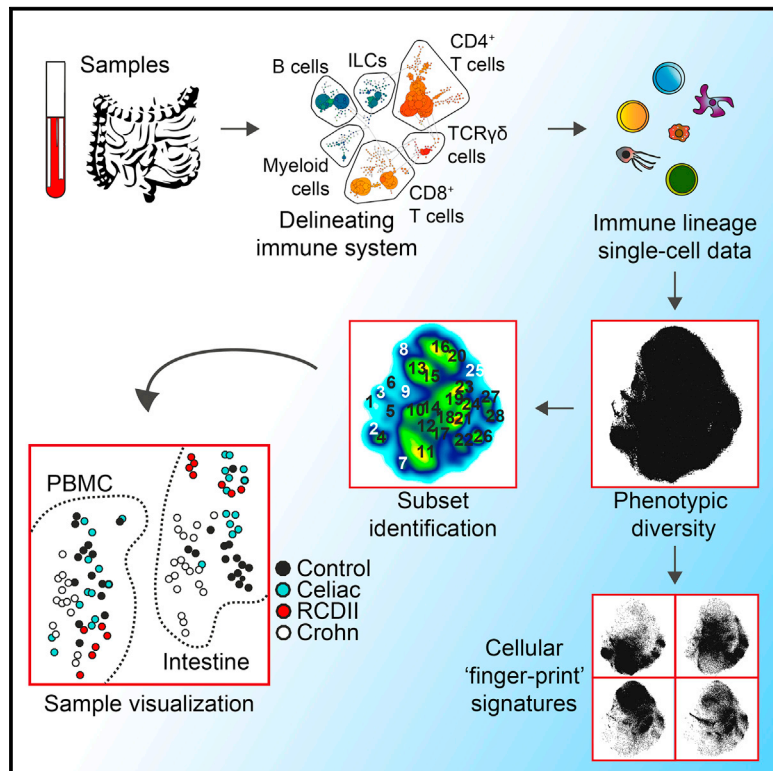# Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets

## Graphical Abstract



## Authors

Vincent van Unen, Na Li,
Ilse Molendijk, ..., Jeroen van Bergen,
Boudewijn P.F. Lelieveldt, Frits Koning

## Correspondence

f.koning@lumc.nl

## In Brief

The role of immune subsets in intestinal pathology has been studied, but a system-wide analysis is lacking. Koning and colleagues use mass cytometry to dissect the human mucosal immune system in health and disease. They identify immune subsets with tissue- and disease-specificity with implications for diagnostic procedures and individualized therapeutics.

## Highlights

- Performed high-dimensional analysis of human mucosal immune system by mass cytometry

- Data-driven approaches revealed previously unrecognized immune cell heterogeneity

- Identified mucosal lymphoid malignancies and their cellular precursors

- Data visualizations identified tissue- and disease-associated immune subsets

CrossMark

CellPress

# Resource

# Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets

Vincent van Unen,[1] Na Li,[1] Ilse Molendijk,[2] Mine Temurhan,[3] Thomas Höllt,[7] Andrea E. van der Meulen-de Jong,[2] Hein W. Verspaget,[2] M. Luisa Mearin,[4] Chris J. Mulder,[6] Jeroen van Bergen,[1] Boudewijn P.F. Lelieveldt,[5,7] and Frits Koning[1,*]

[1]Department of Immunohematology and Blood Transfusion
[2]Department of Gastroenterology
[3]Department of Center for Proteomics
[4]Department of Pediatrics
[5]Department of LKEB Radiology
Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
[6]Department of Gastroenterology, VU Medical Center, 1081 HZ Amsterdam, the Netherlands
[7]Department of Pattern Recognition and Bioinformatics Group, Delft University of Technology, 2600 GA Delft, the Netherlands
*Correspondence: f.koning@lumc.nl
http://dx.doi.org/10.1016/j.immuni.2016.04.014

## SUMMARY

Inflammatory intestinal diseases are characterized by abnormal immune responses and affect distinct locations of the gastrointestinal tract. Although the role of several immune subsets in driving intestinal pathology has been studied, a system-wide approach that simultaneously interrogates all major lineages on a single-cell basis is lacking. We used high-dimensional mass cytometry to generate a system-wide view of the human mucosal immune system in health and disease. We distinguished 142 immune subsets and through computational applications found distinct immune subsets in peripheral blood mononuclear cells and intestinal biopsies that distinguished patients from controls. In addition, mucosal lymphoid malignancies were readily detected as well as precursors from which these likely derived. These findings indicate that an integrated high-dimensional analysis of the entire immune system can identify immune subsets associated with the pathogenesis of complex intestinal disorders. This might have implications for diagnostic procedures, immune-monitoring, and treatment of intestinal diseases and mucosal malignancies.

## INTRODUCTION

The intestinal immune system protects us from bacterial, viral, and parasitic infections. Disruption of intestinal homeostasis, however, can lead to a variety of autoinflammatory intestinal diseases, including celiac disease (CeD) and Crohn's disease (CD), which together have a prevalence of 1,500 per 100,000 adults in the Western world (Kappelman et al., 2013; Rubio-Tapia et al., 2012). Both diseases are multifactorial and encompass a broad spectrum of clinical phenotypes and ages of onset. CeD is a disease of the small intestine caused by pro-inflammatory CD4[+] T cell responses specific for dietary gluten and concomitant destruction of the epithelium due to activation of intraepithelial CD8[+] T cells. The introduction of a strict gluten-free diet constitutes a highly effective treatment for CeD but nevertheless 2%–5% of patients develop refractory CeD type II (RCDII) with persistent inflammation. RCDII is characterized by a monoclonal outgrowth of aberrant intra-epithelial lymphocytes (IELs) from which an aggressive enteropathy-associated T cell lymphoma (EATL) evolves in 40% of patients (Al-Toma et al., 2007). In contrast, CD affects the terminal ileum and/or colon and results from aberrant immune responses against the microbiota (Pascual et al., 2014). CD is usually treated with the use of lifelong pharmacotherapy (Randall et al., 2015), including biologicals (e.g., anti-TNF) to reduce chronic inflammation and to accomplish sustained remission. Despite achieving states of remission, perianal fistulas occur in 25% of CD patients and this is accompanied by multiple relapses and a poor prognosis due to insufficient healing (Molendijk et al., 2014; Schwartz et al., 2002).

Although the role of several immune subsets in driving intestinal pathology has been studied in CeD (Jabri and Sollid, 2009), RCDII (Verbeek et al., 2008), and CD (Geremia et al., 2014), a system-wide approach that simultaneously interrogates immune subsets across all major lineages on a single-cell basis is currently lacking. High-dimensional mass cytometry (cytometry by time-of-flight; CyTOF) now offers the possibility to analyze many cellular markers simultaneously, providing an opportunity to analyze the mucosal immune system with unprecedented resolution (Bandura et al., 2009). Novel computational tools have been developed to handle the high-dimensional single-cell datasets that originate from mass cytometry (Amir et al., 2013; Bendall et al., 2011; Shekhar et al., 2014). In the current study, we applied mass cytometry to analyze the composition of the immune compartment present in intestinal biopsies and paired peripheral blood mononuclear cell (PBMC) samples of patients with inflammatory intestinal diseases and controls. We identified 142 distinct immune cell subsets and through computational
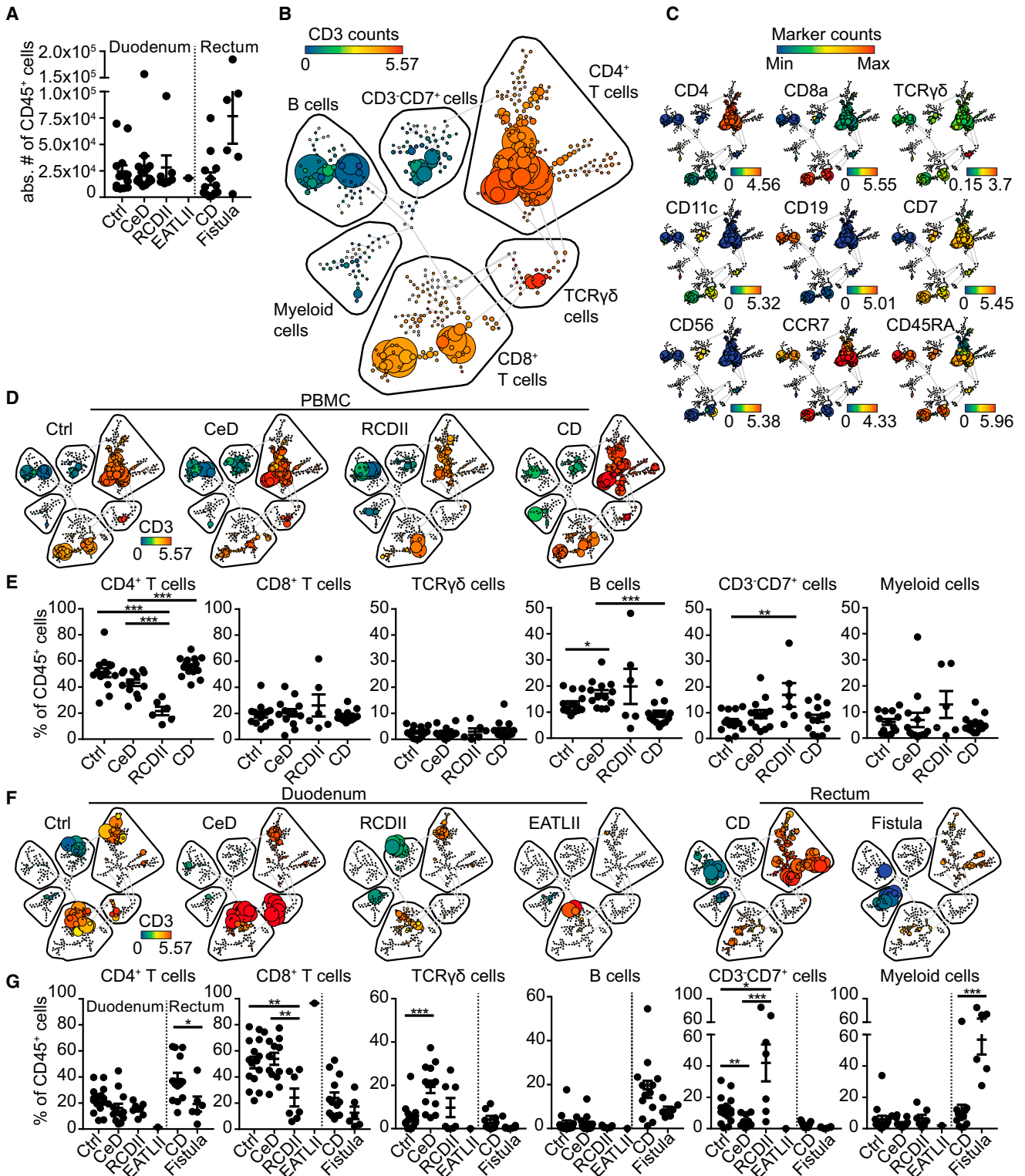
**Figure 1. Collective SPADE Analysis Distinguishes Major Immune Lineages in Peripheral Blood and Intestine**

(A) Live single CD45[+] absolute cell number acquired for 55 intestinal biopsies.

(B) A SPADE tree of a PBMC sample after analysis of the combined 102 sample dataset containing $5.2 \times 10^6$ cells. Size of the nodes is proportional to the respective number of clustered cells. Color bars represent ArcSinh5-transformed values for CD3 marker expression. The identities of major immune lineages are annotated on the basis of lineage marker expression.

(C) Color of the PBMC sample represents expression values for each marker as shown.

*(legend continued on next page)*

applications we found immune subsets in PBMCs and intestinal biopsies that distinguished patients with inflammatory diseases from controls. In addition, mucosal lymphoid malignancies were readily detected as well as the precursors from which these likely derived. Thus, mass cytometry unveiled previously unappreciated heterogeneity in the immune system and our observations might help to develop improved diagnostic and therapeutic approaches for inflammatory bowel diseases.

## RESULTS

### SPADE Analysis Identifies Major Immune Lineages in Peripheral Blood and Intestine

We designed a CyTOF panel of 32 metal isotope-tagged monoclonal antibodies, which was designed to obtain a global overview of the heterogeneity of the innate and adaptive immune system (Table S1). For this purpose, the panel contained lineage markers that distinguish the major adaptive and innate immune cell populations. In addition, markers were included to distinguish naive from memory cells, resting from activated cells, and to identify homing properties and potential responsiveness to humoral factors. With this panel, we analyzed single-cell suspensions from biological samples including duodenum biopsies (N = 36), rectum biopsies (N = 13), perianal fistulas (N = 6), and peripheral blood mononuclear cells (PBMC) from control individuals (N = 15) and from patients with inflammatory intestinal diseases (CeD, N = 13; RCDII, N = 5; EATLII, N = 1 and CD, N = 10) (Table S2). The large majority (N = 28) of antibodies allowed clear discrimination of antibody-positive and -negative cells (Figure S1). To monitor the robustness of the measurements, we included a standardized PBMC sample at regular intervals in the acquisition sessions during the entire 9-month study period. These consecutive control samples yielded highly similar results (Figures S2A–S2C), demonstrating the reproducibility of the data acquisition. We discriminated live, single $CD45^+$ cells with DNA stains and event length. (Figure S2D). From the intestinal biopsies we acquired 27,500 duodenal, 17,500 rectal, and 76,500 perianal fistulous $CD45^+$ cells on average (Figure 1A), and 76,500 $CD45^+$ cells from the PBMC samples (data not shown). We visualized the global cellular heterogeneity by pooling all the acquired data on $5.2 \times 10^6$ cells and applying unsupervised hierarchical clustering with minimum spanning tree projection (SPADE), grouping the cells into a pre-defined number of nodes based on phenotypic similarity (Qiu et al., 2011). A dendrogram displayed the corresponding higher-order relatedness between those nodes (Figure 1B). The major branches in this dendrogram corresponded to $CD4^+$ T cells, $CD8^+$ T cells, TCRγδ cells, B cells, innate lymphocytes (referred to as $CD3^-CD7^+$ cells hereafter), and myeloid cells (Figure 1C). The cell frequencies of these major cell lineages obtained through SPADE were confirmed by traditional gating procedures using two-parameter dot plots (Figure S3).

In general, the subset distribution between PBMC samples from controls and patients was quite similar although a decrease in numbers of $CD4^+$ T cells in patients with RCDII and some variability in the numbers of B cells and $CD3^-CD7^+$ cells was detected (Figures 1D and 1E). In contrast, substantial differences were evident between the PBMC and intestinal samples. For example, duodenal $CD4^+$ T cell, $CD8^+$ T cell, $CD3^-CD7^+$ cell, and rectal myeloid cell subsets were distinct from those in peripheral blood (Figures 1D–1G). In addition, the SPADE dendrograms revealed disease-associated signatures (Figures 1D–1G) exemplified by the disappearance of $CD3^-CD7^+$ cells and an increase in TCRγδ cells in CeD relative to the control duodenal biopsies. Also, compared to controls and patients with CeD, an increase in $CD3^-CD7^+$ cells in the duodenum of patients with RCDII was observed. Of note is the dominant presence of a $CD8^+$ T cell cluster in the duodenum of a patient with enteropathy-associated T cell lymphoma type II (EATLII). Finally, a highly diverse $CD4^+$ T cell compartment was found in rectal biopsies of patients with CD and a dominant presence of myeloid cells in perianal fistulas. Thus, this global analysis indicated that there are immune-system-wide differences in subset composition between peripheral blood and intestinal samples, and between duodenal samples from patients and controls.

### t-SNE-ACCENCE Analysis Identifies 142 Phenotypically Distinct Immune Subsets

Although SPADE analysis provides an overview of the heterogeneity and the relatedness of subsets within the major immune lineages it does not allow analysis at the single-cell level and consequently rare cells are difficult to visualize. Therefore, we applied t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis (van der Maaten, 2014; van der Maaten and Hinton, 2008), which generates a two-dimensional map where cells with similar multidimensional phenotypes are placed close to each other, while maintaining single-cell resolution (Amir et al., 2013). To ensure a similar impact of the cells from PBMC and intestinal samples on the t-SNE analysis the number of cells incorporated from those two compartments were matched. We applied the t-SNE approach for every major lineage individually, here showing the $CD4^+$ T cell compartment where over 440,000 cells were incorporated in the analysis (Figure 2).

The t-SNE analysis revealed expected types of marker distributions on the $CD4^+$ T cells, such as broadly expressed markers (CD7, IL-7Rα), markers that were expressed by subpopulations of the cells (CD56, PD-1, CD25) and markers that were co-expressed (CD45RA and CCR7), but also unanticipated distributions were revealed, like largely mutually exclusive marker expression patterns (i.e., CD27 and CD161) (Figure 2A). Next, we incorporated a kernel density-peak detection algorithm on the t-SNE map (ACCENSE) (Shekhar et al., 2014), which automatically identified 28 $CD4^+$ T cells subsets (Figure 2B) where each subset was defined by its marker expression profile (Figure 2C).

(D) Representative SPADE trees showing an individual PBMC sample from a control and three patients with intestinal diseases. Color represents CD3 marker expression as described in (B).

(E) Comparisons of cellular frequencies for major immune lineages from 47 PBMC samples.

(F) Representative SPADE trees showing an individual intestinal biopsy from a control and five patients with intestinal diseases.

(G) Comparisons of cellular frequencies for major immune lineages from 55 intestinal biopsies. Data are plotted as single values (each data point represents an individual sample). *p < 0.05; **p < 0.01; ***p < 0.001, using Mann-Whitney U test. Error bars show means ± SEM.

(legend on next page)
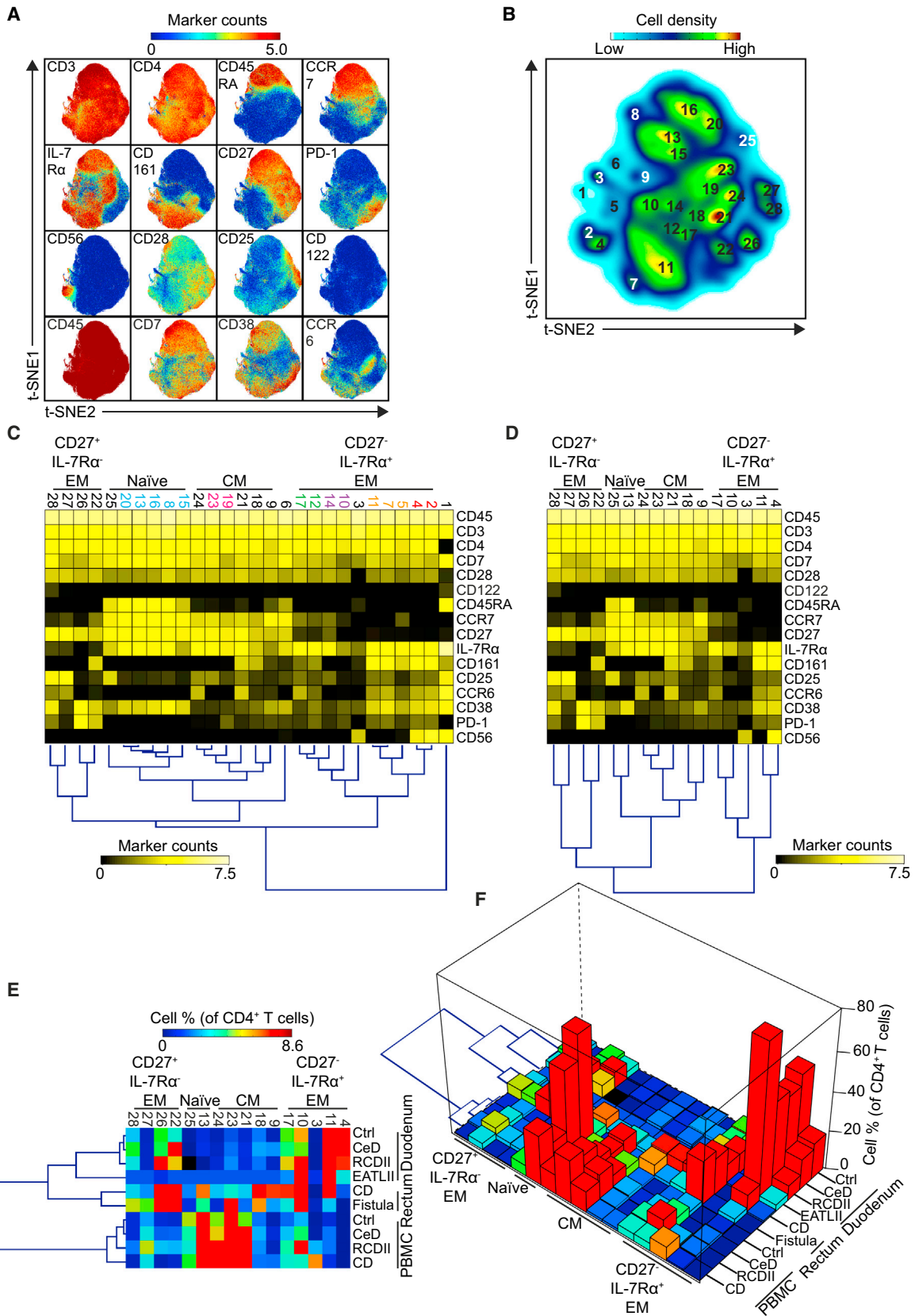
To reduce the complexity implied by the ACCENSE analysis, we merged computationally-derived subsets with highly similar expression profiles, which resulted in 16 cell clusters that each express a distinct set of markers (Figure 2D). They fell within four major CD4$^+$ T cell categories: naive (CD45RA$^+$CCR7$^+$), CD27$^+$IL-7R$\alpha^-$ effector memory (EM; CD45RA$^-$CCR7$^-$), CD27$^-$IL-7R$\alpha^+$ EM, and central memory (CM; CD45RA$^-$CCR7$^+$), and within those categories additional heterogeneity was present. For example, the highly similar CD161$^+$CD27$^-$IL-7R$\alpha^+$ EM cell subsets 4 and 11 were distinguished from each other by the expression of CD56 (Figure 2D). We next analyzed the subset distribution of the CD4$^+$ T cells in the various tissues included in the analysis by plotting the relative frequencies of the subsets for all samples analyzed (Figures 2E and 2F). In line with the SPADE analysis, the CD4$^+$ T cell subsets of the duodenum, rectum, and PBMC samples clustered to distinct locations in the cell frequency heatmap, also when examining the 102 samples individually (Figure S4). Thus, we could effectively delineate cell populations in a data-driven manner, and this approach revealed distinct signatures in the cellular composition of the CD4$^+$ T cells in biopsy material and peripheral blood.

By applying the t-SNE-ACCENSE analysis to all 6 major cell lineages individually, we identified 142 subsets in the entire immune system (Figure 3A), 23 of which contained rare cells that did not fulfill lineage phenotypic criteria (data not shown). The distinct phenotypes of the remaining 119 subsets are summarized in a heatmap (Figure 3B) where the subsets are clustered according to their phenotypic hierarchy within their lineage and clustered based on marker expression. The analysis identifies relatively few subsets within the B cell and myeloid compartments, likely due to the composition of the antibody panel, which was designed to capture the heterogeneity of the CD7$^+$ lymphoid cells. In this context, our antibody panel identified many distinct subsets within the TCR$\gamma\delta$ and CD3$^-$CD7$^+$ immune lineages, a diversity that was even greater than that detected in the CD4$^+$ and CD8$^+$ T cell lineages despite the fact that the latter were generally more abundant in the samples included. Collectively, the combined t-SNE-ACCENSE approach on high-dimensional cytometry data can effectively identify phenotypic distinct subsets in an unbiased and data-driven manner, and the analysis indicates that the heterogeneity of the immune system is far greater than previously appreciated.

## Visualization of Cellular "Fingerprint" Signatures across Tissues and Disease States

To visualize the distribution of immune cells based on tissue-origin and disease state, we used the t-SNE maps to deduce cellular "fingerprint-like" signatures of immune cells in the six major immune lineages (Figure 4A). The "fingerprint" gives a visual representation of the position of a collection of cells from particular (tissue) samples in the t-SNE map of the collective dataset. As such, it gives a unified overview of the distinctness of cells within the samples analyzed stratified for tissues and disease states. In all six immune lineages, the duodenum, rectum, and PBMC samples displayed a distinct cellular signature (Figure 4A). Moreover, on the basis of these signature maps, we were able to identify phenotypically distinct cell clusters that were either specifically present or abundant in certain diseases (highlighted by red boxes and arrows in Figure 4A). For example, the lineage (Lin)$^-$ CD3$^-$CD7$^+$ cells that expanded monoclonally in patients with RCDII (purple arrow; Figure 4A) were distinguished by the expression of CD45RA in six out of seven patients (Figures 4B and 4C). Moreover, these aberrant Lin$^-$ CD3$^-$CD7$^+$ cells were also detectable in PBMC samples of three out of six patients (Figures 4A and 4D), indicating a systemic spread of the pre-malignant cells. Similarly, a massive expansion of CD56$^+$CD161$^+$ CD8$^+$ T cells was observed in a patient with an established lymphoma of EATL-type 2 (green arrow; Figure 4A). Rare cells (0.1%–0.2% of CD8$^+$ T cells) displaying an identical phenotype were detected in 25% of the other duodenum samples, and these might therefore represent the precursor from which the lymphoma arose. In addition, two distinct IL-7R$\alpha^+$ innate lymphoid cell type 3-like (ILC3-like) (Spits and Cupedo, 2012) cell clusters were identified that were either chemokine receptor CCR6$^+$ or CCR6$^-$ (blue arrow; Figures 4A and 4B). While the CCR6$^+$ ILC3-like cells represented 50% of the CD3$^-$CD7$^+$ cells in the rectum of CD patients, its CCR6$^-$ ILC3-like counterpart, which is associated with an inflammatory phenotype in CD (Geremia et al., 2011), was more abundant in the fistulas (Figure 4C). Also, 94% of the myeloid cells within the inflammatory perianal fistulas (red arrow; Figure 4A) displayed a CD11b$^+$CD11c$^+$ dendritic cell-like phenotype (Figure 4B) and they comprised 50% of the accumulated immune infiltrate (Figure 4C).

We next used the Jensen-Shannon (JS) divergence to quantify similarities and dissimilarities between pairs of t-SNE maps (Figure 4E). In these plots, the JS divergences between samples from the intestine and peripheral blood were high for every major lineage (Figure 4E), indicative of dissimilarity of cellular signatures. In addition, the JS divergences reveal disease-associated cellular profiles exemplified by similarity of RCDII and CeD duodenal myeloid cells compared with those in duodenal controls, distinct CD8$^+$ T cells in RCDII patient blood and distinct B cells, CD3$^-$CD7$^+$ cells, and myeloid cells in CD patient blood compared with other blood samples.

Furthermore, we visualized the immune composition as an immune landscape where the distribution of the subsets in the various tissue and blood samples is shown (Figure 5). A

---

**Figure 2. t-SNE-ACCENSE Analysis Pipeline Identifies Tissue-Specific CD4$^+$ T Cell Subsets**

(A) Collective t-SNE dimensionality reduced CD4$^+$ T cell single-cell data (4.7 × 10$^5$ cells) derived from 102 samples analyzed are plotted. Every dot represents a single cell and the color of the cells indicates ArcSinh5-transformed expression values for a given marker analyzed.

(B) A density map depicting the local probability density of cells as embedded in (A), computed using a kernel based transformation. Numbers in this map represents centers of phenotypic subsets and were identified using a peak detection algorithm.

(C) A heatmap summary of median ArcSinh5-transformed expression values of T cell markers expressed by 28 CD4$^+$ T cell subsets identified and hierarchical clustering of subsets with description of four categories. Colored subsets were merged together resulting in 16 CD4$^+$ T cell subsets (D).

(E) A heatmap summary of average subset frequencies across tissues and disease states. Mean frequencies obtained from 102 samples analyzed.

(F) A 3D heatmap as described in (D). EM = effector memory and CM = central memory. A detailed heatmap showing 102 samples is described in Figure S4.
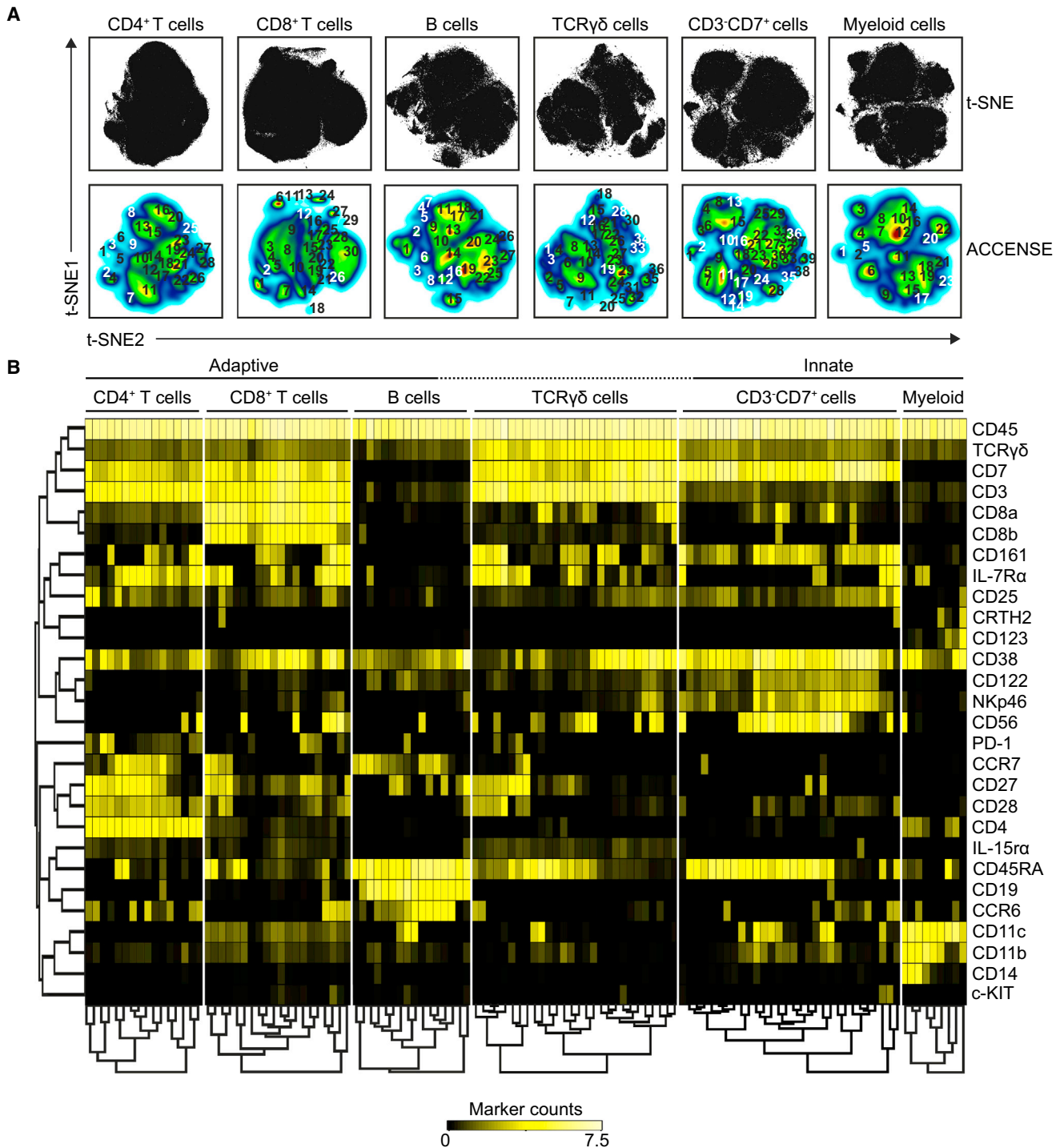
**Figure 3. t-SNE-ACCENSE Analysis Delineates Phenotypic Distinct Immune Subsets in Peripheral Blood and Intestine**

(A) Identification of 142 cell subsets within the six major immune lineages through t-SNE-ACCENSE analysis. Cells from PBMCs were randomly sampled to match cell numbers with those from intestinal biopsies for each immune lineage individually. t-SNE plots are showing $4.7 \times 10^5$ CD4[+] T cells, $9.3 \times 10^5$ CD8[+] T cells, $1.8 \times 10^5$ B cells, $1.8 \times 10^5$ TCR$\gamma\delta$ cells, $1.9 \times 10^5$ CD3[-]CD7[+] cells, and $2.2 \times 10^5$ myeloid cells of the combined 102 sample dataset.

(B) Heatmap showing characterization of 119 cell clusters (16 CD4[+] T cell subsets, 20 CD8[+] T cell subsets, 16 B cell subsets, 28 TCR$\gamma\delta$ cell subsets, 30 CD3[−]CD7[+] cell subsets and 9 myeloid cell subsets). Shown are median ArcSinh5-transformed values of marker expression (black-to-yellow scale) and hierarchical clustering of markers and subsets within their major immune lineage

two-dimensional representation of the immune landscape along with the phenotypes of the associated immune subsets is shown in Figure S5, visualizing the distinct cellular phenotypes of the immune subsets and their occurrence in the various tissues. Together these findings illustrate that in most of the major immune lineages, cellular subsets could be identified that were exclusively present or enriched in defined tissue samples only. In peripheral blood, cells were mainly defined by expression of interleukin-7 receptor α (IL-7Rα), CCR7, CD27, and CD28 for CD4$^+$ and CD8$^+$ T cells, CCR6 for B cells, CD56 for CD3$^-$CD7$^+$ cells, and CD14 for myeloid cells. Similarly, many mucosal cells were defined by expression of CD161 for CD4$^+$ and CD8$^+$ T cells, CD38 for TCRγδ cells, and IL-7Rα for CD3$^-$CD7$^+$ cells.

Together, these analyses demonstrate that by deducing cellular "fingerprint" signatures of immune cells, we were able to visualize and quantify the immune subset distribution in the samples analyzed. In addition, we were able to identify immune subsets that are associated with disease states.

### An Integrated System-Wide View of the Immune System Reveals Disease-Associated Networks of Immune Subsets

Finally, we investigated whether the identified immune-system-wide cellular patterns could be integrated collectively and used to characterize samples according to tissue location or disease state by visualizing them in relation to several clinical variables. For this purpose, we visualized the immune composition of all identified subsets from all included biological samples in a single graph by applying the t-SNE algorithm on cell frequency values. As expected, the PBMC and intestinal biopsy samples formed two distinct clusters (Figure 6A). In addition, the rectum and peri-anal fistula biopsies separated from the duodenal biopsies (Figures 6A and 6B) and the duodenal biopsies from patients with CeD and RCDII clustered away from the duodenal control biopsies (Figures 6A and 6B). Also, the duodenal biopsies of the four RCDII patients with the most severe inflammation (Table S2) mapped far from the other duodenum samples (Figures 6B and 6C). Furthermore, three RCDII biopsies, two of which were from patients in remission, clustered with the other CeD biopsies, suggesting a persisting CeD immune profile. The inflammation state of the biopsies was reflected in the cluster structure of intestinal samples in general (Figure 6C), while this was not the case for gender (Figure 6D). Moreover, the age of the patients and controls from which the samples were derived is reflected in the clustering of the samples, particularly in peripheral blood (Figure 6E). The PBMC and intestinal samples from six patients that were biopsied twice, with a 3- to 6-month time interval, clustered tightly together (Figure 6F) highlighting the reproducibility and robustness of this unbiased approach.

In order to reveal which cellular subsets were associated with the disease-associated patterns (Figure 6G), we performed a second t-SNE analysis on the subsets of the same dataset (instead of the samples) visualizing networks of cellular subsets that determine disease-specificity (Figures 6H and 6I) and identified the top five ranked subsets contributing to these clustering patterns (Figure 6J). In this context, expected types of health- and disease-associated subsets were identified in the intestinal mucosa, such as CD45RA$^-$Lin$^-$ CD3$^-$CD7$^+$ cells in control individuals, CD8a$^+$ and CD8a$^-$ TCRγδ cells in CeD, CD45RA$^+$Lin$^-$

CD3$^-$CD7$^+$ cells in RCDII, CCR6$^+$ILC3 in CD and CD14$^-$CD11b$^+$ myeloid cells in fistulas. In addition, previously unidentified subset associations were revealed as well, such as CD45RA$^-$ cNK cells and CD56$^-$CD27$^-$EM CD4$^+$ T cells in controls, CD27$^-$EM CD8$^+$ T cells in CeD, CRTH2$^+$ myeloid cells in RCDII (ranked 6), CCR6$^-$CM and PD-1$^+$CD27$^+$EM CD4$^+$ T cells in CD, and CCR6$^-$CCR7$^+$ B cells in the fistulas. Thus, by integrating data-driven approaches highly specific disease-associated immune signatures across innate and adaptive major lineages in the intestine were readily identified. Figure S6 gives an overview of the developed integrated analysis pipeline developed in the current study.

### DISCUSSION

Mass cytometry offers the opportunity to simultaneously analyze dozens of single-cell markers on complex cellular samples resulting in highly complex datasets. Conventional approaches for flow cytometry data analysis are not suitable for such datasets, suffer from individual user bias, and require prior knowledge of the cell type of interest. SPADE (Bendall et al., 2011) was originally applied to handle mass cytometric data. More recently, an unbiased analysis pipeline has been developed combining t-SNE (Amir et al., 2013; van der Maaten and Hinton, 2008) and ACCENSE (Shekhar et al., 2014) with mass cytometry to visualize and delineate phenotypically distinct subsets (Becher et al., 2014). In the current study we used a 32 antibody panel that was specifically designed to detect heterogeneity within the major adaptive and innate cell lineages. We applied this antibody panel to a variety of PBMC and intestinal biopsy samples and combined this with the newly available unbiased computational approaches to unravel the complexity of the human mucosal immune system. We used the Barnes-Hut implementation of t-SNE, a recently developed, computationally efficient t-SNE optimization algorithm (van der Maaten, 2014) to accommodate the large datasets. In addition, we provide novel applications of the t-SNE-based analysis allowing the visualization of cellular "fingerprint" signatures of immune cells, and by clustering samples based on their immune composition while visualizing the cluster-contributing subsets in parallel to highlight tissue- and disease-associated patterns.

Our results demonstrate that the mass cytometry-based analysis was robust and reproducible as identical control PBMC samples that were included during the entire 9-month study period provided highly similar results. Moreover, we obtained biopsy and PBMC specimens from a number of patients twice with a time interval between 3 to 6 months and in the final sample visualization analysis, these specimens clustered close together, demonstrating a high degree of reproducibility. Also, we readily observed changes in the composition of the immune compartment that are known to correlate with disease such as the increase in TCRγδ cells in CeD (Spencer et al., 1989).

Both the SPADE and the combined t-SNE-ACCENSE analysis demonstrate that duodenal, rectal, and PBMC samples grouped into different clusters, to a large extent due to substantial differences in the CD4$^+$ and CD8$^+$ T cell compartments. In total, 142 subsets in the immune system were defined, of which 119 displayed distinct marker expression profiles on the basis of which
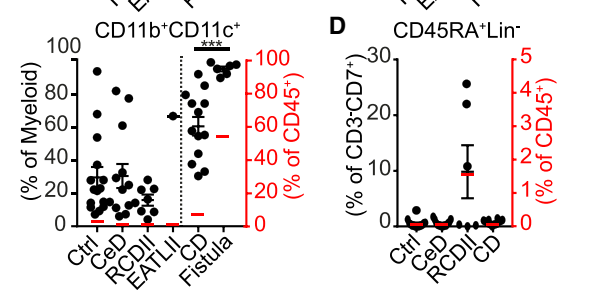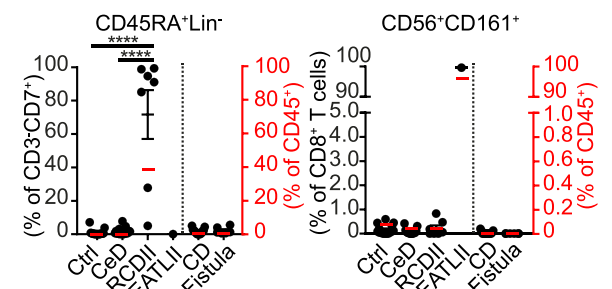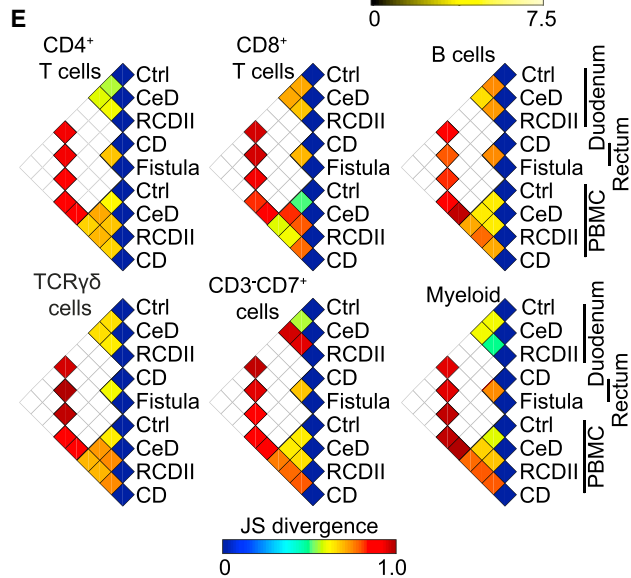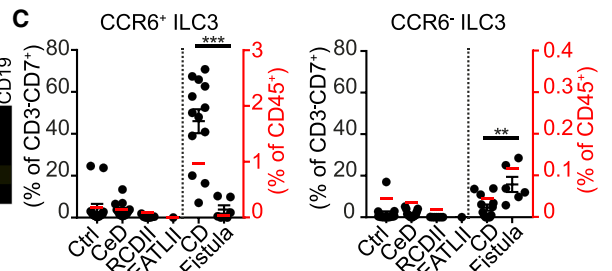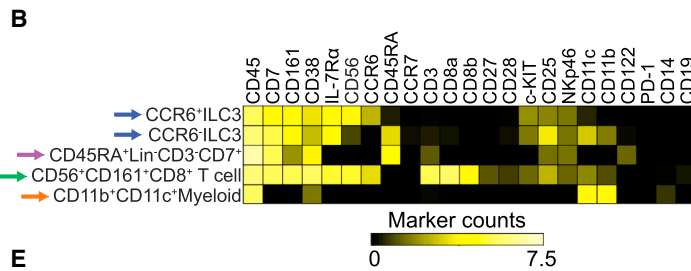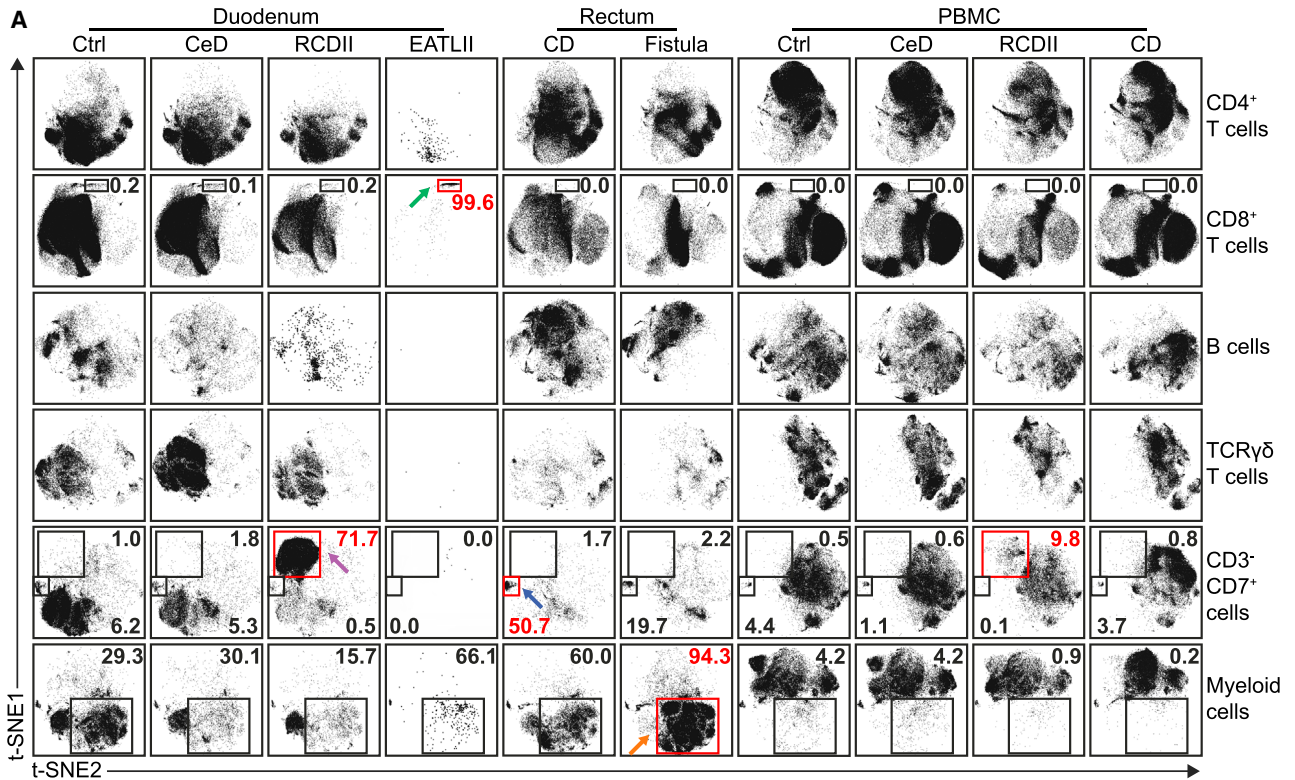
(legend on next page)

these subsets were delineated. Compared to the CD4$^+$ and CD8$^+$ T cell lineages, a larger degree of heterogeneity was detected in the TCRγδ and CD3$^-$CD7$^+$ immune lineages. While we cannot exclude the possibility that by the use of another antibody panel additional heterogeneity within the CD4$^+$ and CD8$^+$ lineages might be revealed, the distinct marker expression profiles of the TCRγδ and CD3$^-$CD7$^+$ subsets might correlate with distinct and potentially location-specific functional properties. This will be the subject of future investigations. Thus, on the basis of 28 markers, mass cytometry visualized system-wide cellular differences in subset composition in the samples obtained from the various anatomical locations. Further studies including control rectum samples are required to determine how these differences relate to the anatomical site and/or disease state.

Recent studies have described a crucial role for ILCs in CD (Hazenberg and Spits, 2014). They reside mainly in mucosal tissues and are functionally specialized cells characterized by the expression of lineage-defining transcription factors. Even without these markers in our antibody panel we were able to distinguish ILC-like subsets on the basis of cell surface phenotype through machine-learning cell cluster detection approaches. We identified a CRTH2$^+$ ILC2-like subset in PBMC samples and in line with previous reports such cells were not found in intestinal samples (Hazenberg and Spits, 2014). Moreover, we detected an ILC3-like subset that was exclusively present in rectum biopsies of CD patients in remission. This subset corresponds with the IL-22-producing anti-inflammatory CD25$^-$CD56$^+$ ILC subset previously found in the intestine of CD patients (Geremia et al., 2011; Hazenberg and Spits, 2014). Moreover, we observed that their CD56$^-$ ILC counterpart was enriched in fistulas, and these cells were previously shown to produce the inflammatory cytokine IL-17A (Geremia et al., 2011). The CD56$^-$-ILC3-like cells in fistulas unexpectedly showed expression of CD11c, a marker that has been used as "dump channel" in the ILC field. Those cells might thus have been discarded from datasets of previous studies, highlighting the importance of unbiased data-driven approaches as used in the current study. In addition, the CD4$^+$ T cell compartment in the CD biopsies was highly heterogeneous. Further studies are required to determine a possible relationship with the highly variable disease symptoms in CD.

Mucosal lymphoid malignancies were readily detected in patients with RCDII and in a patient with an EATL-type 2 lymphoma, along with the distinct cellular phenotypes that distinguish these lymphomas. The latter information could be used to identify the potential precursor cells in the healthy mucosa from which these

malignancies are likely to originate. In the case of RCDII these precursors are Lin$^-$CD3$^-$CD7$^+$ cells, confirming previous results (Schmitz et al., 2013). Moreover, our current analysis indicates that these precursors can be distinguished from their malignant counterpart by the lack of expression of CD45RA. Conversion of this subset to a CD45RA positive phenotype might thus predict development of RCDII and could constitute a novel prognostic marker. Also, extremely low numbers of CD56$^+$ CD161$^+$CD8a$^+$CD8b$^+$ T cells were found in mucosal biopsies of healthy individuals, a phenotype that matches that of the CD8$^+$ T cell malignancy in a patient with EATLII. Thus, the t-SNE analysis is highly suitable for the identification of mucosal malignancies and their likely precursors in healthy individuals, information that might be used to develop therapeutic approaches based on cellular characteristics.

Compared to control duodenal samples, we observed the disappearance of CD3$^-$CD7$^+$ cells and the increase in TCRγδ cells in CeD, both well-described disease hallmarks (Schmitz et al., 2013; Spencer et al., 1989). In the global analysis of the entire cell frequency dataset this resulted in the formation of clusters that distinguish duodenal biopsies derived from CeD patients from those of controls. By transposing the cell frequency dataset, the disease cluster-associated subsets and their relative contribution to the clustering could be visualized and quantified. This tSNE application thus provided detailed information on the disease-associated networks of immune subsets. The identification of mucosal immune signatures that correlated with health and disease might potentially lead to the development of unbiased diagnostic procedures based on a single mass cytometric analysis.

Perianal fistulas in CD remain a substantial clinical challenge, causing pain, discharge, and abscess formation (Kamm and Ng, 2008). Achieving complete fistula healing is difficult and accompanied by multiple relapses, and despite the best available therapies durable remission rates of perianal fistulas remain disappointingly low (Molendijk et al., 2014). In this respect it will be important to unravel the function of the CD11b$^+$CD11c$^+$ myeloid cells that dominate in the fistula, where the immune composition is distinct from that in the adjacent rectum biopsies. This will be addressed in future studies.

In such follow-up studies, our approach can be further refined as four antibodies included in the original 32 antibody panel (CD103, IL21-R, CD34, and TCRαβ) were not informative. In particular, the inclusion of antibodies specific for lineage-defining transcription factors and cytokines might provide further insight into the relationship between cell lineages and their function in the mucosal immune system. Moreover, the inclusion of

---

**Figure 4. Stratified t-SNE Plots Identify Distinct Cellular "Fingerprint" Signatures across Tissues and Disease**

(A) Collective t-SNE dimensionality reduced single-cell data from all 102 samples analyzed are plotted showing six major immune lineages stratified for tissues and disease states. Red boxes and arrows indicate t-SNE location of phenotypically distinct disease-associated clusters of cells.

(B) Heatmap summary of median ArcSinh5-transformed expression values of cellular markers expressed by gated subsets and annotation for each subset. Arrows as in (A).

(C) Comparisons of cellular frequencies for subsets from intestinal biopsies.

(D) Comparisons of cellular frequencies for the CD45RA$^+$Lin$^-$ CD3$^-$CD7$^+$ subset from PBMC samples.

(E) Pairwise Jensen-Shannon (JS) divergence plots of the collective t-SNE maps from all 102 samples analyzed showing six major immune lineages. A higher JS divergence value indicates more dissimilarity between a pair of t-SNE maps as shown in (A). White squares indicate invalid comparisons. Data are plotted as single values. Red lines indicate mean value as percentage of CD45$^+$ cells. (each data point represents an individual sample). *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001, using Mann-Whitney U test. Error bars show means ± SEM.
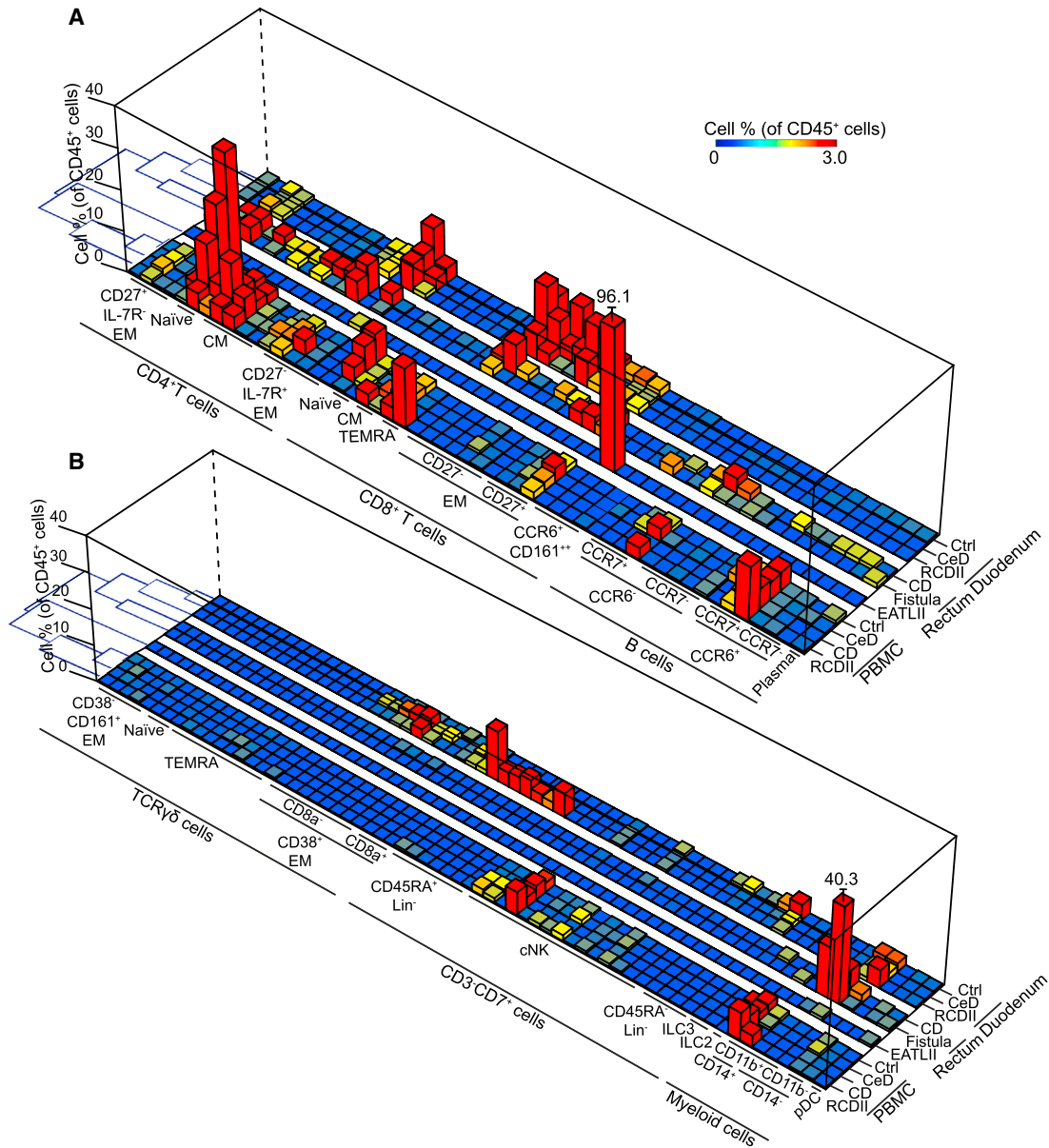
**Figure 5. Immune System Landscape Visualizes Subset Composition**
(A and B) A 3-dimensional heatmap summary showing average frequencies of 119 immune subsets in the 102 biological samples combined with hierarchical clustering of samples and description of tissue type, disease state, and biological assignment of the subsets. Color scale and z axis indicate percentage of CD45$^+$ cells. EM = effector memory, CM = central memory, TEMRA = terminally differentiated, Lin = lineage, cNK = conventional NK cells, ILC = innate lymphoid cells, and pDC = plasmacytoid dendritic cells.

an alternative metal reporter for the CD103-specific antibody in future studies might allow discrimination between cells derived from the epithelium and the lamina propria. Also, with regard to CD and perianal fistula, we could not draw definitive conclusions regarding disease-specific changes because we lacked healthy control rectum samples. It would thus be highly valuable to characterize mucosal biopsies obtained from various intestinal locations within the same patients and controls, allowing direct comparisons. Finally, by combining the analysis of the mucosal immune system with an analysis of the stromal cell compartment, a more integrated view of disease-specific changes might

be obtained, optimizing opportunities to develop more effective personalized treatment modalities.

In conclusion, the mass cytometric analysis of the mucosal immune system revealed heterogeneity that was greater than previously appreciated. Also, our results indicate that disease-specific leukocytes reside mainly in the affected organ and are much less readily detectable in PBMC. The identification of disease-associated changes in immune composition offers opportunities to determine cellular parameters that correlate with disease and predict response to treatment, an important step toward personalized and cost-effective treatment.
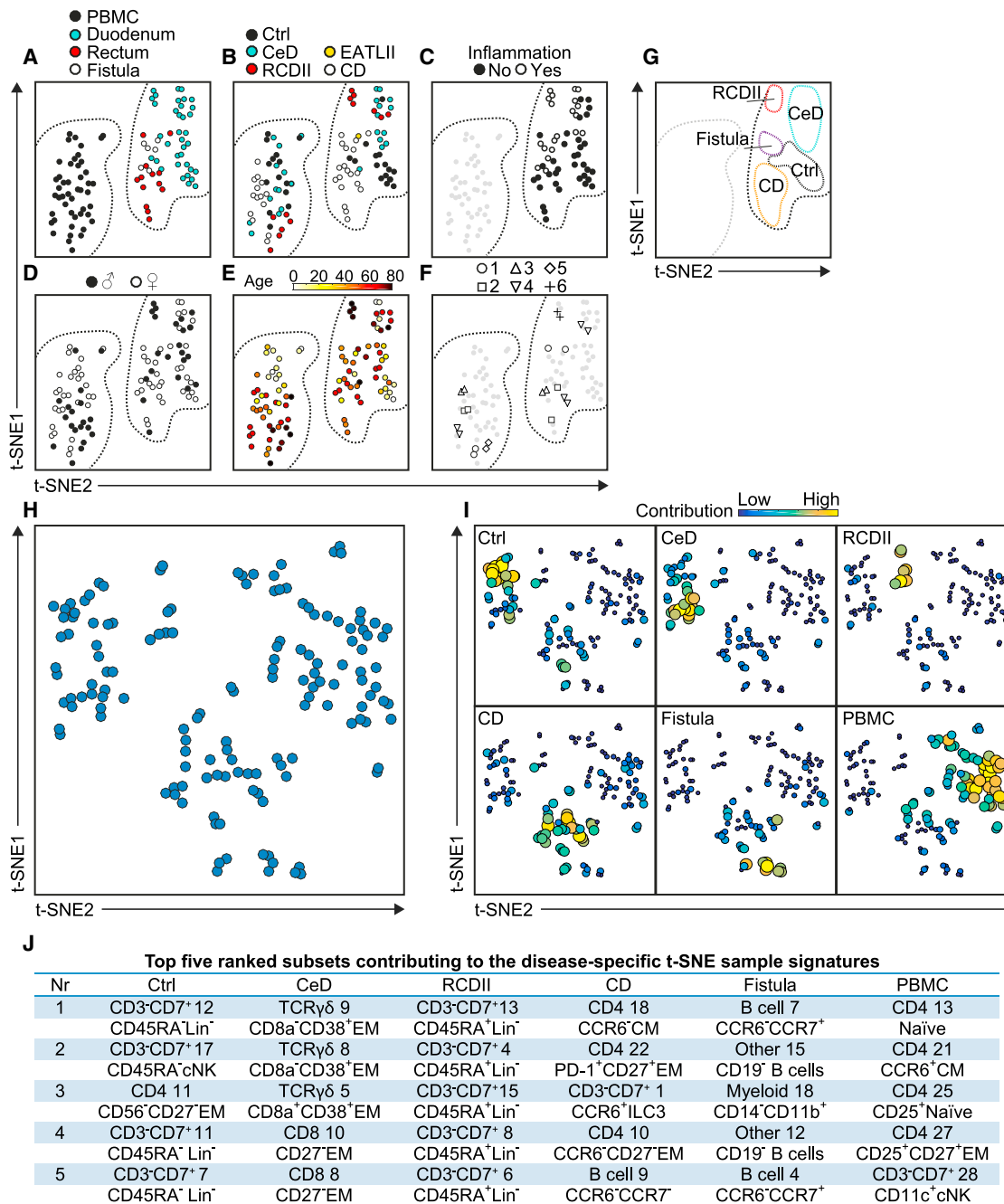
**Figure 6. Integrated Analysis of Immune Subset Composition Reveals Tissue- and Disease-Associated Clustering of Biological Samples**

(A–F) Collective t-SNE dimensionality reduced cell percentage data (as percentage of CD45[+] cells) of 142 subsets for 102 samples analyzed are plotted. Every dot represents a single sample and the color of the samples shows the corresponding clinical information: (A) tissue, (B) disease, (C) biopsy inflammation, (D) gender, (E) age in years, and (F) samples from six patients that were sampled twice. The left dashed border represents the PBMC cluster and the right dashed border represents the intestinal cluster.

(G) Deducing disease-specific signatures in the t-SNE map based on the clustering patterns of the samples.

(H) Collective t-SNE dimensionality reduced cell percentage data (as percentage of CD45[+] cells) of 142 subsets for 102 samples analyzed are plotted. Every dot represents a single immune subset. The closer the subsets are together the more similar the cell frequency values are across the samples.

(I) Disease-specific subset signatures (average subset values per disease cluster in G, encoded with varying dot color and size).

(J) Table showing top five ranked subsets contributing to the disease-specific t-SNE sample signatures (as shown in I) displaying both major lineage subset number and biological assignation. Lineage (Lin), conventional NK cell (cNK), effector memory (EM), central memory (CM), and innate lymphoid cells (ILC). Average cell frequency values are shown in Figure 5.

## EXPERIMENTAL PROCEDURES

### Human Samples

Samples were collected from patients who were undergoing routine diagnostic endoscopies, and the curettage material of perianal fistulas were obtained at surgical intervention. The clinical characteristics of the patients are shown in Table S2. All samples were obtained after informed consent, medical ethical commission approval, in accordance with the local ethical guidelines of the VU Medical Center in Amsterdam (adult duodenal biopsies) and the Leiden University Medical Center (pediatric duodenal biopsies, CD rectum biopsies, and perianal fistulas), and in accordance with the Declaration of Helsinki.

### Isolation of Cells from Intestinal and PBMC Samples

Cells from the epithelium were isolated from two or three intestinal biopsies by treatment with 10 ml of HBSS (Sigma-Aldrich) containing 1 mM EDTA (Merck) under rotation for 2 hr at 37°C. To obtain cells from the lamina propria, we washed the biopsies with PBS containing 0.5% fetal calf serum (FCS) and incubated with 5 ml of a collagenase mix containing IMDM culture medium (Lonza) with 20% FCS, 1,000 U/mL collagenase IV (Worthington), and 10 mg/mL DNaseI grade II (Roche Diagnostics) for 2 hr at 37°C. The cell suspension was then filtered through a 70 μm nylon cell strainer and centrifuged in 0.5% FCS/PBS. Curettage material of perianal fistulas obtained at surgical intervention were minced with fine scissors and incubated with 10 ml of HBSS containing 1 mM EDTA under rotation for 2 hr at 37°C. Peripheral blood mononuclear cells (PBMC) were isolated from up to 5 ml of freshly drawn heparin anticoagulated blood using Ficoll-Paque™ density-gradient centrifugation. PBMC samples from CD patients were cryopreserved and stained after thawing. Cell suspensions were washed with 0.5% FCS/PBS and kept at 4°C until antibody staining.

### Antibodies

Antibodies, manufacturers, and concentrations are listed in Table S1. Primary antibody metal-conjugates were either purchased or conjugated using a total of 100 μg of carrier-free formulations of purified antibody combined with the MaxPar X8 antibody labeling kit (Fluidigm Sciences) according to the manufacturer's instruction. Following conjugation, antibodies were diluted to 200 μl in Candor PBS Antibody Stabilization Buffer (Candor Bioscience GmbH) and stored at 4°C.

### Antibody Staining and Data Acquisition

Procedures for mass cytometry antibody staining and data acquisition were carried out as previously described (Bendall et al., 2011). Briefly, directly after biopsy processing, we resuspended cells in cell staining buffer (CSM; 1× PBS with 0.5% bovine serum albumin and 0.02% sodium azide, Fluidigm Sciences) and incubated with 1 ml of 1:500 diluted 500 μM rhodium DNA intercalator (Fluidigm Sciences) for 15 min to stain dead cells at room temperature (rT). Cells were washed with CSM and surface stained for 45 min at rT with a mixture of metal isotope-conjugated antibodies using predetermined concentrations (Table S1). Antibody staining reactions were performed in 100 μl final volume. After staining, cells were washed twice with CSM and then resuspended in 1 ml of 1:1,000 diluted 125 μM iridium DNA intercalator (DVS Sciences) in Fix and Perm Buffer (PBS with 1.6% paraformaldehyde, Fluidigm Sciences) for 45 min at rT to discriminate single cells. Cells were stored overnight at 4°C. Finally, cells were washed twice in CSM and once in distilled water at rT. Prior to data acquisition, cell pellets were diluted in distilled water containing 1:10 diluted EQ Four Element Calibration Beads (Fluidigm Sciences) to the concentration of $0.4 \times 10^6$ cells/mL to achieve an acquisition rate of 500 events/s on the CyTOF 2™ mass cytometer (Fluidigm Sciences) (Ornatsky et al., 2008). CyTOF data were acquired and analyzed on-the-fly, using dual-count mode and noise-reduction on. All other settings were either default settings or optimized with tuning solution, as instructed by Fluidigm Sciences. After data acquisition, the mass bead signal was used to normalize the short-term signal fluctuations with the reference EQ passport P13H2302 during the course of each experiment and the bead events were removed (Finck et al., 2013).

### Data Analysis

SPADE analyses were performed as described (Qiu et al., 2011) with 500 target number of nodes and 10% of target down-sampled events using the implementation in Cytobank (Kotecha et al., 2010). Data from exported FCS files of major immune lineages as delineated by SPADE (Figure 1) were transformed using hyperbolic arcsin with a cofactor of 5. Because the number of cell events varied greatly between PBMC and intestinal biopsies, the contribution of intestinal and PBMC cells were normalized to a 1:1 ratio for each immune lineage and up to 10,000 events per sample was used. After down-sampling, the cumulative dataset per immune lineage was subjected to t-SNE dimensionality reduction. t-SNE was carried out using the Barnes-Hut implementation of t-SNE, a recent developed, computationally efficient t-SNE optimization algorithm (obtained from L.J.P. van der Maaten) (van der Maaten, 2014) to accommodate the large volumes of our clinical data. t-SNE was run with a default perplexity of 30. Cellular "fingerprint" signatures of immune cells and marker expression color overlays of t-SNE maps were generated with Cyt (Amir et al., 2013). We used the Jensen-Shannon (JS) divergence to quantify the similarity between t-SNE maps. After converting t-SNE maps into two-dimensional probability density functions, the similarity between two maps is quantified as the JS divergence between their corresponding probability density functions. We used the base 2 logarithm in the JS divergence computation, which results in a continuous range of JS divergence values between 0 (for identical distributions) and 1 (for fully disjoint distributions). The density-peak detection algorithm to identify phenotypically distinct subsets was carried out with ACCENSE (Shekhar et al., 2014), using the two coordinates of the t-SNE map for each cell as input. The density-based clustering first searches for the optimal bandwidth, followed by estimating the kernel density that allows the detection of density peaks. The respective amount of subpopulations identified per major lineage was based on the calculated optimal kernel bandwidth. Two-dimensional gating analysis was done using Cytobank (Kotecha et al., 2010). Median intensity values of markers were calculated and visualized via plotting heatmaps. Hierarchical clustering dendrograms of heatmaps were produced using Pearson Correlation and average linkage clustering with MultiExperiment Viewer (http://www.tm4.org). Numbers of cells in different immune subsets were counted for each sample and percentages of each subset were calculated. t-SNE coordinates, ACCENSE subset number, and sample coding tags were added to FCS files as additional parameters to allow aggregate data analysis and visualization. The sample t-SNE map (Figure 6) was computed with the fractions of the total cell count per subtype (as percentages of CD45+ cells) as input variables. Standard t-SNE pre-processing was applied: the data matrix was normalized by centering each variable to zero mean, and scaling to unit vector length. In the sample map, a reprojection of the data on a reduced set of high-variance principal components (PCs) was performed. The component scores of the ten highest variance PCs were used as input to the t-SNE. To reduce sensitivity to local optima, we repeated map construction 100 times with different randomly generated initial maps and the map with the minimal t-SNE error metric (Kullback Leibler divergence per data point) was selected. The subset t-SNE map in Figures 6H and 6I was computed by transposing the normalized datamatrix. This switches the role of samples and variables, hence subsets with similar profiles across the population end up close together in the map. The t-SNE perplexity parameter was set to 10% of the number of data points in each map, i.e., 10 for the sample maps in Figures 6A–6F and 15 for the subset map in Figures 6H and 6I. Average subset values were computed per disease subgroup as identified in the sample t-SNE map in Figure 6G and displayed per subgroup.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.immuni.2016.04.014.

## REFERENCES

Al-Toma, A., Verbeek, W.H., Hadithi, M., von Blomberg, B.M., and Mulder, C.J. (2007). Survival in refractory coeliac disease and enteropathy-associated T-cell lymphoma: retrospective evaluation of single-centre experience. Gut 56, 1373–1378.

Amir, el-A.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol 31, 545–552.

Bandura, D.R., Baranov, V.I., Ornatsky, O.I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J.E., and Tanner, S.D. (2009). Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal. Chem. 81, 6813–6822.

Becher, B., Schlitzer, A., Chen, J., Mair, F., Sumatoh, H.R., Teng, K.W., Low, D., Ruedl, C., Riccardi-Castagnoli, P., Poidinger, M., et al. (2014). High-dimensional analysis of the murine myeloid cell system. Nat. Immunol. 15, 1181–1189.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, A.D., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science 332, 687–696.

Finck, R., Simonds, E.F., Jager, A., Krishnaswamy, S., Sachs, K., Fantl, W., Pe'er, D., Nolan, G.P., and Bendall, S.C. (2013). Normalization of mass cytometry data with bead standards. Cytometry A 83, 483–494.

Geremia, A., Arancibia-Cárcamo, C.V., Fleming, M.P., Rust, N., Singh, B., Mortensen, N.J., Travis, S.P., and Powrie, F. (2011). IL-23-responsive innate lymphoid cells are increased in inflammatory bowel disease. J. Exp. Med. 208, 1127–1133.

Geremia, A., Biancheri, P., Allan, P., Corazza, G.R., and Di Sabatino, A. (2014). Innate and adaptive immunity in inflammatory bowel disease. Autoimmun. Rev. 13, 3–10.

Hazenberg, M.D., and Spits, H. (2014). Human innate lymphoid cells. Blood 124, 700–709.

Jabri, B., and Sollid, L.M. (2009). Tissue-mediated control of immunopathology in coeliac disease. Nat. Rev. Immunol. 9, 858–870.

Kamm, M.A., and Ng, S.C. (2008). Perianal fistulizing Crohn's disease: a call to action. Clin. Gastroenterol. Hepatol. 6, 7–10.

Kappelman, M.D., Moore, K.R., Allen, J.K., and Cook, S.F. (2013). Recent trends in the prevalence of Crohn's disease and ulcerative colitis in a commercially insured US population. Dig. Dis. Sci. 58, 519–525.

Kotecha, N., Krutzik, P.O., and Irish, J.M. (2010). Web-based analysis and publication of flow cytometry experiments. Curr Protoc Cytom Chapter 10, Unit10 17.

Molendijk, I., Nuij, V.J., van der Meulen-de Jong, A.E., and van der Woude, C.J. (2014). Disappointing durable remission rates in complex Crohn's disease fistula. Inflamm. Bowel Dis. 20, 2022–2028.

Ornatsky, O.I., Kinach, R., Bandura, D.R., Lou, X., Tanner, S.D., Baranov, V.I., Nitz, M., and Winnik, M.A. (2008). Development of analytical methods for multiplex bio-assay with inductively coupled plasma mass spectrometry. J. Anal. At. Spectrom. 23, 463–469.

Pascual, V., Dieli-Crimi, R., López-Palacios, N., Bodas, A., Medrano, L.M., and Núñez, C. (2014). Inflammatory bowel disease and celiac disease: overlaps and differences. World J. Gastroenterol. 20, 4846–4856.

Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Jr., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat. Biotechnol. 29, 886–891.

Randall, C.W., Vizuete, J.A., Martinez, N., Alvarez, J.J., Garapati, K.V., Malakouti, M., and Taboada, C.M. (2015). From historical perspectives to modern therapy: a review of current and future biological treatments for Crohn's disease. Therap. Adv. Gastroenterol. 8, 143–159.

Rubio-Tapia, A., Ludvigsson, J.F., Brantner, T.L., Murray, J.A., and Everhart, J.E. (2012). The prevalence of celiac disease in the United States. Am J Gastroenterol 107, 1538–1544, quiz 1537, 1545.

Schmitz, F., Tjon, J.M., Lai, Y., Thompson, A., Kooy-Winkelaar, Y., Lemmers, R.J., Verspaget, H.W., Mearin, M.L., Staal, F.J., Schreurs, M.W., et al. (2013). Identification of a potential physiological precursor of aberrant cells in refractory coeliac disease type II. Gut 62, 509–519.

Schwartz, D.A., Loftus, E.V., Jr., Tremaine, W.J., Panaccione, R., Harmsen, W.S., Zinsmeister, A.R., and Sandborn, W.J. (2002). The natural history of fistulizing Crohn's disease in Olmsted County, Minnesota. Gastroenterology 122, 875–880.

Shekhar, K., Brodin, P., Davis, M.M., and Chakraborty, A.K. (2014). Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). Proc. Natl. Acad. Sci. USA 111, 202–207.

Spencer, J., Isaacson, P.G., Diss, T.C., and MacDonald, T.T. (1989). Expression of disulfide-linked and non-disulfide-linked forms of the T cell receptor gamma/delta heterodimer in human intestinal intraepithelial lymphocytes. Eur. J. Immunol. 19, 1335–1338.

Spits, H., and Cupedo, T. (2012). Innate lymphoid cells: emerging insights in development, lineage relationships, and function. Annu. Rev. Immunol. 30, 647–675.

van der Maaten, L. (2014). Accelerating t-SNE using Tree-Based Algorithms. J. Mach. Learn. Res. 15, 3221–3245.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Verbeek, W.H., Goerres, M.S., von Blomberg, B.M., Oudejans, J.J., Scholten, P.E., Hadithi, M., Al-Toma, A., Schreurs, M.W., and Mulder, C.J. (2008). Flow cytometric determination of aberrant intra-epithelial lymphocytes predicts T-cell lymphoma development more accurately than T-cell clonality analysis in Refractory Celiac Disease. Clin. Immunol. 126, 48–56.