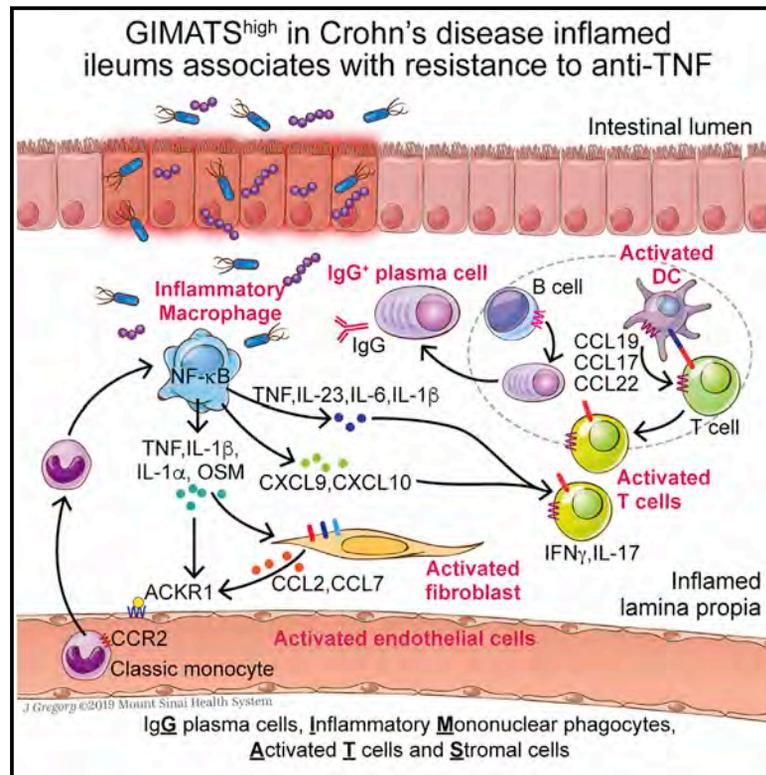


# Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy

## Graphical Abstract



## Authors

Jerome C. Martin, Christie Chang, Gilles Boschetti, ..., Miriam Merad, Judy H. Cho, Ephraim Kenigsberg

## Correspondence

miriam.merad@mssm.edu (M.M.), judy.cho@mssm.edu (J.H.C.), ephraim.kenigsberg@mssm.edu (E.K.)

## In Brief

Single-cell analysis of inflamed tissues from Crohn's patients demonstrates the existence of two qualitatively distinct subsets of disease, with distinct responses to anti-TNF therapy.

## Highlights

- Identification of a cellular module called GIMATS in a subset of CD patients
- The GIMATS module refers to IgG PCs, inf. MNPs, and activated T and stromal cells
- GIMATS organization is driven by a unique MNP-dependent cytokine/chemokine network
- GIMATS associates with failure to achieve durable remission upon anti-TNF therapy



# Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy

Jerome C. Martin,<sup>1,2,3,18</sup> Christie Chang,<sup>1,2,3</sup> Gilles Boschetti,<sup>1,2,3</sup> Ryan Ungaro,<sup>4</sup> Mamta Giri,<sup>5</sup> John A. Grout,<sup>1,2,3</sup> Kyle Gettler,<sup>5</sup> Ling-shiang Chuang,<sup>5</sup> Shikha Nayar,<sup>5</sup> Alexander J. Greenstein,<sup>6</sup> Marla Dubinsky,<sup>7</sup> Laura Walker,<sup>8</sup> Andrew Leader,<sup>1,2,3</sup> Jay S. Fine,<sup>9</sup> Charles E. Whitehurst,<sup>9</sup> M Lamine Mbow,<sup>9</sup> Subra Kugathasan,<sup>10</sup> Lee A. Denson,<sup>11</sup> Jeffrey S. Hyams,<sup>12</sup> Joshua R. Friedman,<sup>13</sup> Prerak T. Desai,<sup>13</sup> Huaibin M. Ko,<sup>4,14</sup> Ilaria Laface,<sup>1,2,8</sup> Guray Akturk,<sup>1,2,8</sup> Eric E. Schadt,<sup>15,16</sup> Helene Salmon,<sup>1,2,3</sup> Sacha Gnjjatic,<sup>1,2,8,17</sup> Adeeb H. Rahman,<sup>1,8,15</sup> Miriam Merad,<sup>1,2,3,8,19,20,\*</sup> Judy H. Cho,<sup>5,18,\*</sup> and Ephraim Kenigsberg<sup>1,15,16,19,\*</sup>

<sup>1</sup>Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>2</sup>Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>3</sup>Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>4</sup>The Dr. Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York City, NY 10029, USA

<sup>5</sup>Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>6</sup>Department of Colorectal Surgery, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>7</sup>Department of Pediatrics, Susan and Leonard Feinstein IBD Clinical Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>8</sup>Human Immune Monitoring Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>9</sup>Boehringer Ingelheim Pharmaceuticals, Immunology and Respiratory Diseases Research, Ridgefield, CT 06877, USA

<sup>10</sup>Division of Pediatric Gastroenterology, Emory University School of Medicine, Atlanta, GA, USA

<sup>11</sup>Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Cincinnati Children's Hospital Medical Center and the University of Cincinnati College of Medicine, Cincinnati, OH, USA

<sup>12</sup>Division of Digestive Diseases, Hepatology, and Nutrition, Connecticut Children's Medical Center, Hartford, CT, USA

<sup>13</sup>Janssen Research and Development, Spring House, PA, USA

<sup>14</sup>Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>15</sup>Department of Genetics and Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>16</sup>Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>17</sup>Department of Medicine, Division of Hematology Oncology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>18</sup>Present address: CRTI UMR1064, ITUN, CHU Nantes Hôtel Dieu, 30, Boulevard Jean Monnet, 44093 Nantes Cedex 1, France

<sup>19</sup>These authors contributed equally

<sup>20</sup>Lead Contact

\*Correspondence: [miriam.merad@mssm.edu](mailto:miriam.merad@mssm.edu) (M.M.), [judy.cho@mssm.edu](mailto:judy.cho@mssm.edu) (J.H.C.), [ephraim.kenigsberg@mssm.edu](mailto:ephraim.kenigsberg@mssm.edu) (E.K.)

<https://doi.org/10.1016/j.cell.2019.08.008>

## SUMMARY

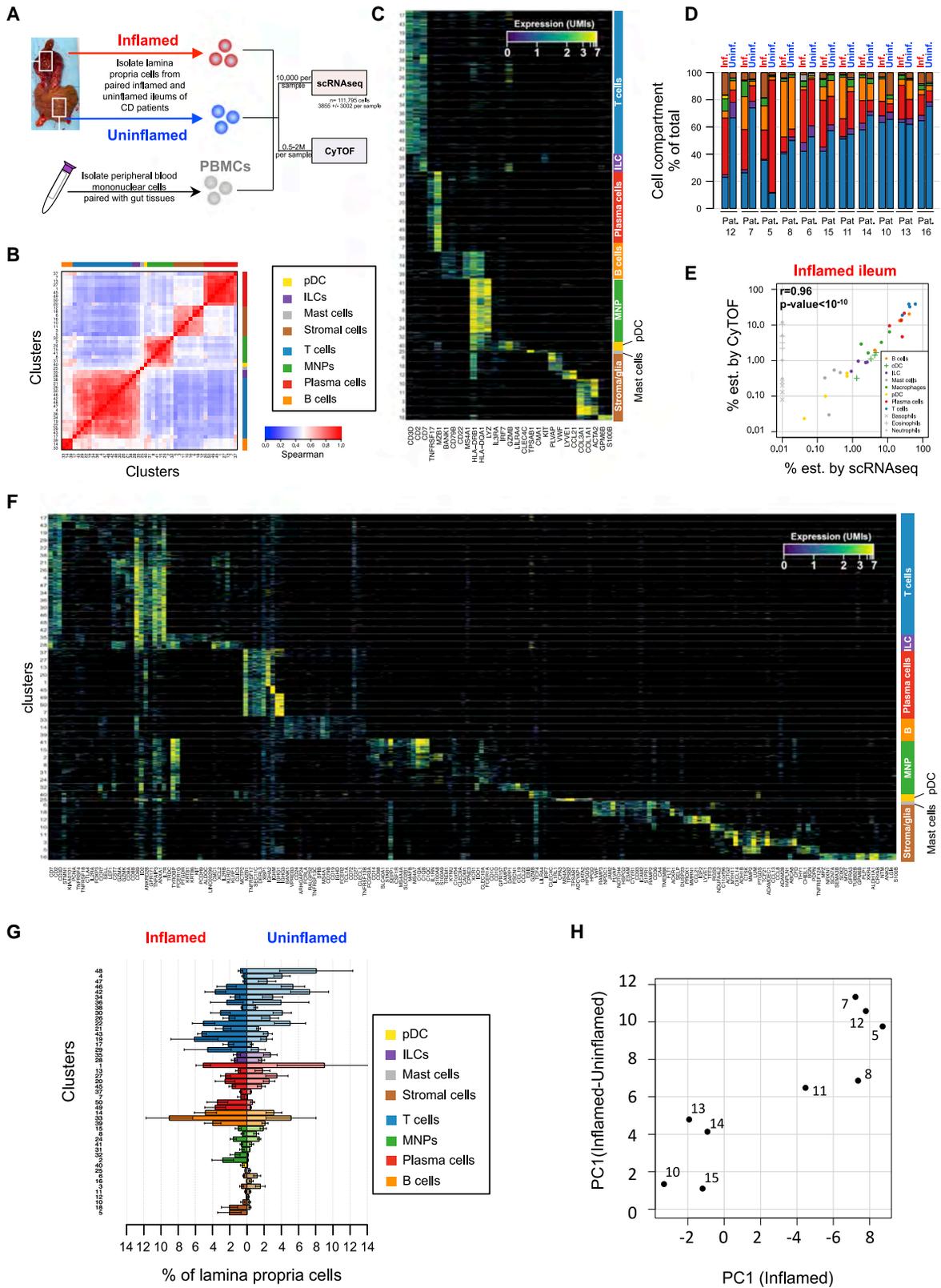
Clinical benefits of cytokine blockade in ileal Crohn's disease (iCD) are limited to a subset of patients. Here, we applied single-cell technologies to iCD lesions to address whether cellular heterogeneity contributes to treatment resistance. We found that a subset of patients expressed a unique cellular module in inflamed tissues that consisted of IgG plasma cells, inflammatory mononuclear phagocytes, activated T cells, and stromal cells, which we named the GIMATS module. Analysis of ligand-receptor interaction pairs identified a distinct network connectivity that likely drives the GIMATS module. Strikingly, the GIMATS module was also present in a subset of patients in four independent iCD cohorts (n = 441), and its presence at diagnosis correlated with failure to achieve durable corticosteroid-free remission upon anti-TNF therapy. These results emphasize the limitations of current diagnostic assays and the potential for single-cell mapping tools to identify novel bio-

markers of treatment response and tailored therapeutic opportunities.

## INTRODUCTION

Inflammatory bowel disease (IBD), which comprises Crohn's disease (CD) and ulcerative colitis, is characterized by intermittent chronic inflammation of the gastrointestinal tract, leading to bowel damage and disabilities (Torres et al., 2017). IBD results from the complex interplay of Westernized lifestyle-associated environmental factors and genetic susceptibilities, culminating in uncontrolled immune responses against luminal triggers (Kaser et al., 2010). Genome-wide association studies (GWASs) have identified more than 200 IBD-associated loci that can be organized into regulatory networks enriched for immune and inflammatory processes (Jostins et al., 2012; Liu et al., 2015). In order to design new drugs targeting immune mediators specifically involved in IBD lesions, numerous efforts combining human tissue analyses and rodent colitis models have attempted to dissect the key cellular and molecular modules of intestinal inflammation (Neurath, 2017; de Souza and Focchi, 2015). The observation that therapeutic responses to immune biotherapies





(legend on next page)

have been limited to a subset of patients, however, suggests that similar clinical phenotypes can emerge from distinct inflammatory signatures (Abraham et al., 2016; Danese et al., 2016). Current approaches restricted to well-established antibody panels based on prior knowledge preclude the identification of novel pathogenic cell populations in the diseased intestine. Recent significant advances of single-cell sequencing technologies allow the characterization of human lesional tissues at high resolution (Jaitin et al., 2014; Macosko et al., 2015; Klein et al., 2015; Zheng et al., 2017a; Azizi et al., 2018). In this study, we sought to map the cellular landscape of inflamed ileum lesions, adjacent non-inflamed ileum, and matched circulating blood cells of ileal Crohn's disease (iCD) patients to help dissect disease heterogeneity among patients and identify the underlying cellular and molecular events that may control disease outcome and response to treatment.

## RESULTS

### High-Resolution Cell-type Mapping of Inflamed and Uninflamed Ileum in Crohn's Disease

Lamina propria cells were isolated from paired uninflamed and inflamed biopsies obtained from surgically resected ileal tissues from 11 iCD patients (Figure 1A; Table S1). Single-cell transcriptomes were isolated from 22 ileal specimens, and unique molecular identifier (UMI) counts matrices were generated (Zheng et al., 2017b) (Table S2, sheet 1; STAR Methods). After exclusion of epithelial and red blood cells as well as cells not passing quality controls (Figures S1A–S1C), 82,417 lamina propria cells from the 22 samples (Figure S1D) were clustered jointly. Based on our previous work, we used an expectation maximization (EM)-like clustering algorithm, which iteratively learns the gene expression profiles of the different cell populations while estimating batch-specific background noise rates (Figures S1E–S1I; STAR Methods) (Jaitin et al., 2014; Paul et al., 2015). The clustering analysis revealed 47 clusters with variable number of cells (157–6,944 cells) (Figure S1J) and UMI counts per cell (Figure S1K). All clusters included cells from multiple patients, suggesting that cells were grouped according to shared lamina

propria-induced program rather than patient specificity (Table S2, sheet 2). Expression profiles and raw single-cell RNA sequencing (scRNA-seq) data are publicly available through an online application for data analysis allowing the interactive multi-dimensional exploration of the different transcriptional programs (<https://scDissector.org/martin>).

The 47 expression profiles were grouped by hierarchical clustering (Figure 1B), leading to the characterization of major cellular compartments of stromal-glia cells and seven distinct immune cell lineages consisting of T cells, innate lymphoid cells (ILCs), B cells, plasma cells (PCs), mononuclear phagocytes (MNPs), plasmacytoid dendritic cells (pDCs), and mast cells (Figure 1C; Table S2, sheet 3). The fractions of the eight major compartments were estimated by the scRNA-seq data in individual uninflamed and inflamed samples (Figure 1D). Importantly, estimated frequencies of the seven immune cell lineages in individual samples were confirmed in four patients by paired cytometry by time-of-flight (CyTOF) analysis of inflamed ( $r = 0.96$ ;  $p < 10^{-10}$ ; Figure 1E) and uninflamed ( $r = 0.93$ ;  $p < 10^{-10}$ ; Figure S1L) tissues. Notably, scRNA-seq sensitivity to estimate rare subpopulations with frequencies as low as 0.2% in individual samples such as pDCs was verified (Figures 1E, S1L, and S1M). In agreement with previous studies, granulocyte populations were not captured by 10× Chromium scRNA-seq, while they were identified by CyTOF (Figure S1N) (Ordovas-Montanes et al., 2018). Deeper transcriptomic characterization revealed remarkable heterogeneity within the T, B, ILCs, PCs, MNPs, and stromal compartments corresponding to distinct cell subtypes and/or cell states (Figure 1F), with frequencies varying between 0.2% and 9% of inflamed or uninflamed samples (Figure 1G). Patients 6 and 16 were excluded from comparative analyses of cell-subtype frequencies because of low cell number recovery (patient 6) and high similarity between uninflamed and inflamed scRNA-seq data (patient 16) (Figure S1O). Principal-component analysis (PCA) of normalized subtype frequencies demonstrated that, while inflamed tissues followed a similar trend of changes from their paired uninflamed counterpart ( $t = -4.96$ ;  $p = 0.001$ ), a second layer of variability existed between inflamed ileums (Figures 1H and S1P).

### Figure 1. High-Resolution Cell-type Mapping of Inflamed and Uninflamed Ileum in Crohn's Disease

- (A) Workflow showing the processing of freshly collected blood and surgical resections, including paired inflamed and uninflamed tissues of ileal Crohn's disease patients for scRNA-seq and CyTOF.
- (B) Grouping scRNA-seq cell clusters by similarity. Heatmap showing Pearson correlation coefficients between the log-averaged expression profiles of clusters. Ordering was determined by hierarchical clustering (see STAR Methods), which unbiasedly grouped the clusters by cellular lineage (color-coded bar).
- (C) Single-cell expression of key cell-type markers across the eight cellular compartments. Heatmap visualization color-coding the mRNA (UMI) counts per single cells (stacked rows) for selected marker genes (columns). Visualized are 100 randomly selected cells per cluster, which were downsampled to 2,000 UMIs/cell. Clusters are separated by gray bars and ordered by lineage (color-coded labels as in B).
- (D) Percentages of cellular lineages in individual samples included in the scRNA-seq analysis. For each sample, a bar graph depicts the percentage of cells in clusters associated with each lineage (lineages are labeled with the same color code as in B and C).
- (E) CyTOF validation of frequencies estimated by scRNA-seq. Frequencies (log-scale) of the seven immune lineages as estimated by CyTOF versus scRNA-seq in four inflamed iCD ileums. Frequencies of granulocyte populations as determined by CyTOF are also shown but were excluded from the correlation analysis.
- (F) High-resolution characterization of cellular subtypes and states. Heatmap for visualization of single-cell expression data, similar to (C) but showing an extended gene list demonstrating the distinct cellular states within the lineages captured by scRNA-seq (lineages are labeled with the same color code as in B, C, and D).
- (G) Percentages of individual scRNA-seq clusters in nine inflamed and nine uninflamed ileums (mean  $\pm$  SEM) (patients 5, 7, 8, 10, 11, 12, 13, 14, and 15).
- (H) Principle-component analysis of cellular-subtype composition. First component values were extracted by principal-component analysis of adjusted cell-subtype frequencies of variable subtypes (see STAR Methods). Shown is the PC1 difference between matched inflamed and uninflamed samples (y axis) versus PC1 inflamed (x axis). Positive y axis values indicate shared inflamed versus uninflamed component. Patient identification numbers are indicated.
- See also Figure S1 and Tables S1 and S2, sheets 1–3.



### High-Resolution Analyses of Immune and Stromal Cells Reveal Distinct Cellular Organization in Inflamed Ileums of Crohn's Disease Patients

High-resolution characterization of immune and stromal cells of iCD lesions revealed major differences between inflamed and uninfamed ileal tissues that involved subtypes in different cellular compartments (Figures 1G, 1H, and S1Q). These included MNPs comprising conventional DCs and macrophages. The former shared a gene program that included the DC hallmark receptor *FLT3*, while the latter expressed the “core macrophage gene signature,” which contains *MERTK*, *CTSC*, *CTSD*, *GLUL*, and *PLD3* (Gautier et al., 2012), in addition to known membrane markers not expressed by DC (*CD14*, *CD68*, and *FCGR3A*) (Figure 2A). By analyzing gene modules defined by strong gene-to-gene correlation of expression within total myeloid cells (Figure S2A; Table S2, sheet 4), we further identified two subtypes of macrophages (Figure 2B). The first expressed high levels of the macrophage transcription factor (TF) *MAFB*, M-CSF receptor (*CSF1R*) (Lavin et al., 2014), and genes required for apoptotic cell clearance (*DNASE1L3*, *C1Qs*) (Sisirak et al., 2016). We annotated these cells as gut-resident macrophages as they lacked pro-inflammatory genes and expressed markers of tissue residency, notably the macrophage mannose receptor *CD206* (*MRC1*) (Mowat et al., 2017). The second subtype was defined as inflammatory macrophages (infl. macs), as supported by the increased expression of nuclear factor  $\kappa$ B (NF- $\kappa$ B) and a number of inflammatory molecules (Figure 2B).

We identified two clusters of conventional DCs, including *CD1c*<sup>+</sup>/*DC2* (expressing *CD1C*, *CD1D*, *CLEC10A*, *FCER1A*, *FCGR2B*) and *CD141*<sup>+</sup>/*DC1* (expressing *BATF3*, *CLEC9A*, *XCR1*, *CADM1*, *CLNK*, *C1orf54*) (Figure 2C), consistent with high-resolution single-cell analyses in recent studies (Guilliams et al., 2016; See et al., 2017; Villani et al., 2017). A third DC cluster annotated as activated conventional DCs expressed transcripts enriched in migratory DCs on their way to tissue draining lymph node (DLN) and included the activation marker DC-LAMP

(*LAMP3*), co-stimulatory molecules (*CD40*, *CD86*), checkpoint molecule PD-L1, and the chemokine receptor *CCR7* (Figure 2C). A fourth MNP cluster expressed DC genes, including *FLT3* and the DC2 markers *CD1c* and *FCER1A* together with *CD207* (lan-gerin), the TF *MAFB*, and *MRC1* (*CD206*) as well as additional genes induced in monocyte-derived DCs (moDCs) cultured with GM-CSF and IL-4 such as *CD209* (DC-SIGN), *SEPP1*, *STAB1*, and *DAB2* (Sander et al., 2017) (Figures 2A and 2C). moDC-like cells were present in uninfamed tissues, suggesting that their differentiation was not driven by an inflammatory milieu but rather referred to a population present in the intestine under homeostasis. Infl. macs and activated conventional DCs expressed distinct cytokine and chemokine patterns (Figure 2D). Infl. macs preferentially expressed neutrophil-attracting chemokines (*CXCL2*, *CXCL3*, *CXCL8*) and cytokines essential for lymphocytes and stromal-cell activation like IL-23, IL-6, IL-1 $\beta$ , IL-1 $\alpha$ , TNF, and OSM (Neurath, 2014; Powell et al., 2015; West et al., 2017), while activated DCs were enriched in adaptive lymphocyte-attracting chemokines *CCL17*, *CCL22*, and *CCL19*. MNP subsets were confirmed by CyTOF analysis (Figures S2B–S2G). Within *CD3*<sup>−</sup> *CD19*<sup>−</sup> *HLA-DR*<sup>+</sup> lamina propria cells, macrophages were distinguished from DCs based on the expression of *CD14* and further separated into *CD206*<sup>high</sup>-resident macrophages and *CD206*<sup>low/−</sup> infl. macs. Activated DCs were identified by their strong co-expression of *CD86* and PD-L1 within total *CD14*<sup>−</sup> *HLA-DR*<sup>+</sup> *CD11c*<sup>high</sup> DCs. Within non-activated DCs, *DC1* were identified based on the expression of *CD141* (Merad et al., 2013). Finally, *CD1c*<sup>+</sup> DCs were separated into *CD206*<sup>−</sup> *DC2* and *CD206*<sup>+</sup> moDC.

B lymphocytes comprised naive and memory B cell populations, as well as plasmablasts and PCs producing different classes of immunoglobulins including IgA, IgM, and IgG (Figures S3A and S3B; STAR Methods). To group and annotate T cell clusters, we calculated scores summarizing the mRNA fraction per cell of genes associated with cell cycling, immunoregulation, naive and/or central memory, CD8-cytotoxic, and resident

#### Figure 2. High-Resolution Analyses of Immune and Stromal Cells Reveal Distinct Cellular Organization in Inflamed Ileums of Crohn's Disease Patients

(A) Expression profiles distinguish macrophages and dendritic cells. Heatmap showing relative expression values of genes (columns) enriched in macrophages or dendritic cells clusters (rows). Relative expression was defined as log ratio (2-based) of the expression of the gene in a cluster and its average across all shown clusters.

(B and C) Distinct transcriptional programs among MNP subtypes. Heatmaps showing color-coded downsampled UMI counts of highly variable genes between macrophage clusters (B) and between dendritic cell clusters (C). Clusters are demarcated by gray bars. For each cluster, 300 cells were randomly selected and downsampled to 2,000 UMIs/cell.

(D) Heatmap showing relative expression of cytokines and chemokines (columns) detected at increased levels in MNP clusters (rows) across all inflamed and uninfamed samples.

(E) Capturing T cell identities by transcriptional scores. Heatmap showing the relative expression of five transcriptional scores (rows) in single T cells (columns). Scores integrate expression of multiple highly correlated genes (STAR Methods). Cell clusters are demarcated by gray bars.

(F) Heatmap showing relative expression of cytokines and chemokines (columns) increased in T cell and ILC clusters (rows) across all inflamed and uninfamed samples.

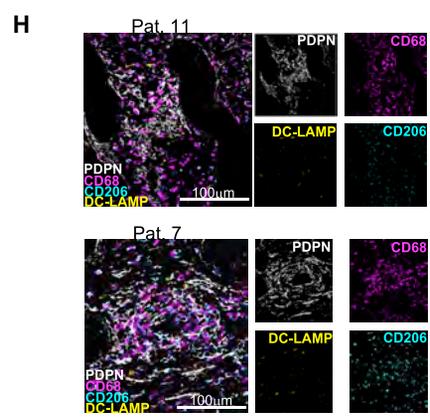
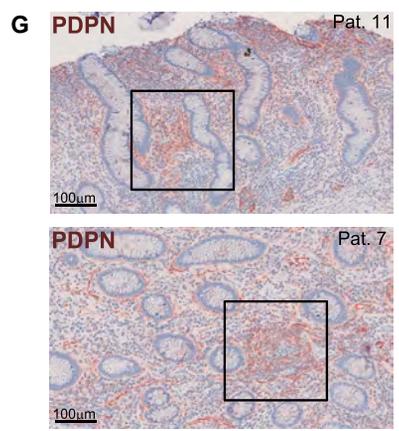
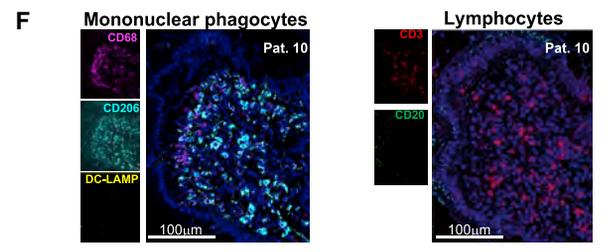
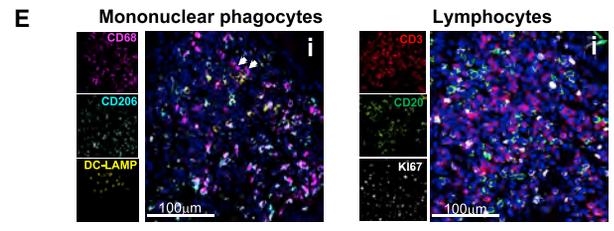
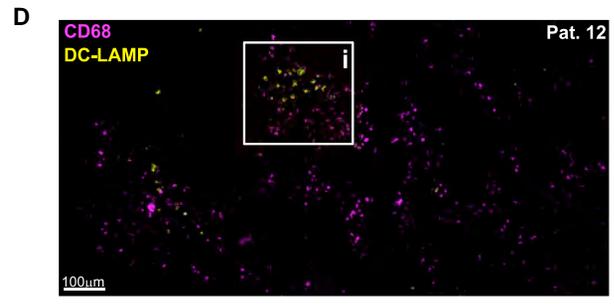
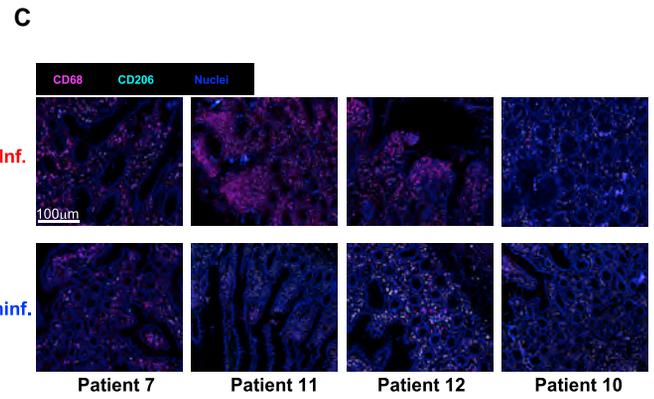
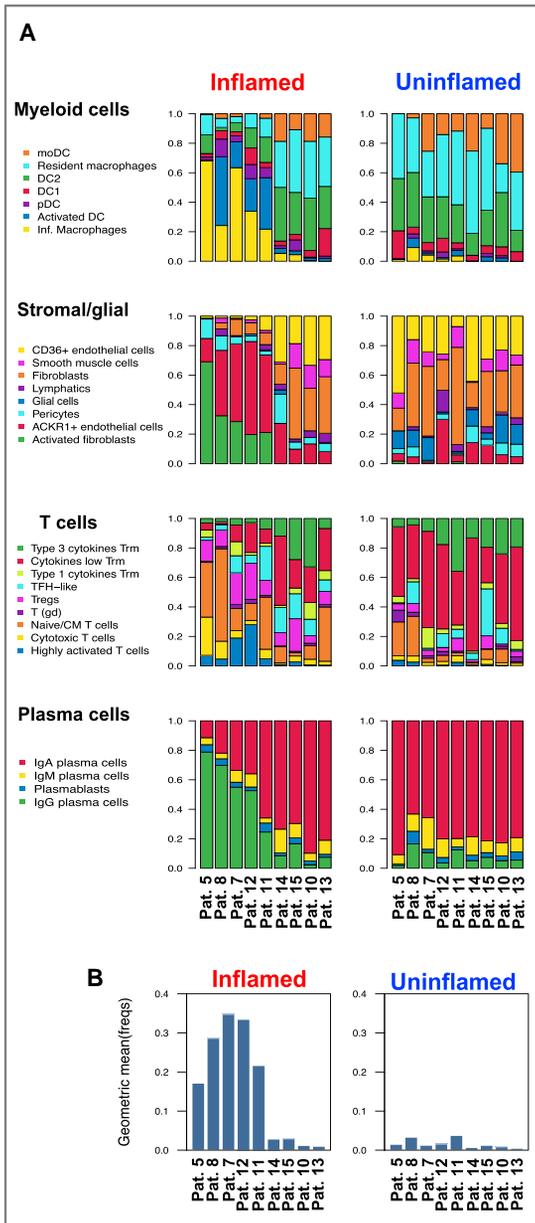
(G) Stromal and glial diversity in inflamed and uninfamed ileum. Heatmap showing color-coded downsampled UMI counts of highly variable genes (columns) between single cells (rows) of different stromal subtypes. For each cluster, 100 cells were randomly selected and downsampled to 2,000 UMIs/cell.

(H) Heatmap showing relative expression of cytokines and chemokines (columns) increased in stromal-cell clusters (rows) across all inflamed and uninfamed samples.

(I) Diversity of myeloid cells, T cells, plasma cells, and stromal cells in inflamed versus uninfamed regions. Shown are stacked frequencies of each subtype divided by the respective total compartment frequency estimated by scRNA-seq in uninfamed or inflamed ileum.

(J) Co-segregation of cell subtypes across inflamed tissues. Pearson correlation between the normalized frequencies of the lamina propria cell subtypes in CD-inflamed ileums ( $n = 9$ ). Subtypes are reordered by hierarchical clustering (STAR Methods).

See also Figures S2–S4 and Table S2, sheets 4 and 5.



(legend on next page)

memory states (Figures 2E, S3C, and S3D; Table S2, sheet 5; STAR Methods). Detailed transcriptomic examination further refined T cell subtypes within these states (Figure S3E; STAR Methods). Almost no cytokine-chemokine expression was detected in naive T cells (cluster 19), and transcripts for type 1 (TNF, IFN $\gamma$ ) and type 3 (IL-17A, IL22) cytokines were detected at low levels in central memory (CM) T cells (cluster 29) (Figure 2F). A cluster of highly activated T cells strongly enriched in cycling genes, HLA-DR, and CD38 transcripts and expressing low levels of cytokine-chemokine transcripts was present (Figures 2D&E). Tissue-resident memory T cells (Trm) expressing high levels of IL-17A and IL-22 transcripts, or TNF and IFN $\gamma$ , also high in cytotoxic T lymphocytes (CTLs), were also identified (Figure 2F; STAR Methods). Cytokine-chemokine expression profiles were dominated by CXCL13 in T follicular helper-like (TFH-like) cells and by IL-10 in regulatory T cells (Tregs), which also expressed low levels of IL-17A, as observed previously in iCD lesions (Hovhannisyan et al., 2011). Group 1 and group 3 ILCs (Figures S3F and S3G; STAR Methods) expressed IFN $\gamma$  and IL-22, respectively (Figure 2F).

Stromal cells included endothelial cells, fibroblasts, pericytes, and smooth muscle cells (Figure 2G). We identified two subsets of endothelial cells that shared the expression of the vascular endothelial growth factor receptor 2 (*KDR*) and included lymphatic endothelial cells, defined by high levels of the lymphatic vessel endothelial hyaluronan receptor 1 (*LYVE1*) and the CCR7 ligand *CCL21*, and blood endothelial cells co-expressing *PECAM1* (CD31) and Von Willebrand factor *VWF*. One subset of blood endothelial cells expressed high levels of the atypical chemokine receptor 1 (*ACKR1*) and adhesion molecules *SELE* (E-selectin) and *SELP* (P-selectin) and was annotated as activated endothelial cells (Aird, 2007). Two clusters expressed high levels of *MCAM* (CD146) and were respectively identified as pericytes (expressing *CSPG4* [NG2] and *RGS5*) (Turley et al., 2015) and smooth muscle cells, as suggested by their expression *ACTG2* (actin gamma-2 enteric smooth muscle), *MYH11* (myosin heavy chain 11), and *DES* (desmin). Finally, fibroblasts were characterized by their expression of the two sub-

units of the platelet-derived growth factor receptors (*PDGFRA*, *PDGFRB*), the prostaglandin-H2 D-isomerase *PTGDS*, and several genes encoding extracellular matrix (ECM) proteins such as lumican (*LUM*) and ECM-remodeling proteins (e.g., matrix metallo-proteinase-2 [*MMP2*]). One of the two subtypes of fibroblasts expressed a hallmark activation program including strong expression of *THY1* (CD90), *PDPN* (podoplanin), *CTHRC1*, and *CHI3L1* (Turley et al., 2015). With increased *IL6*, pro-fibrotic *IL11*, and neutrophil-attracting *CXCL2*, *CXCL8*, *CXCL1*, and *CXCL5*, as well as CCR2<sup>+</sup> monocyte-recruiting *CCL2* and *CCL7*, activated fibroblasts exhibited a marked shift in their cytokine-chemokine expression profile compared to other fibroblasts (Figure 2H). The last cluster of non-hematopoietic cells (cluster 16) corresponded to *S100B* expressing glial cells (Joseph et al., 2011).

By combining clusters with shared annotations (e.g., the two clusters of resident macrophages or the three clusters of PCs expressing IgGs), we defined a total of 33 different cellular subtypes in the ileal CD lamina propria (Figures 2I and S4). To study the association between the different transcriptional profiles, we quantified the subtypes' frequencies within each cellular compartment. Correlation analysis revealed the presence of a group of cell subtypes with highly correlated frequencies across inflamed ileums. This cellular module included infl. macs, activated DCs, highly activated T cells, IgG PCs, activated fibroblasts, and ACKR1<sup>+</sup>-activated endothelial cells, which we named the GIMATS module (IgG PCs, inflammatory MNPs, and activated T and stromal cells) (Figure 2J).

### Identification of a Unique Cellular Signature in a Subset of Ileal Crohn's Disease Patients

Examination of the cellular-subtype proportions within the different cellular compartments revealed that the GIMATS module was present only in a subgroup of patients (patients 5, 7, 8, 11, 12) (Figure 3A). We demonstrated the presence of a distinct cellular response in this subgroup of patients by comparing the GIMATS module intensity score, defined as the geometric mean of the frequencies of the GIMATS module cellular

#### Figure 3. Identification of a Unique Cellular Signature in a Subset of Ileal Crohn's Disease Patients

- (A) Independent analyses of cell-subtype frequencies within the different cellular compartments. The number of cells in each cellular subtype is divided by the total number of cells in its cellular compartment and visualized in a stacked bar for each sample. Each of the four cellular compartments (MNPs, T cells, plasma cells, stromal-glial cells) is shown separately in inflamed and uninfamed samples of each patient. Patients are ordered by hierarchical clustering (see STAR Methods) of frequency data in inflamed ileum. This unbiased visualization exposes the GIMATS module signature in a subgroup of patients.
- (B) Striking differences in GIMATS module intensity between patient groups. Shown is the geometric mean of the normalized frequencies of the GIMATS module subtypes (IgG plasma cells, infl. macs, activated DCs, activated T cells, activated fibroblasts, and ACKR1<sup>+</sup> endothelial cells) in each sample.
- (C) Representative pictures of macrophage staining in the inflamed (top) and uninfamed (bottom) lamina propria of patients enriched (nos. 7, 11, 12) or not enriched (no. 10) for the GIMATS module.
- (D) Representative picture of CD68 and DC-LAMP staining in the inflamed lamina propria of a patient enriched for the GIMATS module.
- (E) Higher magnification of white square (i) in (D) showing an activated DC-associated lymphocyte aggregate. Left: MNPs were stained with CD68, CD206, and DC-LAMP. White arrows indicate CD68<sup>+</sup> CD206<sup>-</sup> macrophages in close contact to CD68<sup>-</sup> DC-LAMP<sup>+</sup>-activated DCs. Right: staining of T cells (CD3), B cells (CD20), and the proliferation marker Ki67.
- (F) T cell-dense area surrounded by subepithelial CD68<sup>+</sup> CD206<sup>-</sup>-resident macrophages. Left: MNPs were stained with CD68, CD206, and DC-LAMP. CD68<sup>+</sup> CD206<sup>-</sup> macrophages and CD68<sup>-</sup> DC-LAMP<sup>+</sup>-activated DCs were not detected. Right: T (CD3) and B (CD20) lymphocytes were stained. CD20<sup>+</sup> B cells were not detected.
- (G) Representative podoplanin (PDPN) staining by immunohistochemistry in the inflamed lamina propria of two patients enriched for the GIMATS module. Black squares depict areas enriched in PDPN<sup>+</sup>-activated fibroblasts in the lamina propria.
- (H) CD68, CD206, DC-LAMP, and PDPN MICSSS imaging of areas enriched in activated fibroblasts (black squares in G), showing preferential localization of activated fibroblast in the vicinity of CD68<sup>+</sup> CD206<sup>-</sup> infl. macs over CD68<sup>-</sup> DC-LAMP<sup>+</sup>-activated DCs.

See also Figure S5.



subtypes. The intensity score was significantly higher in patients 5, 7, 8, 11, and 12 compared to patients 10, 13, 14, and 15 inflamed samples (Figure 3B;  $t$  test  $t = 7.36$ ;  $p = 0.001$ ). CyTOF analysis of seven iCD patients, including four with paired scRNA-seq, confirmed the bimodal distribution of the GIMATS module in iCD patients (Figure S5A). Importantly, we found that the presence of the GIMATS module in patients was independent of pathology severity, disease duration, and systemic markers of inflammation (Figure S5B). Imaging of gut tissues using multiplexed immunohistochemical consecutive staining on a single slide (MICSSS), a multiplex imaging technique we recently developed (Remark et al., 2016), confirmed the accumulation of CD68<sup>+</sup> CD206<sup>-</sup> infl. macs throughout the inflamed lamina propria of patients enriched for the GIMATS module (Figure 3C), as well as the presence of CD68<sup>-</sup> DC-LAMP<sup>+</sup> HLA-DR<sup>+</sup>-activated DCs, which formed dense aggregates with T and B lymphocytes (Figures 3D, 3E, S5C, and S5D for immunohistochemical [IHC] images). Ki-67<sup>+</sup> cycling lymphocytes were enriched in activated DC-centered aggregates in patients with the GIMATS module, suggesting a role for activated DCs in the recruitment, local activation, expansion, and spatial organization of adaptive immune responses in iCD lesions. In contrast, while T cell-dense areas surrounded by subepithelial CD68<sup>+</sup>CD206<sup>+</sup>-resident macrophages were also present in ileal tissues lacking the GIMATS module, these structures were devoid of B cells and activated DCs (Figure 3F). MICSSS also confirmed the presence of activated fibroblasts expressing the mucin-type protein podoplanin (PDPN) in the inflamed lamina propria of patients enriched for the GIMATS module (Figure 3G), while PDPN was restricted to lymphatic vessels in the uninfamed lamina propria (Figure S5E). Interestingly, CD68<sup>+</sup> CD206<sup>-</sup> infl. macs were always in the close vicinity of PDPN<sup>+</sup>-activated fibroblasts (Figure 3H), suggesting a potential role for infl. macs in the activation of fibroblasts.

### The GIMATS Module Is Organized by a Specific Cytokine-Chemokine Network

Taken together, our results identified a unique cellular organization in inflamed tissues of a subset of patients, thus revealing different pathogenic responses between patients despite similar pathological severity and systemic inflammatory markers. We examined whether specific ligand-receptor pairs could inform

the pathophysiological response that drives the cellular organization of the GIMATS module, restricting our analysis to cytokine-chemokine-receptor pairs validated experimentally (Ramilowski et al., 2015). Using scRNA-seq data, we defined a ligand-receptor activity network for patients enriched or lacking the GIMATS module. Edges in the network were defined by the globally normalized expression of the ligand expressed by the source cell type and the averaged gene expression of the receptor expressed by the target cell (STAR Methods). We defined intensity scores of the pairs as the product of the ligand and receptor expression values in patients with or without the GIMATS module (Figure S6A). We visualized the adjacency matrices of the network while demarcating the fold change of intensity scores for each pair between patients with and without the GIMATS module (Figures 4A, S6B, and S6C; Table S2, sheet 6; STAR Methods).

Several ligand-receptor pairs involving receptors on T cells and ligands produced by MNPs were significantly enriched (Benjamini-Hochberg adjusted  $p < 0.01$ ; permutation test, STAR Methods) in patients with the GIMATS module (Figure 4A). DC ligands included the chemokines CCL22 and CCL17, whose receptor CCR4 was expressed by T<sub>CM</sub>, highly activated T cells, TFH-like cells, and Tregs. DC ligands also included CCL19, whose receptor CCR7 was expressed by T<sub>CM</sub>, naive T cells, and B cells. These chemokines were expressed at the highest levels in activated DCs, supporting their potential contribution to the formation of the lymphoid aggregates identified in the inflamed lamina propria of the GIMATS patients (Figures 4B and 4C). Additional chemokine-receptor pairs included CXCL9 and CXCL10 in DCs and/or macrophages, whose receptor CXCR3 was highly expressed on activated T cells, CTLs, and Tregs (Figures 4B and 4C). The intensity score of pairs involving the production of T cell-activating cytokines by DCs (*EBI3*) and macrophages (*EBI3*, IL-1 $\beta$ , IL-6, and IL-23) were significantly higher in GIMATS patients (Figure 4A), suggesting that while the initial steps of lymphocyte aggregates formation may depend on activated DC, infl. macs likely participated in T cell activation *in situ*, in agreement with their presence in DC-lymphoid aggregates (Figure 3E). Interestingly, increased intensities were also revealed between macrophage-derived ligands (IL-6, TNF, IL-1 $\alpha$ , IL-1 $\beta$ , and IL15) and receptors expressed by DCs, and vice versa (*EBI3*,

#### Figure 4. The GIMATS Module Is Organized by a Specific Cytokine-Chemokine Network

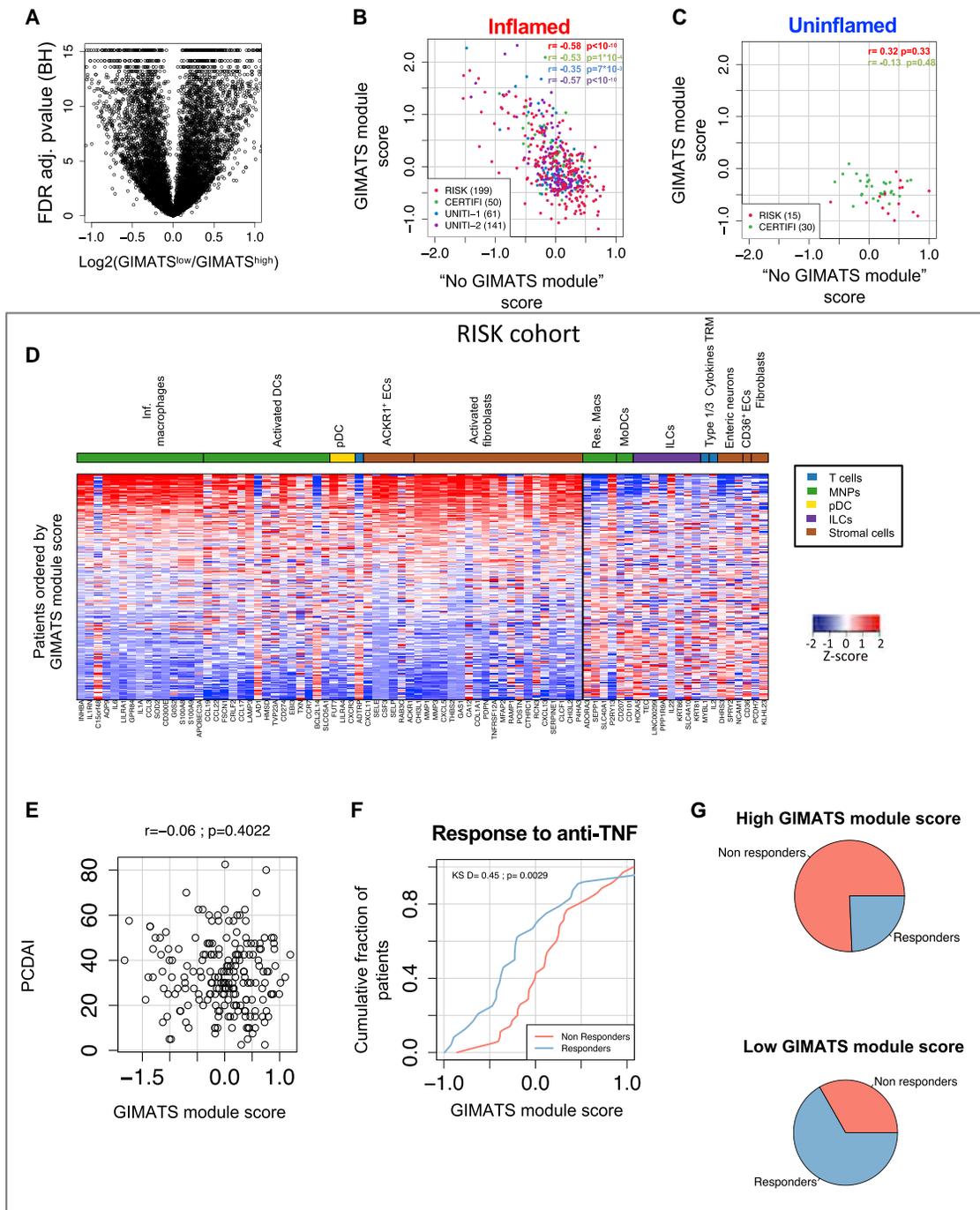
(A) Comparative ligand-receptor network analysis between the two patient groups. Shown is the log ratio between the intensity of the ligand (rows)-receptor (columns) pair in inflamed samples of patients enriched with the module versus patients not enriched with the module (color coded) for different cellular cell-type pairs. The intensity of a ligand-receptor pair is the product between the total expression of the ligand and the expression of the receptor (STAR Methods). Only validated pairs, which are expressed at least in one of the patient groups, are included. Stars indicate Benjamini-Hochberg adjusted  $p < 0.01$  as estimated by permutation test (STAR Methods).

(B and C) Ligand-receptor expression map. Heatmaps showing the expression averaged across all inflamed and uninfamed samples for genes (columns) per cell subtype (rows) of indicated cytokines or chemokines (B) and their receptors (C). Established ligand-receptor pairs are connected with lines. Colored lines highlight pairs relevant to the GIMATS module. Pink, activated DC-T<sub>CM</sub>; yellow, inf. MNP-activated T cells -CTLs -pDCs; green, infl. macs-stromal cells; blue, activated fibroblasts-ACKR1<sup>+</sup> endothelial cells.

(D) Infl. macs express a continuum of monocyte to macrophage genes. Larger heatmap shows downsampled UMI counts of program 1 and program 2 genes as well as infl. mac genes for cells of the infl. mac cluster. Averaged expression of program 1 and program 2 genes shown in parallel heatmap. Cells in both heatmaps were ordered by the ratio between averaged expression of program 1 and program 2 genes. For reference, PBMC cells mapped to cluster "Classical monocytes-1" are shown on top.

(E) Scatterplot showing frequencies of blood classical monocytes (y axis) versus intestinal inflammatory macrophages (x axis) in GIMATS module high (blue dots) and GIMATS module low (green dots) patients.

See also Figure S6 and Table S2, sheet 6.



**Figure 5. Enrichment for the GIMATS Module before Treatment Is Associated with Resistance to Anti-TNF**

(A) Differential expression analysis between pooled scRNA-seq data from inflamed samples of patients stratified by GIMATS module enrichment. Volcano plot showing for each gene the log<sub>2</sub> ratio between averages in GIMATS module high and GIMATS module low patients (x axis) and minus log<sub>2</sub> Benjamini-Hochberg adjusted p value (y axis). See STAR Methods for detailed description of the permutation test that provided the empirical p values.

(B and C) Validation of the GIMATS module in bulk RNA datasets. Scatterplots showing the projected bulk RNA microarray and sequencing data of inflamed (B) and uninfamed (C) biopsies from four Crohn’s disease cohorts onto the signature scores (see STAR Methods). The negative correlation over inflamed samples indicates enrichment of the GIMATS module in a subset of patients in each cohort.

(D) Heatmap of normalized bulk RNA-seq expression values (red, high; blue, low) of selected genes (columns) among CD-inflamed ileal biopsies (stacked rows) ordered according to their GIMATS module signature enrichment.

(E) Pediatric Crohn’s disease index (PCDAI) does not correlate with the GIMATS module score in newly diagnosed patients before anti-TNF therapy. Shown is PCDAI of RISK patients (y axis) versus GIMATS module score at diagnosis (x axis). Pearson correlation test ( $r = -0.06$ ) was not significant.

(legend continued on next page)

IL-15), suggesting bidirectional communications between infl. macs and activated DCs in inflamed tissues (Figure 4A).

Other major pairs identified included receptors on fibroblasts and endothelial cells and production of cytokines such as TNF, IL-1 $\beta$ , and IL-1 $\alpha$  by macrophages (Figure 4A). Accordingly activated fibroblasts in the GIMATS module expressed TNF- and IL-1 $\beta$ -induced genes (e.g., CXCL1, CXCL5, CXCL6, CCL2, and IL-11) (Bamba et al., 2003; Okuno et al., 2002) (Figures 2H and 4B). Oncostatin M, which was recently suggested as a major stromal activator in IBD (West et al., 2017) was also detected, but the OSM-OSMR pair between macrophages and fibroblasts or endothelial cells did not reach statistical significance (Figure 4A). No pairs between DCs and fibroblasts were significantly enriched in patients with the GIMATS module (Figure S6B). Finally, pairs involving fibroblast ligands with receptors on macrophages and DCs also showed higher intensity in patients with the GIMATS module, but no individual pair was statistically significant (Figure S6C).

Activated fibroblasts, however, strongly expressed the chemokines CCL2 and CCL7 (Figure 4B), which are known ligands for CCR2, a receptor mainly expressed by circulating classical monocytes in the blood, prompting our search for myeloid cells with a monocyte signature in patients with the GIMATS module. Deep mining of the infl. macs cluster revealed the presence of cells with two anti-correlated transcriptional programs ( $r = -0.42$ ;  $p < 10^{-10}$ ) (Figure 4D; STAR Methods). One program contained the macrophage markers *C1Qs*, *GLUL*, *APOE*, and *ACP5*, while the second program included the IBD biomarker calprotectin (*S100A8* and *S100A9*) (Konikoff and Denson, 2006), *S100A12*, and versican (*VCAM*) (Figure 4D). These genes were not expressed by cells in the resident macrophage clusters (Figure S6D), but scRNA-seq analysis of 42,983 peripheral blood mononuclear cells (PBMCs) matched to intestinal samples (Figure S6E) revealed their strong expression in blood classical monocytes (Figure 4D). In addition, increased infl. macs frequencies were significantly associated with a depletion of circulating classical monocytes in the blood of patients enriched for the GIMATS module (Figure 4E; Spearman correlation coefficient =  $-0.68$ ;  $p = 0.025$ ). These results indicated that the infl. macs cluster likely includes a continuum of inflammatory cells ranging from recently infiltrated blood monocytes to more differentiated macrophages. Importantly, we found that the pro-inflammatory IL-1 $\beta$  and OSM were most highly expressed by monocyte-like cells within the infl. macs cluster (Figure 4D). Taken together, these results suggest that in patients expressing the GIMATS module, blood classical monocytes contribute to the infl. macs cluster in inflamed tissues, where they promote stromal-cell activation, which then enhance monocyte infiltration through the production of CCR2 ligands by activated fibroblasts and ACKR1 expression by activated endothelial cells, an atypical chemokine receptor binding CCL2 and CCL7 to facilitate mono-

cyte recruitment in tissues (Pruenster et al., 2009), thereby establishing a positive feedback loop.

Remarkably, almost no ligand-receptor pairs between T cell ligands and DCs or macrophage and fibroblast target receptors were enriched in patients with the GIMATS module. Specifically, intensity scores involving major inflammatory ligands produced by T cells, such as IFN $\gamma$  and CSF2, were similar between both subgroups, while GIMATS-negative patients even presented increased score intensities of pairs involving T cell-derived IL-17A and TNF with receptors expressed on macrophages and DCs, respectively (Figure 4A). These cytokines were produced by subsets of Trm and CTLs (Figures 2F and 4B), which populate inflamed tissues of both subgroups of patients (Figure 3A). These results suggested that, while common T cell-driven inflammatory pathways were present in all patients, a distinct pathogenic response associated with activated DCs and monocyte-derived infl. macrophages dominated the lesions enriched in the GIMATS module. These results led us to hypothesize that TNF blockade may not be sufficient to modulate the inflammatory response associated with the GIMATS module.

#### Enrichment for the GIMATS Module before Treatment Is Associated with Resistance to Anti-TNF

Our results above identified a unique cellular signature we called the GIMATS module in a small cohort of surgically resected iCD ileums, prompting us to search for GIMATS expression in larger iCD cohorts ( $n = 441$ ) with accessible gene expression analysis of inflamed ileal biopsies. In line with previous studies (Puram et al., 2017), we developed a scoring function for normalized bulk RNA samples to generate two gene expression scores accounting for the transcriptional variability between lesions with either high or low expression of the GIMATS module in inflamed ileum. First, we defined gene lists associated with the relative abundance of the GIMATS module through a differential expression analysis between pooled gene expression profiles of the two subgroups of patients identified by our scRNA-seq analysis (Figure 5A; Table S2, sheet 7). We then projected bulk RNA microarray and sequencing data of endoscopic inflamed ileum pinch biopsies obtained from four independent cohorts of iCD patients (STAR Methods) onto the two signatures by averaging the gene lists' Z scores for each sample (Figure 5B). For all four cohorts, a significant negative correlation existed between the two scores in the inflamed tissues of iCD patients, whereas no negative correlations were detected in uninflamed ileums (Figures 5C and S7A), confirming the scRNA-seq results. As predicted by the scRNA-seq analysis, GIMATS module genes indicative of different GIMATS cellular subtypes were co-expressed in the bulk RNA data, validating the co-segregation of the GIMATS module subtypes in biopsies (Figures 5D and S7B). Importantly, enrichment of the GIMATS module was also present in the early stages of disease, prior to any biological therapy

(F) High GIMATS module score is associated with non-response to anti-TNF therapy. Cumulative distribution of GIMATS module scores in responders and non-responders to anti-TNF at diagnosis before treatment. Non-responders have significantly higher GIMATS module scores (Kolmogorov-Smirnov  $D = 0.42$ ;  $p = 0.006$ ). Dashed line indicates the score value corresponding to Kolmogorov-Smirnov  $D$  statistic (score =  $-0.2$ ).

(G) Pie charts showing the enrichment for responders and non-responders to anti-TNF in patients classified as GIMATS module high or low according to Kolmogorov-Smirnov  $D$  statistic.

See also Figure S7 and Table S2, sheet 7.

(RISK and UNITE-2;  $n = 340$  patients) (Feagan et al., 2016; Haberman et al., 2014; Kugathasan et al., 2017), suggesting that the GIMATS module was not eliminated by anti-TNF therapy in patients.

The introduction of anti-TNF antibodies has transformed the clinical outcome of IBD patients, but 20%–30% of iCD patients never respond to these drugs and continue to undergo uncontrolled bowel damage requiring surgical intervention within 10 years after the diagnosis (Torres et al., 2017). Since GIMATS was detected at disease initiation as well as in end-stage disease, we asked whether the GIMATS module impacted the response to anti-TNF therapy in iCD patients. We focused our analysis on patients from the RISK cohort, which is a prospective inception cohort of pediatric patients with newly diagnosed Crohn's disease (Kugathasan et al., 2017), for whom long-term clinical follow-up was available.

The RISK cohort enrolled only children older than 2 years at disease onset, and children with autosomal recessive IL-10 or IL10RA/B mutations, who develop anti-TNF-resistant lesions in the first year of life (often the first month), were excluded. Importantly, the pathogenesis of other mutations found more commonly in pediatric cohorts are quite similar to those present in adults. We selected patients who received anti-TNF antibody blockade within the first year of disease and assessed clinical remission between 18 and 24 months after diagnosis. Patients were considered responders if they achieved durable corticosteroid (CS)-free clinical remission. Importantly the GIMATS score did not correlate with the pediatric Crohn's disease activity index (PCDAI) at diagnosis (Figure 5E). Strikingly, however, the GIMATS module score at baseline showed significantly distinct distributions between non-responder and responder populations (Kolmogorov-Smirnov  $D = 0.42$ ;  $p = 0.006$ ) (Figure 5F). Expression of a high GIMATS score at baseline (prior to initiation of anti-TNF treatment) (Figure 5G) was enriched in patients unable to achieve durable CS-free clinical remission at least 6 months after anti-TNF induction (area under the curve [AUC] = 0.69; permutation test  $p = 0.005$ ;  $n = 71$ ) (Figure S7C). As was observed in the discovery cohort, patients with high or low GIMATS module score had similar markers of systemic inflammation, indicating that the GIMATS score conveys information regarding response to biologic therapy that is not provided by standard CD biomarkers (Figure S7D).

## DISCUSSION

Pathophysiological heterogeneity is thought to represent a major cause for the limited success of novel drug trials in IBD patients in the past two decades (Bilsborough et al., 2016). Using single cell technologies to map the ileal lamina propria of surgical ileal resections we identified the GIMATS module in a subset of iCD patients. We confirmed the presence of the GIMATS module in multiple bulk expression datasets and revealed that the presence of the module at diagnosis is associated with failure to achieve durable CS-free remission upon anti-TNF therapy in a pediatric inception cohort (RISK).

The GIMATS module included highly correlated cellular populations consisting of activated DCs, infl. macs, activated T cells, IgG PCs, activated fibroblasts, and endothelial cells. Activated

DCs within the GIMATS module were located within lymphoid aggregates of T and B cells and expressed high levels of the T cell ligands CCL19, CCL17, and CCL22, indicating the potential role for activated DCs in the local recruitment, activation, expansion, and spatial organization of a pathogenic adaptive immune response that included highly activated T cells and IgG PCs. Infl. macs were also present in lymphocyte aggregates containing activated DCs, and the enrichment of receptor-ligand pairs involving inflammatory cytokines in infl. macs and receptors on T cells suggested an additional role for infl. macs in shaping T cell polarization in patients with high GIMATS module scores.

In the healthy intestine, tissue-resident macrophages are epigenetically wired by local microenvironmental cues to acquire a regulatory program (Lavin et al., 2014; Mowat et al., 2017), and part of this program comprises the acquisition of inhibitory mechanisms to limit the production of inflammatory cytokines in response to inflammatory stimuli (Smythies et al., 2005, 2010). In line with these studies, we identified tissue-resident macrophages that lacked inflammatory molecules in all lesions examined. In contrast, the infl. macs cluster identified as part of the GIMATS module expressed both monocyte markers and high levels of inflammatory molecules. The presence of this cluster correlated with reduced circulating classical monocytes, consistent with previous work showing reduced levels of classical monocytes in the blood of CD patients (Thiesen et al., 2014) and suggesting that infl. macs are derived from circulating monocytes. Infl. macs within the GIMATS module also likely contributed to stromal-cell activation through the production of inflammatory cytokines TNF, IL-1 $\beta$  (Kamada et al., 2008; Schenk et al., 2007), and OSM, an IL-6 family member recently shown as a significant inducer of pathogenic stromal-cell activation in IBD tissues (West et al., 2017). Activated stromal cells within the GIMATS module produced the CCR2 ligands CCL2 and CCL7, while activated endothelial cells expressed the atypical chemokine receptor ACKR1, which binds tissue-derived CCL2 and CCL7 to facilitate monocyte transcytosis in inflamed tissues (Pruenster et al., 2009).

These data suggest that activated fibroblasts and endothelial cells within the GIMATS module facilitate the recruitment and transcytosis of blood classical monocytes to the inflamed ileum and their differentiation into pathogenic infl. macs, which in turn further amplify stromal-cell activation through the release of inflammatory cytokines. These results support and extend previous findings identifying hematopoietic-stromal interactions as a central hub in IBD pathology and emphasize the potential key role of infiltrating circulating monocytes over resident macrophages as major drivers of gut inflammatory lesions in a subset of iCD patients (West et al., 2017).

While the pathogenic GIMATS module was discovered in a small group of patients with advanced clinical and histological diseases, we found that the GIMATS module was also present in four independent iCD cohorts ( $n = 441$ ) and showed that presence of this pathogenic module at diagnosis correlated with patients not achieving durable CS-free remission during anti-TNF therapy. Incomplete control of mucosal inflammation in ileal CD lesions represents a significant risk factor for progressive bowel damage and surgical resection (Pariente et al., 2011),

suggesting that combination therapies with anti-TNF could help limit disease progression.

It is interesting that TNF was produced mainly by T cells in patients with low GIMATS module scores, while it was produced both by T cells and infl. MNPs in patients with a high module scores. The mechanisms of action of anti-TNF drugs in IBD are in fact still poorly understood, and not all TNF antagonists are effective in CD (Levin et al., 2016). TNF exists both as transmembrane and soluble forms, the latter resulting from the proteolytic cleavage of the former. TNF receptor expression is protective in dextran sodium sulfate (DSS)-induced mouse colitis models, while transgenic overexpression of TNFR11 by T cells enhanced colitis severity in a transfer model, suggesting that TNF is both protective and pathogenic. Importantly, anti-TNF efficacy has been significantly associated with increased apoptosis of T lymphocytes upon disruption of survival signals provided by mTNF but not soluble TNF, and T cell apoptosis was observed within 24 h following infliximab injection in CD lesions. In contrast, etanercept, a TNFR11-Fc fusion protein (binding only soluble TNF) and Certolizumab, a pegylated anti-TNF F [ab']<sub>2</sub> fragment (unable to bind to FcγR and to lead to antibody-dependent cellular cytotoxicity and complement-dependent cytotoxicity) fail to promote mucosal healing, while both drugs were efficacious in patients with rheumatoid arthritis (RA). These results have led to the hypothesis that anti-TNF blockade promotes CD healing mostly through the induction of apoptosis of mTNF-expressing T cells (Levin et al., 2016). Consequently, it is possible that in lesions with low GIMATS scores, in which comparison with uninflamed ileums suggested T cell-dominant pathogenic responses (Figure 3A), anti-TNF-induced T cell apoptosis is sufficient to improve mucosal healing. However, our ligand-receptor analysis revealed that interactions between members of the GIMATS module are complex and involve many signaling pathways extending beyond TNF. This strongly supports the possibility that TNF signaling alone cannot fully explain the pathogenic cellular organization we identified in GIMATS<sup>high</sup> lesions and raises the hypothesis that, upon TNF blockade, additional infl. mac-derived stimulatory mediators such as IL-1β and OSM are sufficient to trigger pathogenic stromal activation. Similarly, we did not identify TNF among the ligands mediating DC and macrophage interactions with T cells, which constituted another major pathogenic hub in the GIMATS response. These results suggest that therapeutic strategies combining anti-TNF antibodies (Abs) with drug targets that block key nodes in the GIMATS response communication network may provide new opportunities to overcome the anti-TNF resistance associated with GIMATS<sup>high</sup> patients. Because blood classical monocytes appear as the main contributors of infl. macs, which we identified as key drivers of both adaptive and stromal activation in GIMATS<sup>high</sup> lesions (Figure 4A), preventing their infiltration in lesions using anti-CCR2 blocking antibodies could simultaneously reduce major stromal- and T cell-activating mediators in tissues, while preserving tissue-protective gut-resident macrophages. By providing a comprehensive network of the cellular and molecular basis for resistance to anti-TNF blockade, our study thus opens novel opportunities for therapeutic discoveries tailored for combination with anti-TNF antibody blockade.

Our study shows that approaches combining high-resolution single-cell mapping of inflammatory lesions in small numbers of patients with bulk RNA sequencing on large cohorts with extensive clinical characterization leads to generalizable insights, highlighting the potential to broadly transform understanding of human multifactorial immune-mediated inflammatory diseases.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human specimens
- METHOD DETAILS
  - Preparation of lamina propria single cell suspensions
  - Preparation of PBMC
  - CyTOF sample preparation and analysis
  - Multiplexed immunohistochemical consecutive staining on single slide (MICSSS)
  - Processing of MICSSS pictures
  - Molecular profiling
  - Droplet-based scRNaseq
  - Detailed characterization of lymphoid scRNaseq clusters
  - External cohorts
  - Definition of non-response to anti-TNF
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Generation of gene expression matrices
  - Filtering low-quality cells and contaminating cells
  - A batch-aware, multinomial mixture model for single-cell RNaseq data
  - Inference of the model parameters
  - Gene modules
  - Principal component Analysis
  - Differential expression analysis
  - Ligand-Receptor Analysis
  - Defining the GIMATS module high and GIMATS module low gene sets
  - Projecting bulk RNA data onto GIMATS module high and GIMATS module low scores
- DATA AND CODE AVAILABILITY
- ADDITIONAL RESOURCES

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2019.08.008>.

## ACKNOWLEDGMENTS

The authors would like to thank the patients who participated in the study. The authors would like to warmly thank Dr. Jean-Frederic Colombel (co-director of the Feinstein IBD Center, Mount Sinai Hospital, New York, NY) for invaluable discussions and insights on Crohn's disease. The authors would also like to thank Dr. Emilie Grasset (Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY) for critical insights on B cells analysis.

We thank Dr. Andrew Chess for fruitful discussions. J.C.M. is supported by “Prix pour les jeunes chercheurs” de la Fondation Bettencourt-Schueller and by the Philippe Foundation; R.U. is supported by a Career Development Award from the Crohn’s and Colitis Foundation and an NIH K23 Career Development Award (K23KD111995-01A1). J.H.C. is funded by U01 DK062429, U01 DK062422, R01 DK106593, and the Sanford J. Grossman Charitable Trust. M.M. is funded by R56AI137244, U24 AI118644, U19 AI128949, and Boehringer Ingelheim Pharmaceuticals. This work utilized a mass cytometry instrumentation at the Human Immune Monitoring Center supported by S10 OD023547-01. The work was supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

#### AUTHOR CONTRIBUTIONS

Conceptualization, J.C.M., E.K., J.H.C., and M.M.; Methodology, J.C.M., G.B., E.K., G.A., H.S., S.G., and A.H.R.; Software, E.K.; Formal Analysis, E.K.; Investigation, J.C.M., G.B., C.C., M.G., J.A.G., L.W., A.L., H.M.K., and I.L.; Data Curation, M.G., K.G., and R.U.; Resources, L.-S.C., S.N., A.J.G., M.D., J.S.F., C.E.W., M.L.M., S.K., L.A.D., J.S.H., J.R.F., P.T.D., H.M.K., E.S.S., S.G., and A.H.R.; Writing – Original Draft, J.C.M., E.K., and M.M.; Writing – Review & Editing, J.C.M., C.W.E., J.H.C., E.K., M.M., and J.R.F.; Visualization, J.C.M., C.C., G.A., and E.K.; Supervision, J.C.M., M.M., J.H.C., and E.K.; Funding Acquisition, J.H.C. and M.M.

#### DECLARATION OF INTERESTS

This study was funded partially by a Grant from Boehringer Ingelheim to M.M. R.U. has served as an advisory board member or consultant for Janssen, Pfizer, and Takeda and received research grants from Abbvie, Boehringer Ingelheim, and Pfizer. S.G. reports receiving commercial research funding from Immune Design, Pfizer, Janssen R&D, Genentech, Regeneron, and Bristol-Myers Squibb and is a consultant/advisory board member for Merck, OncoMed, Neon Therapeutics, and B4CC. L.A.D. has received grant support from Janssen.

Received: December 5, 2018

Revised: May 6, 2019

Accepted: August 5, 2019

Published: August 29, 2019

#### WEB RESOURCES

scRNA-seq of ileal Crohn’s disease, <https://scDissector.org/martin>

#### REFERENCES

Abraham, C., Dulai, P.S., Vermeire, S., and Sandborn, W.J. (2016). Lessons Learned From Trials Targeting Cytokine Pathways in Patients With Inflammatory Bowel Diseases. *Gastroenterology* *152*, 374–388.e4.

Aird, W.C. (2007). Phenotypic heterogeneity of the endothelium: II. Representative vascular beds. *Circ. Res.* *100*, 174–190.

Anginot, A., Espeli, M., Chasson, L., Mancini, S.J.C., and Schiff, C. (2013). Galectin 1 modulates plasma cell homeostasis and regulates the humoral immune response. *J. Immunol.* *190*, 5526–5533.

Azizi, E., Carr, A.J., Plitas, G., Cornish, A.E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kisieliovas, V., Setty, M., et al. (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* *174*, 1293–1308.e36.

Bamba, S., Andoh, A., Yasui, H., Makino, J., Kim, S., and Fujiyama, Y. (2003). Regulation of IL-11 expression in intestinal myofibroblasts: role of c-Jun AP-1 and MAPK-dependent pathways. *Am. J. Physiol. Gastrointest. Liver Physiol.* *285*, G529–G538.

Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al.

(2017). QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* *7*, 16878.

Baran, Y., Sebe-Pedros, A., Lubling, Y., Giladi, A., Chomsky, E., Meir, Z., Hoichman, M., Lifshitz, A., and Tanay, A. (2018). MetaCell: analysis of single cell RNA-seq data using k-NN graph partitions. *bioRxiv*. <https://doi.org/10.1101/437665>.

Barone, F., Patel, P., Sanderson, J.D., and Spencer, J. (2009). Gut-associated lymphoid tissue contains the molecular machinery to support T-cell-dependent and T-cell-independent class switch recombination. *Mucosal Immunol.* *2*, 495–503.

Bernink, J.H., Peters, C.P., Munneke, M., te Velde, A.A., Meijer, S.L., Weijer, K., Hreggvidsdottir, H.S., Heinsbroek, S.E., Legrand, N., Buskens, C.J., et al. (2013). Human type 1 innate lymphoid cells accumulate in inflamed mucosal tissues. *Nat. Immunol.* *14*, 221–229.

Bilsborough, J., Targan, S.R., and Snapper, S.B. (2016). Therapeutic Targets in Inflammatory Bowel Disease: Current and Future. *Am. J. Gastroenterol. Suppl.* *3*, 27–37.

Björklund, Å.K., Forkel, M., Picelli, S., Konya, V., Theorell, J., Friberg, D., Sandberg, R., and Mjösberg, J. (2016). The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. *Nat. Immunol.* *17*, 451–460.

Cardona, A., Saalfeld, S., Schindelin, J., Arganda-Carreras, I., Preibisch, S., Longair, M., Tomancak, P., Hartenstein, V., and Douglas, R.J. (2012). TrakEM2 software for neural circuit reconstruction. *PLoS ONE* *7*, e38011.

Castigli, E., Scott, S., Dedeoglu, F., Bryce, P., Jabara, H., Bhan, A.K., Mizoguchi, E., and Geha, R.S. (2004). Impaired IgA class switching in APRIL-deficient mice. *Proc. Natl. Acad. Sci. USA* *101*, 3903–3908.

Chattopadhyay, P.K., Yu, J., and Roederer, M. (2005). A live-cell assay to detect antigen-specific CD4+ T cells with diverse cytokine profiles. *Nat. Med.* *11*, 1113–1117.

Cheuk, S., Schlums, H., Gallais Sérézal, I., Martini, E., Chiang, S.C., Marquardt, N., Gibbs, A., Detlofsson, E., Introini, A., Forkel, M., et al. (2017). CD49a Expression Defines Tissue-Resident CD8+ T Cells Poised for Cytotoxic Function in Human Skin. *Immunity* *46*, 287–300.

Danese, S., Fiocchi, C., and Panés, J. (2016). Drug development in IBD: from novel target identification to early clinical trials. *Gut* *65*, 1233–1239.

de Souza, H.S.P., and Fiocchi, C. (2015). Immunopathogenesis of IBD: current state of the art. *Nat. Rev. Gastroenterol. Hepatol.* *13*, 13–27.

Feagan, B.G., Sandborn, W.J., Gasink, C., Jacobstein, D., Lang, Y., Friedman, J.R., Blank, M.A., Johanns, J., Gao, L.-L., Miao, Y., et al.; UNITI-IM-UNITI Study Group (2016). Ustekinumab as Induction and Maintenance Therapy for Crohn’s Disease. *N. Engl. J. Med.* *375*, 1946–1960.

Förster, R., Davalos-Misslitz, A.C., and Rot, A. (2008). CCR7 and its ligands: balancing immunity and tolerance. *Nat. Rev. Immunol.* *8*, 362–371.

Gautier, E.L., Shay, T., Miller, J., Greter, M., Jakubczak, C., Ivanov, S., Helft, J., Chow, A., Elpek, K.G., Gordonov, S., et al.; Immunological Genome Consortium (2012). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* *13*, 1118–1128.

Guilliams, M., Dutertre, C.-A., Scott, C.L., McGovern, N., Sichien, D., Chakarov, S., Van Gassen, S., Chen, J., Poidinger, M., De Prijck, S., et al. (2016). Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. *Immunity* *45*, 669–684.

Haberman, Y., Tickle, T.L., Dexheimer, P.J., Kim, M.-O., Tang, D., Karns, R., Baldassano, R.N., Noe, J.D., Rosh, J., Markowitz, J., et al. (2014). Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J. Clin. Invest.* *124*, 3617–3633.

Hayatsu, N., Miyao, T., Tachibana, M., Murakami, R., Kimura, A., Kato, T., Kawakami, E., Endo, T.A., Setoguchi, R., Watarai, H., et al. (2017). Analyses of a Mutant Foxp3 Allele Reveal BATF as a Critical Transcription Factor in the Differentiation and Accumulation of Tissue Regulatory T Cells. *Immunity* *47*, 268–283.e9.

Hombink, P., Helbig, C., Backer, R.A., Piet, B., Oja, A.E., Stark, R., Brassler, G., Jongejan, A., Jonkers, R.E., Nota, B., et al. (2016). Programs for the

- persistence, vigilance and control of human CD8<sup>+</sup> lung-resident memory T cells. *Nat. Immunol.* **17**, 1467–1478.
- Hovhannisyanyan, Z., Treatman, J., Littman, D.R., and Mayer, L. (2011). Characterization of interleukin-17-producing regulatory T cells in inflamed intestinal mucosa from patients with inflammatory bowel diseases. *Gastroenterology* **140**, 957–965.
- Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779.
- Joller, N., Lozano, E., Burkett, P.R., Patel, B., Xiao, S., Zhu, C., Xia, J., Tan, T.G., Sefik, E., Yajnik, V., et al. (2014). Treg cells expressing the coinhibitory molecule TIGIT selectively inhibit proinflammatory Th1 and Th17 cell responses. *Immunity* **40**, 569–581.
- Joseph, N.M., He, S., Quintana, E., Kim, Y.-G., Núñez, G., and Morrison, S.J. (2011). Enteric glia are multipotent in culture but primarily form glia in the adult rodent gut. *J. Clin. Invest.* **121**, 3398–3411.
- Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al.; International IBD Genetics Consortium (IBDGC) (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124.
- Kamada, N., Hisamatsu, T., Okamoto, S., Chinen, H., Kobayashi, T., Sato, T., Sakuraba, A., Kitazume, M.T., Sugita, A., Koganei, K., et al. (2008). Unique CD14 intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN-gamma axis. *J. Clin. Invest.* **118**, 2269–2280.
- Kaser, A., Zeissig, S., and Blumberg, R.S. (2010). Inflammatory bowel disease. *Annu. Rev. Immunol.* **28**, 573–621.
- Klein, U., Tu, Y., Stolovitzky, G.A., Keller, J.L., Haddad, J., Jr., Miljkovic, V., Cattoretti, G., Califano, A., and Dalla-Favera, R. (2003). Transcriptional analysis of the B cell germinal center reaction. *Proc. Natl. Acad. Sci. USA* **100**, 2639–2644.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201.
- Konikoff, M.R., and Denson, L.A. (2006). Role of fecal calprotectin as a biomarker of intestinal inflammation in inflammatory bowel disease. *Inflamm. Bowel Dis.* **12**, 524–534.
- Kugathasan, S., Denson, L.A., Walters, T.D., Kim, M.-O., Marigorta, U.M., Schirmer, M., Mondal, K., Liu, C., Griffiths, A., Noe, J.D., et al. (2017). Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet* **389**, 1710–1718.
- Kumar, B.V., Ma, W., Miron, M., Granot, T., Guyer, R.S., Carpenter, D.J., Senda, T., Sun, X., Ho, S.-H., Lerner, H., et al. (2017). Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep.* **20**, 2921–2934.
- Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326.
- Levin, A.D., Wildenberg, M.E., and van den Brink, G.R. (2016). Mechanism of Action of Anti-TNF Therapy in Inflammatory Bowel Disease. *J. Crohn's Colitis* **10**, 989–997.
- Lim, A.I., Li, Y., Lopez-Lastra, S., Stadhouders, R., Paul, F., Casrouge, A., Serafini, N., Puel, A., Bustamante, J., Surace, L., et al. (2017). Systemic Human ILC Precursors Provide a Substrate for Tissue ILC Differentiation. *Cell* **168**, 1086–1100.e10.
- Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al.; International Multiple Sclerosis Genetics Consortium; International IBD Genetics Consortium (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986.
- Mackay, L.K., Minnich, M., Kragten, N.A.M., Liao, Y., Nota, B., Seillet, C., Zaid, A., Man, K., Preston, S., Freestone, D., et al. (2016). Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* **352**, 459–463.
- Macosko, E.Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214.
- Mahnke, Y.D., Brodie, T.M., Sallusto, F., Roederer, M., and Lugli, E. (2013). The who's who of T-cell differentiation: human memory T-cell subsets. *Eur. J. Immunol.* **43**, 2797–2809.
- Merad, M., Sathe, P., Helft, J., Miller, J., and Mortha, A. (2013). The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu. Rev. Immunol.* **31**, 563–604.
- Mortha, A., Chudnovskiy, A., Hashimoto, D., Bogunovic, M., Spencer, S.P., Belkaid, Y., and Merad, M. (2014). Microbiota-Dependent Crosstalk Between Macrophages and ILC3 Promotes Intestinal Homeostasis. *Science* **343**, 1249288.
- Mowat, A.M., and Agace, W.W. (2014). Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* **14**, 667–685.
- Mowat, A.M., Scott, C.L., and Bain, C.C. (2017). Barrier-tissue macrophages: functional adaptation to environmental challenges. *Nat. Med.* **23**, 1258–1270.
- Neurath, M.F. (2014). Cytokines in inflammatory bowel disease. *Nat. Rev. Immunol.* **14**, 329–342.
- Neurath, M.F. (2017). Current and emerging therapeutic targets for IBD. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 269–278.
- Okuno, T., Andoh, A., Bamba, S., Araki, Y., Fujiyama, Y., Fujiyama, M., and Bamba, T. (2002). Interleukin-1beta and tumor necrosis factor-alpha induce chemokine and matrix metalloproteinase gene expression in human colonic subepithelial myofibroblasts. *Scand. J. Gastroenterol.* **37**, 317–324.
- Ordovas-Montanes, J., Dwyer, D.F., Nyquist, S.K., Buchheit, K.M., Vukovic, M., Deb, C., Wadsworth, M.H., 2nd, Hughes, T.K., Kazer, S.W., Yoshimoto, E., et al. (2018). Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649–654.
- Pariente, B., Cosnes, J., Danese, S., Sandborn, W.J., Lewin, M., Fletcher, J.G., Chowers, Y., D'Haens, G., Feagan, B.G., Hibi, T., et al. (2011). Development of the Crohn's disease digestive damage score, the Lémann score. *Inflamm. Bowel Dis.* **17**, 1415–1422.
- Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**, 1663–1677.
- Peters, L.A., Perrigoue, J., Mortha, A., Iuga, A., Song, W.-M., Neiman, E.M., Llewellyn, S.R., Di Narzo, A., Kidd, B.A., Telesco, S.E., et al. (2017). A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat. Genet.* **49**, 1437–1449.
- Powell, N., Lo, J.W., Biancheri, P., Vossenkämper, A., Pantazi, E., Walker, A.W., Stolarczyk, E., Ammoscato, F., Goldberg, R., Scott, P., et al. (2015). Interleukin-6 Increases Production of Cytokines by Colonic Innate Lymphoid Cells in Mice and Patients with Chronic Intestinal Inflammation. *Gastroenterology* **149**, 456–467.e15.
- Pruenster, M., Mudde, L., Bombosi, P., Dimitrova, S., Zsak, M., Middleton, J., Richmond, A., Graham, G.J., Segerer, S., Nibbs, R.J.B., and Rot, A. (2009). The Duffy antigen receptor for chemokines transports chemokines and supports their promigratory activity. *Nat. Immunol.* **10**, 101–108.
- Puram, S.V., Tirosh, I., Park, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611–1624.e24.
- R Development Core Team (2008). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
- Ramesh, R., Kozhaya, L., McKeivitt, K., Djuretic, I.M., Carlson, T.J., Quintero, M.A., McCauley, J.L., Abreu, M.T., Unutmaz, D., and Sundrud, M.S. (2014). Pro-inflammatory human Th17 cells selectively express P-glycoprotein and are refractory to glucocorticoids. *J. Exp. Med.* **211**, 89–104.

- Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Sa-tagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., and Forrest, A.R. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866.
- Rao, D.A., Gurish, M.F., Marshall, J.L., Slowikowski, K., Fonseka, C.Y., Liu, Y., Donlin, L.T., Henderson, L.A., Wei, K., Mizoguchi, F., et al. (2017). Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature* **542**, 110–114.
- Remark, R., Merghoub, T., Grabe, N., Litjens, G., Damotte, D., Wolchok, J.D., Merad, M., and Gnjatic, S. (2016). In-depth tissue profiling using multiplexed immunohistochemical consecutive staining on single slide. *Sci. Immunol.* **1**, aaf6925.
- Remedios, K.A., Zirik, B., Sandoval, P.M., Lowe, M.M., Boda, D., Henley, E., Bhattarai, S., Scharschmidt, T.C., Liao, W., Naik, H.B., et al. (2018). The TNFRSF members CD27 and OX40 coordinately limit T<sub>H</sub>17 differentiation in regulatory T cells. *Sci. Immunol.* **3**, eaau2042.
- Sandborn, W.J., Gasink, C., Gao, L.-L., Blank, M.A., Johanns, J., Guzzo, C., Sands, B.E., Hanauer, S.B., Targan, S., Rutgeerts, P., et al.; CERTIFI Study Group (2012). Ustekinumab induction and maintenance therapy in refractory Crohn's disease. *N. Engl. J. Med.* **367**, 1519–1528.
- Sander, J., Schmidt, S.V., Cirovic, B., McGovern, N., Papantonopoulou, O., Hardt, A.-L., Aschenbrenner, A.C., Kreer, C., Quast, T., Xu, A.M., et al. (2017). Cellular Differentiation of Human Monocytes Is Regulated by Time-Dependent Interleukin-4 Signaling and the Transcriptional Regulator NCOR2. *Immunity* **47**, 1051–1066.e12.
- Savas, P., Virassamy, B., Ye, C., Salim, A., Mintoff, C.P., Caramia, F., Salgado, R., Byrne, D.J., Teo, Z.L., Dushyanthen, S., et al.; Kathleen Cuninghame Foundation Consortium for Research into Familial Breast Cancer (kConFab) (2018). Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat. Med.* **24**, 986–993.
- Schenk, M., Bouchon, A., Seibold, F., and Mueller, C. (2007). TREM-1-expressing intestinal macrophages crucially amplify chronic inflammation in experimental colitis and inflammatory bowel diseases. *J. Clin. Invest.* **117**, 3097–3106.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682.
- See, P., Dutertre, C.-A., Chen, J., Günther, P., McGovern, N., Irac, S.E., Gunawan, M., Beyer, M., Händler, K., Duan, K., et al. (2017). Mapping the human DC lineage through the integration of high-dimensional techniques. *Science* **356**, eaag3009.
- Simoni, Y., and Newell, E.W. (2017). Toward Meaningful Definitions of Innate Lymphoid-Cell Subsets. *Immunity* **46**, 760–761.
- Simoni, Y., Fehlings, M., Klöverpris, H.N., McGovern, N., Koo, S.-L., Loh, C.Y., Lim, S., Kurioka, A., Fergusson, J.R., Tang, C.-L., et al. (2017). Human Innate Lymphoid Cell Subsets Possess Tissue-Type Based Heterogeneity in Phenotype and Frequency. *Immunity* **46**, 148–161.
- Sisirak, V., Sally, B., D'Agati, V., Martinez-Ortiz, W., Özçakar, Z.B., David, J., Rashidfarrokhi, A., Yeste, A., Panea, C., Chida, A.S., et al. (2016). Digestion of Chromatin in Apoptotic Cell Microparticles Prevents Autoimmunity. *Cell* **166**, 88–101.
- Smythies, L.E., Sellers, M., Clements, R.H., Mosteller-Barnum, M., Meng, G., Benjamin, W.H., Orenstein, J.M., and Smith, P.D. (2005). Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J. Clin. Invest.* **115**, 66–75.
- Smythies, L.E., Shen, R., Bimczok, D., Novak, L., Clements, R.H., Eckhoff, D.E., Bouchard, P., George, M.D., Hu, W.K., Dandekar, S., and Smith, P.D. (2010). Inflammation anergy in human intestinal macrophages is due to Smad-induced IkappaBalpha expression and NF-kappaB inactivation. *J. Biol. Chem.* **285**, 19593–19604.
- Spencer, J., and Sollid, L.M. (2016). The human intestinal B-cell response. *Mucosal Immunol.* **9**, 1113–1124.
- Thiesen, S., Janciauskiene, S., Uronen-Hansson, H., Agace, W., Högerkorp, C.-M., Spee, P., Håkansson, K., and Grip, O. (2014). CD14(h)HLA-DR(dim) macrophages, with a resemblance to classical blood monocytes, dominate inflamed mucosa in Crohn's disease. *J. Leukoc. Biol.* **95**, 531–541.
- Thome, J.J.C., Bickham, K.L., Ohmura, Y., Kubota, M., Matsuoka, N., Gordon, C., Granot, T., Griesemer, A., Lerner, H., Kato, T., et al. (2015). Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues. *Nat. Med.* **22**, 72–77.
- Torres, J., Mehandru, S., Colombel, J.-F., and Peyrin-Biroulet, L. (2017). Crohn's disease. *Lancet* **389**, 1741–1755.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578.
- Tsai, C.-M., Chiu, Y.-K., Hsu, T.-L., Lin, I.-Y., Hsieh, S.-L., and Lin, K.-I. (2008). Galectin-1 promotes immunoglobulin production during plasma cell differentiation. *J. Immunol.* **181**, 4570–4579.
- Turley, S.J., Cremasco, V., and Astarita, J.L. (2015). Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat. Rev. Immunol.* **15**, 669–682.
- Uo, M., Hisamatsu, T., Miyoshi, J., Kaito, D., Yoneno, K., Kitazume, M.T., Mori, M., Sugita, A., Koganei, K., Matsuoka, K., et al. (2013). Mucosal CXCR4+ IgG plasma cells contribute to the pathogenesis of human ulcerative colitis through FcγR-mediated CD14 macrophage activation. *Gut* **62**, 1734–1744.
- Villani, A.-C., Satiya, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573.
- Walker, J.A., Barlow, J.L., and McKenzie, A.N.J. (2013). Innate lymphoid cells—how did we miss them? *Nat. Rev. Immunol.* **13**, 75–87.
- West, N.R., Hegazy, A.N., Owens, B.M.J., Bullers, S.J., Linggi, B., Buonocore, S., Coccia, M., Görtz, D., This, S., Stockenhuber, K., et al.; Oxford IBD Cohort Investigators (2017). Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. *Nat. Med.* **23**, 579–589.
- Willinger, T., Freeman, T., Herbert, M., Hasegawa, H., McMichael, A.J., and Callan, M.F.C. (2006). Human naive CD8 T cells down-regulate expression of the WNT pathway transcription factors lymphoid enhancer binding factor 1 and transcription factor 7 (T cell factor-1) following antigen encounter in vitro and in vivo. *J. Immunol.* **176**, 1439–1446.
- Wong, M.T., Ong, D.E.H., Lim, F.S.H., Teng, K.W.W., McGovern, N., Narayanan, S., Ho, W.Q., Cerny, D., Tan, H.K.K., Anicete, R., et al. (2016). A High-Dimensional Atlas of Human T Cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures. *Immunity* **45**, 442–456.
- Zemmour, D., Zilionis, R., Kiner, E., Klein, A.M., Mathis, D., and Benoist, C. (2018). Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* **19**, 291–301.
- Zheng, C., Zheng, L., Yoo, J.-K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J.Y., Zhang, Q., et al. (2017a). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* **169**, 1342–1356.e16.
- Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017b). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-human CD45 (clone HI30) – 89Y	Fluidigm	Cat#:3089003B; RRID: AB_2661851
Anti-human CD57 (clone HCD57) – 113In	Biolegend	Cat#:322302; RRID: AB_535988
Anti-human HLABC (clone W6/32) – 115In	Biolegend	Cat#:311402; RRID: AB_314871
Anti-human CD326 (clone 9C4) – 141Pr	Biolegend	Cat#:324202; RRID: AB_756076
Anti-human CD19 (clone HIB19) – 142Nd	Biolegend	Cat#:302202; RRID: AB_314232
Anti-human CD45RA (clone HI100) – 143Nd	Biolegend	Cat#:304102; RRID: AB_314406
Anti-human CD141 (clone M80) – 144Nd	Biolegend	Cat#:344102; RRID: AB_2201808
Anti-human CD4 (clone RPA-T4) – 145Nd	Biolegend	Cat#:300502; RRID: AB_314070
Anti-human CD8 (clone RPA-T8) – 146Nd	Biolegend	Cat#:301002; RRID: AB_314120
Anti-human IgA (clone 9H9H11) – 147Sm	Biolegend	Cat#:409702; RRID: AB_2564448
Anti-human CD16 (clone 3G8) – 148Nd	Biolegend	Cat#:302014; RRID: AB_314202
Anti-human CD127 (clone A019D5) – 149Sm	Biolegend	Cat#:351302; RRID: AB_10718513
Anti-human CD1c (clone L161) – 150Nd	Biolegend	Cat#:331502; RRID: AB_1088995
Anti-human CD123 (clone 6H6) – 151Eu	Biolegend	Cat#:306002; RRID: AB_314576
Anti-human CD66b (clone G10F5) – 152Sm	Biolegend	Cat#:305102; RRID: AB_314494
Anti-human PD-1 (clone EH12.2H7) – 153Eu	Biolegend	Cat#:329912; RRID: AB_940488
Anti-human CD86 (clone IT2.2) – 154Sm	Biolegend	Cat#:305402; RRID: AB_314522
Anti-human CD27 (clone O323) – 155Gd	Biolegend	Cat#:302802; RRID: AB_314294
Anti-human CXCR3 (clone G025H7) – 156Gd	Fluidigm	Cat#:3156004B
Anti-human CD33 (clone WM53) – 158Gd	Biolegend	Cat#:303402; RRID: AB_314346
Anti-human CD103 (clone Ber-Act8) – 159Tb	Biolegend	Cat#:350202; RRID: AB_10639864
Anti-human CD14 (clone M5E2) – 160Gd	Biolegend	Cat#:301810; RRID: AB_314192
Anti-human CD56 (clone B159) – 161Dy	BD Biosciences	Cat#:555513
Anti-human CD64 (clone 10.1) – 162Dy	Biolegend	Cat#:305016; RRID: AB_2103461
Anti-human CD172a/b (clone SIRPa/b) – 163Dy	Biolegend	Cat#:323802; RRID: AB_830701
Anti-human CD69 (clone FN50) – 164Dy	Biolegend	Cat#:310902; RRID: AB_314837
Anti-human FcεR1a (clone AER-37 (CRA-1) – 165Ho	Biolegend	Cat#:334602; RRID: AB_1227649
Anti-human CD25 (clone M-A251) – 166Er	Biolegend	Cat#:356102; RRID: AB_2561752
Anti-human CD11c (clone Bu15) – 167Er	Biolegend	Cat#:337202; RRID: AB_1236381
Anti-human CD3 (clone UCHT1) – 168Er	Biolegend	Cat#:300402; RRID: AB_314056
Anti-human integrin beta7 (clone FIB504) – 169Tm	Biolegend	Cat#:321202; RRID: AB_571975
Anti-human CD38 (clone HB-7) – 170Er	Biolegend	Cat#:356602; RRID: AB_2561794
Anti-human CD161 (clone HP-3G10) – 171Yb	Biolegend	Cat#:339902; RRID: AB_1501090
Anti-human CD206 (clone 15-2) – 172Yb	Biolegend	Cat#:321102; RRID: AB_571923
Anti-human CXCR4 (clone 12G5) – 173Yb	Fluidigm	Cat#:3173001B
Anti-human HLA-DR (clone L243) – 174Yb	Biolegend	Cat#:307602; RRID: AB_314680
Anti-human PD-L1 (clone 29E.2A3) – 175Lu	Fluidigm	Cat#:3175017B
Anti-human CD54 (clone HCD54) – 176Yb	Biolegend	Cat#:322704; RRID: AB_535976
Anti-human CD11b (clone ICRF44) – 209Bi	Fluidigm	Cat#:3209003B; RRID: AB_2687654
Anti-Human CD3 (clone 2GV6)	Ventana	Cat#: 790-4341
Anti-Human DC-LAMP (clone 1010E1.01)	Novus Biologicals	Cat#: DDX0191P

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Anti-Human CD20 (clone L26)	Dako	Cat#: 0755; RRID: AB_2282030
Anti-Human CD68 (clone KP1)	Dako	Cat#: M081401-2; RRID: AB_2750584
anti-human HLA-DR (clone TAL1B5)	Abcam	Cat#: ab20181; RRID: AB_445401
Anti-Human Ki-67 (clone 30-9)	Ventana	Cat#: 790-4286; RRID: AB_2631262
anti-human CD206 (polyclonal)	Abcam	Cat#: ab64693
Anti-human PDPN (clone D2-40)	Ventana	Cat#: 760-4395
<b>Biological Samples</b>		
Inflamed and uninfamed ileums, and venous blood from Crohn's disease patients	Fresh resections and blood from patients with informed consent for an IRB-approved study. Surgeries were performed by the Department of Colorectal Surgery at Mount Sinai Hospital	N/A
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Collagenase, Type IV, Clostridium histolyticum	Sigma-Aldrich	Cat#: C5138
Deoxyribonuclease I	Sigma-Aldrich	Cat#: DN25
Ficoll-Paque PLUS	GE Healthcare	Cat#: 17144003
Rh103 intercalator	Fluidigm	Cat#: 201103A
0.125uM Ir intercalator	Fluidigm	Cat#: 201192A
HEPES 1M	Fisher Scientific	Cat#: MT25060CI
Penicillin/Streptomycin (10,000 U/mL)	Fisher Scientific	Cat#: 15-140-163
0.5 m EDTA, pH 8.0	Corning	Cat#: 46-034-CI
<b>Critical Commercial Assays</b>		
Dead Cell Removal Kit	Miltenyi Biotec	Cat#: 130-090-101
Cytofix/Cytoperm	BD Biosciences	Cat#: 51-2090KZ
Perm/Wash	BD Biosciences	Cat#: 51-2091KZ
Chromium Single Cell 3' Library & Gel Bead Kit v2, 16 rxns	10x Genomics	Cat#: PN-120237
Chromium Single Cell A Chip Kit, 48 rxns	10x Genomics	Cat#: PN-120236
Chromium i7 Multiplex Kit, 96 rxns	10x Genomics	Cat#: PN-120262
<b>Deposited Data</b>		
Single cell RNaseq data	GEO	GEO: GSE134809
RNaseq data RISK cohort	GEO	GEO: GSE134881
<b>Software and Algorithms</b>		
Cytobank	Mountain View	<a href="https://cytobank.org/">https://cytobank.org/</a>
QuPath 0.1.3	<a href="#">Bankhead et al., 2017</a>	<a href="https://qupath.github.io/">https://qupath.github.io/</a>
Fiji - ImageJ 1.52 h	<a href="#">Schindelin et al., 2012</a>	<a href="https://imagej.net/Fiji/Downloads">https://imagej.net/Fiji/Downloads</a>
Trakem2	<a href="#">Cardona et al., 2012</a>	<a href="https://imagej.net/TrakEM2">https://imagej.net/TrakEM2</a>
CellRanger Version 2.0 Version 2.0	10x Genomics	<a href="https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest">https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest</a>
GraphPad Prism version 7.0	GraphPad Software	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
R	<a href="#">R Development Core Team, 2008</a>	<a href="http://www.R-project.org">http://www.R-project.org</a>
TgIkmmeans (R-package) 0.1.3	Amos Tanay Lab	<a href="https://bitbucket.org/tanaylab/tgIkmmeans">https://bitbucket.org/tanaylab/tgIkmmeans</a>
scDissector (R-package) 1.0	Ephraim Kenigsberg Lab	<a href="https://github.com/effiken/scDissector">https://github.com/effiken/scDissector</a>
Seriation (R-package) 1.2.7	CRAN	<a href="https://github.com/mhahsler/seriation">https://github.com/mhahsler/seriation</a>
XPS (R-package) 1.44.0	Bioconductor	<a href="https://bioconductor.org/packages/release/bioc/html/xps.html">https://bioconductor.org/packages/release/bioc/html/xps.html</a>

(Continued on next page)

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
TopHat2 2.0.13	CCB, Johns Hopkins University	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>
Cufflinks 2.1.1	Cole Trapnell Lab	<a href="http://cole-trapnell-lab.github.io/cufflinks/">http://cole-trapnell-lab.github.io/cufflinks/</a>
R code for clustering and analyses	This study	<a href="https://github.com/effiken/martin_et_al_cell_2019">https://github.com/effiken/martin_et_al_cell_2019</a>
Other		
Surgical biopsy forceps	Endochoice MAKO	Cat#: DBC-206-40

## LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Miriam Merad ([miriam.merad@mssm.edu](mailto:miriam.merad@mssm.edu)). This study did not generate new unique reagents.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Human specimens

Patients eligible for inclusion in the study were identified by screening surgical programs at the Mount Sinai Hospital. Ileum CD tissues and venous blood were obtained after 1<sup>st</sup> surgical resection. Protocols were reviewed and approved by the Institutional Review Board (IRB) at the Icahn School of Medicine at Mount Sinai (Mechanisms of intestinal inflammation following Ileal resection for Crohn's Disease; HSM#13-00998). Clinical characteristics, including Montreal classification, of the patients retained for scRNaseq and CyTOF studies are summarized in [Table S1](#). All macroscopically inflamed tissues included in the study were confirmed by pathological examination as active ileitis with transmural chronic inflammation.

## METHOD DETAILS

### Preparation of lamina propria single cell suspensions

Tissues were collected in ice cold RPMI 1640 (Corning Inc., Corning, NY) and processed within one h after termination of the surgery. To limit biased enrichment of specific immune and stromal populations related to local variations in the intestinal micro-organization ([Mowat and Agace, 2014](#)), we pooled fifteen to twenty mucosal biopsies sampled all along the resected specimens using a biopsy forceps (EndoChoice, Alpharetta, GA) to prepare cell suspensions. Epithelial cells were dissociated by incubating the biopsies in an EDTA-enriched dissociation medium (HBSS w/o Ca<sup>2+</sup> Mg<sup>2+</sup> (Life Technologies, Carlsbad, CA) - HEPES 10mM (Life Technologies) - EDTA 5mM (Life Technologies)) at +37°C under 100 rpm agitation for two cycles of 15 min. After each cycle, biopsies were hand-shaken for 30 s and then vortexed vigorously for another 30 s. Biopsies were then washed in complete RPMI media previously put at RT, and transferred in digestion medium (HBSS with Ca<sup>2+</sup> Mg<sup>2+</sup> - FCS 2% - DNase I 0.5mg/mL (Sigma-Aldrich, St. Louis, MO) - Collagenase IV 0.5mg/mL (Sigma-Aldrich)) for 40 min at +37°C under 100 rpm agitation. After digestion, the cell suspension was filtered through a 70µm cell strainer, washed in DBPS/ 2% FCS/ 1mM EDTA and spun down at 400 g for 10 min. After red blood cell lysis (BioLegend, San Diego, CA), cells were washed as before. Dead cells were depleted from the suspension using the dead cell depletion kit (Miltenyi Biotec, Bergisch Gladbach, Germany), following manufacturer's recommendations. Viability of the final cell suspension was calculated using a hemocytometer and Trypan blue (Corning) exclusion and was routinely > 85%.

### Preparation of PBMC

Venous blood was collected intraoperatively in EDTA BD Vacutainer blood collection tubes (Becton Dickinson, Franklin Lakes, NJ). Blood was diluted in sterile PBS (Sigma-Aldrich), layered above a Ficoll-Paque PLUS density gradient media (GE Healthcare, Little Chalfont, UK) and centrifuged 20 min at 800 g, +20°C, brakes off. PBMC were collected at the interphase and washed. After RBC lysis, dead cells were removed from the suspension using the dead cell depletion kit (Miltenyi Biotec). Viability of the final cell suspension was calculated as above and was routinely > 90%.

### CyTOF sample preparation and analysis

Tissue cell suspensions were first incubated with Rh103 intercalator (Fluidigm, San Francisco, CA) for 20 min at 37°C to label non-viable cells, and then washed and labeled with a panel of metal-labeled antibodies for 30 min on ice. The samples were then fixed and permeabilized with BD Cytofix/Cytoperm (BD Biosciences) and incubated in 0.125µM Ir intercalator (Fluidigm) diluted in PBS containing 2% formaldehyde for 30 min. The samples were then washed and stored in PBS containing 0.2% BSA at 4°C until acquisition.

Immediately prior to acquisition, samples were washed once with PBS, once with de-ionized water and then resuspended at a concentration of 1 million cells/mL in deionized water containing a 1:20 dilution of Equation 4 Element Beads (Fluidigm). The samples were acquired on a CyTOF2 (Fluidigm) equipped with a SuperSampler fluidics system (Victorian Airships) at an event rate of < 500 events/second. FCS files were manually pre-gated on Ir193 DNA<sup>+</sup> events, excluding dead cells, doublets, and DNA<sup>-</sup> negative debris. Samples were analyzed using Cytobank (Mountain View, CA).

### Multiplexed immunohistochemical consecutive staining on single slide (MICSSS)

Four  $\mu\text{m}$ -thick formalin-fixed paraffin-embedded (FFPE) tissue sections on glass slides were backed at 37°C overnight, deparaffinized in xylene, and rehydrated in decreasing concentrations of ethanol. Then, tissue sections were incubated in retrieval solution (Citrate buffer, pH6) for antigen retrieval at 95°C for 30 min. Tissue sections were incubated in 3% hydrogen peroxide and in serum-free protein block solution (Dako, X0909) before adding the primary antibody for 1 h at room temperature. After signal amplification using biotinylated-secondary antibody and streptavidin-horseradish peroxidase, chromogenic revelation was performed using 3-amino-9-ethylcarbazole (AEC). Slides were counterstained with hematoxylin, mounted with a glycerol-based mounting medium and scanned for digital imaging (Hamamatsu Nanozoomer S60 Digital Slide Scanner). The same slides were successively stained as described (Remark et al., 2016). Primary antibodies were: anti-human CD3 (clone 2GV6, Ventana), anti-human CD20 (clone L26, Dako), anti-human DC-LAMP (clone 1010E1.01, Novus biologicals), anti-human CD68 (clone KP1, Dako), anti-human HLA-DR (clone TAL1B5, Abcam), anti-Ki67 (clone 30-09, Ventana), anti-human CD206 (polyclonal, Abcam) anti-PDPN (clone D2-40, Ventana).

### Processing of MICSSS pictures

Images were processed by an experienced pathologist. Whole slide images obtained from scanned immunostains performed on the same tissue sections were opened simultaneously the open source digital pathology image analysis software (QuPath 0.1.3) (Bankhead et al., 2017). Regions of interest (ROI) were defined using the same coordinates for each image. Images were exported as .png formatted images for analysis with the Fiji platform (Fiji - ImageJ 1.52 h) (Schindelin et al., 2012). Images are aligned by using the Trakem2 plugin (Cardona et al., 2012). For each image, color deconvolution was achieved with a H-DAB vector to split images into three 8-bit channels including hematoxylin (blue), DAB (brown) and residual (green) channels. One hematoxylin channel was selected to define a base nuclear channel. All DAB channels were assigned different colors by using the Fiji look up tables (LUT) function. Colored channels were inverted and merged as one composite pseudo-fluorescence image. Brightness and contrast settings were optimized for each channel by comparing with original images. Underlying image pixel values were not changed for brightness and contrast settings.

### Molecular profiling

Patients 5-16 were analyzed by scRNaseq (Table S2, sheet 1). As discussed below, in order to minimize technical biases, the cell subtype clustering model was derived from scRNaseq samples acquired with the Chromium V2 chemistry. Cells from Patient 5, which were processed using the V1 chemistry, were projected onto the clustering model. Low cell number recovery prevented accurate high-resolution estimation of the cellular composition of Patient 6, who was consequently excluded from analyses involving correlations and comparisons of cell subtype frequencies. Because we did not observe significant differences between uninfamed and infamed scRNaseq data from Patient 16, we also excluded this patient from analyses involving correlations and comparisons cell subtype frequencies.

### Droplet-based scRNaseq

Cells were suspended at  $1.10^6$ /mL in PBS and 10,000 cells were loaded onto the Chromium<sup>TM</sup> Controller instrument within 15 min after completion of the cell suspension preparation using GemCode Gel Bead and Chip, all from 10x Genomics (Pleasanton, CA), and following the manufacturer's recommendations. Briefly, cells were partitioned into Gel Beads in Emulsion in the Chromium<sup>TM</sup> Controller instrument where cell lysis and barcoded reverse transcription of RNA occurred. Libraries were prepared using 10x Genomics Library Kits and sequenced on an Illumina NextSeq500 according the manufacturer's recommendations. Read-depth of more than 100 million reads per library, or an approximate average of 30,000 reads per cell was obtained (Table S2, sheet 1).

### Detailed characterization of lymphoid scRNaseq clusters

#### T cells characterization

Sixteen scRNaseq clusters (#17, 43, 19, 29, 22, 38, 21, 26, 4, 47, 34, 36, 30, 46, 48, 42) of T cells were characterized by shared co-expression of T cell receptor (TCR) coreceptor encoding genes from the CD3 complex like *CD3D*, TCR costimulatory molecule *CD2* and maturation marker *CD7* (Figures 1B and S3E). Two clusters of innate lymphoid cells (ILCs) also expressed high levels of *CD7*, together with *CD2*, but no expression of *CD3* genes or other lineage gene such as *MS4A1* (*CD20*), *CD19*, *SDC1*, *TNFRSF17*, *LYZ*, *HLA-DRs*, *IL3RA*, *TPSAB1*, as previously reported (Lim et al., 2017). Gene modules within T cell clusters were defined based on genes with strong gene-to-gene correlation of expression, as described above (Figure S3C; Table S2, sheet 5). Based on the module analysis we defined gene expression programs related to **cell cycling** (*STMN1*, *CLSPN*, *CKS1B*, *NUF2*, *ASPM*, *UBE2T*, *CENPF*, *RRM2*, *DTYMK*, *SGOL1*, *CENPE*, *MAD2L1*, *CCNA2*, *CENPU*, *CCNB1*, *DHFR*, *HMMR*, *MXD3*, *GMNN*, *CENPW*, *MCM7*, *EZH2*, *SMC2*, *CCDC34*, *CDCA5*, *H2AFX*, *ZWINT*, *CDK1*, *HELLS*, *MKI67*, *TMEM106C*, *CDKN3*, *NUSAP1*, *CCNB2*, *KIAA0101*, *PRC1*,

*CENPN, CDT1, AURKB, TOP2A, TK1, BIRC5, TYMS, CDC25B, PCNA, TPX2, UBE2C, ASF1B, GTSE1*), **immunoregulation** (*TNFRSF18, TNFRSF4, TNFRSF9, TNFRSF1B, AC133644.2, CTLA4, TIGIT, ICA1, RHBDD2, MAGEH1, IL2RA, TBC1D4, BATF, IKZF2, FOXP3*), **naive/central memory (CM)** (*SELL, LEF1, SOX4, SC5D, CCR7, TOB1, NOSIP*), **CD8/cytotoxic** (*CD8A, CD8B, DHRS3, GZMA, GFOD1, IFITM3, PRF1, KLRD1, GZMB, CCL5, NKG7, FGR, CD160, FCER1G, XCL2, XCL1, GNLY, EOMES, CMC1, DTHD1, AOA, CLIC3, CTSW, KLRF1, KLRC2, KLRC1, PTGDR, MCTP2, CCL3, CCL4, CCL3L3, CCL4L2, MATK, MAPK1, IL2RB*) and **resident memory T cell** (*JUN, KLF6, FOSB, PTGER2, FOS, SYTL3, SPRY1, ANKRD28, GPR171, PDE4D, JAML, IL7R, GLIPR1, CD69, NFKBIA, PPP1R15A, NFKBIZ, TNFAIP3, PTGER4, ANXA1, ID2, ATF3, MGAT4A, AC092580.4, KLRB1, RORA, IL18R1, STAT4, IFNGR1, PFKFB3, GPR65*) states.

The total UMI counts of each of the five transcriptional programs were calculated for each cell and divided by the total number of UMIs/cell. The five scores were log transformed following division by their 75% percentile. These scores demonstrated the different distributions of the expression programs across T cell states (Figure 2E).

T cell cluster #17 expressed a strong cell cycle score and genes encoding for membrane proteins associated with human T cells activation such as *HLA-DRA, CD38* and the high affinity IL-2 receptor chain *IL2RA* (CD25), and thus was annotated as highly activated T cells. Cells in this cluster included both CD4 and CD8 T cells. Concordantly, these cells expressed genes associated with CD4<sup>+</sup> T helper cells functions i.e., *CD40LG* or CD8<sup>+</sup> T cells cytotoxicity (*GZMA, GZMB, PRF1*), indicating that T cells clustered according to their highly activated state over the CD4/CD8 lineage specification. Accordingly, both CD4<sup>+</sup> and CD8<sup>+</sup> HLA-DR<sup>+</sup> activated T cells were identified by CyTOF analysis (Figure S3D).

T cell cluster #21 expressed high levels of the cytotoxic program including granzymes and perforin encoding genes, as well as low levels of the resident memory, naive/CM and cell cycling programs, and low expression of the IL-7 receptor gene *IL7R* (CD127), and thus was referred to as effector memory cytotoxic T lymphocytes (CTLs) (Mahnke et al., 2013), while T cell cluster (#38) expressed lower level of the CD8/cytotoxic program together with strong expression of genes of the  $\gamma\delta$  TCR-like *TRDC*, and was annotated as  $\gamma\delta$  T cells (Figures 2E and S3E).

T cell cluster #43 expressed high levels of Treg-associated genes including *FOXP3, IL2RA* (CD25), *CTLA4*, together with genes which high expression was consistently associated with Tregs by recent scRNaseq studies, including *TNFRSF18, TNFRSF4*, and *LAYN* (Savas et al., 2018; Zemmour et al., 2018; Zheng et al., 2017a) and *BATF*, recently identified as a critical TF for tissue Tregs (Hayatsu et al., 2017) (Figures 2E and S3E), and was thus referred to as Tregs. Interestingly, Tregs strongly expressed the co-inhibitory receptor *TIGIT*, which was shown to inhibit pathogenic Th1 and Th17 responses during murine intestinal inflammation (Joller et al., 2014), as well as the co-stimulatory receptor *CD27*, recently suggested as an important repressor of Th17 function acquisition by Tregs during inflammation (Remedios et al., 2018).

Cells in cluster #22 shared several genes with Tregs including *TIGIT, BATF, ICA1* and *TNFRSF4* (OX-40), but lacked Tregs-defining transcripts *FOXP3, IL2RA, TNFRSF18* and *IKZF2* (HELIOS) while they expressed high levels of *CD40LG* and the chemokine *CXCL13*, and thus were referred to as T follicular helper-like (TFH-like) cells (Rao et al., 2017) (Figure S3E).

Two cell clusters (#19, 29) expressed high levels of the naive/CM signature, including the chemokine receptor *CCR7* and the selectin *CD62L* (*SELL*), known to promote the recruitment of naive and central memory (CM) T cells to the T cell areas of secondary lymphoid organs through the high endothelial venule (Förster et al., 2008) (Figures 2E and S3E). In agreement with the naive and/or CM state of T cells in these clusters, high expression of IL-7 receptor *IL7R* (CD127) was also detected (Mahnke et al., 2013).

The first of these two clusters (#19) exhibited higher levels of the TF *LEF1*, preferentially expressed by naive T cells, before antigen exposure (Willinger et al., 2006) and low-to-no expression of genes associated with CD4 and CD8 effector T cell functions or tissue-directing chemokine receptors. As for activated T cells, CD4<sup>+</sup> and CD8<sup>+</sup> T cells were mixed in this cluster, supporting dominance of the naive state at driving the clustering, an observation already reported for blood naive T cells (Zheng et al., 2017b). In contrast, T cells in cluster #29 expressed additional chemokine receptors like *CCR6*, together with the lectin *KLRB1* (CD161), inflammatory cytokines and chemokines (*IFNG, TNF, CCL20, CCL5*), IFN $\gamma$ -induced genes (*CD74, HLA-G, ISG15, IFITM1*) and genes associated with inflammatory responses, thus suggestive of a population of pro-inflammatory Th1/Th17 CM T cells (Ramesh et al., 2014) (Figures 2F, 4C, and S3E).

Nine clusters (#26, 4, 47, 34, 36, 30, 46, 48, 42) shared strong expression of the tissue resident memory T cells (Trm) program, which included the tissue-retention molecule *CD69*, the TFs like *ATF3, FOS, FOSB, KLF6* and the integrin *ITGA1* (CD49a), and the TF *PRDM1*, previously shown to be associated with gut Trm (Hombrink et al., 2016; Kumar et al., 2017; Mackay et al., 2016; Thome et al., 2015; Wong et al., 2016). In support of the resident memory state of these cells, the Trm clusters accounted for the large majority of the T cells in uninflamed tissues. Additional expressed genes included the prostaglandin E2 receptor (*PTGER4*), *ANKRD28, SPRY1*, G-protein coupled receptor *GPR171*, the TF *ID2, TNFAIP3*, an NF- $\kappa$ b inhibitor induced by TNF protecting T cells against TNF-induced apoptosis and the *IL18* receptor subunit *IL18RAP*, none of which were previously described in human intestinal Trm. Five clusters (#26, 4, 47, 34, 36) corresponded to CD8 Trm. Two of these clusters were highly enriched for the cytotoxic program (#26, 4), thus reminiscent of the Trm poised for cytotoxic functions recently described in the skin (Cheuk et al., 2017). Remaining Trm clusters (#30, 46, 48, 42) expressed multiple genes associated with Th17 T cells, (*KLRB1, CCL20, IL17, IL22, MAF, CCR6*), high levels of *CD40LG* (CD154) and the cytokine IL-2, thus suggesting antigen-dependent activation (Chattopadhyay et al., 2005). The presence of naive, CM, resident memory, highly activated and regulatory T cells in inflamed CD tissues was validated by CyTOF analysis, which also provided better granularity for CD4 and CD8 populations of naive, CM and highly activated T cells (Figures S3D).

### Innate lymphoid cells characterization

The two ILC clusters (#35, 28) expressed *KLRB1* (CD161), *TYROBP* (DAP12) and *FCER1G* and lacked transcript of genes specific for ILC2, like *PTGDR2*, *IL17RB*, *IL1RL1*, *HPGDS* or *HPGD*, and thus were referred to as Group 1 ILC1/NK and ILC3 subtypes (Figures S3F and S3G). Cluster #35 displayed a gene signature highly compatible with ILC3, the most abundant ILC population in human healthy gut (Simoni et al., 2017), including *KIT* (CD117), *IL23R*, *IL1R1*, the TFs *ID2*, *RORC* (ROR $\gamma$ t), *RORA* (ROR $\alpha$ ) and the cytokines *IL22* and *CSF2* (GM-CSF) (Figure 2F) (Björklund et al., 2016; Mortha et al., 2014). This cluster also expressed lower level of *CD45* (*PTPRC*), a well-defined feature of both mouse and human ILC3 (Walker et al., 2013). Cluster #28 expressed cytotoxic genes (*GZMs*, *PRF1*), the cytokines *XCL1*, *XCL2* and *IFN $\gamma$* , and *CCL3/4/5* chemokines, as well as additional NK cell markers including *IL2RB* (CD122). Because some cells also expressed *IL7R*, we chose to annotate this cluster as Group 1 ILCs to encompass NK cells, CD127<sup>+</sup> ILC1 and potential IFN $\gamma$ -producing “ex-ILC3s” whose origin remains unclear (Bernink et al., 2013; Simoni and Newell, 2017). The presence of ILCs was confirmed by CyTOF analysis in CD tissues but our Ab panel did not allow clear separation between the two subtypes defined by scRNaseq (data not shown).

### B cells characterization

The B cell compartment was largely dominated by plasma cells (PC) clusters (#27, 13, 20, 1, 45, 49, 50, 7), which expressed genes encoding PC-associated membrane proteins including *CD138* (*SDC1*), *CD38* and *CD27* as well as genes associated with protein folding, trafficking and secretion (*SSR4*, *FKBP2*, *FKBP11*, *HSP90B1*, *SEC11C*), degradation of misfolded protein (*SDF2L1*, *DERL3*) and endoplasmic reticulum stress, including the IBD susceptibility TF gene *XBP1*. PC also expressed *LGALS1*, encoding for galectin-1, a downstream gene of the PC TF BLIMP-1 important for PC differentiation, survival and Igs secretion (Anginot et al., 2013; Tsai et al., 2008). These clusters were referred as PC expressing IgA (#27, 13, 20, 1), IgM (#45) and IgG PC (#49, 50, 7). We did not consider light chain isotype differences to annotate the clusters (Figure S3A). In agreement with a previous study in colonic IBD, IgG PC expressed *CXCR4*, the receptor for the chemokine *CXCL12* (Uo et al., 2013), while IgM and IgA PC expressed higher levels of *BCMA* (*TNFRSF17*), the receptor for a proliferation-inducing ligand (APRIL) and B cell activating factor of TNF family (BAFF), important for PC survival in the gut (Barone et al., 2009; Castigli et al., 2004). We validated the increased levels of IgA<sup>-</sup> *CXCR4*<sup>+</sup> PC in inflamed tissues by CyTOF analysis (Figure S3B). IgM and IgA-producing PC also expressed high levels of the J-CHAIN encoding gene (*IGJ*), necessary to assemble multimeric IgA and IgM and their epithelial transcytosis and secretion into the intestinal lumen through binding to the poly-Ig receptor (Spencer and Sollid, 2016). Along with PC, a small cluster of plasmablasts (#37) was identified, which together with genes involved in immunoglobulin expression and secretion described above, also expressed cell cycle genes. Aside from PC and plasmablasts, three clusters of B lymphocytes (#33, 14, 39) were identified (Figure S3A). B cells in these clusters shared the expression of important B cell genes including *CD19*, *CD20*, BCR co-receptors *CD79a* and *CD79b*, *BANK1*, *MCH-II* encoding genes and the TFs *IRF8* and *PAX5* for instance. Differential enrichment of co-expressed genes reflecting naive (*IGHD*, *FCER2*, *CD72*) (#33) or switched memory (*CD27*, *IGHG1*) (#14, 39) existed between the clusters (Klein et al., 2003). As for T cells, in analyses involving PC and B cells subtypes frequencies, the clusters of IgA- and IgG-producing PC were pooled, grouping together clusters expressing different subclasses of a main class, and referred as IgA PC (#27, 13, 20, 1) and IgG PC (#49, 50, 7) respectively; the two clusters enriched in memory B cells were pooled and referred as memory B cells (#14, 39) (Figures 2I, 2J, and 3A).

### External cohorts

RNAseq data from the RISK cohort (n = 199) were generated from ileal biopsies of children and adolescents younger than 17 years old, newly diagnosed with ileal CD and with baseline colonoscopy and confirmation of characteristic chronic active colitis/ileitis by histology prior to diagnosis and treatment, as previously described (Haberman et al., 2014). The RISK cohort enrolled only children older than 2 years at disease onset and children with autosomal recessive *IL-10* or *IL10RA/B* mutations, who develop anti-TNF resistant lesions in the first year of life (often the first month), were excluded. Importantly, the pathogenesis of other mutations found more commonly in pediatric cohorts are quite similar to those present in adults.

Microarray data from the CERTIFI cohort (GSE100833) were generated in the phase 2b clinical trial of Stelara (ustekinumab) for moderate-to-severe Crohn's disease patients with at least 3-month history of disease, and who met the criteria for either primary or secondary non-response to anti-TNF, or had unacceptable side effects after receiving a TNF antagonist at an approved dose (Peters et al., 2017; Sandborn et al., 2012). Data from biopsies in involved ileums were analyzed (n = 50).

Microarray data from the UNITI cohorts were obtained from the phase 3 clinical trial of Stelara for moderate-to-severe Crohn's disease (Feagan et al., 2016). Patients in the UNITI-1 cohort met the criteria for either primary or secondary non-response to anti-TNF, or had unacceptable side effects after receiving a TNF antagonist at an approved dose, while patients from the UNITI-2 cohort had failed conventional therapies but had not experienced failure or intolerance of biologics at time of inclusion. Data from biopsies in involved ileums before ustekinumab introduction were analyzed (UNITI-1: n = 61; UNITI-2: n = 141).

### Definition of non-response to anti-TNF

Analyses involving response or non-response to anti-TNF were conducted in patients from the RISK cohort. Inclusions were limited to patients who received anti-TNF within the first year after diagnosis. We defined strict criteria for response by considering patients

achieving durable (6 months) corticosteroid-free clinical remission (pediatric Crohn's disease index (PCDAI) < 10) at months 18 and 24 post-diagnosis, as responders.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Parameters such as sample size, measures of center, dispersion, and precision (mean  $\pm$  SD or SEM), and statistical significance are reported in text, Figures and Figure Legends. Results were considered statistically significant when  $p < 0.05$ , by the appropriate test, as indicated in text and Figure Legends.

### Generation of gene expression matrices

FASTQ were demultiplexed using Cell Ranger v2.0 and aligned to the Grch38 human reference genome. Cell barcodes and unique molecular identifiers (UMIs) were extracted and "Raw" UMI matrix generated for each sample, providing with the number of UMIs for each gene that are associated with individual cell barcode. Cells enriched with mitochondrial, epithelial, or hemoglobin mRNA were excluded (see below Filtering low-quality cells and contaminating cells).

### Filtering low-quality cells and contaminating cells

We extracted cell-barcodes associated with at least 800 UMIs from the "Raw" output UMI matrices of CellRanger (Figure S1A), and filtered out cell-barcodes with at least 0.25% of mitochondrial mRNA (Figure S1B). Because collagenase digestion severely alters epithelial cell integrity, we also filtered out epithelial cells by excluding cell-barcodes with more than 1% UMIs associated with genes from the list below (Figure S1B). Potential residual red-blood cells were filtered out by excluding cell-barcodes with more than 10% UMIs associated with hemoglobin genes. Cell filtering statistics for the different samples are summarized in Table S2, sheet 1. Because of low cell counts recovery which prevented an accurate estimation of the frequencies of rare cellular subtypes, Patient 6 was excluded from analyses conducted after the clustering. The variability in cell-counts or UMI counts between the other samples was not confounding downstream analyses.

Epithelial gene list: *PLA2G2A, CLCA1, REG4, S100A14, ITLN1, ELF3, PIGR, EPCAM, REG1B, REG1A, REG3A, FABP1, RBP2, SST, FABP2, SPINK1, FABP6, AGR2, AGR3, CLDN3, CLDN4, DEFA6, DEFA5, SPINK4, ALDOB, LCN2, MUC2, KRT8, KRT18, TSPAN8, OLFM4, GPX2, IFI27, PHGR1, MT1G, CLDN7, KRT19, FXYD3, LGALS4, FCGBP, TFF3, TFF1*

### A batch-aware, multinomial mixture model for single-cell RNaseq data

Analysis of background noise across samples showed that noise profiles were proportional to the average gene expression profile of the sample as well as to the number of UMIs in the receiving cell. Our previous studies indicated that this type of cell-to-cell contamination could be attributed to molecular switching events that could occur during the amplification of the pooled cDNA library. We updated our previously published batch-aware mixture model (Jaitin et al., 2014; Paul et al., 2015) in order to accommodate the noise distribution observed here and defined the probability of observing gene  $i$  in cell  $j$  in the revised model as:

$$p_{ji} = \frac{1}{Z} [K_{reg} + (1 - \eta_{b^j}) \cdot \alpha_{i, map^j} + \eta_{b^j} \cdot \beta_{i, b^j}]$$

Where  $map^j$  and  $b^j$  are assignments of cells  $j$  to cell-type and batch(sample) respectively;  $\eta_{b^j}$  is the fraction of UMIs contributed by background noise;  $\alpha_{i, map^j}$  is the probability that a molecule drawn from celltype  $map^j$  is of gene  $i$  (assuming no background noise) and  $\beta_{i, b^j}$  is the probability that a noise UMI drawn from batch  $b^j$  will be of gene  $i$ .

Given this model and assuming hard association of cells with types, we can compute the log likelihood of the entire dataset as:

$$LL(U) = \sum_j \sum_i U_{ij} \log(p_{ji})$$

Based on the analysis of background noise distribution, we estimated the noise multinomial parameters by the average expression of the batch including all cell barcodes before filtering:

$$\beta_{i, B} = \frac{\sum_{j \in D^B} U_{ij}}{\sum_{i \in G_j \in D^B} U_{ij}}$$

where  $D^B$  is a set of all barcodes with more than 100 UMIs in batch  $B$  and  $G$  is a set of all genes.

### Inference of the model parameters

We revised our previously described pseudo expectation-maximization (EM) algorithm to infer the assignment of cells to clusters  $map^j$ , gene probability within cluster  $\alpha_{i, map^j}$  and background noise fraction  $\eta_{b^j}$ . We implemented a cross-validation approach to

test the robustness of the model' parameters at each EM iteration and avoid over-fitting of the data. The updated algorithm outline was as follows:

1. Sample the single-cell sets:
  - A. Randomly sample without replacement 1000 cells from each batch to the learning set and 1000 cells to the test set. The resulting UMI matrices would be  $U^L$  and  $U^{Test}$  respectively.
  2. Initialize the model (repeat A-F 10000 times):
    - A. Randomly select a value  $N_{ds\_umis}$  from the  $(P_1, P_2)$  percentiles of the empirical distribution of the number of UMIs/cell of  $U^L$ .
    - B. Downsample  $U^L$  to  $N_{ds\_umis}$  UMIs/cell. The downsampled matrix is denoted as  $U^{1L}$ .
    - C. Select highly variable genes (see below)
    - D. Cluster the cells in  $U^{1L}$  based on the genes selected in step C using kmeans++ (<https://tanaylab.bitbucket.io/tglkmeans/>), with  $k$  seeds, following  $\log_2(X+K_{reg\_ds})$  transformation.
    - E. Estimate  $\alpha, \eta$  given kmeans++ assignments  $map$  for the cells in  $U^{1L}$  (see description below).
    - F. Given the current values of  $\alpha, \eta$  parameters calculate the MAP assignment for each cell in  $U^L$  and update the assignments of cells to clusters  $map^j$ .
    - G. Compute the log-likelihood of  $U^L$  to each of the current initialized types.
    - H. Select model parameters that correspond to the randomized seed that maximized the log-likelihood of  $U^L$ .
  3. Estimate  $\alpha, \eta$  given  $map$  for the cells in  $U^L$ .
  4. Given the values of  $\alpha, \eta$  calculate the MAP assignment for each cell in  $U^L$  and update the assignments of cells to clusters  $map$ .
  5. Return to step 3 and repeat until the likelihood converges, or the maximum number of iterations is reached.
  6. Cross validation:
    - A. Given the values of  $\alpha, \eta$  calculate the MAP assignment for each cell in  $U^{Test}$  and update the assignments of cells to clusters  $map$ .
    - B. Estimate  $\alpha^{Test}, \eta^{Test}$  given  $map$  for the cells in  $U^{Test}$  and compare with  $\alpha, \eta$ .
    7. Estimate  $\eta, map$  given  $\alpha$  for  $U$

### Selecting highly variable genes

Similarly as in other publications (Baran et al., 2018; Jaitin et al., 2014; Paul et al., 2015), we selected genes with variability that could not be explained by multinomial sampling variance. We calculated a loess curve for the  $\log(\text{variance}/\text{mean})$  versus  $\log(\text{mean})$  distribution and binned the  $\log(\text{variance}/\text{mean})$  values by intervals of 0.2 of  $\log(\text{mean})$ . We selected genes with more than 50 UMIs in  $U^{1L}$  from the 8<sup>th</sup> percentile of each bin and also required that their  $\log(\text{variance}/\text{mean})$  is 0.1 or higher above the loess curve (Figure S1E).

### Estimation of the model parameters

- A. Estimate the model parameters. Repeat i-ii until convergence:
  - i) Subtract noise contribution from multinomial parameters (Figures S1F and S1G):

$$\alpha_{i,T} = \frac{1}{Z} \left[ \max \left\{ 0, \sum_{j \text{ s.t. } map^j = T} U_{ij} - \zeta_{bj} \cdot \beta_{i,bj} \cdot N_j^{UMIs} \right\} \right]$$

where  $N_j^{UMIs} = \sum_{i \in G} U_{ij}$ ;  $G$  contains all genes.  $Z$  is a normalization factor.

- ii) Estimate the noise parameters (Figures S1H and S1I) by iterating over each batch  $b$  and selecting  $\eta_b$  that will optimize the likelihood of the cells in the batch  $LL(U^{L,B}) = \sum_{j \in B} \sum_i U_{ij}^L \log(p_{ji})$  given constant  $\alpha$  and  $map$ . Optimization process is simple since the likelihood components are additive.

For the joint clustering of the inflamed and uninfamed ileum samples we included cell-barcodes with more than 1000 UMIs and used  $K_{reg\_ds} = 0.1$ ;  $(P_1, P_2) = (10^{\text{th}}, 40^{\text{th}})$  percentiles;  $K_{reg} = 5 \cdot 10^{-6}$ ;  $k = 50$ .

In the clustering analysis of PBMC samples we included all barcodes with more than 600 UMIs and used  $K_{reg\_ds} = 0.1$ ;  $(P_1, P_2) = (10^{\text{th}}, 50^{\text{th}})$  percentiles;  $K_{reg} = 5 \cdot 10^{-6}$ ;  $k = 20$ .

Permissive and fixed cutoffs on the number of UMIs were selected to allow subtypes with low UMI counts (Figure S1K) such as mast cells and specific subtypes of T cells to be included in the analysis across all samples. Findings of the study were not affected by modification of these thresholds within a range of 500-2000 UMIs/cells.

In order to improve the initiation of the model, genes with lateral expression across different subtypes were excluded from the k-means clustering (step 2.D.). These genes consisted of the following groups: cell-cycle, stress, ribosomal and mitochondrial, metallothionein genes, immunoglobulin genes, HLA class I and II genes and 3 specific genes with variable/noisy expression: *MALAT1*, *JCHAIN* and *XIST*.

To minimize technical variability, scRNAseq data from patient 5 samples, which were generated using Chromium V1 chemistry and showed significant systematic biases, were not used to generate the clustering model in steps 1-6. Patient 5 data were well projected accurately onto the model generated with the V2 chemistry and showed high similarity to data of other patients.

Cells associated with two clusters (#23 and 44) contained marker genes accounting for distinct cellular lineages and were excluded as they likely corresponded to doublets. One cluster of PC (#9) had a similar expression profile to other plasma cell clusters but was excluded as very few cells were projecting in it.

Batch corrected gene expression profiles are obtained by subtracting the expected number of UMIs associated with background noise (“noise UMIs”) from the observed number of UMIs per cluster. Gene expression profiles per cellular subtypes previously defined by grouping clusters with high transcriptional similarities (see below) were estimated using the same approach (Table S2, sheet8). Ordering of clusters as well as samples was performed by hierarchical clustering with GW adjustment as implemented by the R package Seriation.

### Gene modules

We downsampled the cells to 2000 UMIs/cells and selected variable genes similarly to the seeding step of the clustering. In order to focus on biologically relevant gene-to-gene correlation, we calculated a Pearson correlation matrix between genes for each sample. For that purpose expression values were log transformed  $\text{Log}(1+\text{UMI}(\text{gene},\text{cell}))$  while genes with less than 5 UMIs were excluded. Correlation matrices were averaged following z-transformation. The averaged z matrix was then transformed back to correlation coefficients. We grouped the genes into gene “modules” by complete linkage hierarchical clustering.

### Principal component Analysis

Principal component analysis was performed for adjusted subtype frequencies with difference larger than 4-fold between the 15 and 85 percentiles across inflamed and uninfamed samples. The first principal component explained 65.5% of the variability of these subtypes. PCA results are shown in Figures 1H, S1P and S1Q.

### Differential expression analysis

We tested for differential expression between two sets of cells by estimating the gene expression per set (similarly to the estimation of the model multinomial parameters) and calculated the observed log-fold-change between the two sets for each gene. We then randomly shuffled the cells of the two sets for  $10^5$  permutations while maintaining the sizes of the sets and calculating the log-fold-change between the permuted sets for each permutation. The empirical p value was then defined as based on the rank of the absolute value of observed the log-fold-change of each gene within its empirical fold-change distribution.

When comparing between groups of samples, we only tested genes with at least 100 cells with non-zero UMI counts and capped the number of cells per sample to equally represent the different samples within each group. Specifically, for the analysis that compared the total gene expression between the two patient groups we used a maximum of 1000 cells/samples.

### Ligand-Receptor Analysis

We examined specific ligand-receptor pairs by restricting our analysis to cytokine/chemokine-receptor pairs validated experimentally (Ramilowski et al., 2015). Averaged expression of receptor R in a target subtype T was defined as:

$$M_{R,T} = \frac{\sum_{j \in T} U_{Rj}}{\sum_{j \in T \in \text{Genes}} \sum U_{ij}}$$

While globally normalized ligand expression expressed by source subtype S was defined as:

$$C_{L,S} = \frac{\sum_{j \in S} U_{Lj}}{\sum_{j \in \text{Sample} \in \text{Genes}} \sum U_{ij}}$$

Normalizing the ligand expression globally allows rough approximation of the contribution of the subtype to the ligand levels in the tissue. The intensity of the ligand-receptor pair (L,R) between the source subtype S and target subtype T is defined as:

$$I(L, R, S, T) = M_{R,T} C_{L,S}$$

The ligand receptor analysis present in Figures 4A, S5B, and S5C included pairs for which  $M_{R,T} > 10^{-5}$   $C_{L,S} > 10^{-6}$  in at least two inflamed samples of either patients enriched or not enriched with the module. The score of each pair was separately averaged across inflamed samples of patients enriched or not enriched with the modules and the log ratio between them was calculated (see Table S2, sheet 6). Permutation test for empirical estimation of significance was conducted similarly to the test described above while restricting the analysis to ligand and receptor gene list and comparing  $I(L, R, S, T)$  between the groups instead of expression of a single genes. If less than 20 cells in the source or target subtypes associated with UMIs for the ligand or receptor respectively p value was not estimated.

### Defining the GIMATS module high and GIMATS module low gene sets

We sought to define a scoring method that could capture the two modalities observed in our scRNA-seq ileum dataset in bulk RNA-seq or microarray data from ileal biopsy samples. We screened for genes with differential total expression between scRNA-seq inflamed samples of GIMATS-high or GIMATS-low groups and from these two lists we selected only genes that showed higher expression in cellular subtypes associated with the corresponding group. These criteria provided GIMATS-module associated or not-associated lists of genes differentially expressed on the bulk level in the scRNA-seq dataset.

This analysis the GIMATS-module-high group included patients {5,7,8,11,12}, while the GIMATS-module-low group included patients {10,13,14,15}. Based on the scRNA-seq results we defined two sets of cellular subtypes enriched in GIMATS-module-high and GIMATS-module-low patient groups respectively, as follows:

SubtypeSet<sup>high</sup> = {infl.macs, mature DC, pDC, IgG-plasma cells, central memory T cells, cytotoxic T cells, Tregs, ACKR1+ endothelial cells, activated fibroblasts}

SubtypeSet<sup>low</sup> = {resident macrophages, moDC, IgA-plasma cells, ILC3, CD8 TRM cells, TRM, cytotoxic CD8 TRM, enteric neurons, CD36+ endothelial cells, fibroblasts}

We first ran a differential expression analysis (as described above) between inflamed samples of the two groups of patients defined above, and calculated the maximum expression across all cellular subtypes in SubtypeSet<sup>high</sup> and SubtypeSet<sup>low</sup> for each gene, denoted  $m^{\text{high}}$  and  $m^{\text{low}}$  respectively.

We then screened for genes for which all the following conditions were true:

1. Frequency of  $10^{-6}$  or higher in at least one of the patient groups
2.  $|\log_2(X^{\text{high}}(g)/X^{\text{low}}(g))| > 1$  and FDR adjusted (BH) p value  $< 10^{-3}$
3.  $|\log_2(m^{\text{high}}(g)/m^{\text{low}}(g))| > 3$

where  $X^{\text{high}}(g)$ ,  $X^{\text{low}}(g)$  are the UMI fraction of gene  $g$  in GIMATS-module-high and GIMATS-module-low groups respectively, as provided by the differential expression analysis.

The final genes were then assigned to gene Set<sup>high</sup> if  $X^{\text{high}}(g) > X^{\text{low}}(g)$  or to Set<sup>low</sup> otherwise.

### Projecting bulk RNA data onto GIMATS module high and GIMATS module low scores

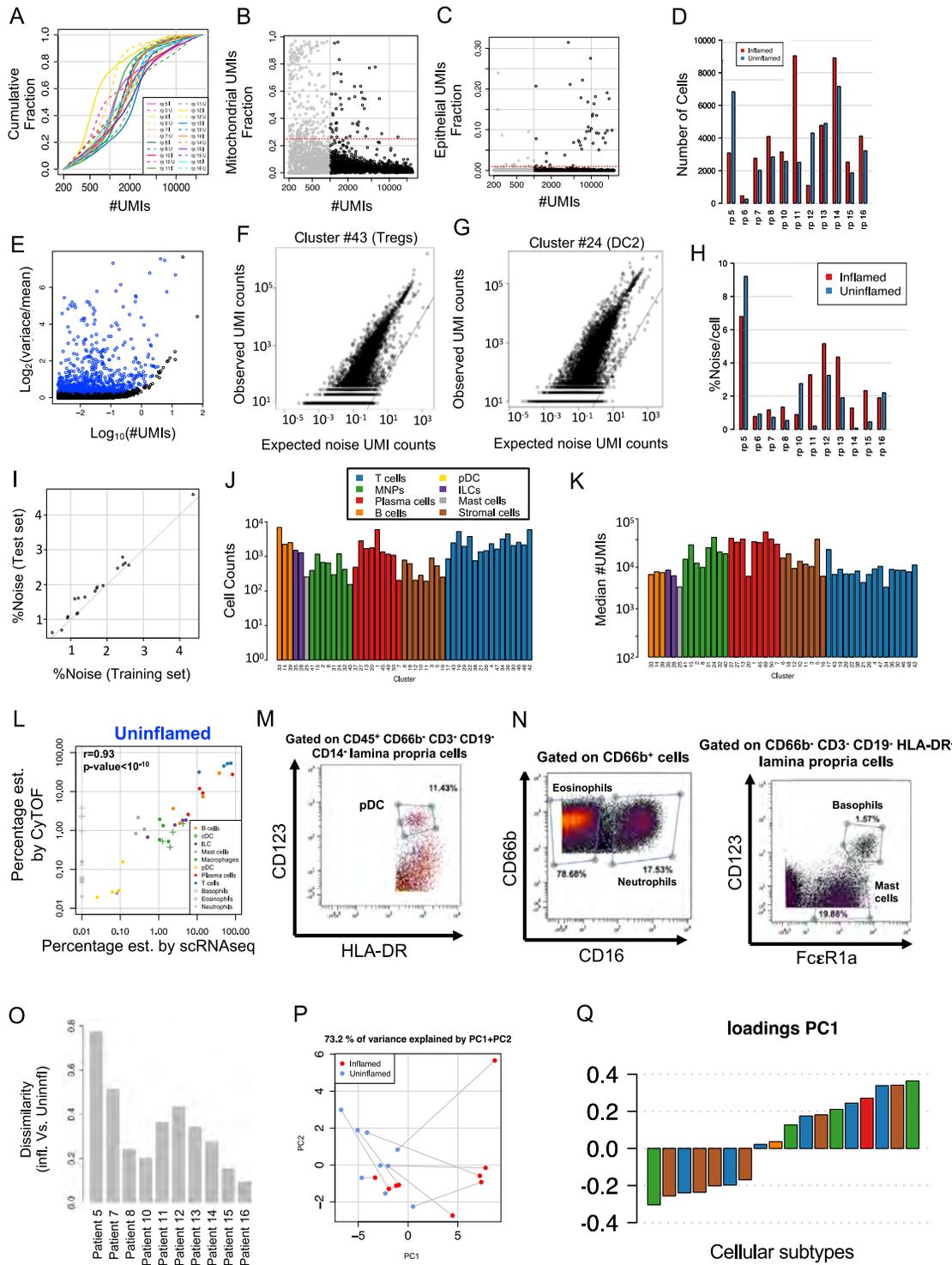
Microarray data of ileal tissues from Crohn's disease patients from CERTIFI (all) (GEO: GSE83687), UNITI-2 (GEO: GSE112366) and UNITI-1 (before treatment or placebo) cohorts were normalized using the R Bioconductor package XPS with standard parameters, applying RMA background correction, quantile normalization and farms summarization. RISK RNA-seq reads were mapped using TopHat2 to the human reference genome version 19. Following RNA-seq mapping, expression levels at the gene was determined and expression quantified using Cufflinks to generate FPKM estimates (Trapnell et al., 2012). Each dataset was then transformed to z-scores and projected onto the GIMATS-module and "No module" scores by averaging over the z-scores of genes in Set<sup>high</sup> and in Set<sup>low</sup> respectively.

### DATA AND CODE AVAILABILITY

Raw sequencing reads of scRNA-seq samples as well as UMI tables are available on the Gene Expression Omnibus under accession number GEO: GSE134809. Bulk RNaseq data from the RISK cohort is available under accession number GEO: GSE134881. The R code developed for clustering and analyses in this study is available in our Github page: [https://github.com/effiken/martin\\_et\\_al\\_cell\\_2019](https://github.com/effiken/martin_et_al_cell_2019).

### ADDITIONAL RESOURCES

Expression profiles and raw scRNaseq data are publicly available through an online application for data analysis allowing the interactive multidimensional exploration of the different transcriptional programs (<https://scdissector.org/martin/>).



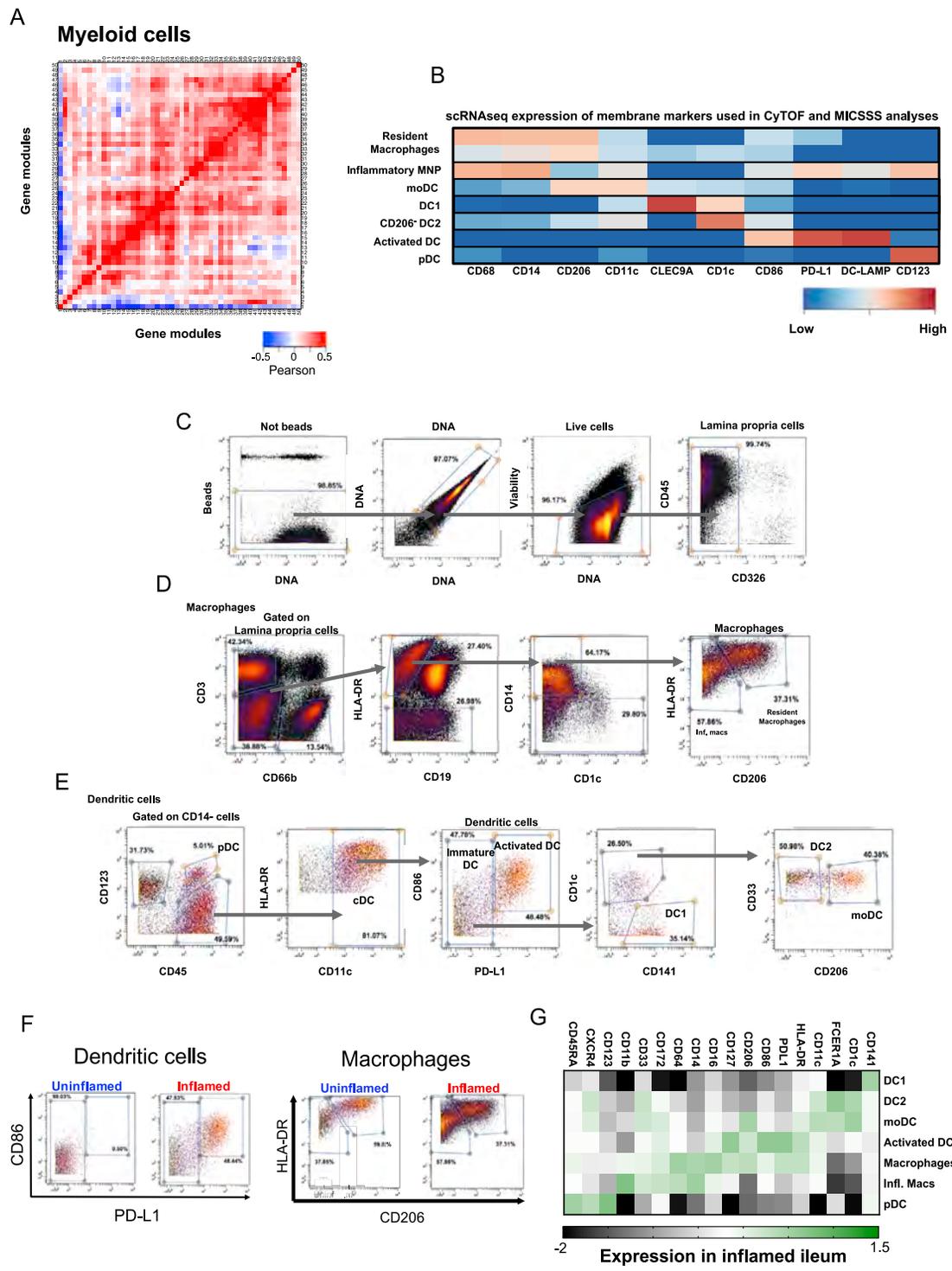
**Figure S1. Clustering Analysis of 22 Inflamed and Uninflamed Ileum CD Tissues, Related to Figure 1**

(A) Cumulative fraction (y axis) of UMI counts per cell (x axis) for inflamed (solid lines) and uninflamed (dashed lines) samples. (B) Filtering low-quality cell-barcodes. Shown are UMIs fraction of mitochondrial genes (y axis) versus the number of UMIs (x axis) per cell-barcode for a representative sample (patient 7 inflamed ileum). Cell-barcodes below 1000 UMIs were filtered (gray dots) while cell-barcodes with mitochondrial fraction lower than 0.25 (red dashed line) were included for further analysis. (C) Similar to B but showing filtering of epithelial cells. Cell-barcodes with more than 1% epithelial UMI fraction were filtered. (D) Number of cells per patient per tissue recovered after QC steps. (E) Selection of variable genes for clustering initiation. Shown is the variance divided by the mean versus the mean for genes included in the initiation step of the clustering. UMI matrix was down-sampled to 557 UMIs. (F-G) Estimation of background noise.

(legend continued on next page)

---

Shown are observed UMI counts versus the expected number of noise UMIs per gene for Treg (F) and DC2 (G) clusters with the identity line (solid). Genes not expressed in these cell subtypes, serve as endogenous “spike-ins” and almost all their UMIs in these clusters are predicted to be noise UMIs. (H) Estimated background noise percentages per inflamed (red bars) and uninfamed (blue bars) sample. (I) Cross validation of noise parameters. Noise percentage per sample as estimated for the test set (y axis) versus the training set (x axis). (J) Number of cells per cluster. Clusters are grouped by cellular compartment. (K) Median number of UMIs/cells for each cluster. (L) Scatterplot showing lineage frequencies estimated based on CyTOF (y axis) versus scRNaseq (x axis) data of uninfamed ileums (similarly to [Figure 1E](#)) of CD patients (n = 4 patients). Frequencies of granulocyte populations as determined by CyTOF are also shown but were excluded from the correlation analysis. (M) Representative gating of CD45<sup>+</sup> CD66b<sup>-</sup> CD3<sup>-</sup> CD19<sup>-</sup> CD14<sup>-</sup> HLA-DR<sup>+</sup> CD123<sup>+</sup> plasmacytoid dendritic cells (pDCs). (N) Representative gating strategies to identify granulocyte populations by CyTOF analysis of CD ileal tissues. Left: Representative gating of CD45<sup>+</sup> CD66b<sup>+</sup> CD16<sup>+</sup> neutrophils and CD45<sup>+</sup> CD66b<sup>+</sup> CD16<sup>-</sup> eosinophils. Right: Representative gating of CD45<sup>+</sup> CD66b<sup>-</sup> CD3<sup>-</sup> CD19<sup>-</sup> HLA-DR<sup>-</sup> FcεR1a<sup>+</sup> CD123<sup>+</sup> basophils and CD123<sup>-</sup> mast cells. (O) Dissimilarity between inflamed and uninfamed samples. Euclidian distance between the subtype frequencies in matched inflamed and uninfamed samples. (P) Principle component analysis of cellular subtypes composition. First two components provided by principal component analysis of adjusted cell subtype frequencies (frequencies were normalized by cellular lineages). Blue and red dots correspond to uninfamed and inflamed ileums respectively. Paired uninfamed and inflamed ileal tissues are connected by a line. Patient identification numbers are indicated next to their corresponding inflamed tissue (red dots). (Q) Loadings of principal components 1 (color-coded by cell-lineage) corresponding to the PCA analysis in [Figures 1H](#) and [S1P](#).



**Figure S2. Validation of Myeloid Cells Diversity at the Protein Level, Related to Figure 2**

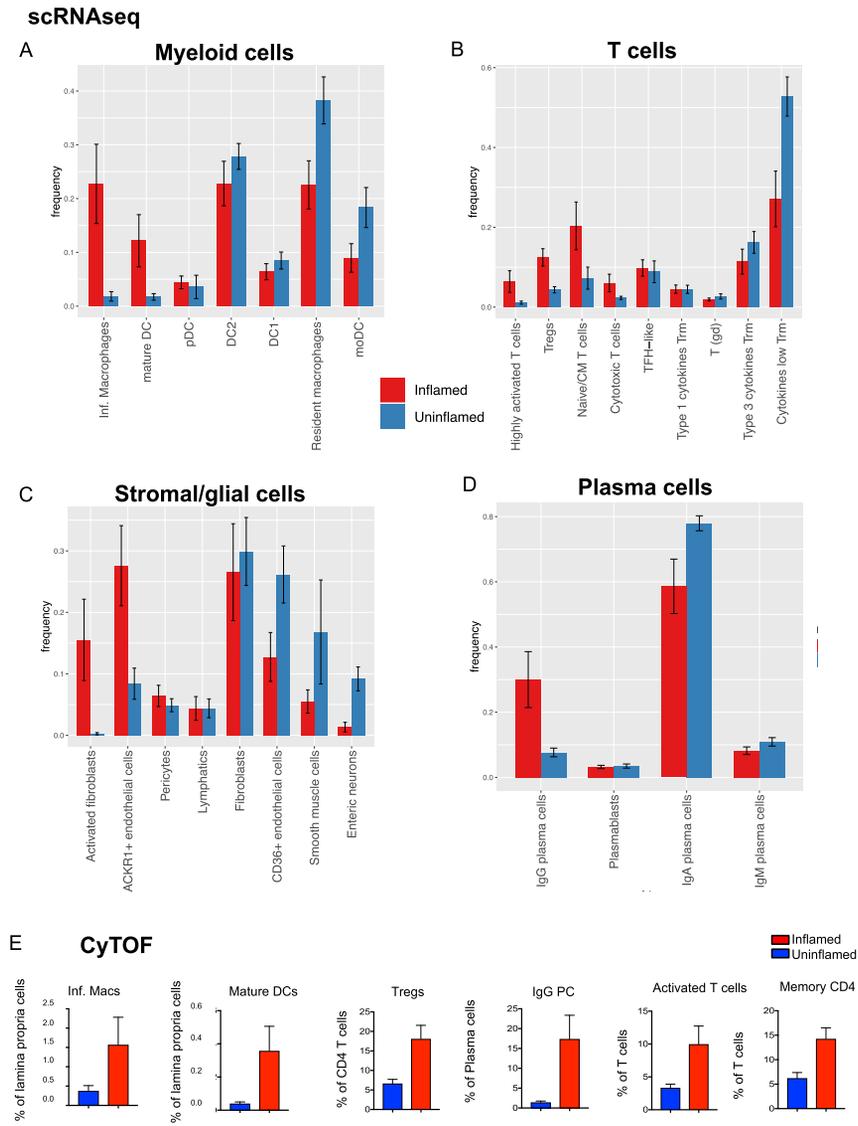
(A) Correlation matrix of gene modules defined by gene-to-gene correlation analysis of single cells within the myeloid clusters. Mapping between modules and genes is detailed in Table S2, sheet 4. (B-G) Validation of the scRNAseq analysis of MNP at the protein level. Heat-map showing relative expression values of genes encoding cell membrane markers (columns) enriched in macrophages or dendritic cells clusters (rows) (B). Step-by-step gating strategy of lamina propria cells (C). Step-by-step gating strategy of macrophages in total lamina propria cells (D). Step-by-step gating strategy of dendritic cells within CD14<sup>+</sup> cells (E). Representative dot plots of immature and activated DCs (left) and macrophages (right) in uninflamed and inflamed ileums (F). Heatmap showing the color-coded log-normalized expression of membrane markers between myeloid cell populations analyzed by CyTOF (n = 7 CD patients) (G).



---

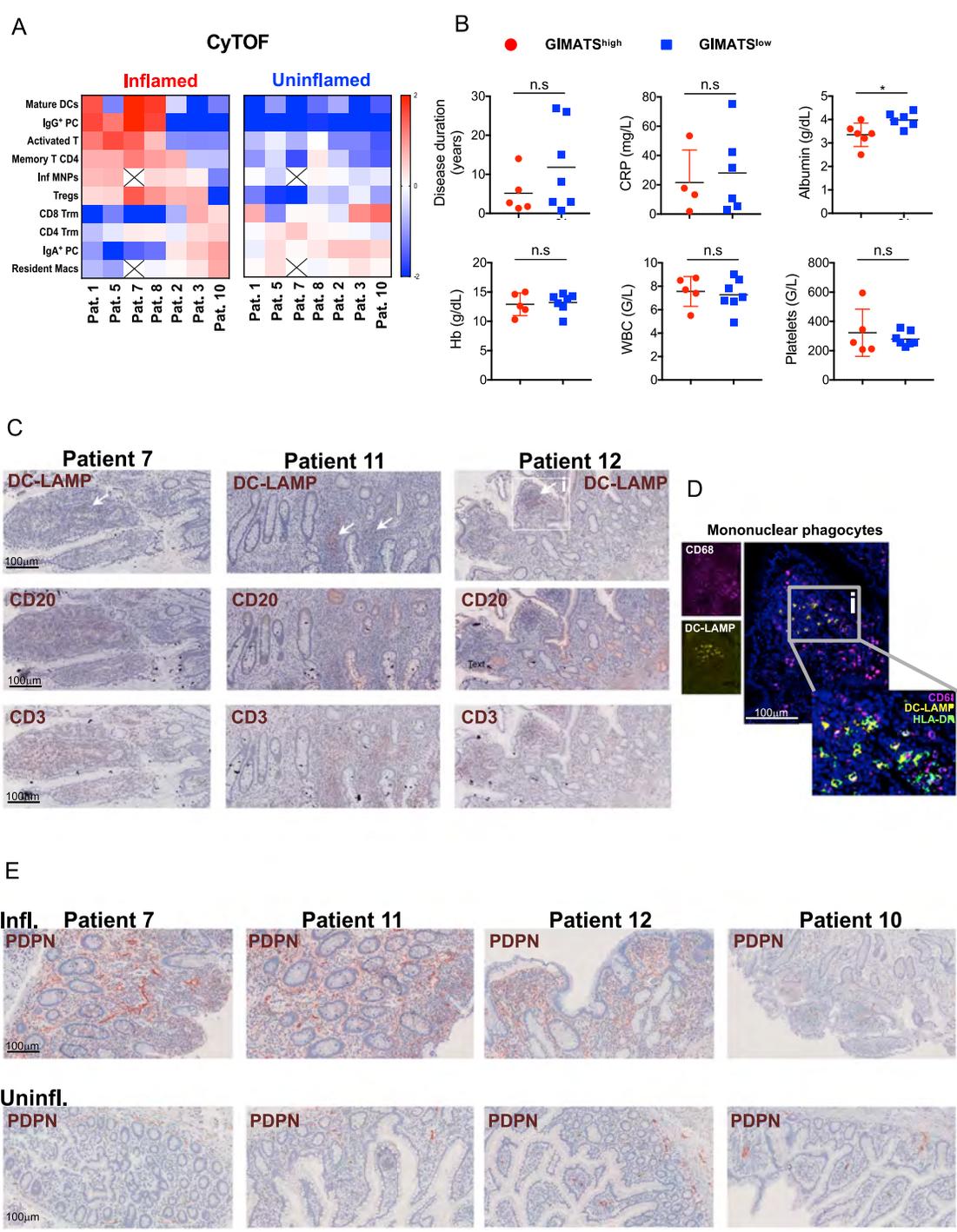
**Figure S3. Characterization of Lymphocyte Populations iCD, Related to Figure 2**

(A) B cell clusters. Stacked columns correspond to single-cells and rows to selected differentially expressed genes for which down-sampled UMI counts are color-coded. Clusters are demarcated by gray bars. (B) Representative gating of IgA<sup>-</sup> CXCR4<sup>+</sup> plasma cells corresponding to IgG producers, in uninflamed and inflamed ileum. (C) Correlation matrix of gene modules defined by gene-to-gene correlation analysis of single-cells within the lymphocyte clusters. Mapping between module and genes is detailed in [Table S2](#), sheet5. (D) CyTOF analysis of T cells. Left: Heatmap showing the color-coded normalized expression of membrane markers between T cell populations analyzed by CyTOF (n = 7 CD patients). Right: Representative gatings of HLA-DR<sup>+</sup> activated T cells and CD25<sup>high</sup> CD127<sup>low</sup> Tregs in uninflamed and inflamed ileum. (E) T cell clusters. Stacked columns correspond to single-cells and rows to selected differentially expressed genes for which down-sampled UMI counts are color-coded. Clusters are demarcated by gray bars. (F) Innate lymphoid cell clusters. Left: Stacked rows correspond to single-cells and columns to selected differentially expressed genes for which down-sampled UMI counts are color-coded. (G) Relative enrichment of group 1 ILCs and ILC3 in uninflamed and inflamed ileums. Shown are stacked frequencies of the two ILC subtypes divided by the total compartment frequency estimated by scRNaseq in uninflamed or inflamed ileum.



**Figure S4. Distribution of Cell-Subtype Frequencies in Uninflamed and Inflamed Ileums, Related to Figure 2**

(A-D) Bar graphs of average frequencies of indicated cellular subtypes of myeloid cells (A), T cells (B), stromal/glial cells (C) and plasma cells (D) in uninflamed and inflamed ileums (n = 9 iCD patients: #5, 7, 8, 10, 11, 12, 13, 14, 15; mean ± SEM). (E) Bar graphs of average frequencies of indicated populations as determined by CyTOF analysis of uninflamed and inflamed ileums (n = 7 CD patients: #1, 2, 3, 5, 7, 8, 10; mean ± SEM).

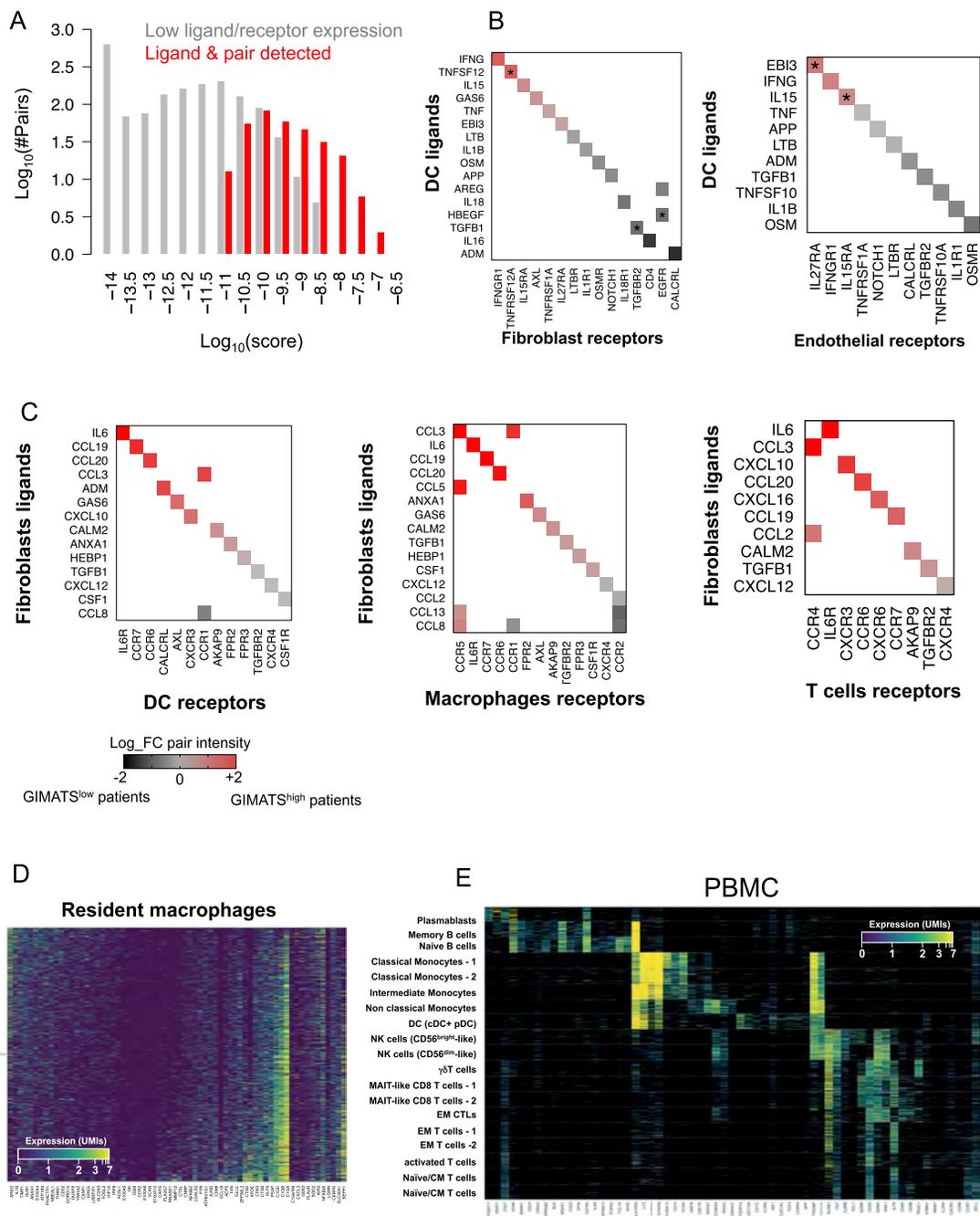


**Figure S5. CyTOF and MICSSS Analysis of the GIMATS, Related to Figure 3**  
 (A) CyTOF data confirm the differential distribution of the GIMATS module between subgroups of patients. Heat-map showing the log<sub>2</sub>-normalized frequency of indicated cell subtypes (rows) in individual inflamed and uninflamed patient tissues (columns). (B) Comparison of clinical parameters between GIMATS module high and GIMATS module low patients. Patients analyzed by scRNAseq and/or CyTOF are included in the comparison. CRP: C-reactive protein; WBC: white blood count; ESR: erythrocyte sedimentation rate; Hb: hemoglobin; Alb: albumin. \* p<0.05. (C) Representative pictures of immunohistochemistry stainings of DC-LAMP, CD20 and CD3 in the inflamed lamina propria of 3 patients enriched for the GIMATS module. DC-LAMP+ cells were not scattered throughout the lamina propria and always detected in T and B cell rich areas forming lymphocyte aggregates. (D) Higher magnification of an activated DC-associated lymphocyte aggregate shown in white

(legend continued on next page)

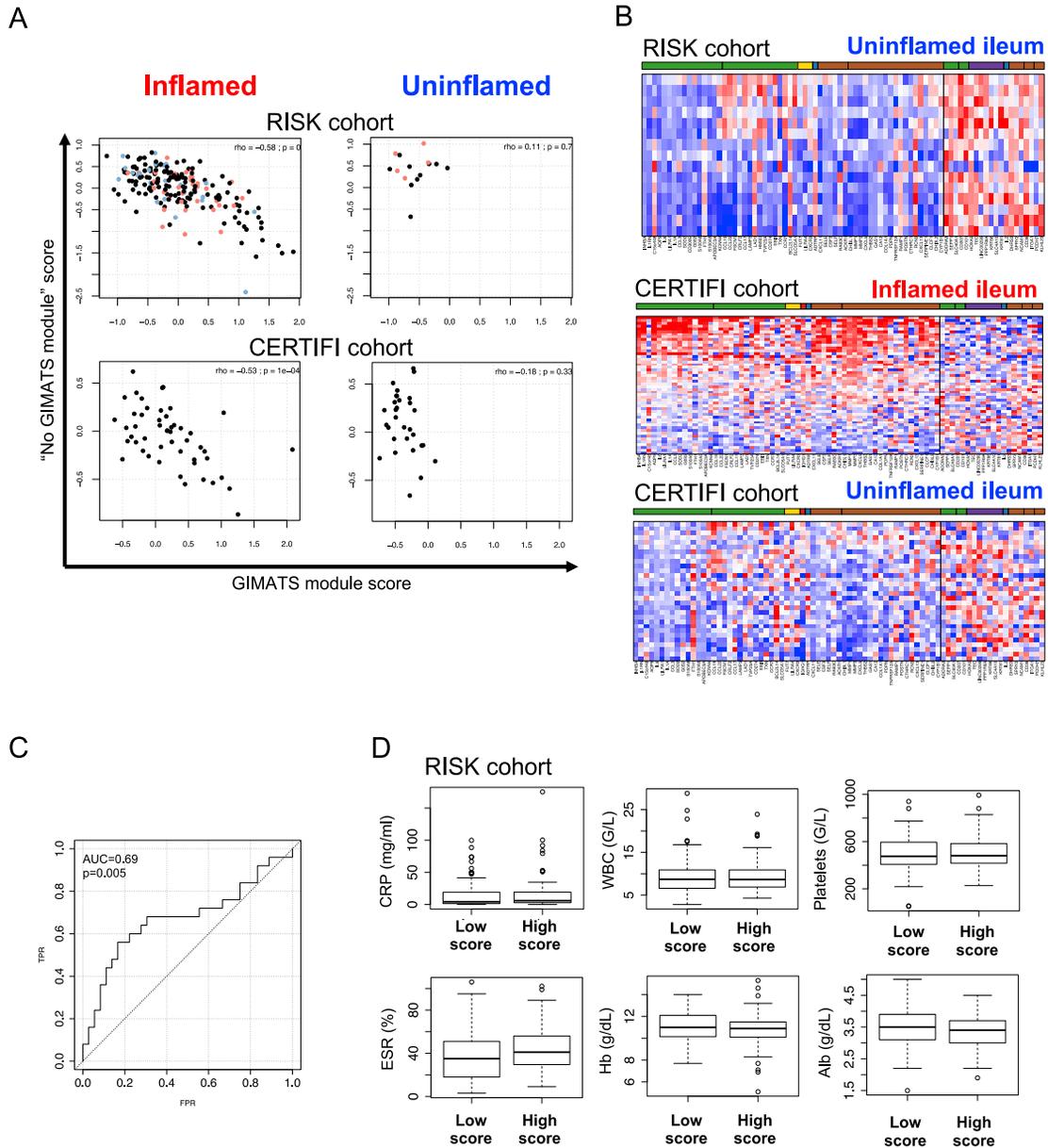
---

square (j) of [Figures 3D-E](#). MNP were stained with CD68, DC-LAMP and HLA-DR to confirm the expression of HLA-DR by CD68- DC-LAMP+ activated DC. (E) Representative podoplanin (PDPN) staining by immunohistochemistry in the inflamed (top) and uninfamed (bottom) lamina propria of patients enriched (Patients 7, 11 and 12) or not enriched (Patient10) for the GIMATS module. In uninfamed tissues as well as in patient 10 inflamed lamina propria, PDPN expression is limited to lymphatics, while PDPN<sup>+</sup> activated fibroblasts are abundant in the inflamed lamina propria of patients enriched for the GIMATS module.



**Figure S6. Ligand-Receptor Analysis, Related to Figure 4**

(A) Ligand-receptor score distribution. Distributions of ligand-receptor intensity scores for validated ligand-receptor pairs where both the ligand and receptor expression were above threshold (red bars) and for pairs with ligand or receptor below threshold (gray bars). (B-D) Comparative ligand-receptor network analysis between the two patient groups. Shown is the relative intensity (color-coded) of the ligand (rows)-receptor (columns) pair in inflamed samples of patients enriched with the module versus patients not enriched with the module for different cellular cell-types pairs. The intensity of a ligand receptor pairs is the product between the total expression of the ligand and the expression of the receptor (see methods). Only validated pairs, which are expressed at least in one of the patient groups, are included. Stars correspond to FDR adjusted (Benjamini-Hochberg) p value < 0.01. (D) Resident macrophages lack strong monocyte signature – related to figure 4D. Heat-map showing down-sampled UMI counts of monocyte and macrophage genes as well as resident macrophage marker genes for cells of the two resident macrophage clusters. Cells are ordered by the ratio between averaged expression of the monocyte and macrophage genes. (E) PBMC clustering analysis. Stacked columns correspond to single-cells and rows to selected differentially expressed genes for which down-sampled UMI counts are color-coded. Clusters are demarcated by gray bars.



**Figure S7. Blood Markers of Inflammation Do Not Predict GIMATS Module Scores, Related to Figure 5**

(A) Scatterplot showing the projected bulk RNA microarray and sequencing data of inflamed biopsies from the RISK (top) and CERTIFI (bottom) cohorts onto the signature scores defined after the scRNaseq analyses to detect enrichment of the GIMATS module. (B) Heatmap of normalized bulk RNA-seq gene expression values (read-high; blue-low) for selected genes (columns) for different CD patients' ileal samples (stacked rows), ordered according to their enrichment for the GIMATS-high score. Shown are uninflamed ileum from the RISK cohort, inflamed and uninflamed ileum from the CERTIFI cohort. (C) Receiver operating characteristic (ROC) curve of the GIMATS module expression at diagnosis to distinguish patients achieving or not achieving durable corticosteroid-free remission upon anti-TNF therapy. (AUC = 0.69, permutation test  $p$  value = 0.005). (D) Whisker plots comparing markers of systemic inflammation in GIMATS module high and GIMATS module low patients from the RISK before anti-TNF therapy. CRP: C-reactive protein; WBC: white blood count; ESR: erythrocyte sedimentation rate; Hb: hemoglobin; Alb: albumin.