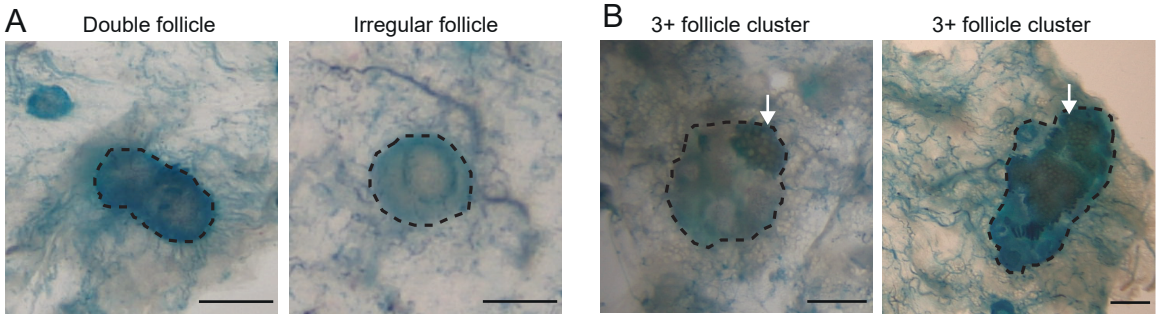
Figure S1



C

Site	Double/irregular follicles	3+ follicles
Cecum	11 of 13 (84.6%)	3 of 13 (23%)
Ascending colon	15 of 19 (78.9%)	3 of 19 (15.8%)
Transverse colon	13 of 20 (65%)	3 of 20 (15%)
Sigmoid colon	9 of 16 (56.3%)	0 of 16 (0%)
Rectum	0 of 3 (0%)	0 of 3 (0%)

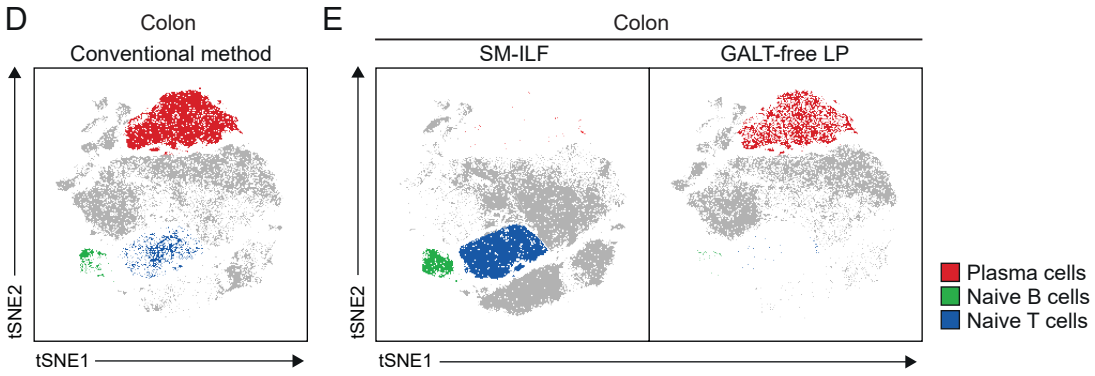


Figure S1. Related to Figure 1.

(A and B) GALT in the submucosa were occasionally found to consist of (A) two follicles or have irregular shapes or (B) clusters of three or more follicles. Scale bars represent 1 mm. (C) Proportion of intestinal site that contained irregular or multi-follicular SM-GALT. (D) Representative tSNE plot derived from flow cytometric analysis of $CD45^+$ cells isolated from digested colon mucosa with attached submucosa using conventional protocols. Surface markers used for tSNE: CD45, CD3, CD8, CD20, CD38, CD138, CD45RA, TCRαβ, IgD, TCRαβ, CD117, CCR6, IgA. (E) Representative tSNE plots derived from flow cytometric analysis of $CD45^+$ cells isolated from colon SM-ILF and GALT-free colon LP. GALT, gut-associated lymphoid tissue; SM-ILF, submucosal follicle; LP, lamina propria.

Figure S2

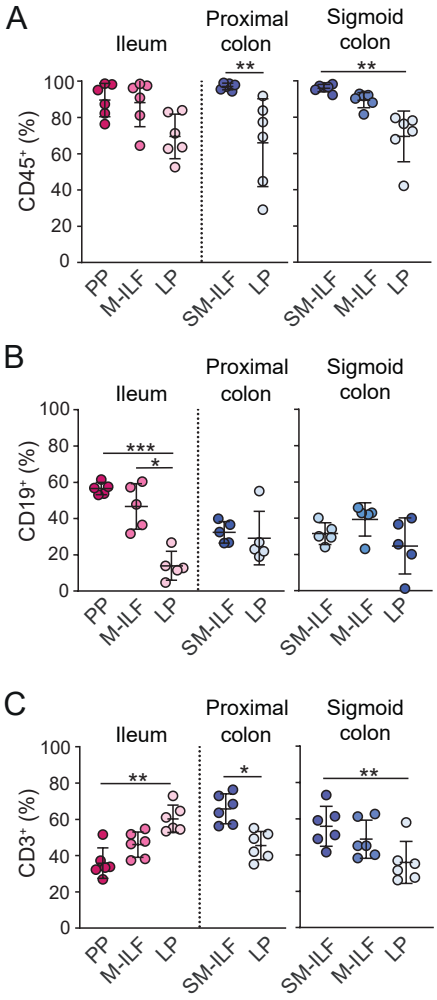


Figure S2. Related to Figure 2.

(A-C) Flow cytometric analysis of indicated immune niches. (A) Proportion of viable cells expressing CD45, and proportion of $CD45^+$ cells expressing (B) CD19 and (C) CD3. Each circle represents an individual donor with paired samples from 4 patients. SM-ILF and M-ILF data are from at least 5 pooled follicles. Bars represent mean ± 1SD and statistical significance between paired samples was determined using the paired Friedman test with Dunn's multiple comparisons, *$p<0.05$, **$p<0.01$, ***$p<0.001$. PP, Peyer's patch; SM-ILF, submucosal follicle; M-ILF, mucosal follicle; LP, lamina propria.

Figure S3

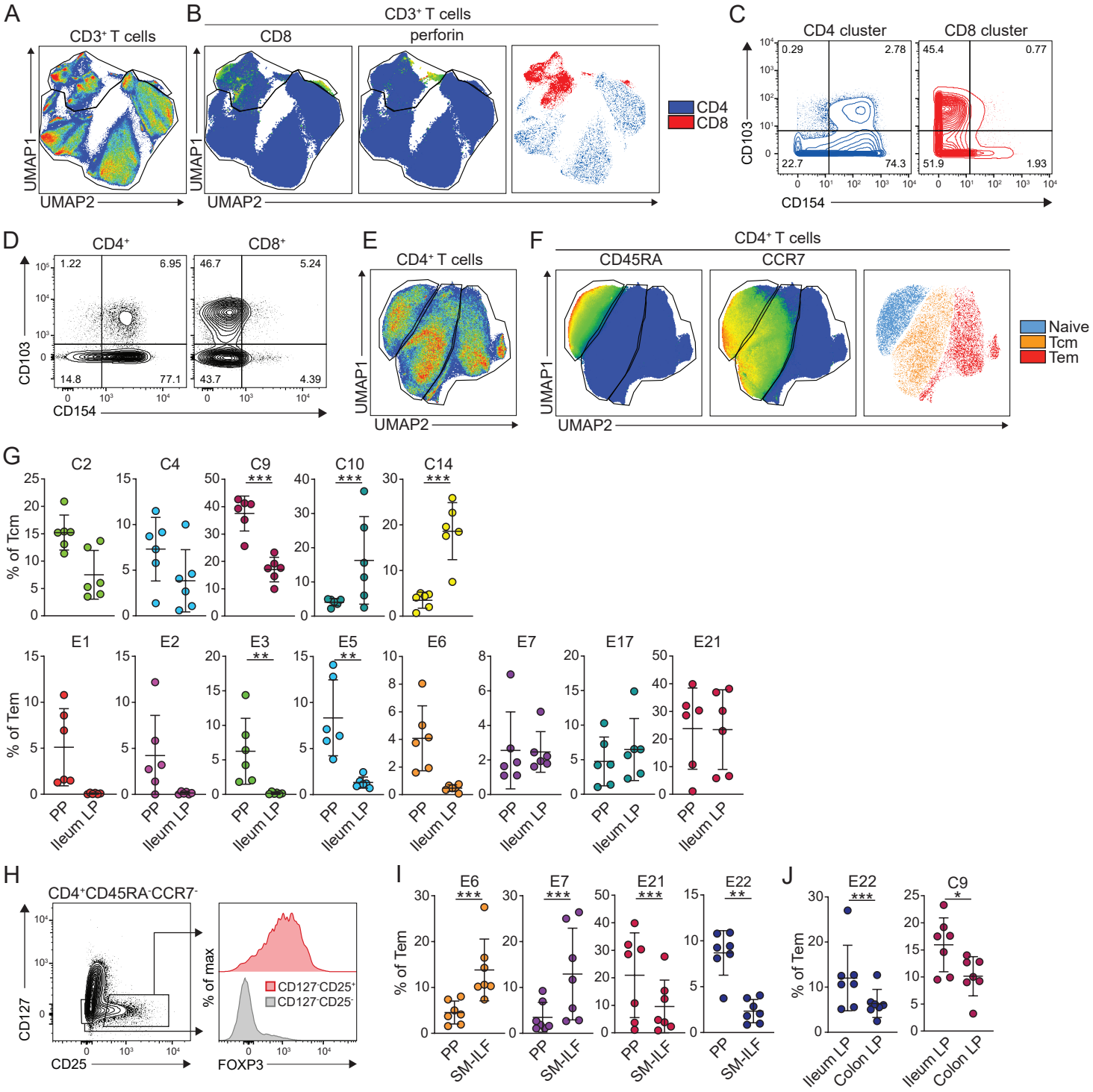

A
CD3⁺ T cells
UMAP1
UMAP2
B
CD3⁺ T cells
CD8
perforin
CD4
CD8
C
CD4 cluster
CD8 cluster
0.29
2.78
22.7
74.3
45.4
0.77
51.9
1.93
CD103
CD154
D
CD4⁺
CD8⁺
1.22
6.95
14.8
77.1
46.7
5.24
43.7
4.39
CD103
CD154
E
CD4⁺ T cells
UMAP1
UMAP2
F
CD4⁺ T cells
CD45RA
CCR7
UMAP1
UMAP2
Naive
Tcm
Tem
G
C2
C4
C9
C10
C14
% of Tcm
E1
E2
E3
E5
E6
E7
E17
E21
% of Tem
PP
Ileum LP
H
CD4⁺CD45RA⁻CCR7⁻
CD127
CD25
% of max
CD127⁻CD25⁺
CD127⁻CD25⁻
FOXP3
I
E6
E7
E21
E22
% of Tem
PP
SM-ILF
J
E22
C9
% of Tem
Ileum LP
Colon LP

Figure S3. Related to Figure 3.

(A-H) Single-cell suspensions of PP, proximal colon SM-ILF and ileum and colon LP were stimulated with PMA and ionomycin and the T cell populations were analyzed by CyTOF. (A) Pseudocolour density plot of total T cell UMAP used to distinguish $CD4^+$ from $CD8^+$ T cells. (B) CD8 and perforin expression, and putative CD4 (blue) and CD8 (red) gates overlaid on the total T cell UMAP. (C) CD103 and CD154 expression overlaid on the $CD4^+$ (blue) and $CD8^+$ (red) T cell clusters defined in (B). (D) CD103 and CD154 expression by $CD4^+$ and $CD8^+$ T cells in colon LP as assessed by flow cytometry. (E) Pseudocolour density plot of $CD4^+$ T cell UMAP (see STAR methods for details) including manual gates demarcating $CD4^+$ naïve, Tcm, and Tem. (F) CD45RA and CCR7 expression overlaid on $CD4^+$ T cell UMAP plot. (G) Frequency of memory $CD4^+$ T cell clusters in PP and ileum LP. Each circle represents a paired sample from 6 individual patients. Bars represent mean ± 1SD. (H) FOXP3 expression by $CD25^+CD127^-CD4^+$ Tem in SM-ILF. Results are from 1 representative patient of 3, with at least 10 pooled SM-ILF per patient. (I and J) Frequency of memory $CD4^+$ T cell clusters in (I) Peyer's patch and SM-ILF and (J) ileum LP and colon LP. Each circle represents a paired sample from 7 individual patients. Bars represent mean ± 1SD. PP, Peyer's patch; SM-ILF, submucosal isolated lymphoid follicle; LP, lamina propria. Statistical significance was determined using paired 2-way ANOVA, false discovery rate was corrected with the two-stage step-up method of Benjamini, Krieger and Yekutieli, *$p<0.05$, **$p<0.01$, ***$p<0.001$. See also Figure S3.

Figure S4

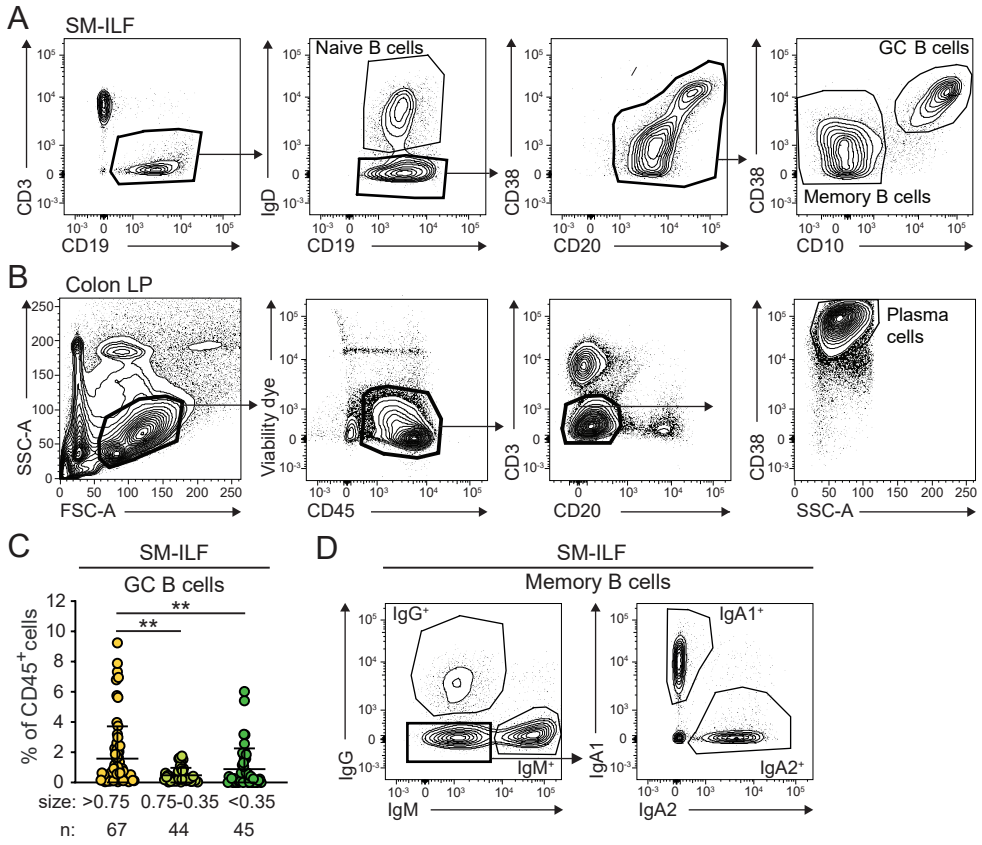


Figure S4. Related to Figure 4.

(A-B) Gating strategy to identify naïve, memory and GC B cells and $CD45^+$ PC by flow cytometry in (A) colon SM-ILF and (B) colon LP. (C) Proportion of $CD10^+CD20^{hi}$ GC B cells amongst total $CD45^+$ cells in colon SM-ILF of the indicated size. Each circle represents a single SM-ILF. Samples are from 6 proximal colon and 5 distal colon resections. Bars represent mean ± 1SD. (D) Gating strategy to identify IgM^+, IgG^+, $IgA1^+$ and $IgA2^+$ memory and GC B cells by flow cytometry. Statistical significance between samples was determined using the Kruskal-Wallis test with Dunn's multiple comparisons, **p<0.01. GC, germinal center; SM-ILF, submucosal follicle; LP, lamina propria.

Figure S5

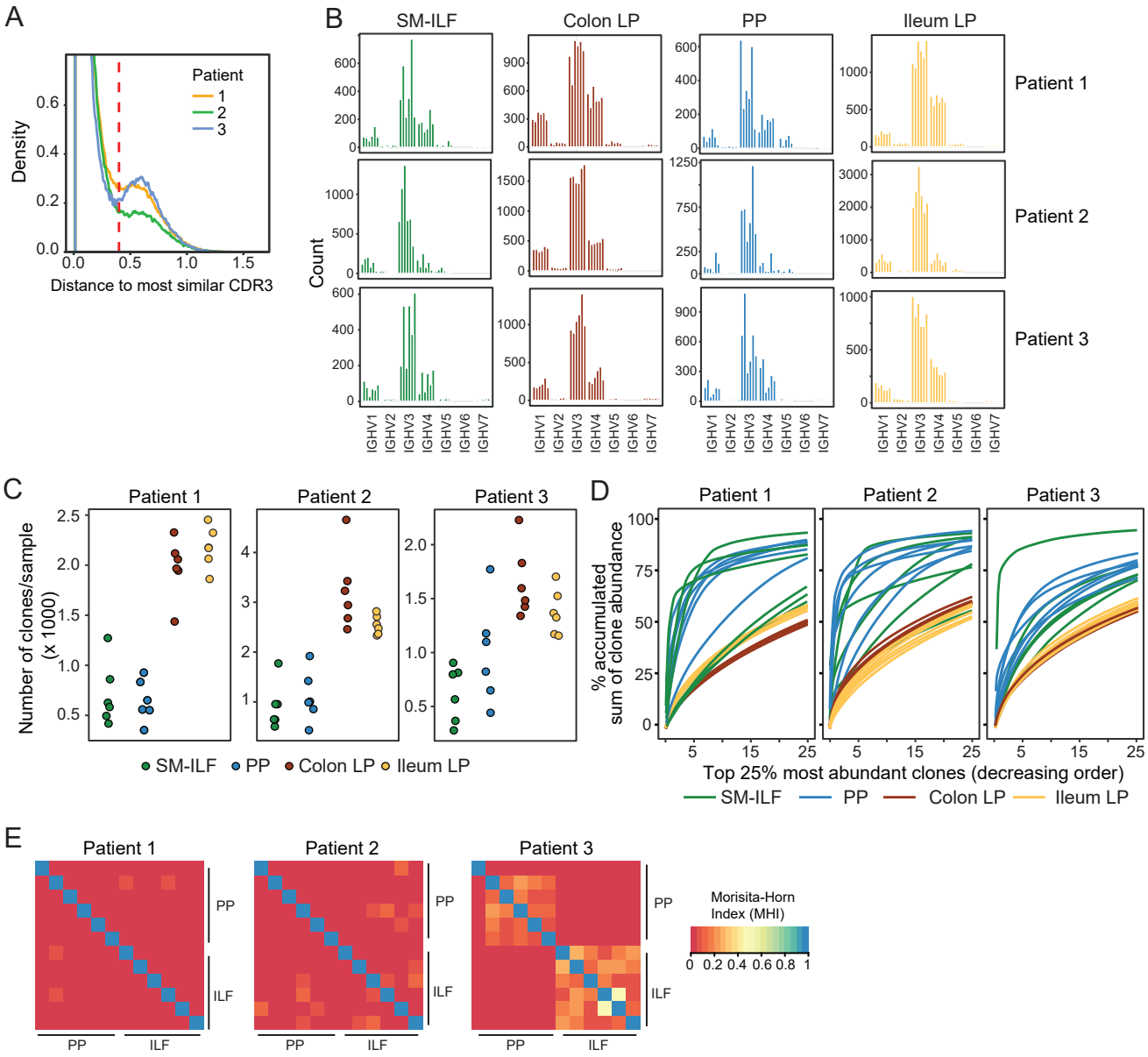


Figure S5. Related to Figure 5.

(A) Distribution of the sequence difference, as given by the SF5 substitution matrix, of each CDR3 to the most similar other CDR3 in the IgA repertoire that have the same V and J-gene combination. (B) Distribution of V gene usage in indicated tissues and patients. (C and D) The IgA repertoire in isolated ileum and proximal colon GALT and LP was determined in 3 individual donors. (C) Total number of clones with ≥ 10 reads in the indicated tissue in each donor and (D) the cumulative abundance of the 25% most abundant clones within each sample. (E) Morisita-Horn similarity index (MHI) between PP and ILF (SM-ILF) samples taking into account only the top 100 clones in each compartment. PP, Peyer's patch; SM-ILF, submucosal follicle; LP, lamina propria.

Table S1

	Total	GALT count	B cells	T cells	CyTOF	IgA	FOXP3	B cell Migration	IHC
Number of patients	99	68	24	12	11	3	3	4	25
Age	70.7 ± 9.6	70.7 ± 9.4	72.8 ± 7.3	67.5 ± 11.7	69.8 ± 13.1	73 ± 13.2	70.7 ± 2.1	64.8 ± 5.7	77.7 ± 4.7
Sex									
Female	51	37	9	6	4	2	1	1	20
Male	48	31	15	6	7	1	2	3	5
Site									
Terminal ileum	54	14	18[c]	6[c]	8[c]	3[c]	-	4[c]	15
Proximal colon	73[a]	49[a]	18[c]	6[c]	8[c]	3[c]	1	4[c]	14
Sigmoid colon/ rectum	26[b]	20[b]	6	6	3	-	2	-	4
Proximal colon site									
Cecum	16	13	5	4	2	1	-	-	4
Ascending	25	19	7	1	2	-	-	2	6
Transverse	36	20	6	1	2	2	1	2	4
Combined proximal	2	-	-	-	2	-	-	-	-
Diagnosis									
Volvulus	4	2	1	1	2	-	-	-	-
Diverticulitis	3	2	1	2	-	-	-	-	-
Stenosis	2	1	-		-	-	1	-	-
Adenocarcinoma	90	63	22	9	9	3	2	4	25
UICC stage									
0	15	11	4	-	1	1	-	-	5
1	14	12	3	1	1	-	-	1	2
2	33	20	7	5	6	2	-	1	9
3	22	16	8	1	-	-	1	2	7
4	6	4	-	2	1	-	1	-	2

[a] Samples taken from multiple proximal sites per patient

[b] One patient had samples taken from both sigmoid colon and rectum

[c] Samples taken from both ileum and colon per patient allowing paired data analysis

Table S1. Related to experimental model and subject details section of STAR methods

Table S2

Gene name	Primers name	Forward	Reverse
Actin γ 1	ACTG1	CCGAGCCGTGTTTCCTTCC	GCCATGCTCAATGGGGTACT
Ribosomal protein S27	RPS27	ATGCCTCTCGCAAAGGATCTC	TGAAGTAGGAATTGGGGCTCT
Activation Induced Cytidine Deaminase	AICDA.1	GAGGCGTGACAGTGCTACATC	CAGGGTCTAGGTCCCAGTCC
Activation Induced Cytidine Deaminase	AICDA.2	AAGGGCTGCATGAAAATTCAGT	AAGGGCTGCATGAAAATTCAGT
Syndecan-1 (CD138)	SDC1.1	CTGCCGCAAATTGTGGCTAC	TGAGCCGGAGAAGTTGTCAGA
Syndecan-1 (CD138)	SDC1.2	TTGTGGCTACTAATTTGCCC	TATCTTGCAAAGCACCTGC

Table S2. Related to Figure 4.