

Published in final edited form as:

Nat Immunol. 2020 March ; 21(3): 343–353. doi:10.1038/s41590-020-0602-z.

Distinct microbial and immune niches of the human colon

Kylie R James^{1,*}, Tomas Gomes¹, Rasa Elmentaite¹, Nitin Kumar¹, Emily L Gulliver^{2,3}, Hamish W King⁴, Mark D Stares¹, Bethany R Bareham⁵, John R Ferdinand⁶, Velislava N Petrova¹, Krzysztof Polański¹, Samuel C Forster^{1,2,3}, Lorna B Jarvis⁷, Ondrej Suchanek⁶, Sarah Howlett⁷, Louisa K James⁴, Joanne L Jones^{7,8}, Kerstin B Meyer¹, Menna R Clatworthy^{1,6}, Kourosh Saeb-Parsy⁵, Trevor D Lawley¹, Sarah A Teichmann^{1,9,10,*}

¹Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom, CB10 1SA

²Centre for Innate Immunity and Infectious Diseases, Hudson Institute of Medical Research, Clayton, Victoria, Australia

³Department of Molecular and Translational Sciences, Monash University, Clayton, Victoria, Australia

⁴Centre for Immunobiology, Blizard Institute, Queen Mary University of London, United Kingdom, E1 2AT

⁵Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, United Kingdom, CB2 0QQ

⁶Molecular Immunity Unit, Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, United Kingdom, CB2 0QQ

⁷Department of Haematology, Clifford Allbutt Building, United Kingdom, CB2 0AH

⁸Cambridge University Hospitals NHS Foundation Trust, United Kingdom, CB2 0QQ

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Corresponding authors: kj7@sanger.ac.uk, st9@sanger.ac.uk.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw sequencing data files are available at ArrayExpress (accession numbers: E-MTAB-8007, E-MTAB-8474, E-MTAB-8476, E-MTAB-8484 & E-MTAB-8486). Sequencing data for the microbiome are available at MGnify (ERA numbers are listed in Supplementary Table 6). Processed single-cell RNA sequencing data is available for online visualization and download at gutcellatlas.org.

Author contributions

K. R. J. initiated this project, designed and performed scRNA-seq and microbiome experiments, analyzed data and wrote the manuscript. T. G. analyzed bulk BCR-seq data and contributed extensively to scRNA-seq data analysis. R. E. contributed to data interpretation and data analysis. N. K. and E. L. G. analyzed 16S ribosomal sequencing and metagenomics data. M. D. S. assisted in microbiome related experiments. H. W. K. and L. K. J. analyzed the 10x Genomics VDJ datasets and contributed to generation of figures. B. R. B. and K. S. P. carried out tissue collection. J. R. F. designed flow-sorting panel and assisted in flow-sorting. V. P. assisted in bulk BCR library preparation and analysis. L. B. J., O. S., S. H. and J. L. J. dissociated tissues from donor 390c. K. P. carried out scRNA-seq read alignment and quality control. S. C. F., K. B. M. and M. R. C. designed experiments and interpreted data. T. D. L. and S. A. T. initiated and supervised the project and interpreted data. All authors edited the paper.

Competing interests

S.C.F. and T.D.L. are either employees of, or consultants to, Microbiotica Pty Ltd.

⁹Theory of Condensed Matter, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, United Kingdom, CB3 0HE

¹⁰European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, United Kingdom, CB10 1SA

Abstract

Gastrointestinal microbiota and immune cells interact closely and display regional specificity, but little is known about how these communities differ with location. Here, we simultaneously assess microbiota and single immune cells across the healthy, adult human colon, with paired characterization of immune cells in the mesenteric lymph nodes, to delineate colonic immune niches at steady-state. We describe distinct T helper cell activation and migration profiles along the colon and characterize the transcriptional adaptation trajectory of T regulatory cells between lymphoid tissue and colon. Finally, we show increasing B cell accumulation, clonal expansion and mutational frequency from cecum to sigmoid colon, and link this to the increasing number of reactive bacterial species.

The colon, as a barrier tissue, represents a unique immune environment where immune cells display tolerance towards a diverse community of microbes - collectively known as the microbiome. The microbiome is critical for many aspects of health and an imbalance of commensals and pathogenic microbes is linked with many disease states ¹. Thus, understanding what constitutes a healthy, homeostatic relationship between host immune cells and the microbiome of the human colon is of critical importance.

The composition of the microbiota at any location in the intestines is determined by the availability of nutrients and oxygen, the transit rate of luminal content and compartmentalized host immune activity, and as such, is spatially distinct ². Regional differences are most evident when comparing the small intestine and distal colon in humans or other mammals ². Within the colon, increasing bacterial diversity from proximal (including the cecum, ascending and transverse colon) to distal regions (comprising the descending and sigmoid colon connecting to the rectum) has been reported ³.

The intestinal immune system has a symbiotic relationship with the microbiome and is central to the maintenance of epithelial barrier integrity. The lamina propria and associated lymphoid tissues contain one of the largest and most diverse communities of immune cells - including both lymphocytes and myeloid cells ⁴. There is marked regional variation in immune cells along the gastrointestinal tract, with T helper (T_H) 17 cells decreasing in number from duodenum to colon, and T regulatory (T_{reg}) cell numbers being highest in the colon ⁵. Immune cells can respond to environmental cues including the microbiota. Mouse studies have demonstrated that specific bacterial species can fine-tune intestinal immune responses, including T_H17 ^{6,7}, T_{reg} ⁸, or T_H1 ^{9,10} and B cell activation ¹¹. However, the extent to which there is regional variation in the mucosal microbiome within an individual, and how this might influence local immune cell niches along the colon, has not been investigated to date.

Here, we catalogued the mucosal microbiome in different regions of the human colon, a gastrointestinal organ with the most diverse and dense microbiome content and region-restricted disease states². In parallel, we applied single-cell RNA-seq (scRNA-seq) to make a census of steady-state immune cell populations in the adjacent tissue and in draining mesenteric lymph nodes (mLN), results of which are available at www.gutcellatlas.org. We demonstrate previously unappreciated changes in the proportions and activation status of T and B cells in distinct regions of the healthy human colon from proximal to distal, and relate these differences to the changing microbiota.

Results

Microbiome composition differs along distinct colon regions

To create a map of bacterial composition at the mucosal surface of the colon, we performed 16S ribosomal RNA (16S rRNA) sequencing of swabs from the mucosa surface of the cecum, transverse colon and sigmoid colon of twelve disease-free Caucasian deceased transplant donors (Methods, Figure 1a and Supplementary Table 1). The major gut phyla - *Bacteroidetes*, *Firmicutes*, *Proteobacteria* and *Actinobacteria*, were present throughout the gut of each donor (Figure 1b and Supplementary Table 2). While diversity of operational taxonomic units (OTUs) was consistent across the colon, and considerable variability existed between donor as previously reported¹² (Supplementary Figure 1a), we did observe changes in the composition of the microbiome. Most notably at the level of phyla, *Bacteroidetes* was more prevalent in sigmoid colon (Figure 1c and Supplementary Figure 1b). This was mostly attributable to an increase in *Bacteroides*, which dominates the colonic microbiome of individuals on high protein and fat diets typical in Western countries¹³ (Figure 1c and Supplementary Figure 1c). Additionally, *Enterococcus* was more prevalent in the proximal colon, and *Coprobacillus* and *Escherichia/Shigella* were more abundant in the distal colon, although these proportions varied considerably between donors (Figure 1b,c and Supplementary Figure 1c).

Past studies characterizing the colonic microbiome typically rely on stool samples, which do not accurately recapitulate the composition of bacteria at the mucosal surface¹⁴. Our catalog of mucosal bacteria throughout the colon demonstrates heterogeneity in the microbiome at the mucosal surface from proximal to distal colon, and reveals specific genera with preference for colonizing certain colon regions.

Immune cell heterogeneity in steady-state colon

Next, we sought to determine whether the heterogeneity we observed in the colonic microbiome was accompanied by differences in the adjacent host immune cells. To this end, we generated high-quality transcriptional data from over 41,000 single immune cells from the mesenteric lymph nodes (mLN) and lamina propria of cecum, transverse colon and sigmoid colon (Figure 1a). We acquired tissue biopsies from five deceased transplant donors (Methods and Supplementary Table 1). Samples were dissociated to release cells from the mLN and lamina propria of colonic tissue. Immune cells were enriched either by flow-sorting of CD4⁺ T cells (live CD45⁺CD3⁺CD4⁺) and other immune cells (live CD45⁺CD3⁻CD4⁻), or by CD45⁺ magnetic bead selection or Ficoll gradient (Methods and

Supplementary Figure 2a). Each fraction was then subjected to scRNA-seq (Figure 1a). Despite the enrichment for immune cells, we captured epithelial cells and fibroblasts and these were computationally removed from further analysis (Methods).

Pooled analysis and visualization with Uniform Manifold Approximation and Projection (UMAP) of all four tissues from all donors revealed distinct clusters in the lymphoid and colonic tissues (Figure 2a and Supplementary Figure 2b). We identified 25 cell types and states in the intestinal lamina propria and mLN (Figure 2a, Supplementary Table 3 and Supplementary Table 4), consistent with the immune populations described in recent reports¹⁵. Among these were follicular and memory B cells, IgA⁺ and IgG⁺ plasma cells, effector and memory CD4⁺ T cells, T regulatory (T_{reg}) cells, CD8⁺ T cells, $\gamma\delta$ T cells, innate lymphoid cells (ILCs), natural killer (NK) cells, mast cells and myeloid cells (Figure 2a). Sub-clustering of myeloid cells showed two distinct populations of conventional dendritic cells: cDC1 expressing *XCR1*, *CADMI*, *CLEC9A*, *BATF3* and *IDO1*, and cDC2 expressing *CLEC10A* and *CD86* (Figure 2a,b). In addition, we identified monocytes expressing *CD14* and *CD68*, macrophages expressing *FCGR3A* (gene encoding CD16), *LYVE1*⁺ macrophages¹⁶ and plasmacytoid DCs (pDCs) expressing *IRF4* and *SELL*. B cells, DCs and $\gamma\delta$ T cells had *MKI67*⁺ cycling populations suggesting higher rates of proliferation compared to other colonic immune cell populations (Figure 2a,b).

To determine how immune cells differ along the colon from proximal to distal regions, we investigated the relative proportions of cell types within mLN and colon regions (Figure 2c,d). Since CD4⁺ T cells and all other immune cells were sorted separately during initial tissue processing, our analyses here were also kept separate. As noted from our visual inspection of the UMAP plot (Figure 2a), there were major differences in the cell types present between mLN and the colon. In particular, there were marked differences in the activation and memory status of T and B cells between the colon and lymph nodes, suggesting that these cell types are molded by their environment (Figure 2c,d). In the mLN, CD4⁺ T cells were typically *CXCR5*⁺, *ICOS*^{hi} follicular helper cells, and *SELL*⁺ (encodes CD62L), *CCR7*⁺ central memory cells (Figure 2b,c). In contrast, colonic CD4⁺ T cells had a more effector phenotype, expressing high levels of the tissue residency marker *CD69*¹⁷, falling into the T_H17 (*CCR6*⁺, *IL22*⁺, *CCL20*⁺) or T_H1 (*CXCR3*⁺, *IFNG*⁺) subtypes. There was an inverted gradient in the relative proportion of T_H17 and T_H1 cells, with the cecum dominated by T_H17 cells that reduced in frequency in the transverse colon, and still further in the sigmoid colon, and T_H1 cells following the opposite trend, being more abundant in the sigmoid colon (Figure 2c). This distinct distribution of colonic T_H1 and T_H17 cells is concordant with spatial variation of the microbiome, hinting at a relationship between the two populations.

B cells in the mLN were predominantly *CD19*⁺, *MS4A1*⁺ (encodes CD20), *CD40*⁺, *TNFRSF13B*⁺ (encodes TACI), *CD38*⁻ memory or *CXCR5*⁺, *TCL1A*⁺, *FCER2*⁺ (encodes CD23) follicular B cells (Figure 2b,d). In contrast, in the three regions of the colon, the main population of B cells were *SDCI1*⁺ (encodes CD138), *CD38*⁺, plasma cells. Plasma cells were enriched in the sigmoid colon relative to both the cecum and transverse colon, whereas the proportion of memory B cells was lower in the sigmoid colon (Figure 2d). This suggests that conditions in the sigmoid colon may favor the generation of plasma cells rather than

memory B cells from germinal center responses in this region, or that the tissue contains more plasma cell niches.

T helper cells disseminate through the colon and adopt region-specific transcriptional profiles

We next investigated CD4⁺ T effector cells across the colon. These were annotated above based on expression of functional markers (*IL17A*, *IL22* and *CCL20* versus *IFNG* and *CXCR3*) (Supplementary Figure 3a) rather than transcriptional regulators (*RORC* and *TBX21*) that are lowly expressed. Correlation analysis of T_{H1} and T_{H17} cells between different colon tissues revealed high transcriptional similarity between these effector cell subtypes (Figure 3a), with mLN versus peripheral tissue signature accounting for the greatest amount of variability (Spearman's corr = 0.88). Within the effector T cells of the colon, transverse colon and cecum cells cluster by T helper subtype. T_{H1} and T_{H17} cells of the sigmoid colon did not cluster with their respective effector subtypes from other regions. Differential gene expression analysis between the effector cells in the sigmoid colon versus those in cecum and transverse colon revealed higher expression of activation-related molecules including *TANK* (TRAF family member associated NF-κB activator) (adjusted P < 10⁻¹⁰), *CD83* (adjusted P < 10⁻¹⁰) and *PIM3* (adjusted P < 10⁻⁸) (Figure 3b). Expression of *CCL20*, encoding the ligand for CCR6 that is expressed by epithelial and myeloid cells more highly in small intestines than colon¹⁸, was also slightly increased by T helper cells of the proximal colon (Figure 3b). Although this is likely due to higher abundance of CCL20⁺ T_{H17} cells at this site (Figure 2c and Supplementary Figure 3a). Conversely, sigmoid colon effector T cells showed higher expression of *KLF2* (adjusted P < 10⁻¹⁴) (Figure 3b) that encodes a transcriptional factor that transactivates the promoter for Sphingosine-1-phosphate receptor 1 (S1PR1) and is critical for T cell recirculation through peripheral lymphoid tissue¹⁹, *LMNA* (adjusted P < 10⁻⁴⁷) that encodes a molecule that is reported to promote T_{H1} differentiation²⁰ and *EEF1G* (adjusted P < 10⁻¹³) that encodes a driver of protein synthesis.

Next, we looked into the clonal relationships between T helper cells. We performed plate-based Smartseq2 on TCRα/β⁺ flow-sorted cells from colon regions and mLN of a sixth donor to capture paired gene expression and TCR sequences from individual T cells. Clonal groups were shared between T_{H1} and T_{H17} subtypes, supporting the notion that effector fate of CD4⁺ T cells is determined after their initial activation (Figure 3c,d)²¹. Additionally, clonal expansion was observed by T_{H1} cells of the sigmoid colon (Figure 3d), in line with greater abundance in this tissue. Likewise, clonal expansion of T_{H17} cells was greatest in the cecum matching accumulation seen with the droplet-based scRNA-seq analysis (Figure 3d). Several T_{H1} and T_{H17} clonal sisters were shared between clonal sites (Figure 3d), evidence that T helper clones disseminate to distant regions of the colon.

Together these data demonstrate region-specific transcriptional differences relating to activation and tissue migration in T_{H1} and T_{H17} cells of the proximal and sigmoid colon. Identification of clonal sharing between these colon regions supports the idea that these observed transcriptional differences are due to cell-extrinsic rather than intrinsic factors.

Activation trajectory of colonic CD4⁺ T regulatory cells

T_{reg} cells are known to play a role in balancing the immune activity of other CD4⁺ T cell subsets. Documented T_{reg} cell activation by *Clostridium spp.*⁸ and previous descriptions of tissue-specific transcriptional profiles²² inspired us to interrogate these cells in greater detail. Firstly, we noted that the relative proportion of T_{reg} cells did not change significantly from proximal to distal colon (Figure 2c). We then investigated whether the transcriptional profiles of T_{reg} cells from different compartments indicated distinct activation states or functionalities.

As previously observed in the mouse²², sub-clustering of T_{reg} cells from the mLN revealed major populations of central T_{reg} cells and effector T_{reg} cells (Figure 4a). Central T_{reg} cells were defined by highest expression of *SELL*, while effector T_{reg} cells were characterized by genes associated with the TNFRSF-NF- κ B pathway (*TNFRSF9* and *TNF*) (Supplementary Figure 4a). In addition, we observed a population, previously termed non-lymphoid tissue-like T_{reg} cells (NLT-like T_{reg}). These have characteristics of non-lymphoid tissue T_{reg} cells, including high expression of *FOXP3*, *PRDM1*, *RORA*, *IL2RG*, *IL2RA* and *CTLA4* (Figure 4a & Supplementary Figure 4a). A fourth population, termed T_{reg}-like cells, lacked the conventional markers of T_{reg} cells - *FOXP3* and *IL2RA* - but clustered more closely with T_{reg} cells than conventional T cells (Figure 4a and Supplementary Figure 4b). These expressed the highest levels of *PDCDI* (gene encoding PD-1) among T_{reg} cell populations and, uniquely for this cell type, the transmembrane protein-encoding gene *MS4A6A* (Supplementary Figure 4a). This population of T_{reg}-like cells could represent a T_{reg} population that transiently loses FOXP3 expression²³.

In the colon, T_{reg} cells clustered into three populations as previously described in mouse (Figure 4a)²². KLRB1⁺ (also known as CD161) T_{reg} cells were characterized by expression of *LAG3*, *IL2RA*, *CTLA4*, *KLRB1*, *ICOS* and *FOXP3* (Supplementary Figure 4a) suggesting a robust regulatory function, analogous to CD161⁺ T_{reg} cells described in human colon tissue²⁴. Non-lymphoid tissue T_{reg} (NLT T_{reg}) cells express *IKZF2*, *GATA3* and *DUSP4* (Supplementary Figure 4a) consistent with the profile of thymic-derived T_{reg} cells²⁵. The third population exhibited a profile reminiscent of lymphoid tissue T_{reg} cells with expression of *SELL*, *CCR7*, *TCR7*, *CXCR5* and *RGS2*, and were termed Lymphoid Tissue-like T_{reg} (LT-like T_{reg}) cells (Supplementary Figure 4a). LT-like T_{reg} cells are likely newly arrived in the colon tissue from mLN. A small number of *Ki67*⁺ T_{reg} cells were also identified in the colon and mLN (Figure 4a and Supplementary Table 3). The proportions of these subsets varied by donor, but were mostly consistent between colon regions (Figure 4b).

The presence of NLT-like T_{reg} cells in the mLN and LT-like T_{reg} cells in the colon, both with profiles suggesting migration, led us to recreate this migration pathway *in silico*. We ordered all T_{reg} cells along “pseudospace” using Monocle2. This gave rise to a smooth pseudospace trajectory from resting central T_{reg} cells in the mLN to highly regulatory T_{reg} cells in the colon (Figure 4c). As seen in the mouse²², NLT-like T_{reg} cells and LT-like T_{reg} cells blended in the middle of the trajectory, in accordance with these cells representing transitioning and migratory populations between lymphoid and peripheral tissues. In order to understand which gene signatures drive the migration and tissue adaptation of T_{reg} cells in human tissues, we determined the genes changing along the previously calculated

pseudospace (Figure 4c). Genes expressed at the beginning of pseudospace included *SELL*, *CCR7* and *CXCR4*, permitting entry into lymph nodes (Figure 4c). At the end of pseudospace, the most highly expressed genes were *FOXP3*, *IL2RA*, *CTLA4*, *IL10* and *LAG3*. These genes encoding suppressive molecules were co-expressed with TNF receptor genes (*TNFRSF4*, *TNFRSF18*, *TNFRSF1B*) indicating a reliance on the TNFRSF-NF- κ B axis. Chemokine receptor-encoding genes, *CXCR3*, *CXCR6*, *CCR6* and *CCR4*, were also expressed by T_{reg} cells in the periphery, matching previous reports of T_{H1}- and T_{H17}-like T_{reg} cells in the colon (Figure 4c)²².

Together, these results highlight heterogeneity in T_{reg} cell states in mLN and colon, and reveal a possible FOXP3-transiently absent population. We also infer a continuous activation trajectory of these T_{reg} cell states between draining lymph nodes and colon, and highlight genes regulating T_{reg} cell migration between tissues and their adoption of Th-like profiles.

B cells display a proximal-to-distal colon activation gradient

Following from our observations in CD4⁺ T cells across the colon, we next focused on humoral responses by performing a more in-depth analysis of B cells in different colon regions. We compared transcriptional profiles of plasma cells between different colonic regions. This analysis revealed *CCL3* and *CCL4* as highly enriched in cecal plasma cells (log fold change of 0.61 and 0.70 and adjusted P < 10⁻²⁸ and < 10⁻⁹ respectively) (Figure 5a). These chemokine-encoding genes are expressed by B cells in response to BCR activation²⁶ and result in the migration of CCR5-expressing cells such as T cells and monocytes²⁷ to the tissue microenvironment. This suggests that BCR cross-linking and signaling may be more prominent in the proximal colon. Cecal plasma B cells were also enriched for *CXCR4* (log fold change of 0.44, adjusted P < 10⁻¹⁰) (Figure 5a), which encodes a chemokine receptor highly expressed by germinal center (GC) B cells and important for the movement of plasmablasts to the GC-T zone interface post-GC responses²⁸.

Among the genes more highly expressed by plasma cells in the sigmoid colon was *CD27* (Figure 5a; log fold change of 0.24; adjusted P < 10⁻⁴⁰), encoding a member of the TNF receptor family that is expressed by memory B cells and even more highly by plasma cells²⁹. We confirmed differential expression of CD27 protein and additionally observed a proximal-to-distal gradient of increasing CD27 expression by plasma cells in the colon (Figure 5b). Targeted homing of B cells from their site of activation in lymphoid tissues to the colonic lamina propria relies on signaling through CCR10 and its cognate ligand, CCL28, and integrin $\alpha_4\beta_7$ ³⁰. *CCR10* (log fold change of 0.24; adjusted P < 10⁻¹⁶), *ITGA4* (log fold change of 0.58; adjusted P < 10⁻²⁴) and *ITGB4* (log fold change 0.57; adjusted P < 10⁻¹⁰⁸) were also more highly expressed by sigmoid colon IgA⁺ plasma cells (Figure 5a).

To determine whether the B cell clonal repertoire changes across the colon, we took advantage of the paired single-cell VDJ-sequencing data available from two donors for which scRNA-seq libraries were generated using 10x Genomics 5' chemistry. We confirmed the expression of IgM and IgD isotypes by follicular B cells, IgG1 and IgG2 by IgG⁺ plasma cells, IgA1 and IgM expression by memory cells in the mLN and predominantly IgA2 expression by plasma cells of the colon (Figure 5c). The mutation frequency of the heavy chain variable region was greatest in the plasma cells followed by memory B cells,

indicating more somatic hypermutation by these cell types compared to the naive follicular B cells (Figure 5d). Additionally, while mutational frequency was consistent across colon regions and mLN for memory and follicular B cells, it was significantly increased in IgA⁺ plasma cell of the sigmoid colon compared to the other colon regions (Figure 5e). IgG⁺ plasma cells also showed a trend towards increased mutational frequency in the sigmoid colon, however their numbers were limiting (Figure 5e).

We then identified clonally-related B cells to explore clonal expansion dynamics of different cellular populations throughout the gut. Clonal expansion was evident in memory B cells, IgA⁺ plasma cells and IgG⁺ plasma cells (Figure 5f). While the relative abundance of clonal groups did not differ across the colon regions of memory B cells and IgG⁺ plasma cells, again this was greatest for IgA⁺ plasma cells in the sigmoid colon (Figure 5f). This was supported by bootstrapped VDJ sequence diversity analysis of clonally-related IgA⁺ plasma cells, which showed that the diversity of BCR sequences was consistent between donors and that there was a trend for decreased diversity (consistent with higher rates of clonal expansion) of IgA⁺ plasma cells in the sigmoid colon (Supplementary Figure 5a,b). Although some clones were shared between B cell types (i.e. memory and IgA⁺ plasma), indicating that alternate B cell fates can derive from a single precursor cell, most expanded clones within the gut were of the same cell type (Supplementary Figure 5c). Finally, we found many examples of B cell clones shared between all three colonic regions, and to a lesser extent the mLN, for both donors (Figure 5g), indicating dissemination of B cells throughout the colon as previously reported³¹. Our observation of B cell dissemination throughout the colon was replicated with bulk BCR sequencing from whole tissue (Figure 5h). Increased clonal variability of sigmoid colon B cells was reflected in a greater spread of BCR variable chains expressed compared with cecum and transverse colon (Supplementary Figure 5d), altogether suggesting a more active response in the distal versus proximal colon.

These data indicate a highly activated state of plasma cells in the distal colon compared with proximal colon plasma cells, characterized by greater accumulation, somatic hypermutation, clonal expansion and stronger homing to the colon mucosa.

IgA is responsive to a richer microbial community in the sigmoid colon

Previous reports have shown that IgA is secreted by plasma cells in response to the presence of specific bacteria species, rather than as a general response to the microbiota³². In light of this, we examined whether the increased plasma cell activation we observed in the sigmoid colon was linked to the differences in bacteria species. To this end, we assessed IgA-opsinization of bacteria from the donor microbiota samples (Figure 6a). A greater proportion of bacteria from the sigmoid colon was positive for IgA-binding compared with bacteria of the cecum and transverse colon (Figure 6b). Furthermore, shotgun sequencing of IgA-opsinized bacteria revealed a richer community of species in the sigmoid colon (Figure 6c and Supplementary Table 5). Diversity of IgA-bound bacteria, which considers relative abundance of each species, was lower in the sigmoid colon compared to cecum (Figure 6d).

These data suggest that, compared with the cecum, IgA⁺ plasma cells of the sigmoid colon respond to a rich and unevenly represented community of bacterial species, likely

contributing to their increased activation status, strong homing to the colon and trend towards greater clonal diversity (Supplementary Figure 6).

Discussion

In this study we performed the first simultaneous assessment of the colonic mucosal microbiome and immune cells in human donors at steady-state. This enabled us to compare lymphoid and peripheral tissue immunity and explore how immune cells and their neighboring microbiome change along the colon within the same individuals. In doing so, we highlight previously unappreciated regional differences in both cellular communities. Our unique annotated colon immune single-cell dataset is available at www.gutcellatlas.org, where users can visualize their genes of interest.

We describe a shift in the balance of T helper subsets, with a predominance of T_H17 in the cecum and T_H1 in the sigmoid colon. Decreasing abundance of T_H17 cells has similarly been shown from proximal small intestine to colon of mice⁵. Additionally, simultaneous increase of the genus *Bacteroides* and T_H1 numbers in the sigmoid colon are in line with findings that polysaccharide from *Bacteroides fragilis* preferentially induces T_H1 differentiation in the intestine of germ-free mice⁹. An alternative or complementary explanation for skewed T helper proportions and variation in transcriptional profiles is offered in a study by Harbour et al., which demonstrated that T_H17 cells can give rise to a IFN γ ⁺ ‘T_H1-like’ cell in response to IL-23 production by innate cells³³. These findings demonstrate the complexity of external signals shaping colonic Th responses leading to regional changes in their numbers and differentiation.

In contrast to conventional T cells, T_{reg} cells are evenly represented across the colon. T_{reg} cell subpopulations within the mLN and colon tissue are analogous to those we have recently described in mouse²². We also identify an additional population, termed T_{reg}-like cells, reminiscent of a CD25⁻FOXP3^{lo}PD1^{hi} T_{reg} population in the peripheral blood, although the latter was described to also express Ki67³⁴. This population could represent uncommitted T_{reg} cells experiencing transient loss of FOXP3 while retaining regulatory potential²³ or permanent loss of FOXP3 and adoption of a more pro-inflammatory phenotype after repeated stimulation³⁵. Our pseudospace analysis of T_{reg} cells suggests a continuum of activation states from resting cells in the lymphoid tissue through to highly suppressive cells in the periphery. We identify genes underlying this transition including chemokine receptors that are strongly expressed on arrival to the intestine and enable interaction with, and suppression of, T_H1 and T_H17 cells³⁶. Similarly to mouse, pseudospace from lymphoid to peripheral tissue correlates with transcriptional markers of migratory potential and suppression of effector cells.

We find that IgA⁺ plasma cells are more abundant and have greater expression of colon-specific migration markers (*CCR10*, *ITGA4* and *ITGB7*) in the sigmoid colon compared to the proximal colon. This adds finer resolution to previous reports describing increasing abundance of IgA⁺ plasma cells from the small intestine to the colon³⁷. Our B cell repertoire analysis demonstrates extensive clonal expansion within each colon region and to a lesser extent between regions, arguing for colonic dissemination of B cells from the same

precursor pool, followed by local expansion. Furthermore, more clonal sharing existed between regions of the colon than with mLN, consistent with recent work showing that while mLN clones can also be detected in blood, the intestines are host to a unique B cell clonal network³⁸. Sigmoid colon plasma cells, in particular, exhibit greater mutational abundance and clonal expansion. This is consistent with a trend towards reduced plasma cell clonal diversity in the sigmoid colon. While this was not statistically significant, this may have been due to the relatively low number of VDJ sequences obtained from these samples, thus decreasing the power of this analysis. Previous work has shown the mutational frequency of B cells is consistently high between duodenum and colon and are primarily driven by dietary and microbiome antigens respectively³⁹. Thus, we suggest that enhanced plasma cell accumulation, mutation and expansion in the sigmoid colon is in response to continued stimulation from the local microbiome. This may happen through increased engagement of sigmoid colon plasma cells in T cell-mediated germinal center reactions in gut-associated lymphoid structures or in T cell-independent somatic hypermutation in local isolated lymphoid follicles⁴⁰ followed by local expansion. Yet the exact mechanisms require further study.

Finally, we show greater IgA binding to the microbiota in the sigmoid colon compared to proximal sites. One possible explanation for this observation is the accumulation of upstream secreted IgA and IgA-bound bacteria in the sigmoid colon. However, our simultaneous observation of enhanced plasma cell responses in the sigmoid colon suggests that IgA is locally produced as a result of immune poisoning. Possible scenarios contributing to a richer immune-reactive microbiome in the sigmoid colon are bacteria derived externally via the rectum. Alternatively, environmental pressures (i.e. lower water and nutrient levels²) could restrict outgrowth of dominant gastrointestinal species of the proximal colon, providing space for smaller communities of opportunistic species. The IgA response in the colon is antigen-specific rather than a general response to the presence of bacteria³². Thus, the overall increased number of unique species recognized by host IgA antibodies in the sigmoid colon is fitting with the enhanced clonal expansion and mutation of plasma cells at this site.

Together, our simultaneous analyses of microbiome and neighboring immune cells highlight the significance of environmental signals in shaping and maintaining regional adaptive immune cell composition and function in the intestine at steady-state. Dysregulation of T helper cells⁴¹ and plasma cells⁴², has been implicated in susceptibility to inflammatory bowel disease. Observations of the linked compartmentalization of these immune cells and microbial species along the colon at steady-state may provide a platform for understanding the mechanisms underpinning the tropism of different intestinal diseases to specific regions of the gut, such as Crohn's disease and ulcerative colitis.

Methods

Colon and mesenteric lymph node tissue retrieval

Human tissue was obtained from deceased transplant organ donors after ethical approval (reference 15/EE/0152, East of England Cambridge South Research Ethics Committee) and informed consent from the donor family. Fresh mucosal tissue from the cecum, transverse

colon and sigmoid colon, and lymph nodes from the intestine mesentery, were excised within 60 minutes of circulatory arrest and colon tissue preserved in University of Wisconsin (UW) organ preservation solution (Belzer UW® Cold Storage Solution, Bridge to Life, USA) and mLN stored in saline at 4°C until processing. Tissue dissociation was conducted within 2 hours of tissue retrieval. Four individuals (287c, 296b, 403c and 411c) had received antibiotics in the two weeks prior to death (Supplementary Table 1).

Tissue dissociation for flow-sorting separation and MAC separation

Tissue pieces from donors 290b, 298c, 302c, 364b and 411c were manually diced and transferred into 5 mM EDTA (Thermo Fisher Scientific)/1 mM DTT (Sigma-Aldrich)/10 mM HEPES (Thermo Fisher Scientific)/2% FBS in RPMI and incubated in a shaker (~200 rpm) for 20 minutes at 37°C. Samples were briefly vortexed before the media renewed and incubation repeated. Tissue pieces were washed with 10 mM HEPES in PBS and transferred into 0.42 mg/ml Liberase TL (Roche)/0.125 KU DNase1 (Sigma)/10 mM HEPES in RPMI and incubation for 30 minutes at 37°C. The digested samples were passed through a 40 µm strainer and washed through with FBS/PBS.

Flow-sorting

Cells from donor 290b, 298c and 302c were pelleted and resuspended in 40% Percoll (GE Healthcare). This was underlayered with 80% Percoll and centrifuged at 600g for 20 minutes with minimal acceleration and break. Cells at the interface were collected and washed with PBS. Cells were stained for fluorescent cytometry using Zombie Aqua Fixable Viability Dye (Biolegend, cat. 423101; diluted 1:200), CD45-BV605 (clone HI30, Biolegend, cat. 304043; dilution 1:100), CD3-FITC (clone OKT3; Biolegend, cat. 317305; dilution 1:100), CD4-BV421 (clone SK3; Biolegend, cat. 344631; dilution 1:100), CD8-PE-Cy7 (clone SK1, Biolegend, cat. 344711; dilution 1:100), CD19-APC-Cy7 (clone HIB19, Biolegend, cat. 302217; dilution 1:100), IgD-PE Dazzle (clone IA6-2, Biolegend, cat. 348207; dilution 1:100), CD27-BV711 (clone M-T271, Biolegend, cat. 356429; dilution 1:100), HLA-DR-BV785 (clone L243, Biolegend, cat. 307641; dilution 1:100), CD14-APC (clone 63D3, Biolegend, cat. 367117; dilution 1:100) and CD11c-PE (clone 3.9; eBioscience, cat. 12-0116-42; dilution 1:100). Non-CD4⁺ T immune cells were sorted as live singlet CD45⁺, CD3⁻ and CD4⁻. CD4⁺ T cells were sorted as live singlet CD45⁺, CD3⁺ and CD4⁺. Each fraction was manually counted using 0.4% Trypan Blue (gibco) and a haemocytometer and diluted to 500 cells/µl in PBS. Sorting was carried out on a BD FACS ARIA Fusion. Analysis of flow-sorting data was conducted with FlowJo Software package (version 10.4).

MACS cell enrichment

Cells from donor 411c were pelleted for 5 minutes at 300g and resuspended in 80 µl of ice-cold MACS buffer (0.5% BSA (Sigma-Aldrich Co. Ltd), 2 mM EDTA (ThermoFisher) in DPBS (Gibco)) and 20 µl of CD45 Microbeads (Miltenyi Biotech). Cells were incubated for 15 minutes at 4°C before being washed with 2 ml of MACS buffer and centrifuged as above. Cells were resuspended in 500 µl of MACS buffer and passed through a pre-wetted MS column on QuadroMACS magnetic cell separator (Miltenyi). The column was washed 4 times with 500 µl of MACS buffer, allowing the full volume of each wash to pass through the column before the next wash. The column was removed from the magnet and the cells

eluted with force with 1 ml of MACS buffer into a 15 ml tube. Cells were pelleted as above and cell number and viability were determined using a NucleoCounter NC-200 and Via1-Cassette (ChemoMetec). Cells were resuspended at 500 cells/ μ l in 0.04% BSA in PBS.

Tissue dissociation for Ficoll separation

Tissues from donor 390c were manually diced and <5.0 grams was added per Miltenyi C tube with 5mL tissue dissociation media (Liberase TL (0.13 U/mL; Roche) and DNase (10 U/mL Benzonase nuclease; Merck)) in 1% FCS and 20 mM HEPES in PBS (Lonza). Samples were dissociated with GentleMACS Octo for 30 minutes homogenizing/37°C cycle. Enzymatic digestion was stopped with the addition of 2 mM EDTA in tissue dissociation media. Digested samples were then passed through a 70 μ M smart strainer (Miltenyi) before being washed with PBS and pelleted at 500g for 10 minutes. Cells were resuspended in PBS, layered onto FicollPaque Plus (GE Healthcare) and spun at RT 400g for 25 minutes. Mononuclear cells were retrieved from Ficoll layer and washed with PBS. Cells were filtered through 0.2 μ M filter (FLOwmi cell strainers, BelArt). Cells were manually counted using a hemocytometer and diluted to a concentration of 1000 cells/ μ l in 0.04% BSA in PBS.

10x Genomics Chromium GEX library preparation and sequencing

Cells were loaded according to the manufacturer's protocol for Chromium single cell 3' kit (version 2) or 5' gene expression (version 2) in order to attain between 2000-5000 cells/well. Library preparation was carried out according to the manufacturer's protocol. For samples from donors 290b, 298c and 302c, eight 10x Genomics Chromium 3' libraries were pooled sequenced on eight lanes of an Illumina HiSeq 4000. For samples from donors 390c and 417c, sixteen 10x Genomics Chromium 5' libraries were pooled and sequenced on 2 lanes of a S2 flowcell of Illumina Novaseq 6000 with 50 bp paired end reads.

10x Genomics Chromium VDJ library preparation and sequencing

10x Genomics VDJ libraries were generated from the 5' 10x Genomics Chromium cDNA libraries as detailed in the manufacturer's protocol. BCR libraries for each sample were pooled and sequenced on a single lane of Illumina HiSeq 4000 with 150 bp paired end reads.

Plate-based scRNA-seq

Plate-based scRNA-seq was performed with the NEBNext Single Cell/Low Input RNA Library Prep Kit for Illumina (New England Biolabs Inc, E6420L). Cells from donor 364b were snap frozen in 10% DMSO in 90% BSA. Cells were thawed rapidly in a 37°C water bath and diluted slowly with pre-warmed 2% FBS in D-PBS. Cells were pelleted for 5 minutes at 300g and washed with 500 μ l of DPBS and pelleted as before. Cells were resuspended in 200 μ l of CD25-PE (clone M-A251, Biolegend, cat. 356102; diluted 1:200), CD127-FITC (clone eBioRDR5, eBioscience, cat. 11-1278-42; diluted 1:200), CD4-BV421 (clone SK3, Biolegend, cat. 344632; diluted 1:200) and TCR α / β -APC (clone 1p26, Biolegend, cat. 306718; diluted 1:200) and Zombie Aqua Fixable Viability Dye (diluted 1:400) and incubated for 30 minutes in the dark at room temperature. Cells were washed twice with 500 μ l of 2% FBS in D-PBS before being filtered through a 100 μ M filter. Single,

live, TCR β ⁺ cells were FACS sorted into a pre-prepared 384-well plate (Eppendorf, Cat. No. 0030128508) containing 2 μ l of 1X NEBNext Cell Lysis Buffer. FACS sorting was performed with a BD Influx sorter with the indexing setting enabled. Plates were sealed and spun at 100g for 1 minute then immediately frozen on dry ice and stored at -80°C.

cDNA generation was then performed in an automated manner on the Agilent Bravo NGS workstation (Agilent Technologies). Briefly, 1.6 μ l of Single Cell RT Primer Mix (New England Biolabs Inc) was added to each well and annealed on a PCR machine (MJ Research Peltier Thermal Cycler) at 70°C for 5 minutes. 4.4 μ l of Reverse Transcription (RT) mix was added to the mixture and further incubated at 42°C for 90 minutes followed by 70°C for 10 minutes to generate cDNA. 22 μ l of cDNA amplification mix containing NEBNext Single Cell cDNA PCR MasterMix and PCR primer was mixed with the cDNA, sealed and spun at 100g for 1 minute. cDNA amplification was then performed on a PCR machine (MJ Research Peltier Thermal Cycler) with 98°C 45 s, 20 cycles of [98°C 10 s, 62°C 15 s, 72°C 3 mins], 72°C 5 mins. The 384-well plate containing the amplified cDNA was purified with an AMPure XP workflow (Beckman Coulter, Cat No. A63880) and quantified with the Accuclear Ultra High Sensitivity dsDNA kit (Biotium, Cat. No. 31028). ~10 ng of cDNA was stamped into a fresh 384-well plate (Eppendorf, Cat. No. 0030128508) for sequencing library preparation.

Sequencing libraries were then generated on the Agilent Bravo NGS workstation (Agilent Technologies). Purified cDNA was fragmented by the addition of 0.8 μ l of NEBNext Ultra II FS Enzyme Mix and 2.8 μ l of NEBNext Ultra II FS Reaction buffer to each well and incubated on a PCR machine (MJ Research Peltier Thermal Cycler) for 72°C at 15 minutes and 65°C for 30 minutes. A ligation mixture was then prepared containing NEBNext Ultra II Ligation Master Mix, NEBNext Ligation Enhancer and 100 μ M Illumina compatible adapters (Integrated DNA Technologies) and 13.4 μ l added to each well of the 384-well plate. The ligation reaction was incubated on the Agilent workstation at 20°C for 15 minutes and then purified and size selected with an AMPure XP workflow (Beckman Coulter, Cat No. A63880). 20 μ l of KAPA HiFi HS Ready Mix (Kapa Biosystems, Cat. No. 07958927001) was then added to a pre-prepared 384-well plate (Eppendorf, Cat. No. 0030128508) containing 100 μ M i5 and i7 indexing primer mix (50 μ M each) (Integrated DNA Technologies). The indexing primers pairs were unique to allow multiplexing of up to 384 single cells in one sequencing pool. The plate containing the PCR Master Mix and indexing primers was stamped onto the adapter ligated purified cDNA, sealed and spun at 100g for 1 minute. Amplification was performed on a PCR machine (MJ Research Peltier Thermal Cycler) with 95°C for 5 minutes, 8 cycles of [98°C 30 seconds, 65°C 30 seconds, 72°C 1 minute], 72°C 5 minutes. The PCR products were pooled in equal volume on the Microlab STAR automated liquid handler (Hamilton Robotics) and the pool purified and size selected with an AMPure XP workflow (Beckman Coulter, Cat No. A63880). The purified pool was quantified on an Agilent Bioanalyser (Agilent Technologies) and sequenced on one lane of an Illumina HiSeq 4000 instrument.

Single-cell RNA sequencing data alignment

10x Genomics gene expression raw sequencing data was processed using Cell Ranger software version 3.0.2 and 10X human transcriptome GRCh38-3.0.0 as the reference. 10x Genomics VDJ immunoglobulin heavy and light chain were processed using cellranger vdj (version 3.1.0) and the reference cellranger-vdj-GRCh38-alts-ensembl-3.1.0 with default settings. NEB sequencing data was processed using STAR 2.5.1b into HTSeq (version 0.11.2) and mapped to 10X human transcriptome GRCh38-1.2.0.

Single-cell RNA sequencing quality control

Single cell read counts from all samples were pooled and filtered considering number of UMIs - keeping genes expressed in minimum of 3 cells, keeping cells where genes detected are in a range 700-6000. Non-immune cells were excluded from the final analysis based on the absence of PTPRC and presence of markers such as EPCAM (epithelial cells) and COL1A1 (fibroblasts).

Cell type annotation

Cells were clustered using Scanpy (version 1.4) processing pipeline⁴³. In short, the counts were normalized to 10,000 reads per cell (`sc.pp.normalise_per_cell`) and log transformed (`sc.pp.log1p`) to be comparable amongst the cells. The number of UMIs and percentage of mitochondrial genes were regressed out (`sc.pp.regress_out`) and genes were scaled (`sc.pp.scale`) to unit variance. The normalized counts were used to detect highly variable genes (`sc.pp.highly_variable_genes`). Batch correction between the donors was performed using `bbknn` (version 1.3.6)⁴⁴ on 50 PCs and trim parameter set to 100. Clusters were then identified using Leiden graph-based clustering (resolution set to 1). Cell identity was assigned using known markers shown in Figure 2b and the top differentially expressed genes identified using Wilcoxon rank-sum test (`sc.tl.rank_genes_groups` function, Supplementary Table 4). CD4 T cells, myeloid cells, B cells were subclustered for identification of subsets within each cluster. T_{reg} cells annotated above were further subclustered using the “FindClusters” and “RunUMAP” functions from Seurat (version 3.0.1)⁴⁵ (Figure 4). The number of PCs used for T_{reg} cell clustering were estimated by the elbow of a PCA screen plot, in combination to manual exploration of the top genes from each PC. Clustering of mLN T_{reg} cells was performed using 1-14 PCs and resolution of 0.3 and colonic T_{reg} cells with 1-11 PCs and resolution of 0.3.

‘Pseudospace’ analysis

‘Pseudospace’ analysis of T_{reg} cells was performed using Monocle (version 2.10.1)⁴⁶. Data was log normalized and cell ordered based on DDRTree reduction on highly variable genes with donor effect regression. The heatmap in Figure 4c was generated using the “plot_pseudotime_heatmap” function in Monocle (version 2.10.1)⁴⁶. Genes contributing to ‘pseudospace’ were first filtered for exclusion of mitochondria and immunoglobulin genes, expression in at least 15 cells and $qval < 0.001$. Gene expression was smoothed into 100 bins along pseudospace using a natural spline with 3 degrees of freedom. Matching column annotation for each bin was determined as the most prevalent tissue region origin of the cells

within that bin. Genes were grouped into 2 clusters and ordered by hierarchical clustering. Cells were ordered through pseudotime.

Sampling of microbiome

Swabs were taken immediate of the mucosal surface of excised tissue using MWE Transwab Cary Blair (catalogue number: MW168). Swabs were maintained at 4°C. Working in a biosafety cabinet, swabs were washed in 500 µl of anaerobic PBS and mixed with anaerobic 50% glycerol before being snap frozen on dry ice and stored at -80°C until use.

Microbiota profiling and sequencing

The microbiota vials were defrosted on ice. Approximately 100 µl of each was transferred into new Eppendorf tubes. DNA was extracted from microbiome samples using the MP Biomedical FastDNA SPIN Kit for soil (catalogue number 116560200). 16S rRNA gene amplicon libraries were made by PCR amplification of variable regions 1 and 2 of the 16S rRNA gene using the Q5 High-Fidelity Polymerase Kit supplied by New England Biolabs as described in ⁴⁷. Primers 27F AATGATACGGCGACCACCGAGATCTACAC (first part, Illumina adaptor) TATGGTAATT (second part, forward primer pad) CC (third part, forward primer linker) AGMGTTYGATYMTGGCTCAG (fourth part, forward primer) and 338R CAAGCAGAAGACGGCATAACGAGAT (first part, reverse complement of 3' Illumina adaptor) ACGAGACTGATT (second part, golay barcode) AGTCAGTCAG (third part, reverse primer pad) AA (fourth part, reverse primer linker) GCTGCCTCCCGTAGGAGT (fifth part, reverse primer) were used. Four PCR amplification reactions per sample were carried out; products were pooled and combined in equimolar amounts for sequencing using the Illumina MiSeq platform, generating 150 bp reads. Analysis of partial 16S rRNA sequences was carried out using SILVA v132 and mothur MiSeq SOP v1.42.3 ⁴⁸. The 16S rRNA gene alignments were used to determine a maximum likelihood phylogeny using FastTree (version 2.1.10) ⁴⁹. Phylogenetic trees were visualized and edited using iTOL (version 5) ⁵⁰.

T cell clonal sharing analysis

T cell receptor sequences generated using the Smartseq2 scRNA-seq protocol were reconstructed using the TraCeR software as previously described ⁵¹.

Bulk B cell receptor sequencing

Small portions of samples were taken from excised tissues and snap frozen in 1ml of RNAlater (Ambion). RNA extracted from tissue using the QIAshredder and QIAGEN Mini Kit (50). RNA concentration was measured using a Bioanalyzer. B Cell Receptor (BCR) heavy chain sequences of all B lineage subsets present in the tissue were amplified as previously described ⁵². Briefly, RNA was reverse transcribed using a barcoded reverse primer set capturing all antibody (sub)classes. Targeted heavy-chain amplification was performed with a multiplex set of IGHV gene primers to FR1 and a universal reverse primer using HiFi qPCR KAPA Biosystems. After adapter filtering and trimming, BCR sequences were assembled and aligned using MiXCR (version 3.0.1) ⁵³. It is worth noting that detected BCR sequences are biased towards those included in the reference database and while there

is a continuous discovery of novel germline alleles, no database is currently a complete reflection of the human IGH locus diversity. Only in-frame and IGH sequences with at least 3 read counts were kept for the analysis. To calculate the CDR3 nucleotide shared repertoire, the tCR package (version 2.2.1) was used ⁵⁴.

10x Genomics single-cell VDJ data processing, quality control and annotation

Poor quality VDJ contigs that either did not map to immunoglobulin chains or were assigned incomplete by cellranger vdj were discarded. For additional processing, all IgH sequence contigs per donor were combined together. We further filtered IgH contigs as to whether they had sufficient coverage of constant regions to ensure accurate isotype assignment between closely related subclasses using MaskPrimers.py (pRESTO version 0.5.10) ⁵⁵. IgH sequences were then further annotated using IgBlast (version 1.12.0) ⁵⁶ and reassigned isotype classes using AssignGenes.py (pRESTO) prior to correction of ambiguous V gene assignments using TIGGER (version.03.1) ⁵⁷. Clonally-related IgH sequences were identified using DefineClones.py (ChangeO version 0.4.5) ⁵⁷ with a nearest neighbor distance threshold of 0.2, as determined by visual inspection of the output of distToNearest (Shazam version 0.1.11) ⁵⁷. CreateGermlines.py (ChangeO version 0.4.5) was then used to infer germline sequences for each clonal family and observedMutations (Shazam version 0.1.11) was used to calculate somatic hypermutation frequencies for each IgH contig. Estimated clonal abundances and IgH diversity analyses within each donor were performed using estimateAbundance, rarefyDiversity and testDiversity of Alakazam (version 0.2.11) ⁵⁷ with a bootstrap number of 500. Finally, the number of quality filtered and annotated IgH, IgK or IgL chains were determined per unique cell barcode. If more than one contig per chain was identified, metadata for that cell was ascribed as “Multi”. The subsequent metadata table was then integrated with the single-cell RNA-seq gene expression objects for annotation of IgH contigs with B cell types and downstream analysis. Co-occurrence of expanded clone members between tissues and/or cell types was reported as a binary event for each clone that contained a member within two different tissues and/or cell types.

Quantifying IgA binding of bacteria

Microbiome frozen in 50% glycerol were defrosted on ice before being washed in PBS and pelleted for 3 minutes at 8,000 rpm. Bacteria were then stained on ice for 30 minutes with IgG-PE (Biolegend, clone HP6017, cat. 409304; dilution 1:100), IgA1/2 biotin (BD, clone G20-359, cat. 555884; dilution 1:50), followed by 20 minutes with streptavidin-APC (Biolegend, cat. 405207; dilution 1:100). For isotype controls, mouse IgG1 κ Isotype- Biotin (BD, cat. 550615; dilution 1:50) or mouse IgG2a, κ Isotype- PE (Biolegend, cat. 400212; dilution 1:100) were used. The bacteria were washed before and after DNA was stained with Hoechst33342 (Sigma-Aldrich). The stained bacteria were sorted as Hoechst⁺ and IgG⁺IgA⁺ or IgG⁻IgA⁺ into PBS using the BD Influx and then stored at -80°C until DNA extraction.

Genomic DNA Extraction for shotgun sequencing of IgA-bound bacteria

Bacteria in PBS was defrosted on ice before being pelleted at 3900 rpm for 5 minutes at 4°C. The supernatant was removed and the pellet resuspended in 2 ml of 25% sucrose in TE Buffer (10mM Tris pH8 and 1mM EDTA pH8). 50 μ l of 100 mg/ml lysozyme in 0.25M Tris was added and incubated at 37°C for 1 hour. 100ul of Proteinase K (18 mg/ml), 15 μ l of

RNAase A (20 mg/ml), 400 µl of 0.5 M EDTA (pH 8), and 250 µl of 10% Sarkosyl were then added. This was left on ice for 2 hours and then incubated at 50°C. The DNA was then mixed with 5 ml of Tris-EDTA buffer and purified using four cycles of addition of 5 ml Phenol: Chlorophorm: Isoamyl Alcohol in 15 ml MaXtract High Density Phase Lock Gel (PLG) tubes (QIAGEN), centrifuging at 2800 rpm for 5 min and retrieving the aqueous phase into a new PLG tube. DNA in the final aqueous phase was precipitated with 10 ml of 100% ethanol at -20 °C overnight, centrifuged at 3900 rpm for 20 min at 4 °C and washed with 10 ml of 70% ethanol, before being centrifuged at 4500 rpm for 10 min at 4 °C and gently dried at 50 °C overnight. DNA was then resuspended in 200 µl of Tris-EDTA buffer and isolated by running through a gel. Samples were quantitated by qbit and pooled. Pooled DNA was sequenced on the Illumina Hiseq2500 platform. Inverse Simpson (diversity) and chao (richness) of IgA-opsionized bacteria was determined by using R package “microbiome”.

Statistical analysis

Sample sizes for each experimental group and statistical tests used are included in the relevant figure legends. Measurements were taken from distinct donors in all experiments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors gratefully acknowledge the support received from the Wellcome Sanger Institute Flow Cytometry Facility, Sanger Cellular Genetics Department IT team and Core Sanger Sequencing pipeline. We thank Dr F. Rouhani and T. Castro-Dopico for insightful discussions about project design and experimental design and Dr K. Mahubani for help with collection of human tissue. We acknowledge J. Eliasova for graphical images. This research was supported by funding from Wellcome (grant no. WT206194; S. A. T.) and European Research Council (grant no. 646794 ThDefine; S. A. T.). K. R. J. receives financial support from Christ’s College, University of Cambridge. T. G. was funded by the European Union’s H2020 research and innovation programme “ENLIGHT-TEN” under the Marie Skłodowska-Curie grant agreement 675395. H. W. K. was funded by a Sir Henry Wellcome PostDoctoral Fellowship (grant no. 213555/Z/18/Z). B. R. B. was funded by the NIHR Cambridge Biomedical Research Centre (grant no. RG90712). We are grateful to the deceased organ donors, donor families and the Cambridge Biorepository for Translational Medicine for access to the tissue samples. This publication is part of the Human Cell Atlas-www.humancellatlas.org/publications.

References

1. Halfvarson J, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol.* 2017; 2
2. Donaldson GP, Lee S Melanie, Mazmanian SK. Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology.* 2016; 14:20–32. [PubMed: 26499895]
3. Wang X, Heazlewood SP, Krause DO, Florin THJ. Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J Appl Microbiol.* 2003; 95:508–520. [PubMed: 12911699]
4. Mowat AM, Agace WW. Regional specialization within the intestinal immune system. *Nat Rev Immunol.* 2014; 14:667–685. [PubMed: 25234148]
5. Denning TL, et al. Functional Specializations of Intestinal Dendritic Cell and Macrophage Subsets That Control Th17 and Regulatory T Cell Responses Are Dependent on the T Cell/APC Ratio, Source of Mouse Strain, and Regional Localization. *The Journal of Immunology.* 2011; 187:733–747. [PubMed: 21666057]

6. Atarashi K, et al. Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells. *Cell*. 2015; 163:367–380. [PubMed: 26411289]
7. Ivanov II, et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*. 2009; 139:485–498. [PubMed: 19836068]
8. Atarashi K, et al. Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science*. 2011; 331:337–341. [PubMed: 21205640]
9. Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*. 2005; 122:107–118. [PubMed: 16009137]
10. Atarashi K, et al. Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science*. 2017; 358:359–365. [PubMed: 29051379]
11. Fanning S, et al. Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc Natl Acad Sci U S A*. 2012; 109:2108–2113. [PubMed: 22308390]
12. Consortium, T.H.M.P. & The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012; 486:207–214. [PubMed: 22699609]
13. Wu GD, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science*. 2011; 334:105–108. [PubMed: 21885731]
14. Jones RB, et al. Inter-niche and inter-individual variation in gut microbial community assessment using stool, rectal swab, and mucosal samples. *Sci Rep*. 2018; 8
15. Smillie CS, et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell*. 2019; 178
16. Chakarov S, et al. Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches. *Science*. 2019; 363:714–730.
17. Kumar BV, et al. Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep*. 2017; 20:2921–2934. [PubMed: 28930685]
18. Cook DN, et al. CCR6 mediates dendritic cell localization, lymphocyte homeostasis, and immune responses in mucosal tissue. *Immunity*. 2000; 12:495–503. [PubMed: 10843382]
19. Sebzda E, Zou Z, Lee JS, Wang T, Kahn ML. Transcription factor KLF2 regulates the migration of naive T cells by restricting chemokine receptor expression patterns. *Nat Immunol*. 2008; 9:292–300. [PubMed: 18246069]
20. Toribio-Fernández R, et al. Lamin A/C augments Th1 differentiation and response against vaccinia virus and *Leishmania major*. *Cell Death Dis*. 2018; 9:9. [PubMed: 29311549]
21. Lönnberg T, et al. Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria. *Sci Immunol*. 2017; 2
22. Miragaia RJ, et al. Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation. *Immunity*. 2019; doi: 10.1016/j.immuni.2019.01.001
23. Miyao T, et al. Plasticity of Foxp3 T Cells Reflects Promiscuous Foxp3 Expression in Conventional T Cells but Not Reprogramming of Regulatory T Cells. *Immunity*. 2012; 36:262–275. [PubMed: 22326580]
24. Povelori GAM, et al. Human retinoic acid-regulated CD161 regulatory T cells support wound repair in intestinal mucosa. *Nat Immunol*. 2018; 19:1403–1414. [PubMed: 30397350]
25. Schiering C, et al. The alarmin IL-33 promotes regulatory T-cell function in the intestine. *Nature*. 2014; 513:564–568. [PubMed: 25043027]
26. Krzysiek R, et al. Antigen receptor engagement selectively induces macrophage inflammatory protein-1 alpha (MIP-1 alpha) and MIP-1 beta chemokine production in human B cells. *J Immunol*. 1999; 162:4455–4463. [PubMed: 10201982]
27. Mencarelli A, et al. Highly specific blockade of CCR5 inhibits leukocyte trafficking and reduces mucosal inflammation in murine colitis. *Sci Rep*. 2016; 6
28. Zhang Y, et al. Plasma cell output from germinal centers is regulated by signals from Tfh and stromal cells. *J Exp Med*. 2018; 215:1227–1243. [PubMed: 29549115]

29. Caraux A, et al. Circulating human B and plasma cells. Age-associated changes in counts and detailed characterization of circulating normal CD138- and CD138 plasma cells. *Haematologica*. 2010; 95:1016–1020. [PubMed: 20081059]
30. Mora JR, von Andrian UH. Differentiation and homing of IgA-secreting cells. *Mucosal Immunol*. 2008; 1:96. [PubMed: 19079167]
31. Zhang W, et al. Characterization of the B Cell Receptor Repertoire in the Intestinal Mucosa and of Tumor-Infiltrating Lymphocytes in Colorectal Adenoma and Carcinoma. *The Journal of Immunology*. 2017; 198:3719–3728. [PubMed: 28348270]
32. Macpherson AJ. A Primitive T Cell-Independent Mechanism of Intestinal Mucosal IgA Responses to Commensal Bacteria. *Science*. 2000; 288:2222–2226. [PubMed: 10864873]
33. Harbour SN, Maynard CL, Zindl CL, Schoeb TR, Weaver CT. Th17 cells give rise to Th1 cells that are required for the pathogenesis of colitis. *Proc Natl Acad Sci U S A*. 2015; 112:7061–7066. [PubMed: 26038559]
34. Ferreira RC, et al. Cells with Treg-specific FOXP3 demethylation but low CD25 are prevalent in autoimmunity. *J Autoimmun*. 2017; 84:75–86. [PubMed: 28747257]
35. Hoffmann P, et al. Loss of FOXP3 expression in natural human CD4+CD25+ regulatory T cells upon repetitive in vitro stimulation. *Eur J Immunol*. 2009; 39:1088–1097. [PubMed: 19283780]
36. Kyewski, B, Suri-Payer, E. CD4+CD25+ Regulatory T Cells: Origin, Function and Therapeutic Potential. Springer Science & Business Media; 2006.
37. Brandtzaeg P. Function of mucosa-associated lymphoid tissue in antibody formation. *Immunol Invest*. 2010; 39:303–355. [PubMed: 20450282]
38. Meng W, et al. An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol*. 2017; 35:879–884. [PubMed: 28829438]
39. Dunn-Walters DK, Boursier L, Spencer J. Hypermutation, diversity and dissemination of human intestinal lamina propria plasma cells. *Eur J Immunol*. 1997; 27:2959–2964. [PubMed: 9394824]
40. Tsuji M, et al. Requirement for Lymphoid Tissue-Inducer Cells in Isolated Follicle Formation and T Cell-Independent Immunoglobulin A Generation in the Gut. *Immunity*. 2008; 29:261–271. [PubMed: 18656387]
41. Imam T, Park S, Kaplan MH, Olson MR. Effector T Helper Cell Subsets in Inflammatory Bowel Diseases. *Front Immunol*. 2018; 9:1212. [PubMed: 29910812]
42. Castro-Dopico T, et al. Anti-commensal IgG drives intestinal inflammation and type 17 immunity in ulcerative colitis. *Immunity*. 2019; 50:1099–1114. DOI: 10.1016/j.immuni.2019.02.006 [PubMed: 30876876]
43. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. 2018; 19:15. [PubMed: 29409532]
44. Polaski K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park J-E. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics*. 2019; doi: 10.1093/bioinformatics/btz625
45. Stuart T, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019; 177:1888–1902.e21. [PubMed: 31178118]
46. Trapnell C, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*. 2014; 32:381–386.
47. Browne HP, et al. Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature*. 2016; 533:543–546. [PubMed: 27144353]
48. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 2013; 79:5112–5120. [PubMed: 23793624]
49. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One*. 2010; 5:e9490. [PubMed: 20224823]
50. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016; 44:W242–5. [PubMed: 27095192]
51. Stubbington MJT, et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods*. 2016; 13:329–332. [PubMed: 26950746]

52. Petrova VN, et al. Combined Influence of B-Cell Receptor Rearrangement and Somatic Hypermutation on B-Cell Class-Switch Fate in Health and in Chronic Lymphocytic Leukemia. *Front Immunol.* 2018; 9:1784. [PubMed: 30147686]
53. Bolotin DA, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* 2015; 12:380–381. [PubMed: 25924071]
54. Nazarov VI, et al. tcR: an R package for T cell receptor repertoire advanced data analysis. *BMC Bioinformatics.* 2015; 16
55. Heiden JAV, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics.* 2014; 30:1930–1932. [PubMed: 24618469]
56. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Research.* 2013; 41:W34–W40. [PubMed: 23671333]
57. Gupta NT, et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics.* 2015; 31:3356–3358. [PubMed: 26069265]

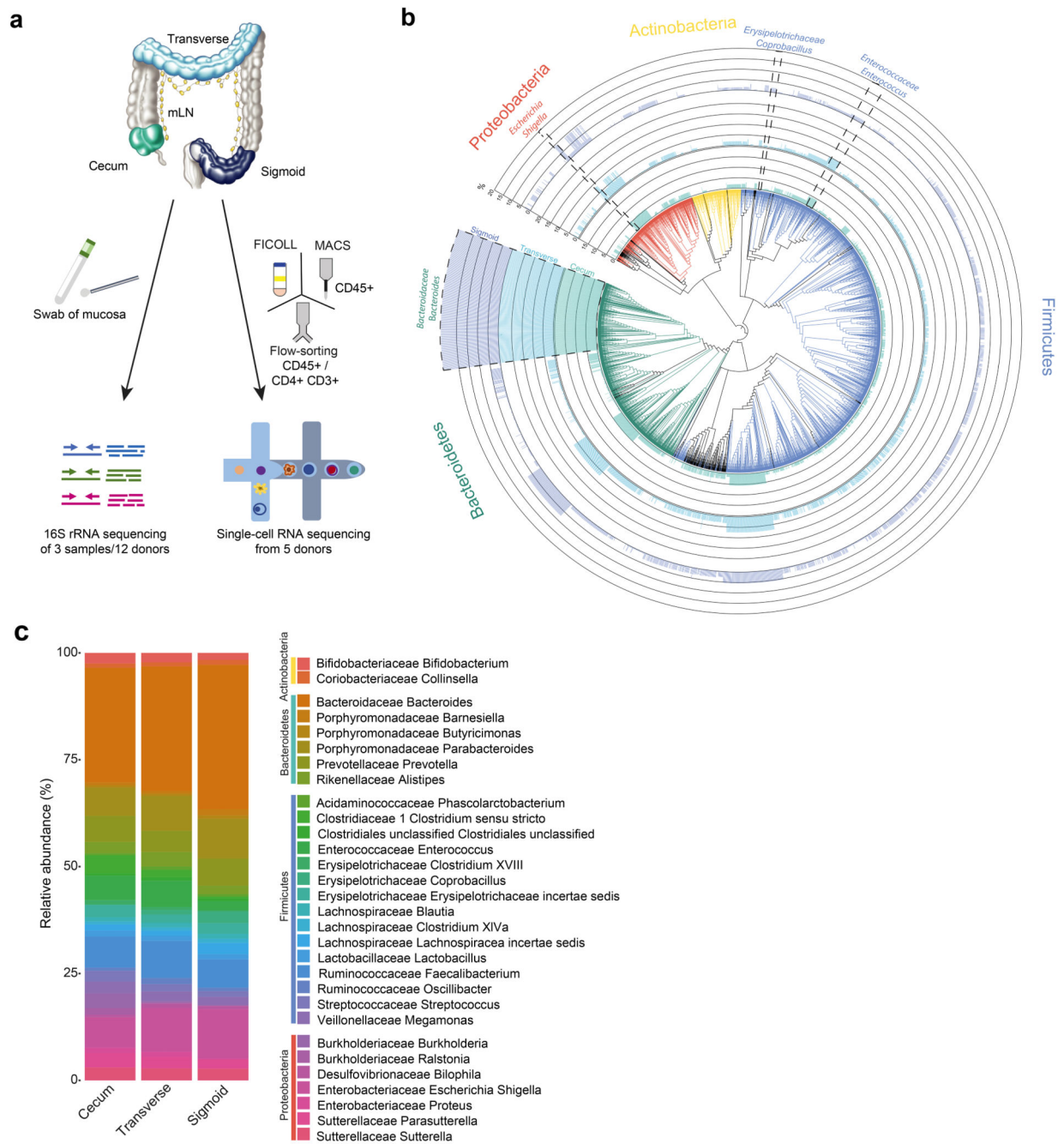


Figure 1. Variation in the microbiome from proximal to distal colon.

a) Workflow for 16S ribosomal sequencing of matching mucosal microbiomes and scRNA-seq profiling of immune cells from mesenteric lymph node (mLN), and lamina propria of cecum, transverse colon and sigmoid colon. b) Phylogenetic tree representing diversity and mean abundance of bacterial species in the cecum, transverse colon and sigmoid colon. Mean abundance was calculated as the percentage of operational taxonomic units (OTUs) for each species from total as determined by 16S rRNA sequencing and averaged for twelve donors (black scale). Unassigned OTUs are shown as black branches. Bacteria groups of

interest are highlighted. c) Relative abundances of OTUs at genus level of bacteria species in colon regions as in (b).

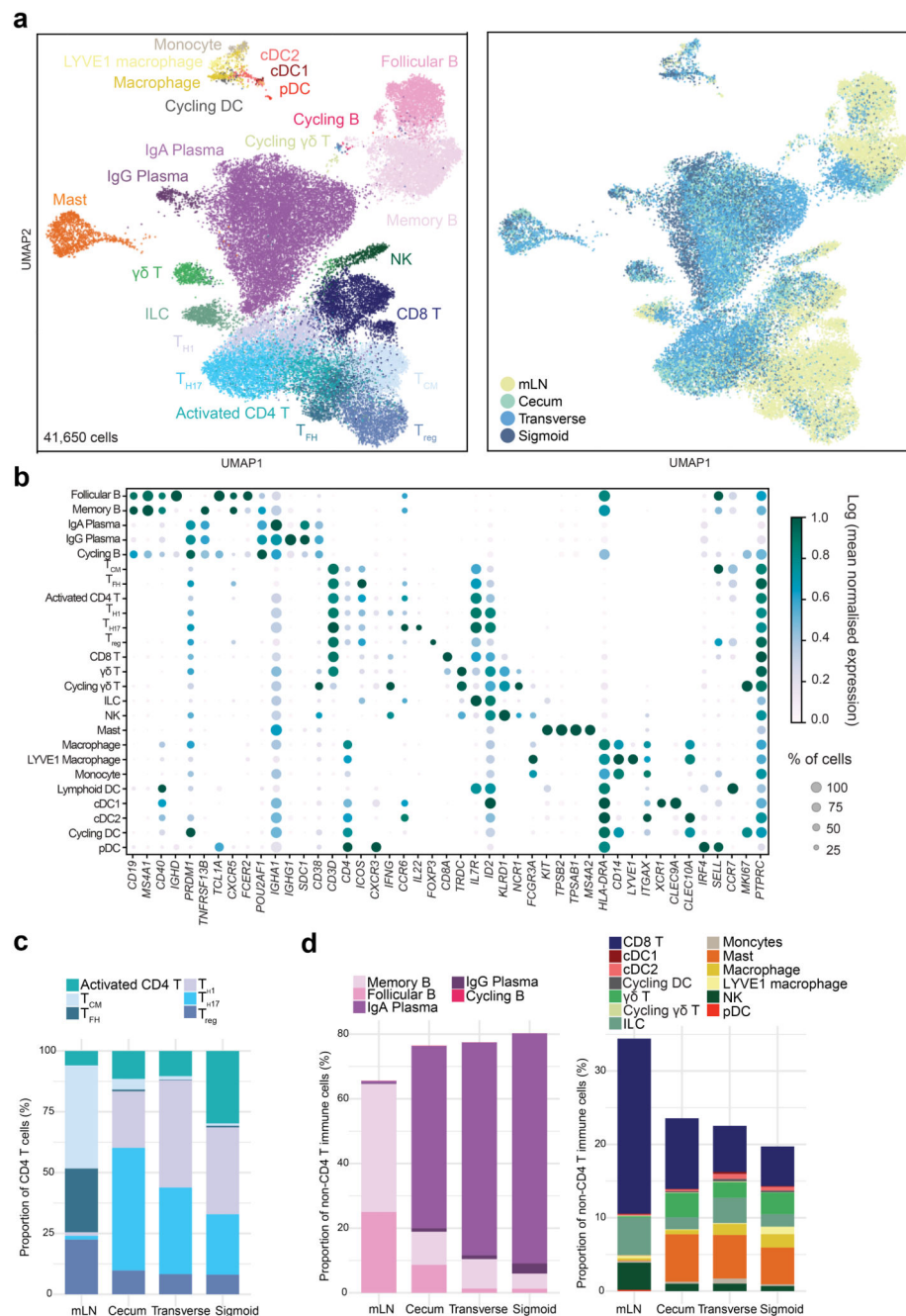


Figure 2. Profiling immune cells along the steady-state colon.

a) UMAP illustration of pooled scRNA-seq data of immune cells of mLN, cecum, transverse colon and sigmoid colon from five donors colored by cell type annotation (left) and tissue of origin (right). b) Heatmap of mean expression of marker genes used to annotate cell types in (a). Point size shows the fraction of cells with nonzero expression. c) Relative percentages of CD4⁺ T subtypes within all CD4⁺ T cells for each tissue as in (a). d) Relative percentages of cell types within all non-CD4⁺ T immune cells for each tissue as in (a), with B lineage cells shown in the left panel and all other cell types in the right panel.

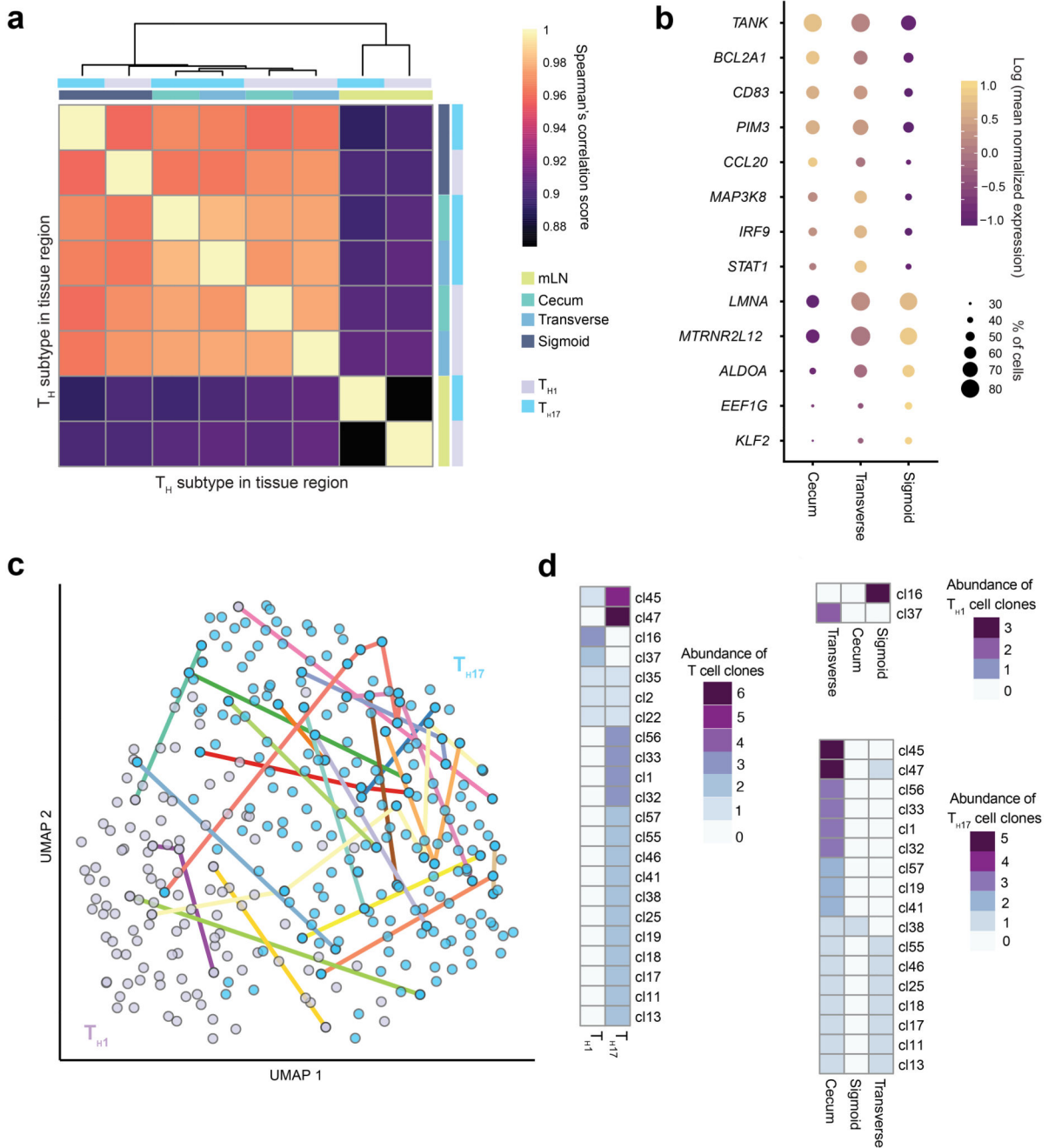


Figure 3. Dissemination of T helper cells in colon and region-determined transcriptional profiles.

a) Correlation matrix of mean transcriptional profiles of T_{H1} and T_{H17} cells from cecum, transverse colon, sigmoid colon and mLN (n=5 donors). b) Mean expression level of differentially expressed genes of pooled T_{H1} and T_{H17} between cecum, transverse colon and sigmoid colon as in (a). Point size shows the fraction of cells with nonzero expression. c) UMAP projection of Smartseq2 profiled flow-sorted T cells annotated as T_{H1} and T_{H17} cells of the cecum, transverse colon and sigmoid colon (n=1 donor). Colored lines connect

cells sharing the same CDR3 sequence. d) Heatmap of numbers of members within clonal families in T_H1/T_H17 subsets (left) and colon region (right) as in (c).

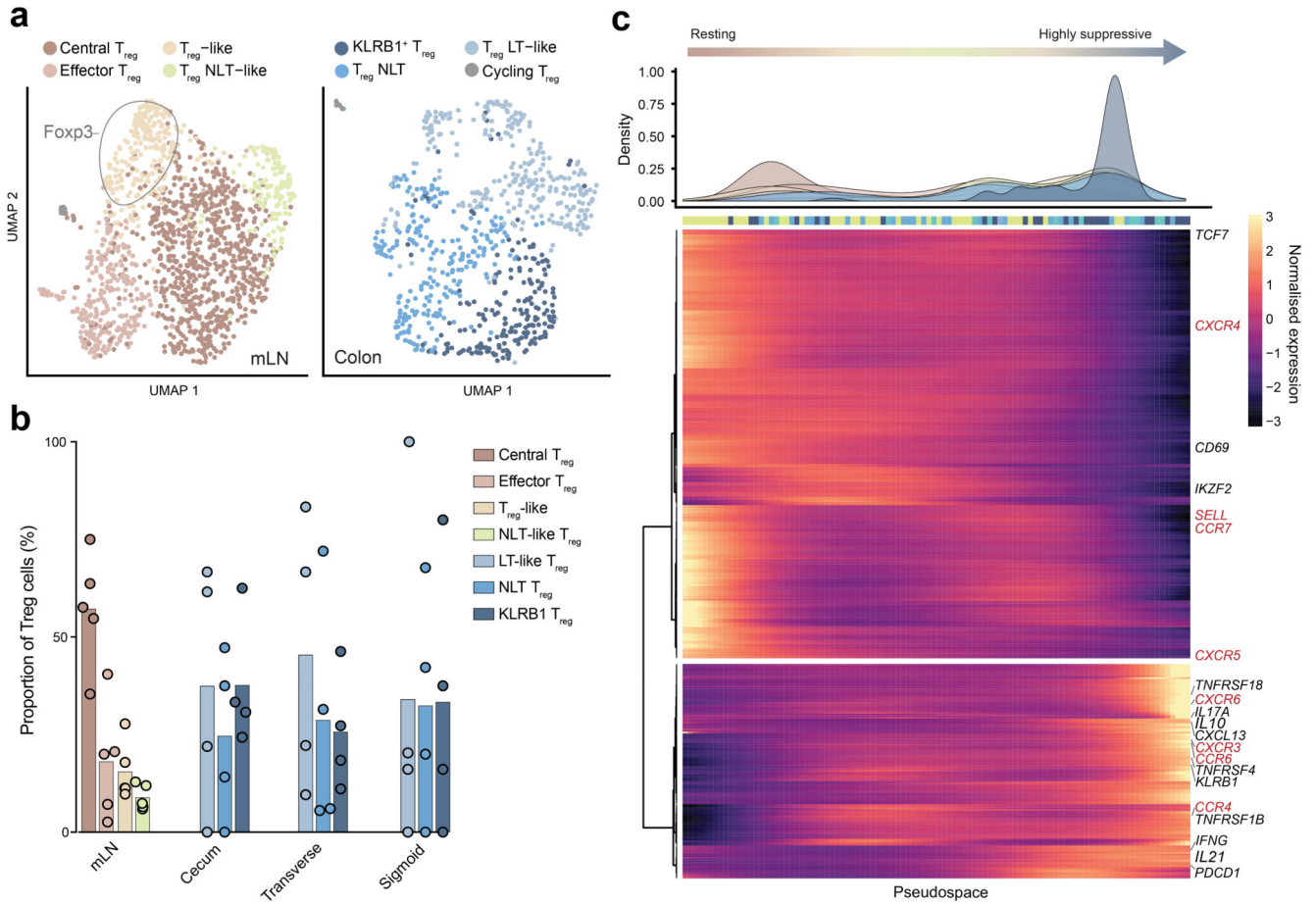


Figure 4. T_{reg} activation pathway from lymphoid to peripheral tissue.

a) UMAP visualization of T_{reg} subtypes in mLN (left) and pooled from cecum, transverse colon and sigmoid colon (right) (n=5 donors). b) Relative proportions of T_{reg} subsets within all T_{reg} cells from mLN and colon tissue regions as in (a). Bars show mean proportion across all donors (circles). c) Density of T_{reg} subclusters as in a across ‘pseudospace’ (top) and expression kinetics of genes contributing to pseudospace smoothed into 100 bins (filtered by $qval < 0.001$ and expression in >15 cells). Top bar shows the most represented tissue within each bin. Various dynamically expressed immune-related molecules are annotated, with key genes colored red.

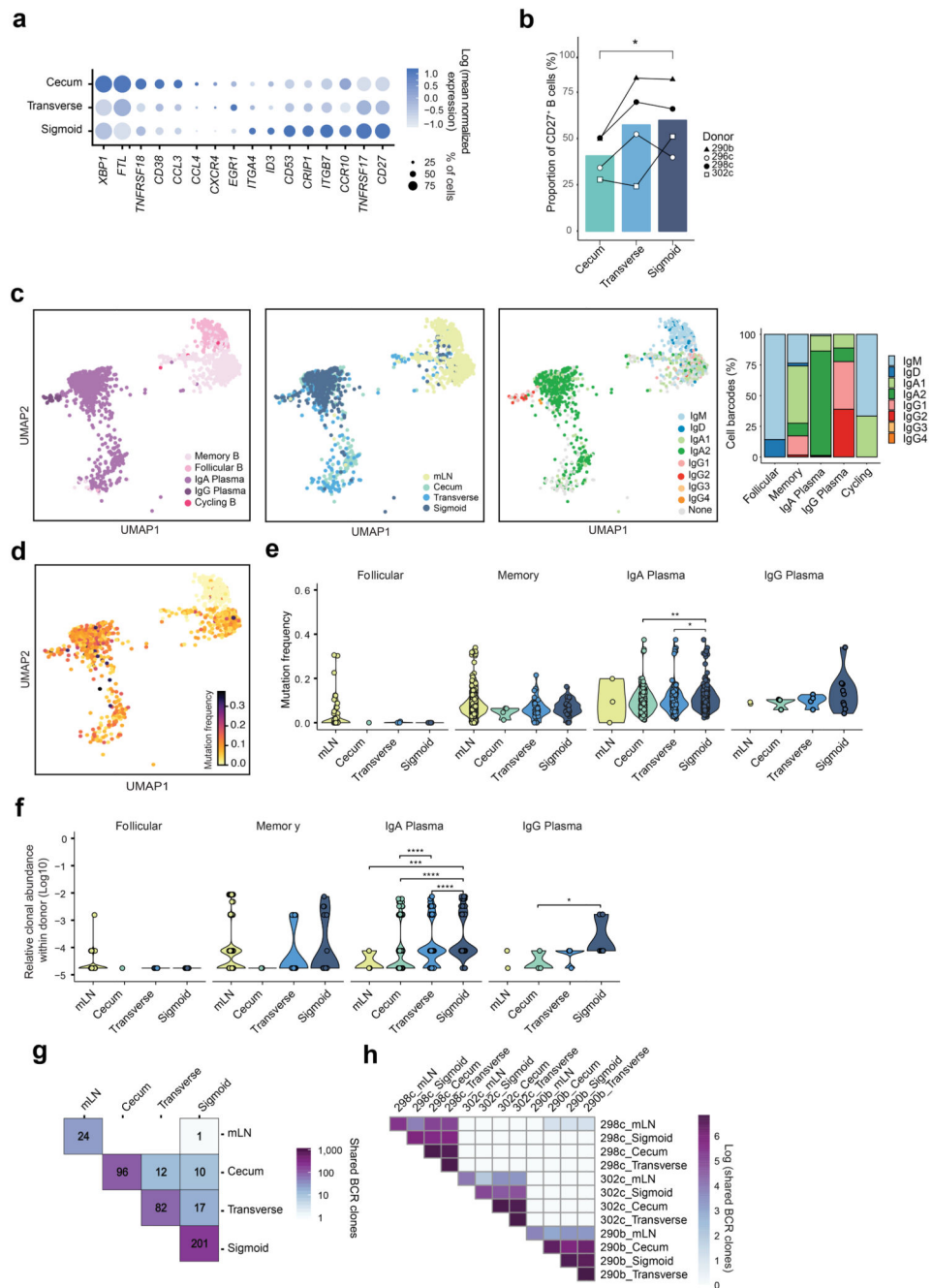


Figure 5. B cells are more abundant, clonally expanded and mutated in the sigmoid colon.
 a) Mean expression of key differentially expressed genes by IgA⁺ plasma cells in cecum, transverse colon and sigmoid colon (n=5 donors). Point size shows the fraction of cells with nonzero expression. b) Proportion of CD27⁺ B cells of total B cells from cecum, transverse colon and sigmoid colon determined by flow cytometry (n=4 donors). Bar represents the mean and connected points represents values of each donor. Analysis is a two-tailed paired t test. c) UMAP visualization of B cells for which matched single-cell VDJ libraries were derived using 10x Genomics 5' scRNA-seq (n=2 donors) colored by cell type annotation

(left), tissue (middle) and antibody isotype (right). Bar plot of antibody isotype frequencies per annotated cell type (far right). d) UMAP visualization of somatic hypermutation frequencies of IgH sequences as in (c). e) Quantitation of somatic hypermutation frequencies of IgH sequences from B cell types and gut regions as in (c). f) Estimated clonal abundances per donor for members of expanded B cell clones in B cell types and gut regions as in (c). g) Binary count of co-occurrence of expanded B cell clones identified by single-cell VDJ analysis shared across gut regions as in (c). h) Co-occurrence of expanded B cell clones identified by bulk B cell receptor sequencing across gut regions (n=3 donors). Statistics in (e) and (f) are calculated with two-sided Wilcoxon signed-rank tests. Rows and columns in (g) and (H) are ordered by hierarchical clustering. * P <0.05; ** P <0.01; *** P <0.001; **** P <0.0001

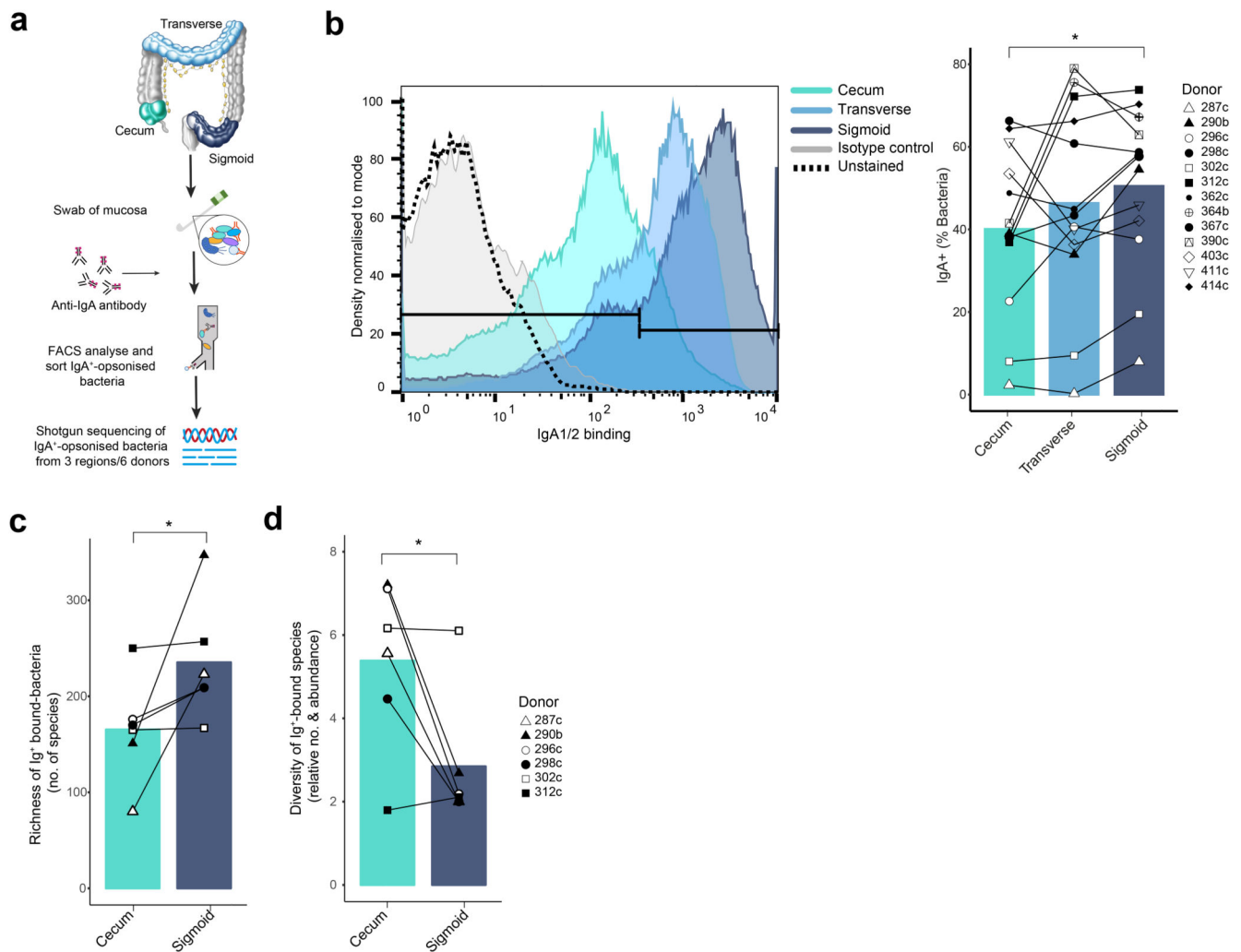


Figure 6. Increasing number of microbiota species recognized by antibodies in the sigmoid colon. a) Experimental workflow for assessing Ig-opsonized colon bacterial species. b) Representative histogram of IgA1/2-bound Hoechst⁺ bacteria and summary plot of bound bacteria as a proportion of total bacteria (n=13 donors). Positive binding is set against an isotype control. c) Richness of bacteria species determined as the number of unique species and d) diversity of species identified from shotgun sequencing of Ig-opsonized bacteria from (b) (n=6 donors). P values were calculated using one-tailed paired t tests. * P < 0.05