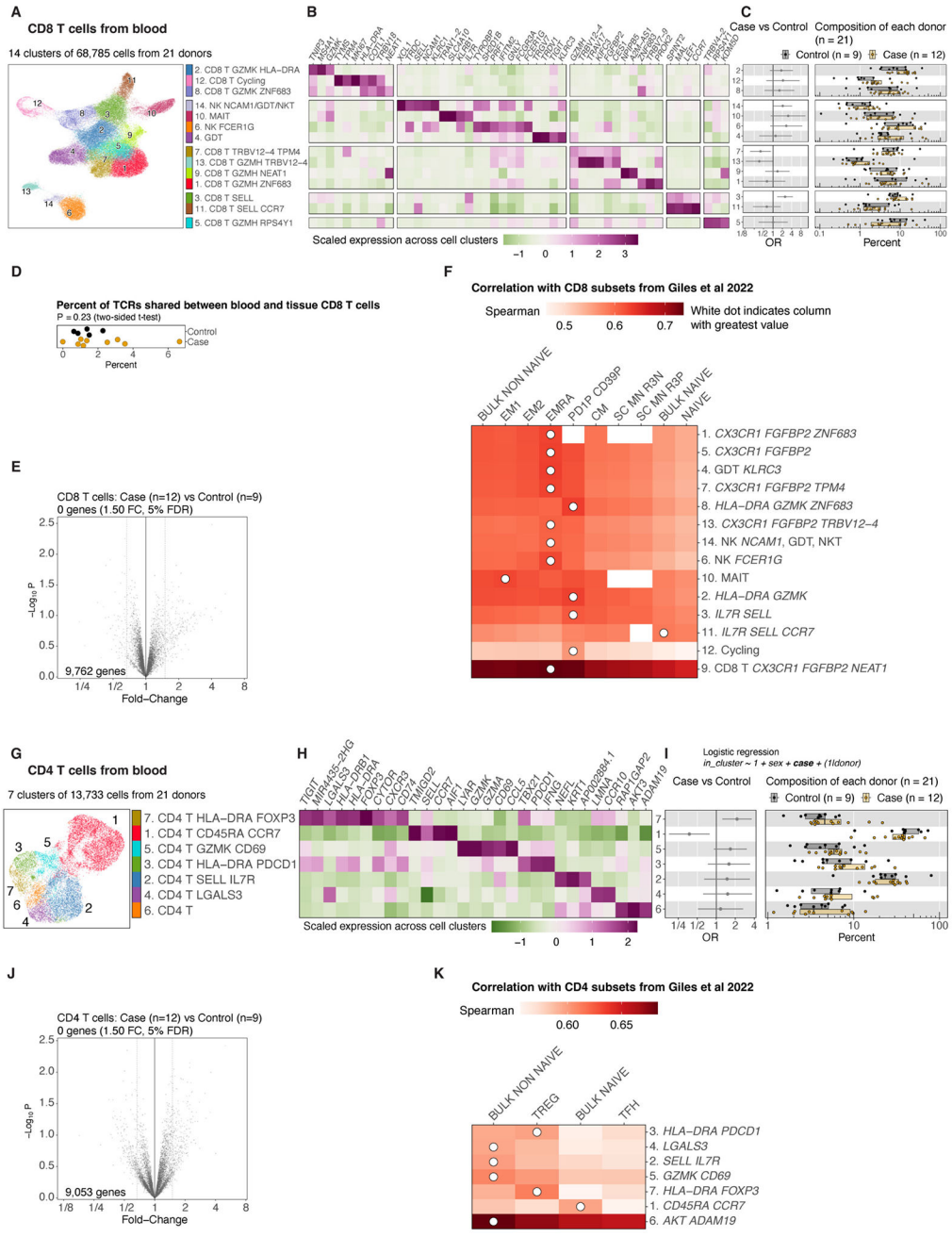
colonoscopy). Each dot represents a patient. Composition of each donor is reported as a percent of cells from each patient in each cluster, and box plots show median and interquartile range. Error-bars indicate logistic regression OR 95% CI for differential abundance of cells from controls on anti-PD-1 versus healthy controls not on ICI therapy for each cell cluster. Unadjusted likelihood ratio test *p*-values are shown. MNP cell cluster 8 was detected in only one patient in the control anti-PD-1 group, so the logistic regression model was not fit for this cluster. **(B)** Volcano plots show log fold-change (x-axis) and negative log10 P-value (two-sided) (y-axis) for each gene. Number of differentially expressed genes (FC > 1.5 and FDR < 10%) is shown, and some of the top genes are labeled. *ADAR* expression is shown for cluster 9 CD4 Tregs. Dots represent individual patients. Feature plots for *HLA-DOB* and *ATP11A* expression in MP cells is shown. **(C)** Number of differentially expressed genes in each cell cluster (FC > 1.5 and FDR < 10%).



Extended Data Figure 3. TCR and BCR detection and diversity analysis for immune cells from colon tissue and paired blood specimens.

UMAP embedding with color depicting cells with TCR (**A, B, E, F**) or BCR (**C, D**) are shown (left most column) for **(A)** CD8 T cells from colon tissue, **(B)** CD4 T cells from colon tissue **(C)** B cells from colon tissue **(D)** B cells from blood, **(E)** CD4 T cells from blood, and **(F)** CD8 T cells from blood. TCR (**A, B, E, F**) or BCR (**C, D**) diversity plots are shown in the remaining three columns including the number of distinct clones per patient (second column from the left), cumulative percent of cells with the top N unique TCR or BCR clones

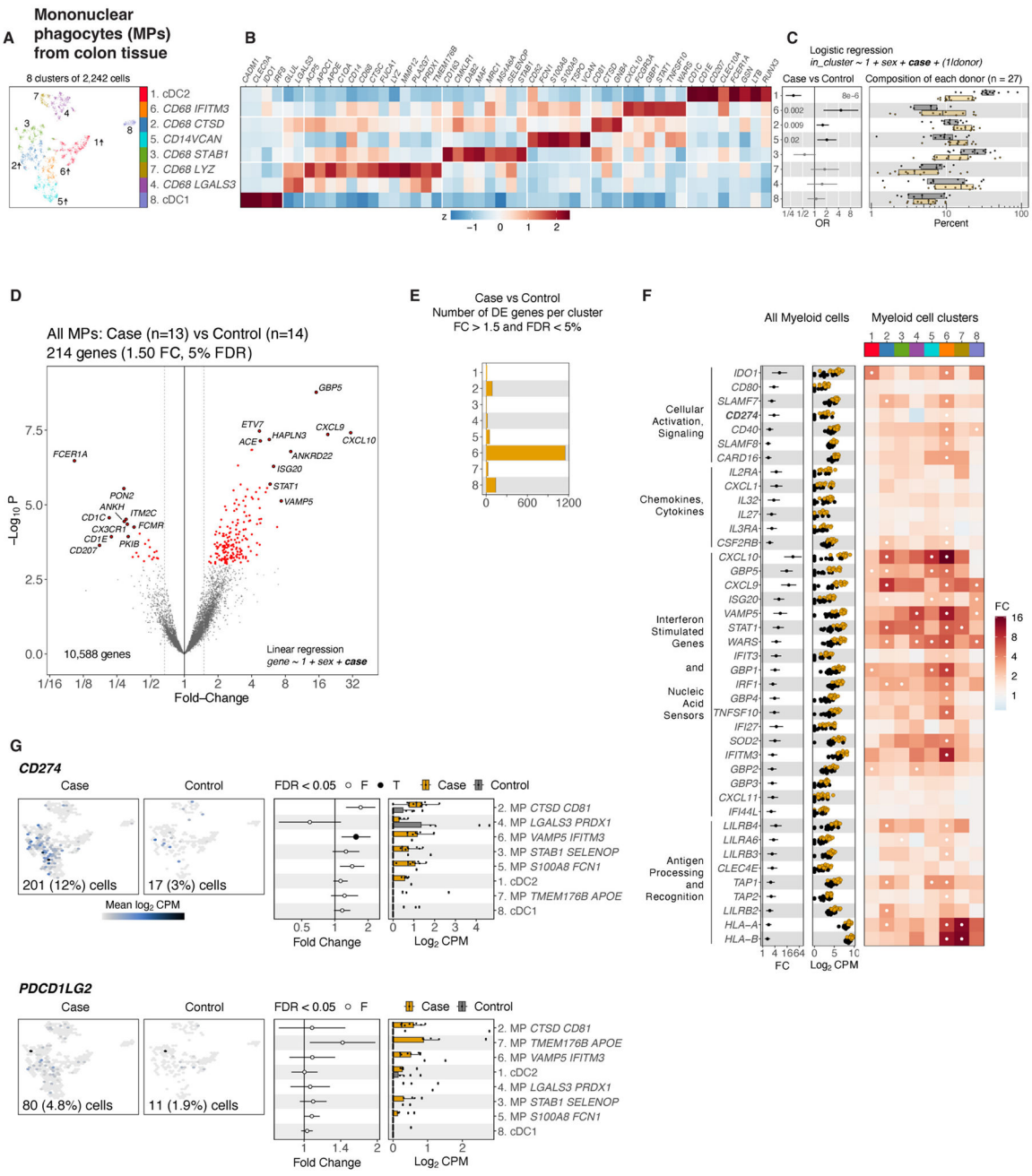
(second column from the right), and Hill diversity index (right most column). Related to Figures 2 and 4.



Extended Data Figure 4. Detailed analysis of CD8 T/Gamma Delta T/NK cells and CD4 T cells from blood

(A) UMAP embedding, **(B)** normalized gene expression, and **(C)** cell subset abundance difference between irColitis cases and controls for CD8 T/GDT/NK cells. **(D)** Percent of alpha beta CD8 T cell TCR clones shared between paired blood and tissue for each individual patient. Dots represent patients. irColitis case versus control two-sided t-test *p*-value is shown. (**E**) Volcano plot of pseudobulk differential gene expression with all cells

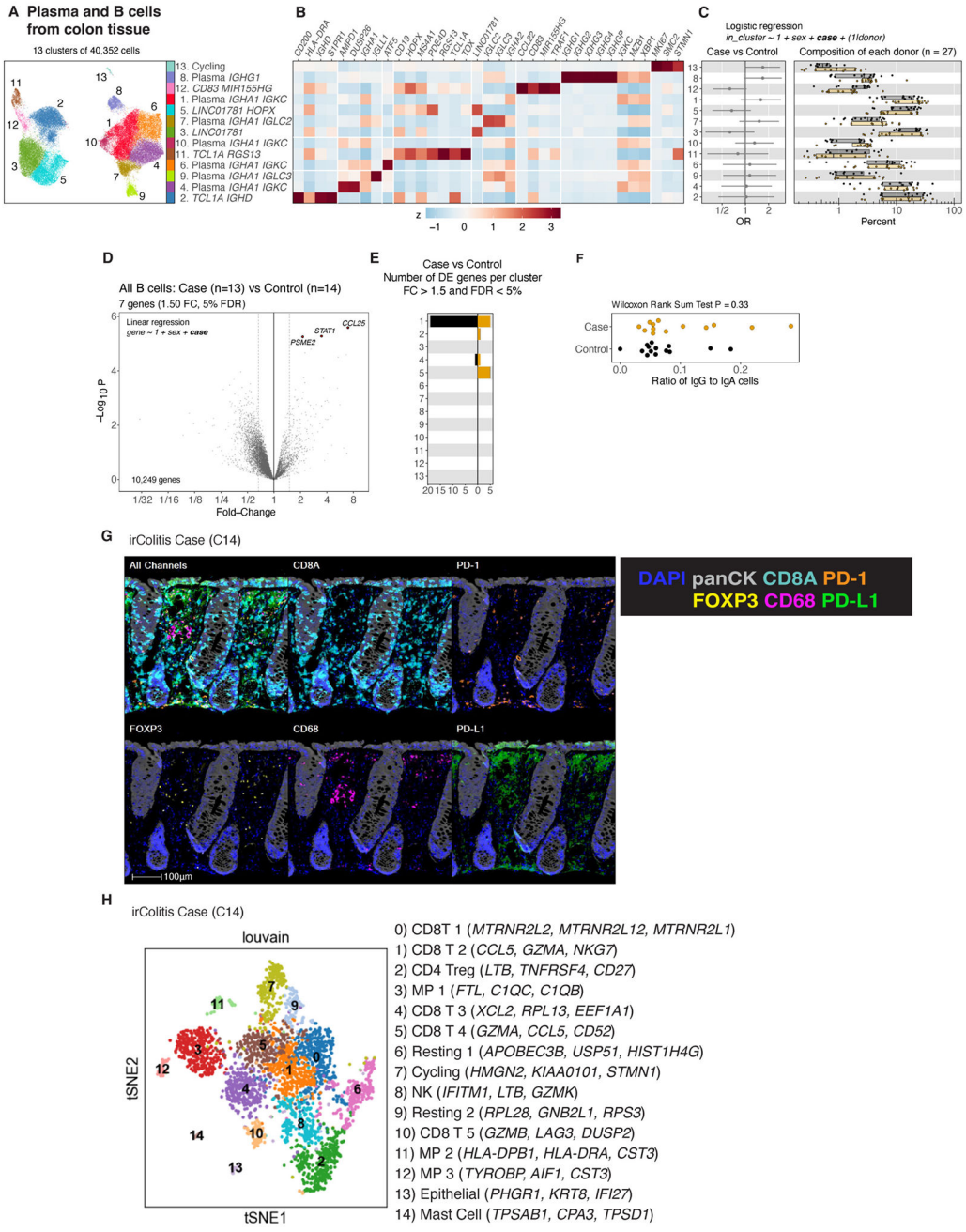
for the contrast of irColitis cases versus controls for CD8 T/GDT/NK cells. The x-axis indicates the fold-change and y-axis indicates negative log10 *p*-value (two-sided) reported by limma. **(F)** Spearman correlation between each cluster and each of the reported CD8 T cell subsets from Giles et al, *Immunity*, 2022. **(G)** UMAP embedding, **(H)** normalized gene expression, and **(I)** cell subset abundance difference between irColitis cases and controls for CD4 T cells. **(J)** Volcano plot of pseudobulk differential gene expression with all cells for the contrast of irColitis cases versus controls for CD4 T cells. **(K)** Spearman correlation between each cluster and each of the reported CD4 T cell subsets from Giles et al, *Immunity*, 2022. Panels **B**, **H** show normalized gene expression (mean zero, unit variance) for selected genes, showing relative expression across cell clusters. Panels **C**, **I** show cell subset abundance differences between irColitis cases in orange and controls in gray, unadjusted likelihood ratio test *p*-values. Boxplots show median and interquartile range of patient cell type compositions where each dot represents a patient. The cellular composition of each patient is reported as the percentage of cells from a patient in each cell cluster. Error-bars indicate logistic regression OR 95% CI for differential abundance of cells from irColitis cases for each cell cluster. GDT gamma delta T cell. NK natural killer. Related to Figure 4.



Extended Data Figure 5. Detailed analysis of MP cells from colon tissue.

(A) UMAP embedding of 2,242 MP cells. Colors indicate cell cluster identities, which are listed on the right. **(B)** Normalized expression (mean zero, unit variance) of selected genes showing relative expression across cell clusters. Heatmap rows are aligned to every cluster defined in the UMAP. **(C)** Cell subset abundance differences between cases in orange (n=13) and controls in gray (n=14) across all MP cell subsets. Boxplots show patient cell type compositions where each dot represents a patient. Mononuclear phagocyte composition of each patient is reported as the percent of cells from a patient in each cell cluster, and box plots show median and interquartile range. Error bars indicate logistic regression 95% CI of

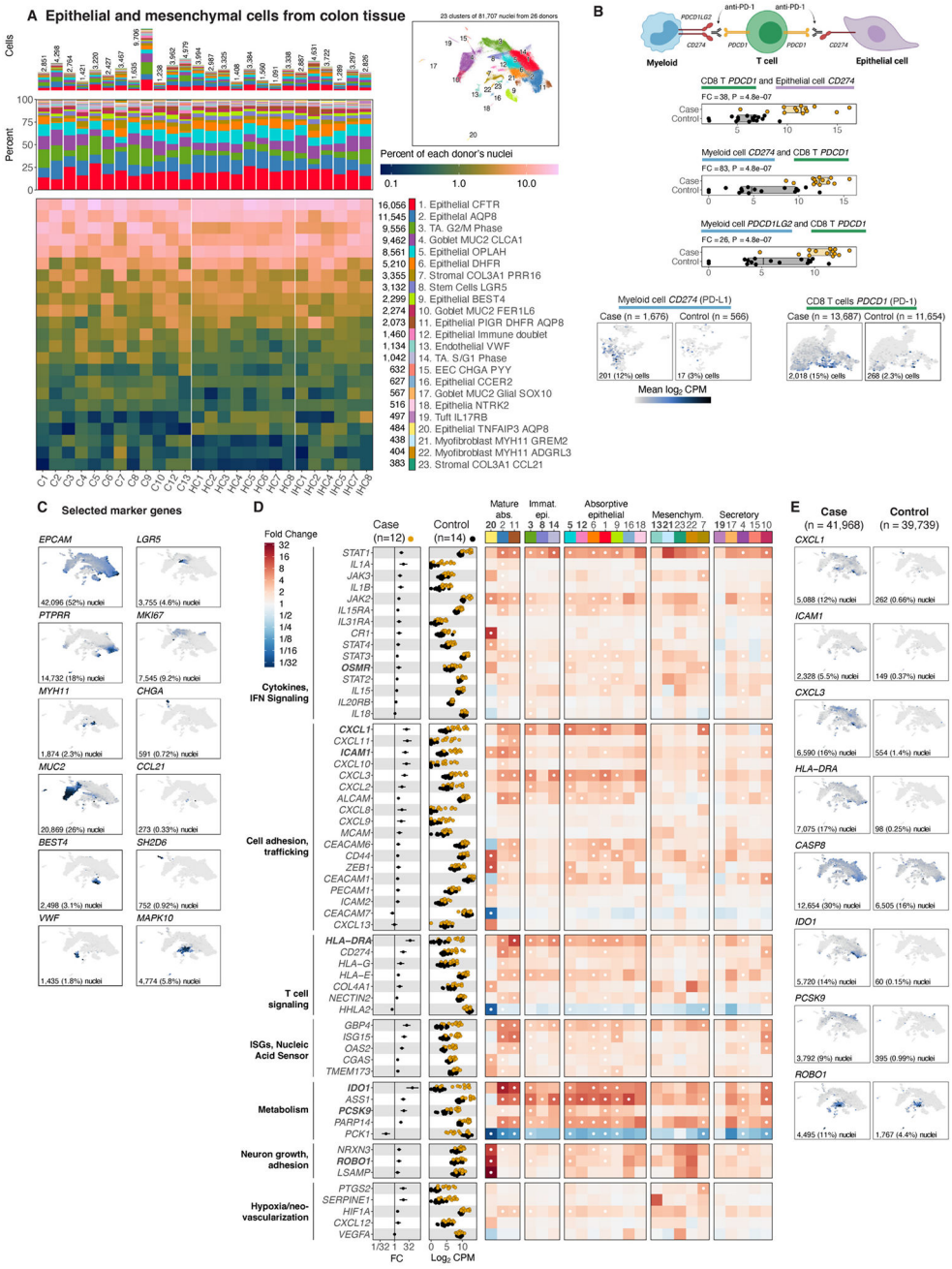
OR for differential abundance of case cells for each cell cluster, and unadjusted likelihood ratio test *p*-values are shown. **(D)** Volcano of pseudobulk differential gene expression with all cells for the contrast of irColitis cases versus controls. The x-axis indicates the fold-change and y-axis indicates negative log10 *p*-value reported by limma (two-sided). **(E)** Bar plots showing differentially expressed (DE) genes per MP cluster (fold-change > 1.5 and FDR < 5%). **(F)** MP cell gene expression fold-change and $\log_2$CPM for cases (orange) and controls (black) is reported for selected genes across T cells (left and middle columns). Heatmap color indicates FC differences between irColitis cases and controls (right column). White dot indicates FDR < 5%. Panels show representative genes across multiple biological themes. **(G)** Gene expression for PD-1 ligands PD-L1 (*CD274*) and PD-L2 (*PDCD1LG2*) in each respective UMAP embedding (left panels). Estimated fold-changes between case and control for each myeloid cell cluster (middle). Gene expression values for the cells from each patient in each myeloid cell cluster where dots represent individual patients (right). Error bars show 95% CI and box plots show median and interquartile range. Related to Figure 1.



Extended Data Figure 6. Detailed analysis of B cells from colon tissue.

(A) UMAP embedding of 40,352 B cells. Colors indicate cell cluster identities, which are listed on the right. **(B)** Normalized expression (mean zero, unit variance) of selected genes showing relative expression across cell clusters. Heatmap rows are aligned to every cluster defined in the UMAP. **(C)** Cell subset abundance differences between cases in orange (n=13) and controls in gray (n=14) across all B cell subsets. Boxplots show patient cell type compositions where each dot represents a patient. B cell composition of each patient is reported as the percent of cells from a patient in each cell cluster. Error-bars indicate 95% CI of logistic regression OR for differential abundance of case cells for each cell cluster, and

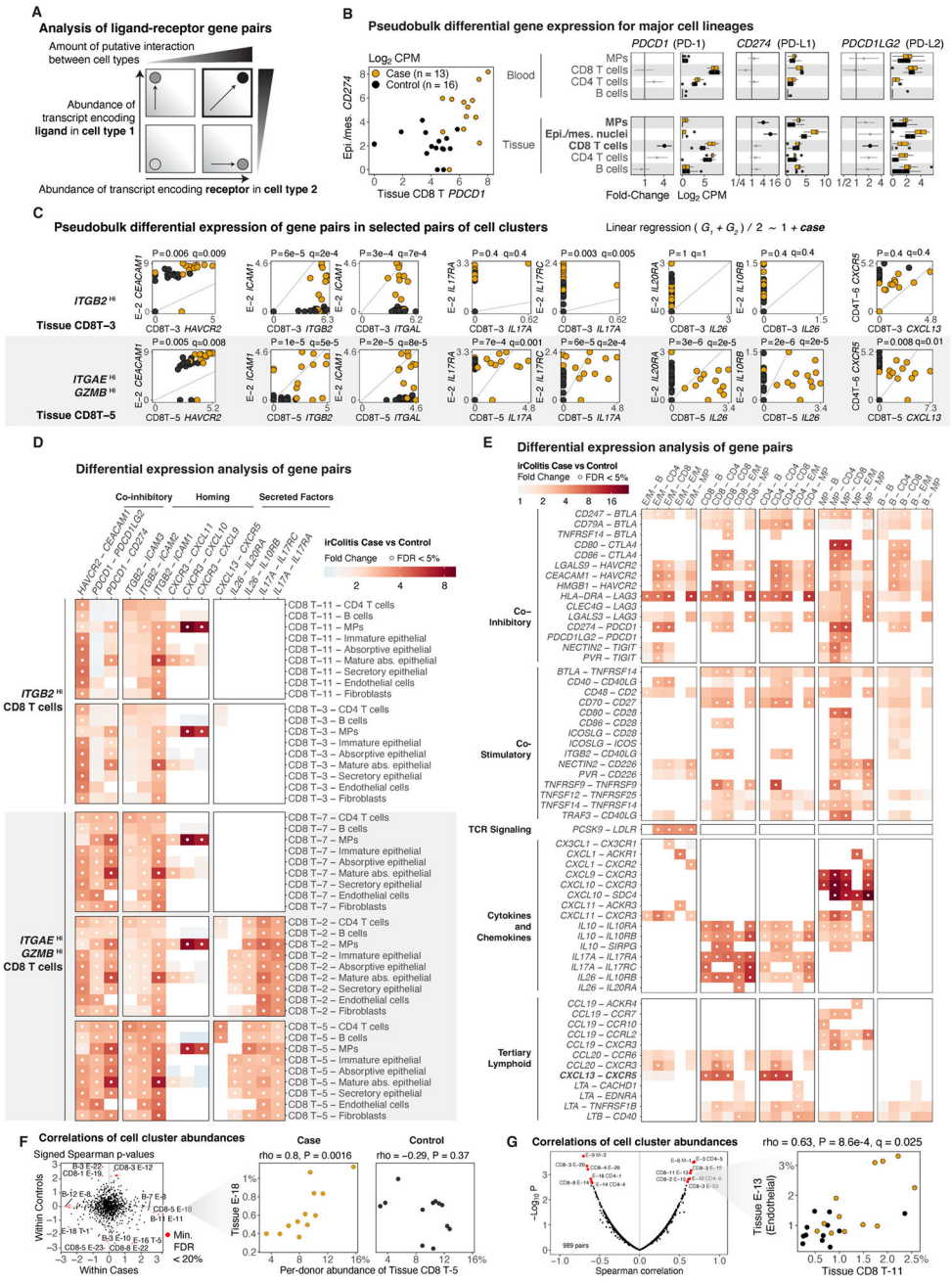
unadjusted likelihood ratio test *p*-values are shown. **(D)** Volcano of pseudobulk differential gene expression with all B cells for the contrast of irColitis cases versus controls. The x-axis indicates the fold-change and y-axis indicates negative log10 *p*-value reported by limma. **(E)** Bar plots showing differentially expressed (DE) genes per B cluster (fold-change greater than 1.5 and FDR less than 5%). **(F)** Ratio of IgG to IgA plasma cells across individual irColitis cases and controls. Each dot represents an individual patient. *P*-value for Wilcoxon rank sum test is shown. **(G-H)** irColitis in a B cell-depleted patient receiving ICI therapy for lymphoma. **(G)** Multispectral immunofluorescence staining of fixed colon mucosal tissue from patient C14* with a 7-color panel: DAPI (blue), panCK (gray), CD8A (aqua), PD-1 (orange), FOXP3 (yellow), CD68 (pink), and PD-L1 (green). **(H)** tSNE-embedding of 3,295 $CD45^{+}$-sorted cells from a patient depleted of B cells. Cell cluster identity and top three AUC genes in parentheses are shown. MP: mononuclear phagocyte. Box plots show median and interquartile range. Related to Figure 1.



Extended Data Figure 7. Detailed analysis of epithelial and mesenchymal cells from tissue.

(A) Detailed composition of each patient across cell clusters for epithelial and mesenchymal nuclei from colon tissue. Heatmap color indicates percent of a patient's cells assigned to each cell cluster. Every column represents one individual donor, referenced at the bottom of the heatmap. Upper right panel shows unique cell clusters depicted with different colors in UMAP embedding, which match the color scheme of the cell subset identities listed on the right side of the heatmap. **(B)** Schematic representing computationally-predicted cell-cell interactions between T cells expressing PD-1 (*PDCD1*) and cells expressing the PD-1 receptors PD-L1 (*CD274*) and PD-L2 (*PDCD1LG2*) (top cartoon). Three middle

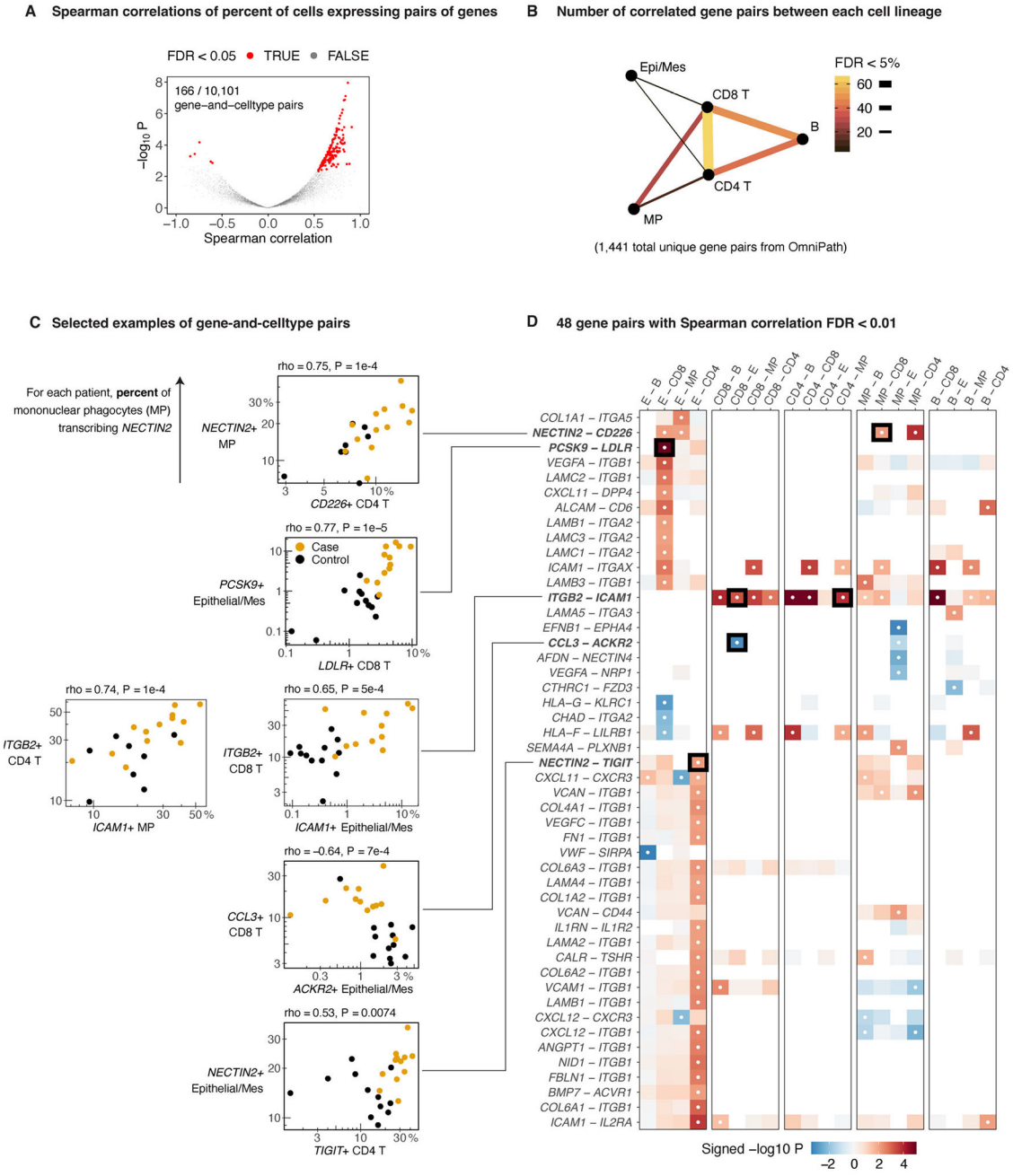
panels show differential expression analysis of means of gene pairs in indicated tissue cell types with cases (orange) and controls (black). Dots represent individual patients. Box plots show median and interquartile range. Limma fold-changes (FC) and two-sided *p*-values are shown. Bottom panels show feature plots of gene expression (Log_2CPM) level in the UMAP embedding for MP cells from Extended Data Fig. 5A (left two panels) and CD8/GDT/NK cells from Fig. 2A (right two panels) presented separately for cells from cases and controls. Number and percentage of cells (from cases or from controls) with detected expression are reported at the bottom of each feature plot. **(C)** Feature plots use color to indicate gene expression (Log_2CPM) level for selected epithelial and mesenchymal genes in the UMAP embedding in panel A. Number and percentage of nuclei with detected expression of each candidate gene are reported at the bottom of each feature plot. **(D)** FC, 95% CI, and gene expression (Log_2CPM) (two left columns) for cases (orange) (n=12) and controls (black) (n=14) is reported for a set of genes organized across 7 biological themes. Heatmap (five right columns) indicates FC differences between cases and controls. White dot indicates FDR < 5%. **(E)** Feature plots show gene expression in nuclei from cases (left) and controls (right) using UMAP embedding in panel A. All feature plots shown in panels B, C, and E use color to indicate gene expression (Log_2CPM). Number and percentage of cells with detected expression of each candidate gene are reported at the bottom of each feature plot. Related to Figure 5.



Extended Data Figure 8. Receptor-ligand interactions predict altered cellular communication in irColitis

(A) Schematic of cell-cell communication inference. Predicted communication between two cell types is defined as proportional to the transcript levels of ligand and receptor genes in the two cell types. **(B)** Left: *PDCD1* (log_2CPM) (encoding PD-1) in tissue CD8 T cells (x-axis) and *CD274* (encoding PD-L1) in epithelial and mesenchymal nuclei (y-axis). Right: Gene expression of *PDCD1* and its ligands *CD274* and *PDCD1LG2* (encoding PD-L2) across blood immune cell types (top panels) and tissue immune, epithelial, and mesenchymal cells/nuclei (epi./mes.) (lower panels). Error-bars show 95% CI of fold-changes (black indicates FDR < 5%), cases (orange) and controls (black). Box plots

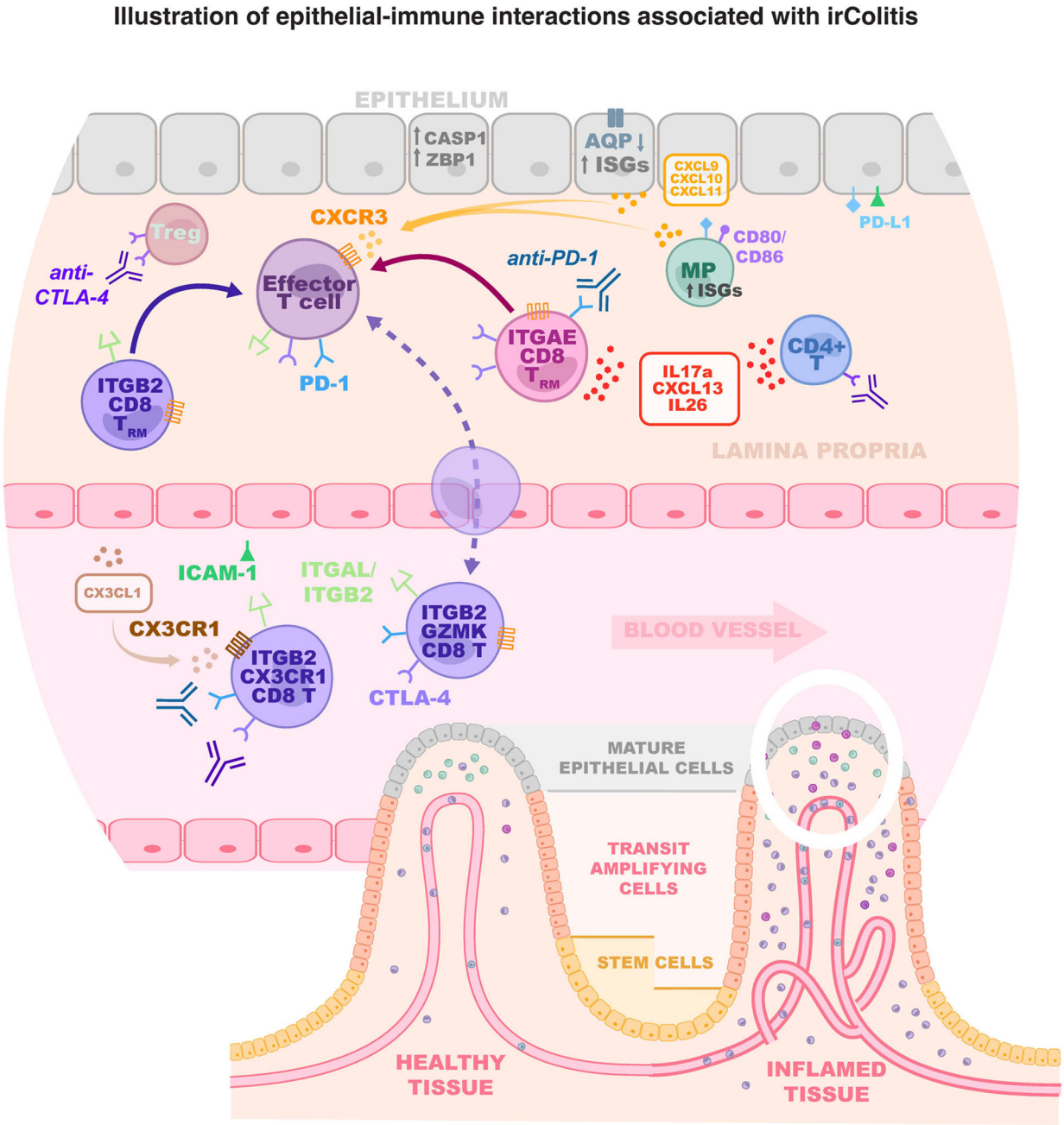
show median and interquartile range. **(C)** Gene expression for ligand-receptor gene pairs, dots represent patients. X-axis indicates expression of *HAVCR2, ITGB2, ITGAL, IL17A, IL26, CXCL13* in CD8 T cluster 3 cells (top row) or cluster 5 cells (bottom row). Y-axis indicates expression in epithelial cluster 2 cells (*CEACAM1, ICAM1, IL17RA, IL17RC, IL20RA, IL10RB*) or CD4 T cluster 6 cells (*CXCR5*). Unadjusted *p*-value and FDR (q-value) from differential expression (t-test) of the mean of each gene pair between cases and controls. **(D)** Differential expression analysis of means of gene pairs, focusing on putative communication between CD8 T cell clusters from Figure 2A–B and other immune, epithelial, and mesenchymal cell lineages (Methods). Top: Tissue $ITGB2^{HI}$ circulatory (CD8-11, 3). Bottom: Activated effector CD8 T_{RM} ($ITGAE^{HI}$ $GZMB^{HI}$) (CD8-7, 2, 5). Heatmap color indicates FC between irColitis cases and controls. White dot indicates FDR < 5%. **(E)** DE analysis with five major tissue cell lineages (columns). Gene pairs are in five biological themes. White dots indicate FDR <5%. **(F)** Left: correlations of cell cluster abundances within cases (x-axis) and within controls (y-axis), signed Spearman *p*-values. Right: abundance of tissue CD8-5 cells (x-axis) vs E-18 cells (y-axis). **(G)** Left: correlations of cell cluster abundances (cases and controls), x-axis Spearman correlation, y-axis −log10 *p*-value. Right: abundance of tissue CD8-11 (x-axis) vs E-13 (y-axis). E/M: epithelial and mesenchymal nuclei, B: B cells, MP: mononuclear phagocytes, CD8: CD8 T/gamma delta T/NK cells, CD4: CD4 T cells. Related to Figures 2–5.



Extended Data Figure 9. Spearman analysis of ligand-receptor gene pairs across multiple colon mucosal cell types.

(A) Volcano plot of Spearman correlations of percent of cells expressing pairs of genes, for all 10,101 gene-and-cell-type pairs (1,441 total unique gene pairs, tested for each pair of cell lineages). X-axis indicates correlation coefficient and y-axis *p*-value. **(B)** Number of gene pairs with FDR < 5% is depicted as an edge connecting each pair of cell lineages. Edge thickness and color is proportional to the number of gene pairs. **(C)** Spearman correlation of percent of cells with gene expression for selected pairs of genes for each pair of major cell lineages. **(D)** Heatmap color depicts the signed Spearman *p*-value for each pair of genes,

for each pair of cell lineages. White dots indicate FDR < 5%. Columns indicate pairs of different cell lineages. E: epithelial and mesenchymal nuclei, B: B cells, MP: mononuclear phagocytes, CD8: CD8 T/GDT/NK cells, CD4: CD4 T cells. Pairs of genes in boldface are shown in panel (C). Related to Figures 2–5.



Extended Data Figure 10. Illustration of epithelial-immune interactions associated with irColitis.
Cartoon illustrating the major findings of our study showing that irColitis is defined by the colon mucosal expansion of *ITGAE*^{HI} CD8 T_{RM} T cells expressing *CXCL13* and Th17 gene programs, *ITGB2*^{Hi} CD8 T cells that may recirculate, Tregs, CD4 T

cells expressing *CXCL13* and *IL17A*, and ISG^{Hi} MP cells. Putative ligand/receptor pathways that recruit circulating cells to the endothelium (CX3CR1-CX3CL1, ITGAL/ITGB2-ICAM-1/2, CXCR3-CXCL9/10/11) and retain expanded CD8 T cells in tissue (ITGAL/ITGB2-ICAM-1, CXCR3-CXCL9/10/11) are shown. Epithelial defects in irColitis include decreased stem cells, increased transit amplifying cells, and top crypt epithelial cells with marked upregulation of interferon-stimulated genes (ISGs), *CASP1*, *ZBP1*, *ICAM1*, *CD274*/PD-L1, and *CXCL10/11* and downregulation of aquaporin (*AQP*) water channel genes. Mesenchymal alterations in irColitis are notable for increased endothelial cells. The white oval at the bottom right shows the part of the crypt depicted in greater detail in the upper part of the cartoon. Related to Figures 1–6.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are deeply grateful to all donors and their families, as well as to the Severe Immunotherapy Complications service at Massachusetts General Hospital. We acknowledge the contributions of Eugene Drokhlyansky, Ramnik Xavier, Moshe Biton, and Hacohen lab members for their feedback on experimental design and data interpretation. We thank Cameron Flayer and Kathleen Gallagher for help with generating Luminex data. This work was supported by several training grants, including a NIAID grant T32AR007258 (to K.S.), a NHLBI grant K08HL157725 and American Heart Association Career Development Award (to. P.S.), two NIDDK training grants – 1K08DK127246-01A1 and T32DK007191 (to M.F.T), the Spanish Society of Medical Oncology (SEOM) grant (to L.Z.), 1T32CA207021(to J.H.C.), 1K08CA273547-01A1 (to J.H.C.), MGH Fund for Medical Discovery (to J.H.C. and to M.F.T.), MGH Krantz Stewardship (J.H.C.), SITC/AstraZeneca Forward Fund (J.H.C.), NCI R00CA259511 (K.P.). This work was made possible by the generous support from the National Institute of Health Director's New Innovator Award (DP2CA247831; to A.C.V.), the Massachusetts General Hospital Transformative Scholar in Medicine Award (to A.C.V.), the Damon Runyon-Rachleff Innovation Award (to A.C.V.), The Melanoma Research Alliance Young Investigator Award (https://doi.org/10.48050/pc.gr.143739; to A.C.V.), the Kraft Foundation Award (to. K.L.R. and A.C.V.), the Arthur, Sandra, and Sarah Irving Fund for Gastrointestinal Immuno-Oncology (to J.H.C, N.H. and A.C.V.), Gordon Pugh (to K.L.R.), the Adelson Foundation (to G.M.B), R01AI169188-01 (to M.D.), the Fariborz Maseeh Award for Innovative Medical Education (to M.D.), The Peter and Ann Lambertus Family Foundation (to M.D. and R.J.S.), Merck (R.J.S.), the American College of Gastroenterology Clinical Research Award and R01AG068390 (to H.K.). This work was also made possible by the generous support of an anonymous donor (to. K.L.R. and A.C.V.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests Statement

From August 9, 2021, B.L. is an employee of Genentech. M.D. has received consulting fees from Genentech, Partner Therapeutics, SQZ Biotech, AzurRx, Eli Lilly, Mallinckrodt Pharmaceuticals, Aditum, Foghorn Therapeutics, Palleon, Sorriso Pharmaceuticals, Generate Biomedicines, Asher Bio, Neoleukin Therapeutics, Moderna, Alloy Therapeutics, Third Rock Ventures, DE Shaw Research, Agenus, and Curie Bio; he is a member of the Scientific Advisory Board for Veravas, Monod Bio, Axxis Bio, and Cerberus Therapeutics. R.J.S. is a consultant to BMS, Marengo, Merck, Novartis, Pfizer, and Replimune. H.K. received research funding from Pfizer and Takeda. H.J. received consulting fees from Aditium Bio, Abbvie, and Takeda. H.J. serves on scientific advisory board of Vivante Health. D.J. reports grants and personal fees from Novartis, Genentech, Syros, Eisai. D.J. reports personal fees from Vibliome, PIC Therapeutics, Mapkure, Relay Therapeutics. D.J. reports grants from Pfizer, Amgen, InventisBio, Arvinas, Takeda, Blueprint, AstraZeneca, Ribon Therapeutics, Infinity that are outside the submitted work. G.M.B. has sponsored research agreements through her institution with: Olink Proteomics, Teiko Bio, InterVenn Biosciences, Palleon Pharmaceuticals. G.M.B. served on advisory boards for: Iovance, Merck, Nektar Therapeutics, Novartis, and Ankyra Therapeutics. G.M.B. consults for: Merck, InterVenn Biosciences, Iovance, and Ankyra Therapeutics. She holds equity in Ankyra Therapeutics. KP is a consultant to Santa Ana Bio. N.H. holds equity in and advises Danger Bio/Related Sciences, is on the scientific advisory board of Repertoire Immune Medicines and CytoReason, owns equity and has licensed patents to BioNtech, and receives research funding from Bristol Myers Squibb and Calico Life Sciences. KLR has received advisory board fees from SAGA diagnostics and institutional research support from BMS. A.C.V. received consulting fees from Merck and BMS. A.C.V. has a financial interest in 10X Genomics. The company designs and manufactures gene sequencing

technology for use in research, and such technology is being used in this research. Dr. Villani's interests were reviewed by The Massachusetts General Hospital and Mass General Brigham in accordance with their institutional policies. M.F.T., K.S., K.M., P.S., N.S., J.T., M.N., L.Z., N.P.S., A.T., S.R., B.Y.A., L.T.N., J.H.C., T.E., Y.S., K.H.X., V.J., C.J.P., T.S., R.G., P.Y.C., do not have competing interest to declare.

Data Availability

The scRNA-seq and snRNA-seq data and results are available for browsing here: https://villani.mgh.harvard.edu/ircolitis. The same processed data files are available on NCBI GEO (GSE206301) and Zenodo (https://zenodo.org/records/10650921). Raw human sequencing data is available at dbGaP (phs003418.v1.p1). The dbGaP database provides authorized access to sequencing data that requires a formal request be made to the appropriate NIH Data Access Committee.

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Alexandra-Chloé Villani (AVILLANI@mgh.harvard.edu).

References

1. Wolchok JD et al. Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. N. Engl. J. Med 377, 1345–1356 (2017). [PubMed: 28889792]
2. Arnaud-Coffin P et al. A systematic review of adverse events in randomized trials assessing immune checkpoint inhibitors. Int. J. Cancer 145, 639–648 (2019). [PubMed: 30653255]
3. Chen JH, Pezhouh MK, Lauwers GY & Masia R Histopathologic Features of Colitis Due to Immunotherapy With Anti-PD-1 Antibodies. Am J Surg Pathol 41, 643–654 (2017). [PubMed: 28296676]
4. Verschuren EC et al. Clinical, Endoscopic, and Histologic Characteristics of Ipilimumab-Associated Colitis. Clin Gastroenterol Hepatol 14, 836–842 (2016). [PubMed: 26748223]
5. Curran MA, Montalvo W, Yagita H & Allison JP PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors. Proc Natl Acad Sci U A 107, 4275–4280 (2010).
6. Heul AV & Stappenbeck T Establishing A Mouse Model of Checkpoint Inhibitor-Induced Colitis: Pilot Experiments and Future Directions. J. Allergy Clin. Immunol 141, AB119–AB119 (2018).
7. Perez-Ruiz E et al. Prophylactic TNF blockade uncouples efficacy and toxicity in dual CTLA-4 and PD-1 immunotherapy. Nature 569, 428–432 (2019). [PubMed: 31043740]
8. Callahan MK Tandon S, X. Y, Subudhi SK, Roman RA, Heine AI, Pogoriler E, Kuk D, Panageas K, Yuan JD, Allison JP, and Wolchok JD, Yang A et al. Evaluation of serum IL-17 levels during ipilimumab therapy: Correlation with colitis. J. Clin. Oncol 29, 2505–2505 (2011).
9. Shahabi V et al. Gene expression profiling of whole blood in ipilimumab-treated patients for identification of potential biomarkers of immune-related gastrointestinal adverse events. J Transl Med 11, 75–75 (2013). [PubMed: 23521917]
10. Das R et al. Early B cell changes predict autoimmunity following combination immune checkpoint blockade. J. Clin. Invest 128, 715–720 (2018). [PubMed: 29309048]
11. Chaput N et al. Baseline gut microbiota predicts clinical response and colitis in metastatic melanoma patients treated with ipilimumab. Ann Oncol 28, 1368–1379 (2017). [PubMed: 28368458]
12. Oh DY et al. Immune Toxicities Elicited by CTLA-4 Blockade in Cancer Patients Are Associated with Early Diversification of the T-cell Repertoire. Cancer Res 77, 1322–1330 (2017). [PubMed: 28031229]
13. Sasson SC et al. Mucosal-associated invariant T (MAIT) cells are activated in the gastrointestinal tissue of patients with combination ipilimumab and nivolumab therapy-related colitis in a pathology distinct from ulcerative colitis. Clin. Exp. Immunol 202, 335–352 (2020). [PubMed: 32734627]

14. Sasson SC et al. Interferon-Gamma–Producing CD8+ Tissue Resident Memory T Cells Are a Targetable Hallmark of Immune Checkpoint Inhibitor–Colitis. Gastroenterology 161, 1229–1244.e9 (2021). [PubMed: 34147519]
15. Luoma AM et al. Molecular Pathways of Colon Inflammation Induced by Cancer Immunotherapy. Cell (2020) doi:10.1016/j.cell.2020.06.001.
16. Brahmer JR et al. Society for Immunotherapy of Cancer (SITC) clinical practice guideline on immune checkpoint inhibitor-related adverse events. J. Immunother. Cancer 9, e002435 (2021). [PubMed: 34172516]
17. Thompson JA et al. NCCN Guidelines Insights: Management of Immunotherapy-Related Toxicities, Version 1.2020. J. Natl. Compr. Cancer Netw. JNCCN 18, 230–241 (2020). [PubMed: 32135517]
18. Faje AT et al. High-dose glucocorticoids for the treatment of ipilimumab-induced hypophysitis is associated with reduced survival in patients with melanoma. Cancer 124, 3706–3714 (2018). [PubMed: 29975414]
19. Bai X et al. Early Use of High-Dose Glucocorticoid for the Management of irAE Is Associated with Poorer Survival in Patients with Advanced Melanoma Treated with Anti-PD-1 Monotherapy. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res (2021) doi:10.1158/1078-0432.CCR-21-1283.
20. FitzPatrick MEB et al. Human intestinal tissue-resident memory T cells comprise transcriptionally and functionally distinct subsets. Cell Rep 34, 108661 (2021). [PubMed: 33472060]
21. Walling BL & Kim M LFA-1 in T Cell Migration and Differentiation. Front. Immunol 9, (2018).
22. Singer M et al. A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. Cell 166, 1500–1511.e9 (2016). [PubMed: 27610572]
23. Li H et al. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. Cell 176, 775–789.e18 (2018). [PubMed: 30595452]
24. Buggert M et al. The Identity of Human Tissue-Emigrant CD8+ T Cells. Cell 183, 1946–1961.e15 (2020). [PubMed: 33306960]
25. Hu W, Wang Y, Fang Z, He W & Li S Integrated Characterization of lncRNA-Immune Interactions in Prostate Cancer. Front. Cell Dev. Biol 9, 271 (2021).
26. Yost KE et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat. Med 25, 1251–1259 (2019). [PubMed: 31359002]
27. Giles JR et al. Human epigenetic and transcriptional T cell differentiation atlas for identifying functional T cell-specific enhancers. Immunity 55, 557–574.e7 (2022). [PubMed: 35263570]
28. Gerlach C et al. The Chemokine Receptor CX3CR1 Defines Three Antigen-Experienced CD8 T Cell Subsets with Distinct Roles in Immune Surveillance and Homeostasis. Immunity 45, 1270–1284 (2016). [PubMed: 27939671]
29. Abu-Sbeih H et al. Outcomes of vedolizumab therapy in patients with immune checkpoint inhibitor–induced colitis: a multi-center study. J. Immunother. Cancer 6, 142 (2018). [PubMed: 30518410]
30. Kim YU, Kee P, Danila D & Teng B-B A Critical Role of PCSK9 in Mediating IL-17-Producing T Cell Responses in Hyperlipidemia. Immune Netw 19, e41 (2019). [PubMed: 31921471]
31. Smillie CS et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell 178, 714–730.e22 (2019). [PubMed: 31348891]
32. Anonymous. Highlights of prescribing information. Food and Drug Administration [Internet] (2015).
33. Rao DA et al. Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. Nature 542, 110–114 (2017). [PubMed: 28150777]
34. Shimizu J, Yamazaki S, Takahashi T, Ishida Y & Sakaguchi S Stimulation of CD25(+)CD4(+) regulatory T cells through GITR breaks immunological self-tolerance. Nat. Immunol 3, 135–142 (2002). [PubMed: 11812990]
35. Vu MD et al. OX40 costimulation turns off Foxp3+ Tregs. Blood 110, 2501–2510 (2007). [PubMed: 17575071]
36. Oh DY & Fong L Cytotoxic CD4+ T cells in cancer: Expanding the immune effector toolbox. Immunity 54, 2701–2711 (2021). [PubMed: 34910940]

37. Menzel K et al. Cathepsins B, L and D in inflammatory bowel disease macrophages and potential therapeutic effects of cathepsin inhibition in vivo. Clin. Exp. Immunol 146, 169–180 (2006). [PubMed: 16968411]
38. Baklien K & Brandtzaeg P Comparative mapping of the local distribution of immunoglobulin-containing cells in ulcerative colitis and Crohn's disease of the colon. Clin. Exp. Immunol 22, 197–209 (1975). [PubMed: 1082398]
39. Li J et al. KIR+CD8+ T cells suppress pathogenic T cells and are active in autoimmune diseases and COVID-19. Science 376, eabi9591 (2022).
40. Chung H & Rice CM T time for ADAR: ADAR1 is required for T cell self-tolerance. EMBO Rep 19, e47237 (2018). [PubMed: 30467234]
41. Yi W et al. Targeted regulation of self-peptide presentation prevents type I diabetes in mice without disrupting general immunocompetence. J. Clin. Invest 120, 1324–1336 (2010). [PubMed: 20200448]
42. Parikh K et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature 567, 49–55 (2019). [PubMed: 30814735]
43. Leung E et al. Polymorphisms in the organic cation transporter genes SLC22A4 and SLC22A5 and Crohn's disease in a New Zealand Caucasian cohort. Immunol. Cell Biol 84, 233–236 (2006). [PubMed: 16519742]
44. Serrano León A et al. Single-nucleotide polymorphisms in SLC22A23 are associated with ulcerative colitis in a Canadian white cohort. Am. J. Clin. Nutr 100, 289–294 (2014). [PubMed: 24740203]
45. Dicay MS et al. Interferon-γ Suppresses Intestinal Epithelial Aquaporin-1 Expression via Janus Kinase and STAT3 Activation. PLOS ONE 10, e0118713 (2015). [PubMed: 25793528]
46. Kim J et al. Constitutive and inducible expression of B7 family of ligands by human airway epithelial cells. Am. J. Respir. Cell Mol. Biol (2005) doi:10.1165/rcmb.2004-0129OC.
47. Xie J et al. Slit2/Robo1 Mitigates DSS-induced Ulcerative Colitis by Activating Autophagy in Intestinal Stem Cell. Int. J. Biol. Sci 16, 1876–1887 (2020). [PubMed: 32398956]
48. Sharpe AH & Pauken KE The diverse functions of the PD1 inhibitory pathway. Nat. Rev. Immunol 18, 153–167 (2018). [PubMed: 28990585]
49. Türei D et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol. Syst. Biol 17, e9923 (2021). [PubMed: 33749993]
50. Kim J, Kim M-G, Jeong SH, Kim HJ & Son SW STAT3 maintains skin barrier integrity by modulating SPINK5 and KLK5 expression in keratinocytes. Exp. Dermatol 31, 223–232 (2022). [PubMed: 34378233]
51. Matsuoka K et al. DOP056 Efficacy and safety of anti-fractalkine monoclonal antibody, E6011, in patients with Crohn's disease who had lost response to anti-TNFα agents: A multicentre, open-label, Phase 1/2 study. J. Crohns Colitis 12, S070–S070 (2018).
52. Xia X et al. Regulation of PCSK9 Expression and Function: Mechanisms and Therapeutic Implications. Front. Cardiovasc. Med 8, (2021).
53. Yang W et al. Potentiating the antitumour response of CD8+ T cells by modulating cholesterol metabolism. Nature 531, 651–655 (2016). [PubMed: 26982734]
54. Lei L et al. Inhibition of proprotein convertase subtilisin/kexin type 9 attenuates 2,4,6-trinitrobenzenesulfonic acid-induced colitis via repressing toll-like receptor 4/nuclear factor-kappa B. Kaohsiung J. Med. Sci 36, 705–711 (2020). [PubMed: 32396274]
55. Zheng L et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science 374, abe6474 (2021).
56. 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). [PubMed: 22595313]
57. Corridoni D et al. Single-cell atlas of colonic CD8+ T cells in ulcerative colitis. Nat. Med (2020) doi:10.1038/s41591-020-1003-4.
58. Buchan SL et al. Antibodies to Costimulatory Receptor 4–1BB Enhance Anti-tumor Immunity via T Regulatory Cell Depletion and Promotion of CD8 T Cell Effector Function. Immunity 49, 958–970.e7 (2018). [PubMed: 30446386]

59. Riudavets M et al. Correlation between immune-related adverse events (irAEs) and efficacy in patients with solid tumors treated with immune-checkpoints inhibitors (ICIs). J. Clin. Oncol 36, 3064–3064 (2018). [PubMed: 30188784]
60. Crotty S Follicular Helper CD4 T Cells (TFH). Annu. Rev. Immunol 29, 621–663 (2011). [PubMed: 21314428]
61. Veatch JR et al. Neoantigen-specific CD4+ T cells in human melanoma have diverse differentiation states and correlate with CD8+ T cell, macrophage, and B cell function. Cancer Cell 40, 393–409.e9 (2022). [PubMed: 35413271]
62. Pelka K et al. Spatially organized multicellular immune hubs in human colorectal cancer. Cell 184, 4734–4752.e20 (2021). [PubMed: 34450029]
63. Yang M et al. CXCL13 shapes immunoactive tumor microenvironment and enhances the efficacy of PD-1 checkpoint blockade in high-grade serous ovarian cancer. J. Immunother. Cancer 9, e001136 (2021). [PubMed: 33452206]
64. Song X et al. Growth Factor FGF2 Cooperates with Interleukin-17 to Repair Intestinal Epithelial Damage. Immunity 43, 488–501 (2015). [PubMed: 26320657]
65. Lee JS et al. Interleukin-23-Independent IL-17 Production Regulates Intestinal Epithelial Permeability. Immunity 43, 727–738 (2015). [PubMed: 26431948]
66. Song X, He X, Li X & Qian Y The roles and functional mechanisms of interleukin-17 family cytokines in mucosal immunity. Cell. Mol. Immunol 13, 418–431 (2016). [PubMed: 27018218]
67. Dubin K et al. Intestinal microbiome analyses identify melanoma patients at risk for checkpoint-blockade-induced colitis. Nat. Commun 7, 10391 (2016). [PubMed: 26837003]
68. Baruch EN et al. Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. Science 371, 602–609 (2021). [PubMed: 33303685]
69. Boland BS et al. Heterogeneity and clonal relationships of adaptive immune cells in ulcerative colitis revealed by single-cell analyses. Sci. Immunol 5, (2020).
70. Böttcher JP et al. Functional classification of memory CD8(+) T cells by CX3CR1 expression. Nat. Commun 6, 8306 (2015). [PubMed: 26404698]
71. Yamauchi T et al. T-cell CX3CR1 expression as a dynamic blood-based biomarker of response to immune checkpoint inhibitors. Nat. Commun 12, 1402 (2021). [PubMed: 33658501]
72. Ricanek P et al. Reduced expression of aquaporins in human intestinal mucosa in early stage inflammatory bowel disease. Clin. Exp. Gastroenterol 8, 49–67 (2015). [PubMed: 25624769]
73. Leber A et al. Activation of NLRX1 by NX-13 Alleviates Inflammatory Bowel Disease through Immunometabolic Mechanisms in CD4+ T Cells. J. Immunol 203, 3407–3415 (2019). [PubMed: 31694910]
74. Lee G et al. P120 BBT-401 IS A SELECTIVE PELLINO-1 PROTEIN-PROTEIN INTERACTION INHIBITOR IN CLINICAL DEVELOPMENT TARGETING A FIRST-IN-CLASS DRUG FOR UC TREATMENT. Inflamm. Bowel Dis 25, S58 (2019).
75. Kim Y-I et al. Local Stabilization of Hypoxia-Inducible Factor-1α Controls Intestinal Inflammation via Enhanced Gut Barrier Function and Immune Regulation. Front. Immunol 11, (2021).
76. Geukes Foppen MH et al. Immune checkpoint inhibition-related colitis: Symptoms, endoscopic features, histology and response to management. ESMO Open (2018) doi:10.1136/esmoopen-2017-000278.
77. Wright AP, Piper MS, Bishu S & Stidham RW Systematic review and case series: flexible sigmoidoscopy identifies most cases of checkpoint inhibitor-induced colitis. Aliment. Pharmacol. Ther 49, 1474–1483 (2019). [PubMed: 31035308]
78. Adam M, Potter AS & Potter SS Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. Dev. Camb. Engl 144, 3625–3632 (2017).
79. Stoeckius M et al. Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods 14, 865–868 (2017). [PubMed: 28759029]
80. Drokhlyansky E et al. The Human and Mouse Enteric Nervous System at Single-Cell Resolution. Cell 182, 1606–1622.e23 (2020). [PubMed: 32888429]

81. Benjamini Y & Hochberg Y Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol 57, 289–300 (1995).
82. Li B et al. Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. Nat. Methods 17, 793–798 (2020). [PubMed: 32719530]
83. Gaublomme JT et al. Nuclei multiplexing with barcoded antibodies for single-nucleus genomics. Nat. Commun 10, 2907 (2019). [PubMed: 31266958]
84. Batson J, Royer L & Webber J Molecular Cross-Validation for Single-Cell RNA-Seq http://biorxiv.org/lookup/doi/10.1101/786269 (2019) doi:10.1101/786269.
85. Korsunsky I et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296 (2019). [PubMed: 31740819]
86. Traag V, Waltman L & van Eck NJ From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep 9, 5233 (2019). [PubMed: 30914743]
87. McInnes L, Healy J & Melville J UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv180203426 Cs Stat (2020).
88. Lun L, A. T, Bach K & Marioni JC Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol 17, 75 (2016). [PubMed: 27122128]
89. Ritchie ME et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47–e47 (2015). [PubMed: 25605792]
90. James KR et al. Distinct microbial and immune niches of the human colon. Nat. Immunol 21, 343–353 (2020). [PubMed: 32066951]
91. Fonseka CY et al. Mixed-effects association of single cells identifies an expanded effector CD4+ T cell subset in rheumatoid arthritis. Sci. Transl. Med 10, eaaq0305 (2018).
92. Gupta NT et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data: Table 1. Bioinformatics 31, 3356–3358 (2015). [PubMed: 26069265]
93. Bates D, Mächler M, Bolker B & Walker S Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw 67, (2015).

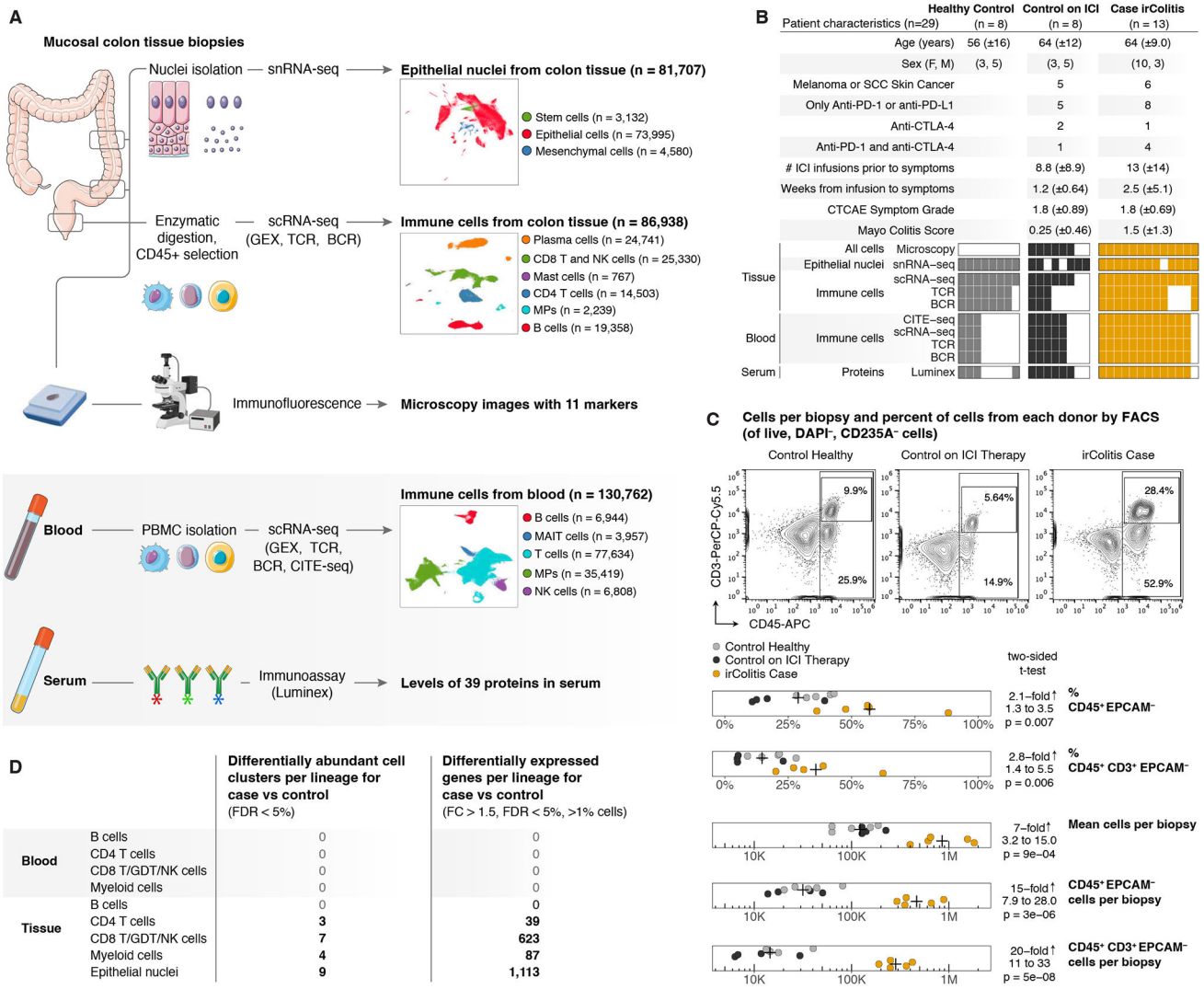


Figure 1. Overview of checkpoint colitis cohort.

(A) Sample processing pipeline overview. Mucosal colonic biopsies along with paired serum and PBMC specimens were collected from patients at the time of endoscopy. From biopsies, we performed single-nuclei RNA-sequencing (snRNA-seq) and single-cell RNA-sequencing (scRNA-seq) with paired T cell receptor (TCR) and B cell receptor (BCR) sequencing. The uniform manifold approximation and projection (UMAP) embedding panels on the right display clustering solutions for indicated cell lineages from tissue and blood sc/snRNAseq datasets. For every collection, biopsy specimens were also fixed for immunofluorescence microscopy of up to 11 markers. Secreted factors were measured by Luminex from the serum, and scRNA-seq with paired TCR, BCR, and CITE-seq was performed on PBMCs. CITE-seq: Cellular Indexing of Transcriptomes and Epitopes by Sequencing. **(B)** Patient cohort overview reporting the breakdown of samples from 29 patients by clinical meta-data and experimental approach. ICI: immune checkpoint inhibitor; irColitis: ICI-related colitis; CTCAE: Common Terminology Criteria for Adverse Events; SCC: Squamous cell carcinoma. **(C)** Representative flow cytometry plots of $CD45^+$ immune cells and $CD45^+$ $CD3^+$ T cells from endoscopic colon mucosal biopsies (upper panels) along with the

frequency of these respective cell populations across three different patient cohorts (second and third panels from the top). The absolute number of all cells/biopsy, $CD45^+$ immune cells/biopsy, and $CD45^+$ $CD3^+$ T cells/biopsy are shown (three bottom rows). Cells per biopsy were normalized by using identically sized endoscopic biopsy forceps for all patient samples and averaging over 6–16 biopsies per patient (Table S5, Methods). Absolute $CD45^+$ and $CD45^+$ $CD3^+$ cell percentages per biopsy were then estimated from total cell numbers using immune and T cell frequencies for each respective population defined by FACS. 95% confidence intervals and two-sided t-test *p*-values for fold-changes are shown. **(D)** Table with number of differentially abundant cell clusters (left) and differentially expressed genes (right) across sc/snRNA-seq blood and colon tissue datasets. GDT: gamma delta T cells; NK: natural killer cells. Parts of the figure were drawn by using pictures from Servier Medical Art. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.

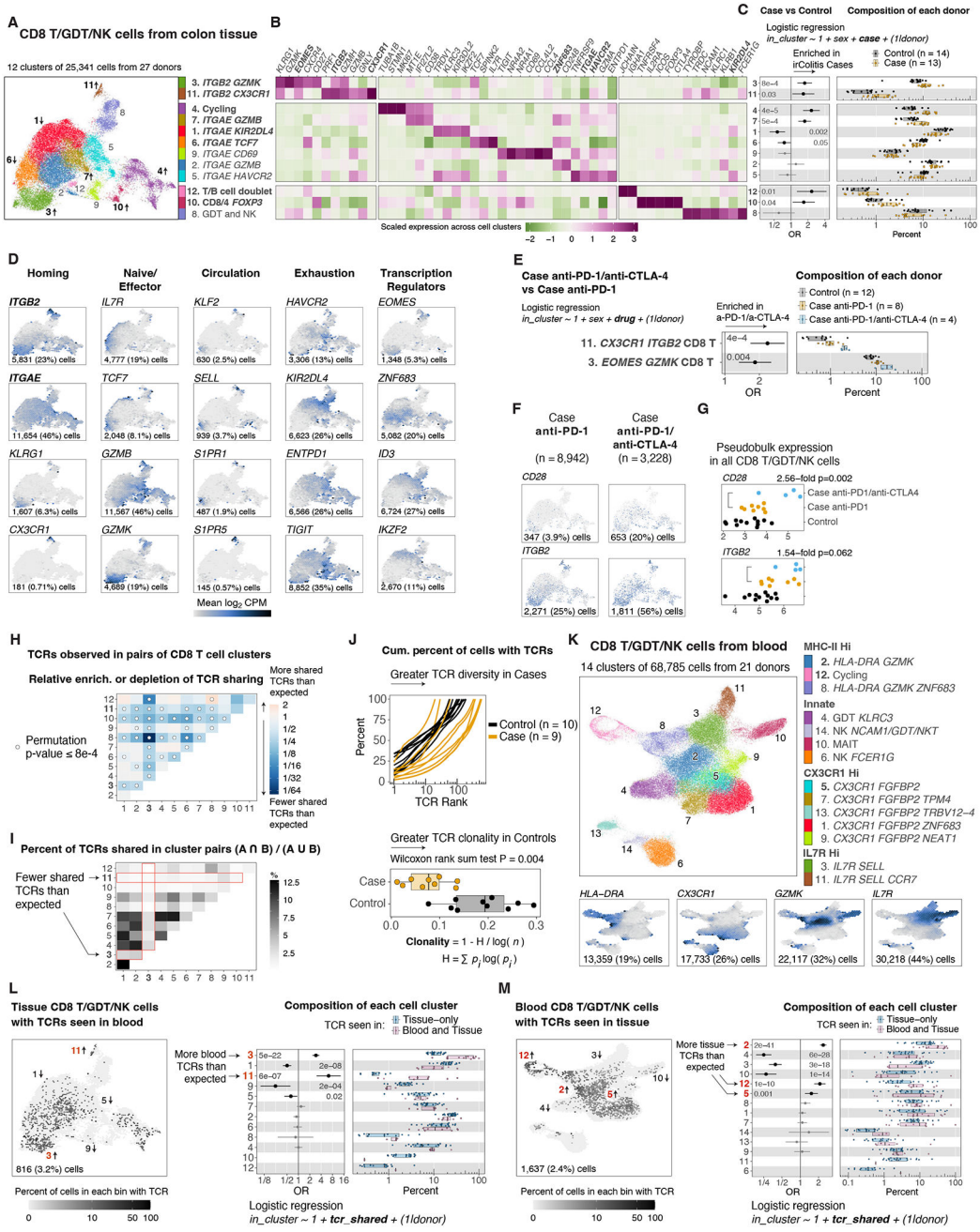


Figure 2. Expanded *ITGAE*^{Hi} tissue-resident memory and *ITGB2*^{Hi} CD8 T cell subsets in irColitis

(A) Tissue UMAP of CD8 T/GDT/NK cells colored by identities and listed across major subsets: *ITGB2*^{HI}, *ITGAE*^{HI}, and mixed lineages. Arrows indicate direction of association. **(B)** Normalized gene expression (columns) across cell clusters (rows). **(C,E)** Cell abundance differences between (**C**) cases (orange; n=13) and controls (gray; n=14), and (**E**) controls (gray; n=12), cases on anti-PD-1 (orange; n=8) or anti-PD-1/CTLA-4 therapy (blue; n=4). **(D,F,K)** Gene expression (Log_2CPM) in UMAP embedding, number and percentage of cells with detected expression. **(F)** Gene expression of DE genes between cases on anti-PD-1

(left) and anti-PD-1/CTLA-4 (right) therapy. **(G)** CD8T/GDT/NK pseudobulk expression level across controls (black dots), cases on anti-PD-1 (orange) or anti-PD-1/CTLA-4 (blue) therapy. Fold-change and unadjusted two-sided *p*-value. **(H)** Heatmap color indicates relative enrichment or depletion of shared TCRs between cluster pairs, normalized to the mean from shuffled data. White dot: permutation *p*-value<8×10^{-4}. **(I)** Color indicates percent of shared TCR clones between cluster pairs. **(J)** Cumulative percent of cells with TCR clones, ranked by descending abundance. **(K)** Blood CD8 T/GDT/NK cells UMAP colored by identities and organized across four subsets, with marker gene UMAP plots (bottom). **(L,M)** Left: (**L**) tissue and (**M**) blood CD8 T/GDT/NK UMAP, arrows indicate enrichment or depletion of (L) blood TCRs in tissue and (M) tissue TCRs in blood. Middle: logistic regression OR for differential abundance of TCR clones observed in (L) tissue-only or in both blood and tissue, and (M) in blood-only or in both blood and tissue, red font indicates associated clusters. Right: percent of cells from each tissue cluster that do or do not share TCRs between tissue and blood. Likelihood ratio test *p*-values shown for clusters with FDR < 5%. (**C,E,J,L,M**) Box plots show median and interquartile range. Each dot represents one patient. Error-bars indicate logistic regression OR and 95% CI for differential abundance of case cells for each cell cluster. Unadjusted likelihood ratio test *p*-values.

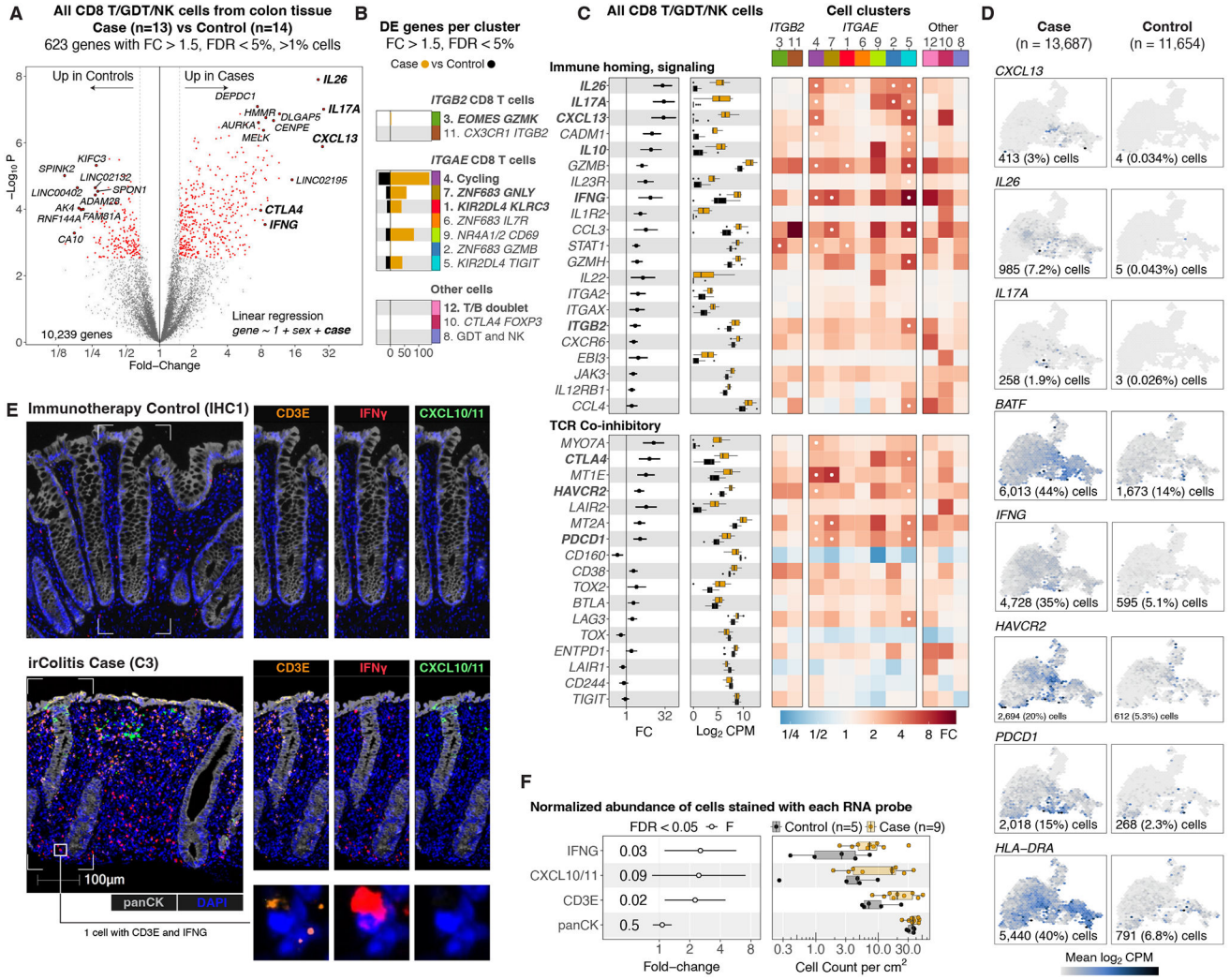


Figure 3. *IL26, IL17A*, and *CXCL13* are upregulated in colon mucosal *ITGAE*^{Hi} CD8 T cells in irColitis

(A) Volcano plot of pseudobulk differential gene expression for all CD8 T, gamma delta T (GDT), and NK cells showing the contrast of irColitis cases (n=13) versus controls (n=14). The x-axis indicates fold-change and y-axis indicates the negative log10 *p*-value (two-sided) reported by limma. **(B)** Bar plots showing differentially expressed (DE) genes per CD8 T/GDT/NK cell cluster defined in Figure 2 and organized across 3 cellular subsets: *ITGB2*^{HI} CD8 T cells, *ITGAE*^{HI} CD8 T cells, and clusters with mixed cell lineages (FC > 1.5 and FDR < 5%). **(C)** FC and gene expression (log_2CPM) for cases (orange) and controls (black) are reported for selected genes across colon tissue CD8 T and innate cytotoxic lymphocyte cell clusters (left and middle columns). Heatmap color indicates FC differences between irColitis cases and controls (right column). White dot indicates FDR < 5%. **(D)** Feature plots use color to indicate gene expression (Log_2CPM) level in the UMAP embedding from figure 2A. Number and percentage of cells with detected expression of each candidate gene are reported at the bottom of each feature plot. **(E)** Multispectral fluorescence panel depicting the RNA expression level of *CD3E* (orange), *IFNG* (red), and *CXCL10/11* (green), protein immunofluorescence of panCK (gray), and DAPI (blue) from one control subject in the

upper panels and one irColitis case in the middle panels. Bottom right boxes highlight representative co-expression of *CD3E* and *IFNG* in the same T cell observed in tissue from one irColitis case. **(F)** Quantification of RNA (*IFNG, CXCL10/11, CD3E*) and protein (PanCK) signals from microscopy images. Fold-change (left column, where black color indicates FDR < 5% and unadjusted linear model *p*-value is shown) and frequency of cells positive for a given marker normalized to the area (cm^2) of tissue (right column) for cases (orange) and controls (black). Dots represent individual patients and box plots show median and interquartile range. Error bars represent 95% CI intervals.

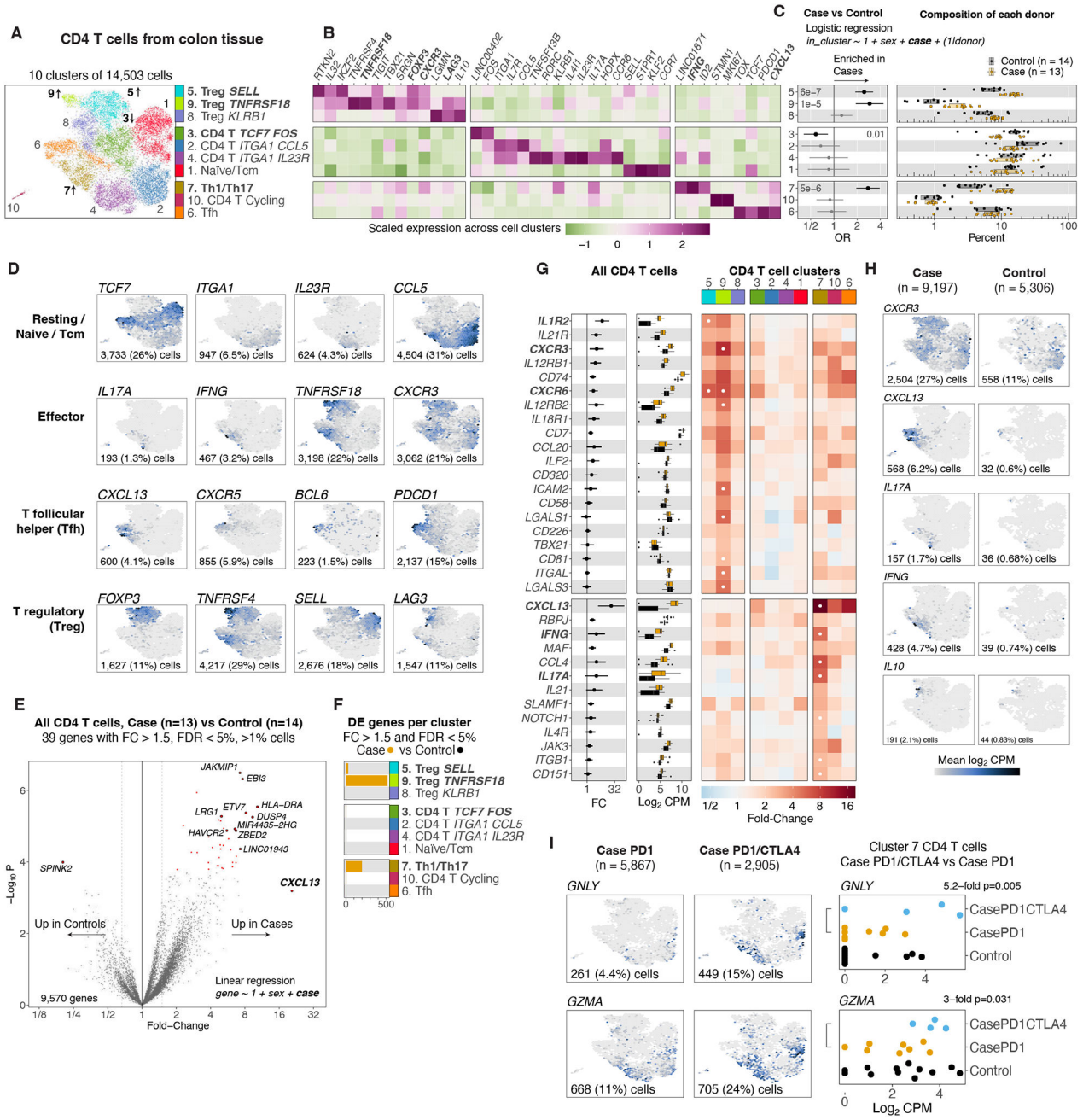


Figure 4. Expanded *IL17A* and *CXCL13*-expressing CD4 T cell effectors and *TNFRSF4*-expressing Tregs in patients with irColitis

(A) UMAP embedding of CD4 T cells color coded according to cell cluster identities, which are bolded for cell subsets showing abundance differences in panel C. Arrows next to cluster numbers show the direction of observed abundance differences. **(B)** Normalized expression (mean zero, unit variance) of selected genes showing relative expression across cell clusters. Rows of the heatmap are aligned to each cluster. **(C)** Cell subset abundance differences between cases in orange (n=13) and controls in gray (n=14) across subsets. Boxplots show patient cell type compositions where each dot represents a patient. Composition of each patient is reported as the percent of cells from a patient in each cell cluster. Error-bars

indicate 95% CI of logistic regression odds ratio for differential abundance of irColitis case cells for each cell cluster. Unadjusted likelihood ratio test two-sided *p*-values are shown. **(D,H)** Feature plots of candidate genes use color to indicate gene expression (Log_2CPM) levels in the UMAP embedding from panel A. Number and percentage of cells with detected expression for each candidate gene are reported at the bottom of each feature plot. **(E)** Volcano plot of pseudobulk differential gene expression for all CD4 T cells showing the contrast of cases versus controls. The x-axis indicates fold-change and y-axis indicates the negative log10 *p*-value from limma. **(F)** Bar plots showing DE genes per CD4 T cell cluster defined in panel A (FC > 1.5 and FDR < 5%). **(G)** FC and gene expression (log_2CPM) levels for cases (orange) and controls (black) is reported for selected genes across CD4 T cell clusters (left and middle columns). Heatmap color indicates FC differences between cases and controls (right column). White dot indicates FDR <5%. **(H)** Feature plots presented separately for cells from irColitis cases and controls. **(I)** *GNLY* and *GZMA* expression levels in cells from cases on anti-PD-1 or dual anti-PD-1/CTLA-4 therapy in the UMAP embedding from panel A. Dot plot shows pseudobulk expression where each dot represents all the CD8 T cells from a patient. Fold-change and *p*-value from limma for cases on dual anti-PD-1/CTLA-4 therapy versus anti-PD-1 monotherapy. Error bars represent 95% CI intervals, box plots show median and interquartile range.

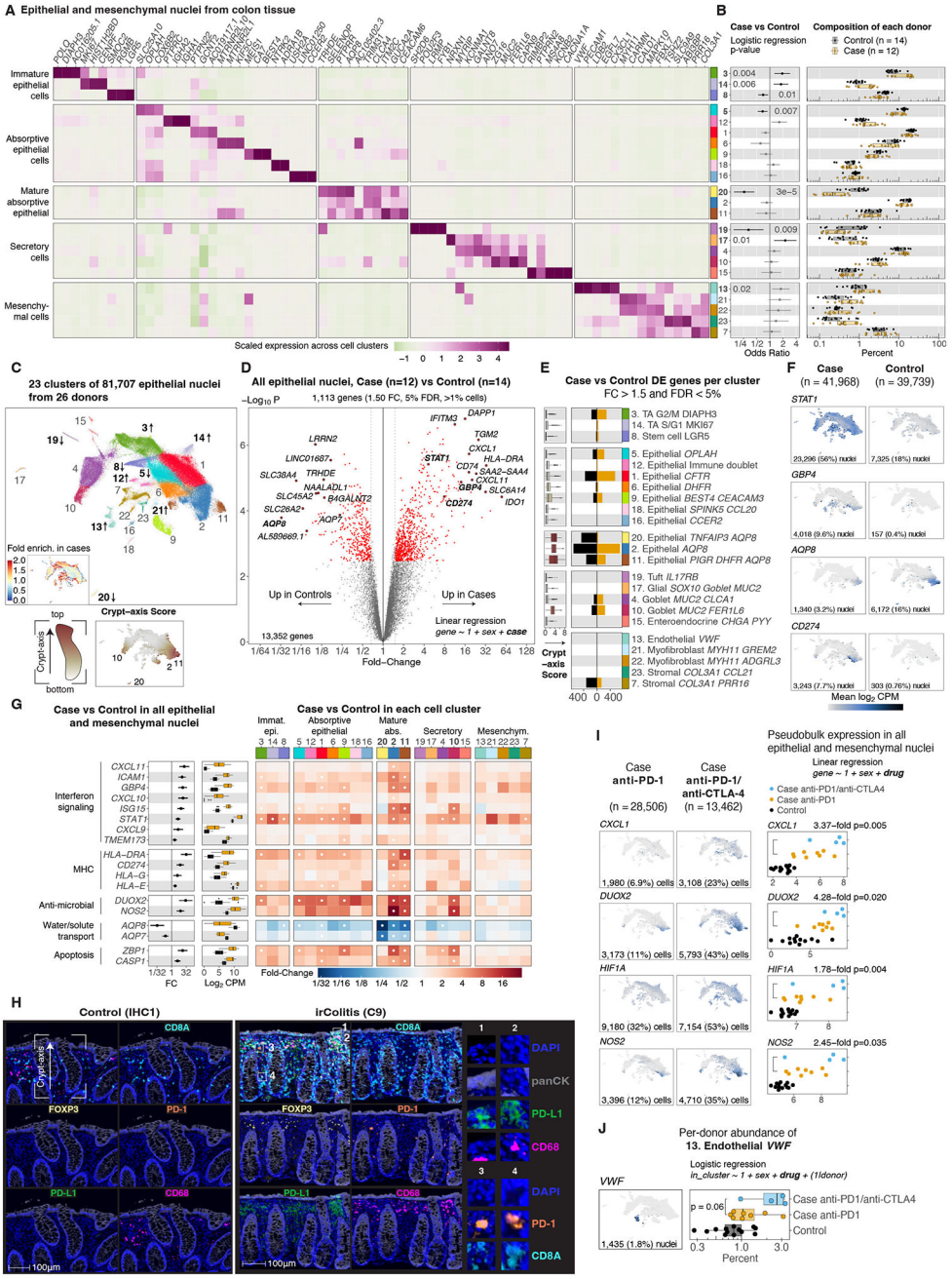


Figure 5. Epithelial and mesenchymal remodeling in IrColitis

(A) Marker genes normalized relative expression for epithelial/mesenchymal subsets across five cellular families (left). **(B)** Error-bars indicate 95% CI of logistic regression OR for differential abundance between cases (orange; n=12) and controls (gray; n=14) for each subset, unadjusted likelihood ratio test *p*-values. **(B,J)** Boxplots show patient cell type composition reported as the percentage of nuclei from a patient in each cluster; each dot represents a patient. **(C)** UMAP of epithelial/mesenchymal subsets colored as in **A**, arrows showing direction of association. Inset shows nuclei binned into hexagons. Bin color represents fold enrichment of cells from cases. Crypt-axis score shown in UMAP (Methods).

(D) Pseudobulk DGE volcano plot for all epithelial/mesenchymal nuclei for cases versus controls. X-axis indicates fold-change and y-axis the negative log10 *p*-value (limma). **(E)** Crypt-axis score boxplot for each cluster (left) (Methods). Bar plot depicts number of DE genes per subset (right) (FC>1.5 and FDR<5%), ordered as in A and B. **(F,I)** Feature plots show gene expression (log_2CPM) level in cases (left) and controls (right) projected on panel C UMAP. Number and percentage of nuclei expressing each gene are reported. **(G)** FC and gene expression (log_2CPM) (two left columns) for cases (orange) and controls (black) of genes organized by themes. Heatmap color indicates case vs. control FC differences. White dot indicates FDR < 5%. **(H)** Staining of colon mucosa from ICI-control (left) and a case (right) for DAPI (blue), panCK (gray), CD8A (aqua), FOXP3 (yellow), PD-1 (orange), PD-L1 (green), and CD68 (pink). Right panels highlight co-detection of PD-L1 +/− CD68 and PD-L1 +/− panCK in individual cells (1, 2), and co-detection of PD-1 +/− CD8A in individual cells (3, 4). **(I) Left:** DE genes between cases on anti-PD-1 (left) versus anti-PD-1/CTLA-4 therapy (right) across subsets. **Right:** Pseudobulk expression level of *CXCL1, DUOX2, HIF1A, NOS2* in all epithelial/mesenchymal subsets stratified across: controls (black), cases on anti-PD-1 (orange) or anti-PD-1/CTLA-4 (blue) therapy. **(J)** Abundance analysis of endothelial nuclei (cluster 13) expressing *VWF*. Controls (black; n=13), cases on anti-PD-1 (orange; n=8) or anti-PD-1/CTLA-4 therapy (blue; n=4). (**B,G,J**) Box plots show median and interquartile range. Error bars represent 95% CI intervals.

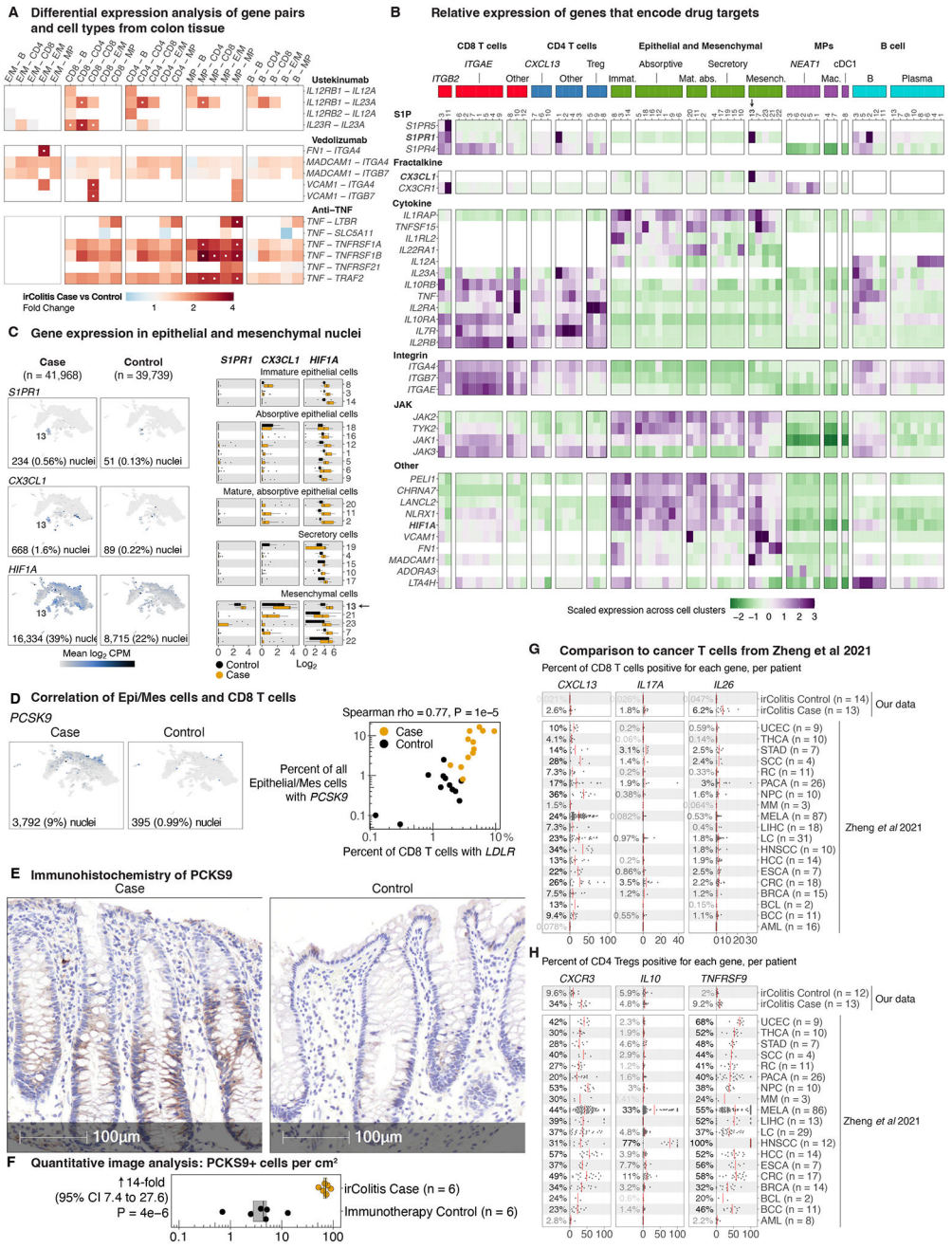


Figure 6. Identification of putative therapeutic targets to treat irColitis and comparison of irColitis to diverse tumor immune microenvironments

(A) Heatmap color depicts case versus control DGE fold-change for pairs of genes encoding targets of monoclonal antibodies for IBD (ustekinumab, anti-TNF, vedolizumab) in pair of cell lineages depicted in each column (E/M: epithelial and mesenchymal nuclei, B: B cells, MP: mononuclear phagocytes, CD8: CD8 T/gamma delta T/NK cells, CD4: CD4 T cells). White dots indicate FDR <5%. **(B)** Relative expression of candidate genes encoding IBD drug across biological themes labeled on the left (Table S13). Columns depict cell clusters organized by color-coded cell lineages – CD8 T/GDT/NK cells (red), CD4 T cells (blue), epithelial and mesenchymal cell types (green), MP cells (purple), and B cells

(turquoise). Color scheme indicates scaled expression across cell types (pooled results from cases and controls). **(C)** Feature plots show gene expression (log_2CPM) levels in cases (left) and controls (right) projected on Fig. 5C UMAP, highlighting endothelial cell cluster 13. Number and percentage of nuclei expressing each gene reported on each plot. Right-most panels show expression (log_2CPM) of indicated genes. Controls in black, cases in orange. Each dot represents a patient, box plots show median and interquartile range. Composition of each donor is reported as a percent of nuclei from each patient in each cluster; box plots show median and interquartile range. **(D)** *PCSK9* expression in nuclei from cases and controls (left). Percent of cells from each patient with *PCSK9* in epithelial/mesenchymal cells and *LDLR* in CD8 T cells; each dot represents a patient (right). Two-sided Spearman *p*-value. **(E)** Immunohistochemistry staining of PCSK9 and hematoxylin in one representative case and one control that is **(F)** quantified in cases and controls, box plots show median and interquartile range. Each dot represents one patient, two-sided t-test *p*-values. **(G)** Expression of *CXCL13, IL17A*, and *IL26* in CD8 T cells and **(H)** *CXCR3*, *IL10*, and *TNFRSF9* in Tregs across 19 tumor types: melanoma (MELA), B-cell lymphoma (BCL), renal carcinoma (RC), nasopharyngeal carcinoma (NPC), thyroid carcinoma (THCA), squamous cell carcinoma (SCC), multiple myeloma (MM), lung cancer (LC), breast cancer (BRCA), head and neck squamous cell carcinoma (HNSCC), uterine corpus endometrial carcinoma (UCEC), hepatocellular carcinoma (HCC), basal cell carcinoma (BCC), pancreatic cancer (PACA), ovarian cancer (OV), stomach adenocarcinoma (STAD), esophageal cancer (ESCA), colorectal cancer (CRC), and cholangiocarcinoma (CHOL).