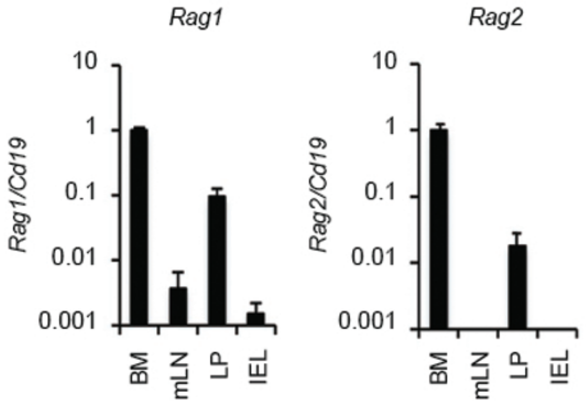
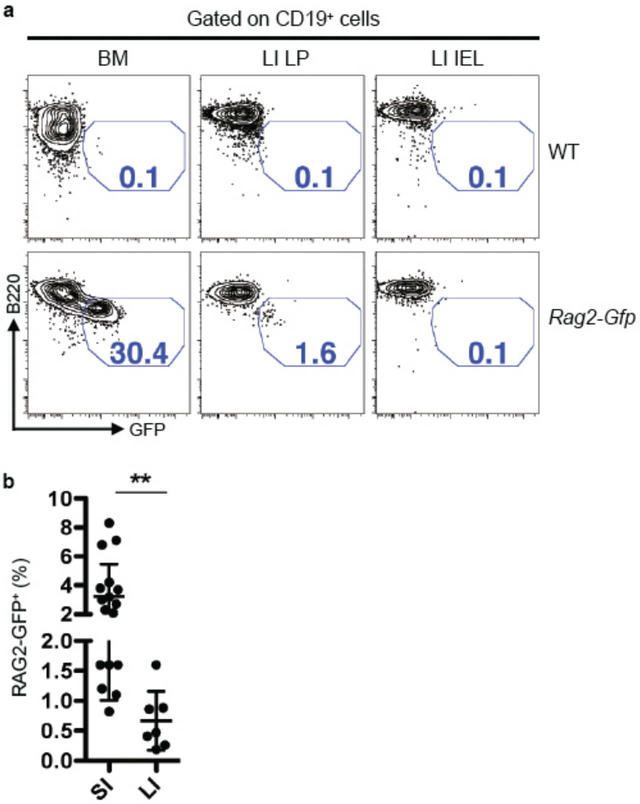
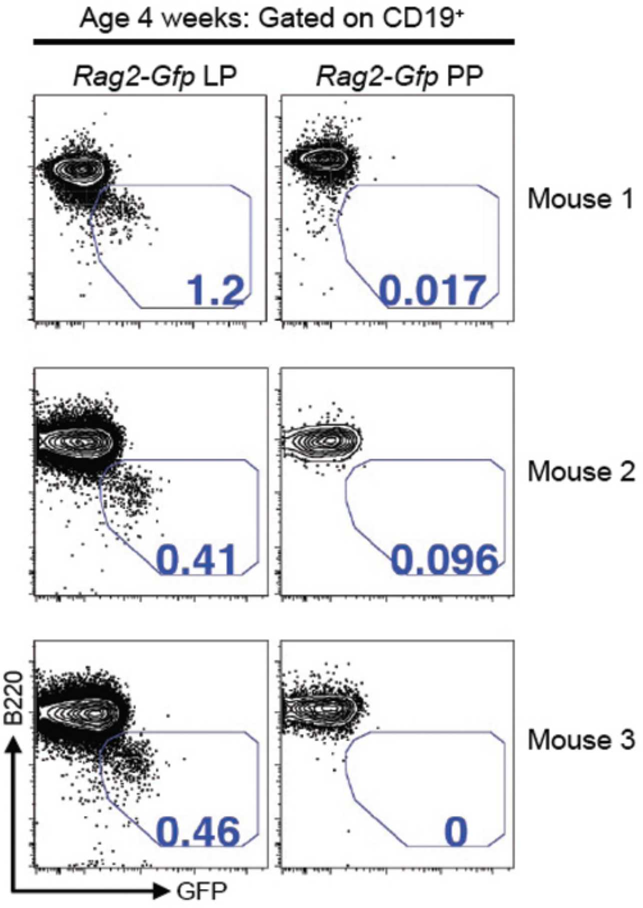
doi:10.1038/nature12496



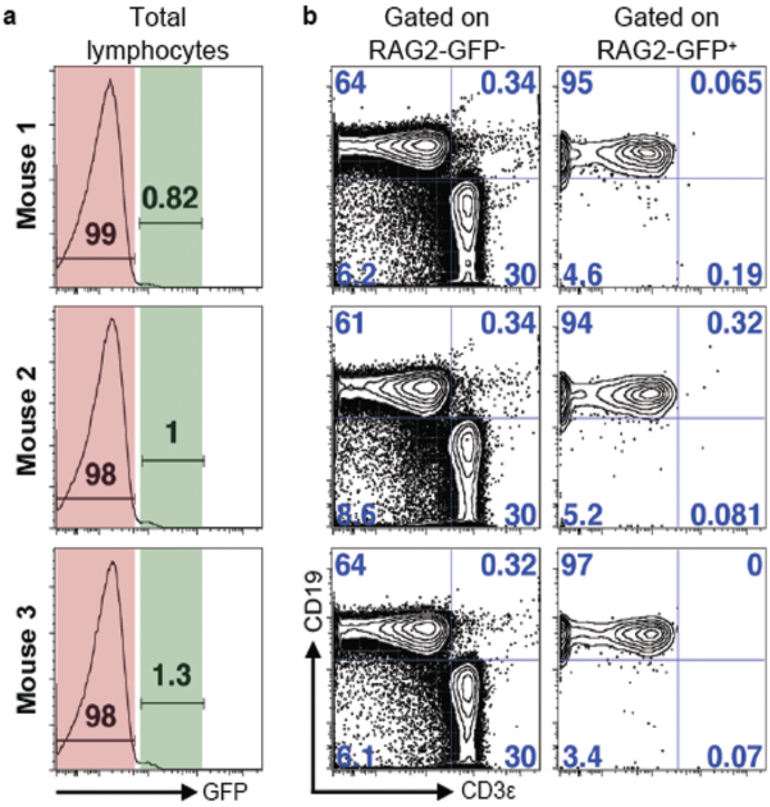
Supplementary Figure 1 | *Rag1* and *Rag2* are expressed in the small intestinal lamina propria of young wild type mice. Quantitative PCR analysis of *Rag1* and *Rag2* expression in the bone marrow (BM), mesenteric lymph nodes (mLN), small intestinal lamina propria (LP), and intraepithelial lymphocytes (IEL) of 3 wk-old wild type Balb/c mice. Relative *Rag1* and *Rag2* expression leves of each tissue sample were normalized to *Cd19* expression. The *y*-axis indicates expression levels relative to BM. Shown are mean values ± s.e.m of three independent experiments.



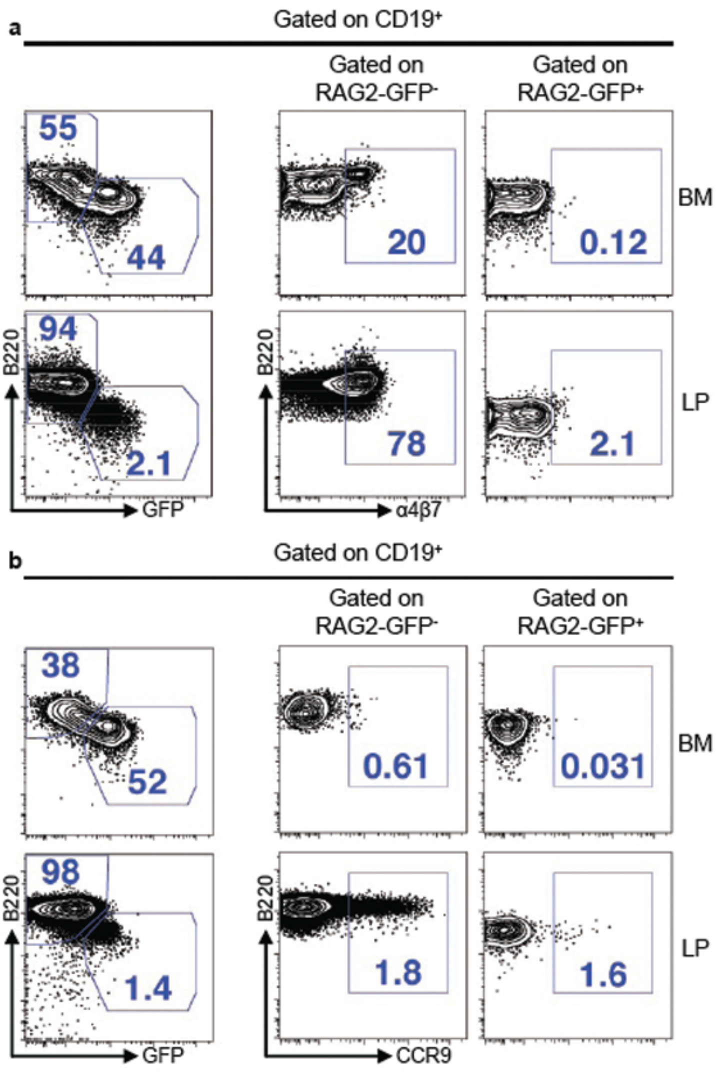
Supplementary Figure 2 | RAG2-GFP$^+$ B cells are found in the large intestine at lower levels compared with the small intestine. a, FACS plots of CD19$^+$ gated cells from bone marrow (BM), large intestinal intraepithelial lymphocytes (LI IEL) and large intestinal lamina propria (LI LP) from wild type (WT) (top plots) or homozygous *Rag2-Gfp* knock-in (bottom plots) mice at post-natal day 18. Plots show B220 expression against GFP fluorescence. Numbers in the plots denote percentage of CD19$^+$ cells that are B220^{low} RAG2-GFP$^+$. Wild type (WT) BM was analyzed as a control to measure background autofluorescence. **b**, Dot plot of cumulative data demonstrating the percentage of RAG2-GFP$^+$ among CD19$^+$ cells in the LP of small intestine (SI) and large intestine (LI). The post-natal age 18-21 RAG2-GFP$^+$ SI LP data from from Figure 1b is plotted here as well for comparison. Each point represents one mouse. Shown are mean values ± s.e.m. ***P*<0.01 (two-tailed Student's *t*-test).



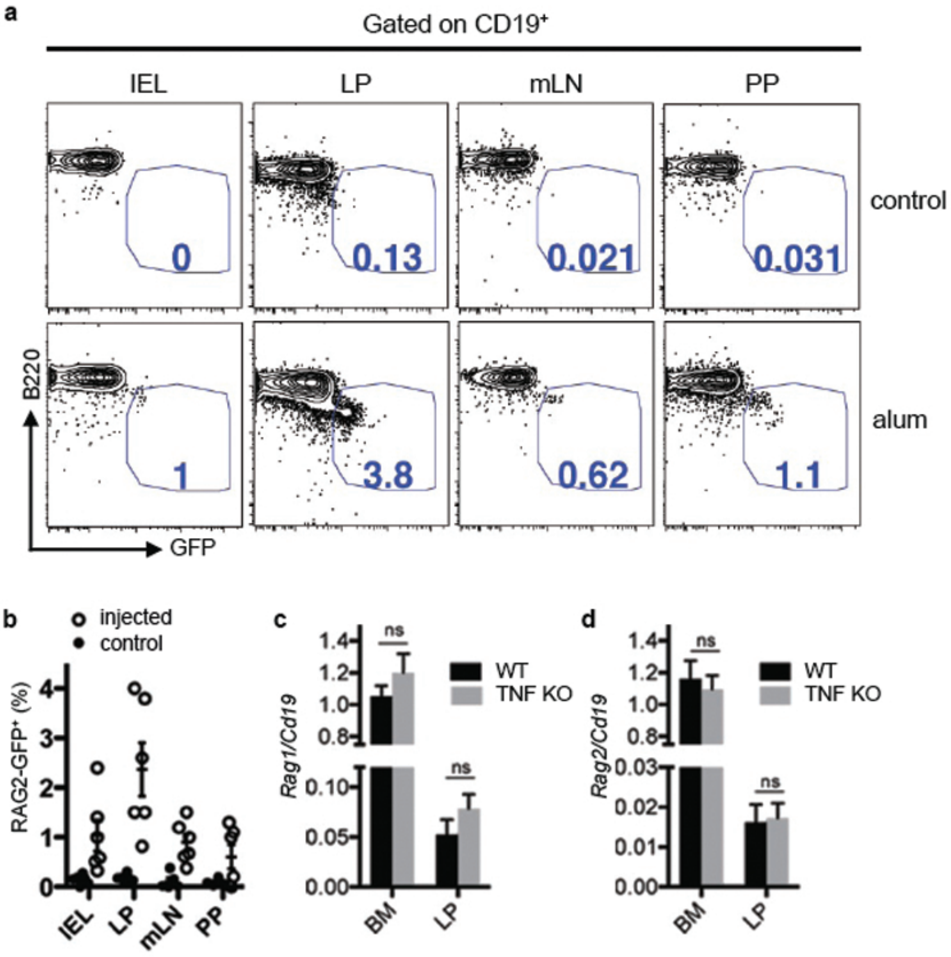
Supplementary Figure 3 | Peyer's patches do not harbor RAG2-GFP$^+$ cells. FACS plots of CD19$^+$ gated cells from the lamina propria (LP) and Peyer's patches (PP) isolated from three independent *Rag2-Gfp* mice on post-natal day 28. The numbers in the gates shown indicate percentage of B220^{low} RAG2-GFP$^+$ cells out of total CD19$^+$ cells. These data indicate that RAG2-GFP$^+$ cells are found in the LP, but not in the PP.



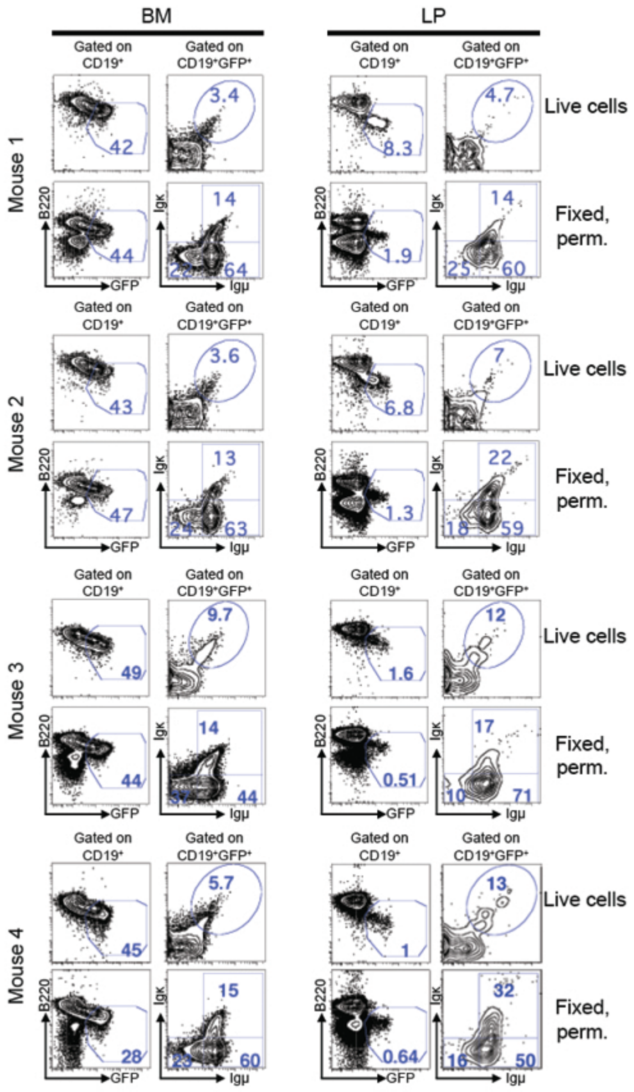
Supplementary Figure 4 | RAG2-GFP is not detected in LP T lineage cells. a, Histogram plots showing flow cytometric measurement of RAG2-GFP$^-$ (red) and RAG2-GFP$^+$ (green) cell populations from total small intestinal LP lymphocytes at post-natal day 18-21. Cells were stained for CD19 and the pan T cell lineage marker, CD3ε. **b**, The RAG2-GFP$^-$ and RAG2-GFP$^+$ gated cells were plotted to reveal the B (CD19$^+$ CD3ε$^-$, upper left quadrant) and T (CD19$^-$ CD3ε$^+$, lower right quadrant) lineage populations present in each gate. These data show that the great majority of RAG2-GFP$^+$ cells are contained within the *Cd19*-expressing population and are not found in the *CD3ε*-expressing cells. Plots from three mice are shown.



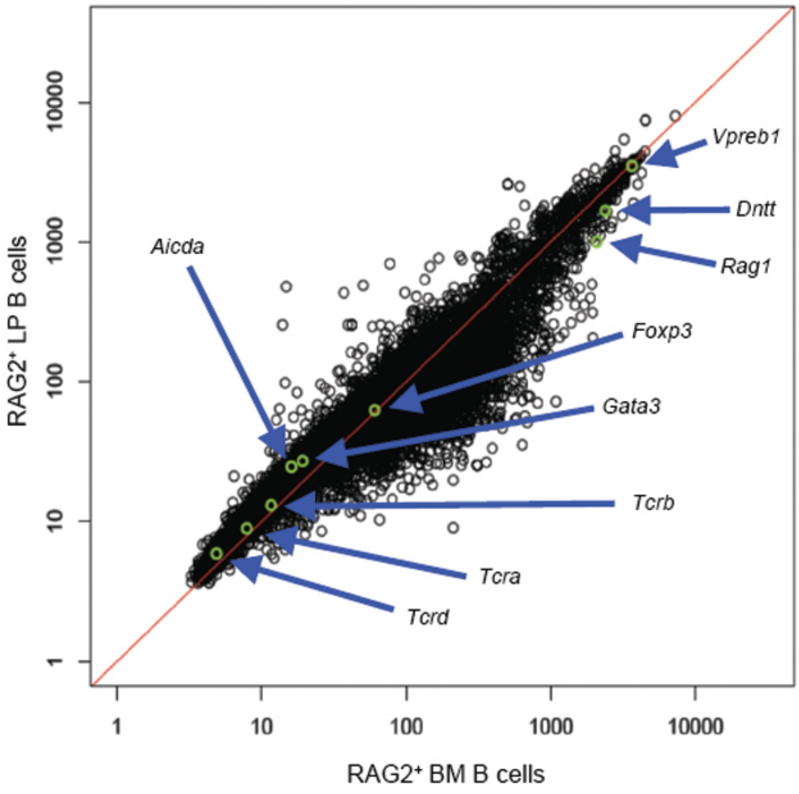
Supplementary Figure 5 | Small intestinal LP-resident RAG2-GFP$^+$ B lineage cells do not express α4β7 or CCR9. a,b, Both α4β7 integrin and CCR9 chemokine receptor have been implicated in lymphocyte homing to the gut[31]. Lymphocytes from *Rag2-Gfp* mice were stained for surface expression of CD19, B220, α4β7 (**a**) and CCR9 (**b**). Lymphocytes were first gated for CD19 expression, then both RAG2-GFP$^-$ and RAG2-GFP$^+$ cells from CD19$^+$ gated cells were analyzed for expression of B220 and α4β7 (**a**) or B220 and CCR9 (**b**). Numbers indicate percentage of cells within the indicated gates of total cells in the plot. RAG2-GFP$^-$ cells (left plots, top left gate). These data show that RAG2-GFP$^+$ cells do not express α4β7 or CCR9. This experiment was repeated once with similar results.



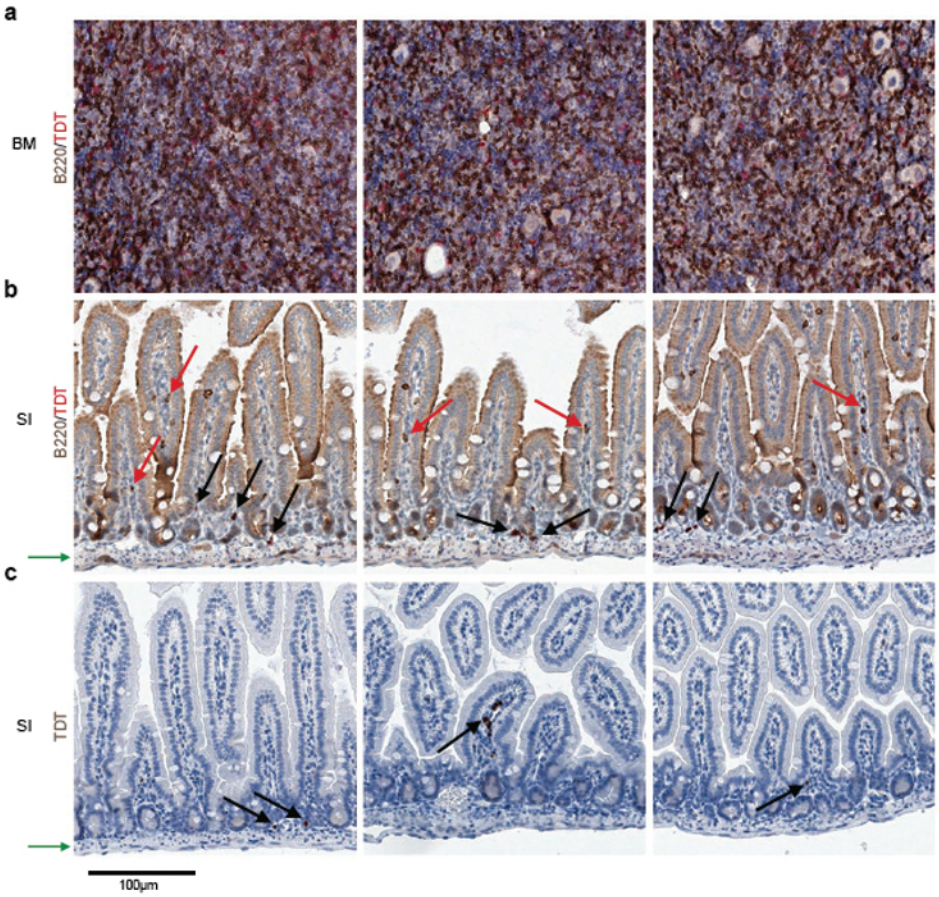
Supplemental Figure 6 | Intraperitoneal alum injection leads to the appearance of RAG2$^+$ cells in the gut mucosa. a, FACS plots of CD19$^+$ gated cells showing the percentages of RAG2-GFP$^+$ B220^{low} cells from uninjected (control) or mice injected with intraperitoneal alum. **b**, Scatter dot plot showing percentage of B220^{low} RAG2-GFP$^+$ of CD19$^+$ gated cells from the indicated tissues in adult (4-6 mo.) mice injected (open circles) or not injected (closed circles). Mean values ± s.e.m. are shown. "PP" denotes Peyer's patches. **c**, Quantitative PCR analysis of *Rag1* and *Rag2* expression from bone marrow (BM) and small intestinal lamina propria (LP) of 3 wk-old wild type (WT) controls or mice deficient in tumor necrosis factor-α (TNF KO). Levels of each were normalized to *Cd19* expression. The *y*-axis indicates levels relative to those of BM. Shown are mean values ± s.e.m of three independent experiments. "ns" denotes non-significant (two-tailed Student's *t*-test).



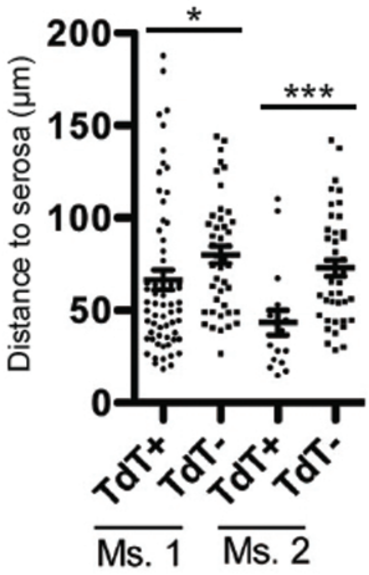
Supplementary Figure 7 │ Analysis of RAG2-GFP$^+$ B lineage developmental subsets. FACS plots showing bone marrow (BM) and small intestinal lamina propria (LP) cells from post-natal day 17-24 *Rag2-Gfp* mice, which were stained live (top plots), or after fixation and permeabilization (bottom plots) for CD19, B220, μ heavy chain (Igμ), and Igκ. The CD19$^+$ gated plots on the left of each set show the polygon gate capturing the RAG2-GFP$^+$ cells, which are analyzed in the plots immediatedly to their right for staining with anti-Igμ (*x*-axis) and anti-Igκ (*y*-axis). Live cells positive for both Igμ and Igκ are surface IgM$^+$, and are identified in the oval gates (top, right plots of each set). Based on prior studies and classification (see text for details). The CD19$^+$, RAG2-GFP$^+$, Igμ$^-$ Igκ$^-$ cells (bottom right plots, bottom left gate) are pro-B cells, Igμ$^+$ Igκ$^-$ (bottom right plots, bottom right gate) are pre-B cells, and Igμ$^+$ Igκ$^+$ cells (bottom right plots, top right gate) are immature B cells undergoing editing. RAG2$^+$ cells expressing surface IgM have also been identified as editing B cells[2]. Percentages are indicated in each gate. Statistical analysis of these data are shown in Figure 2a, where mean values ± s.e.m of pro-B, pre-B, immature-B and surface IgM$^+$ B lineage cells out of total RAG2-GFP$^+$ cells are ploted. Student's *t*-tests failed to detect significant differences between any of the BM and LP subgroups.



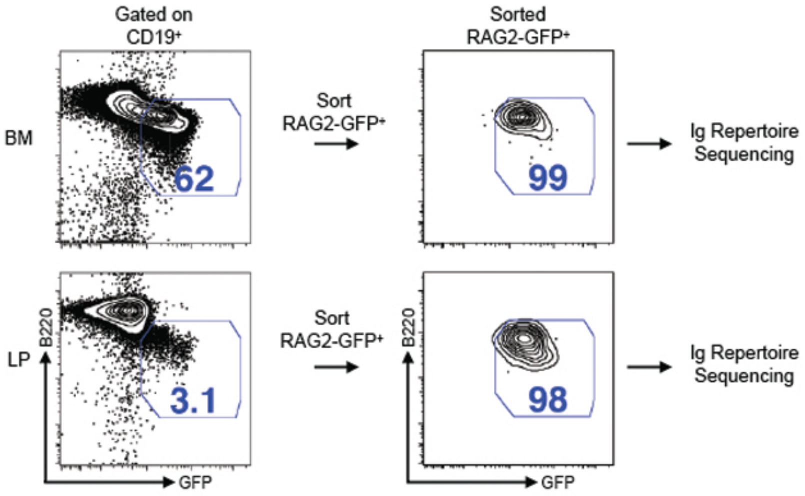
Supplementary Figure 8 | Comparative transcriptome analysis indicates a high degree of similarity between bone marrow (BM) and small intestinal lamina propria (LP) RAG2-GFP$^+$ cells. Scatter plot of affymetrix 1.0 ST microarray data comparing RAG2-GFP$^+$ cells from BM versus RAG2-GFP$^+$ cells sorted from LP. Although statistical *t*-tests showed some nominally significant differentially expressed genes, none passed correction for multiple comparisons testing (Benjamini-Hochberg procedure). Arrows point to selected genes highlighted in green that are not expected to be expressed to significant levels in early lineage cells (*Fox3, Gata3, Tcra, Tcrb, Tcrd, Aicda*), as well as genes that are expected to be expressed at early stages of B lineage development (*Vpreb1, Dntt, Rag1*) for comparison. These data indicate a high degree of transcript similarity between these two populations.



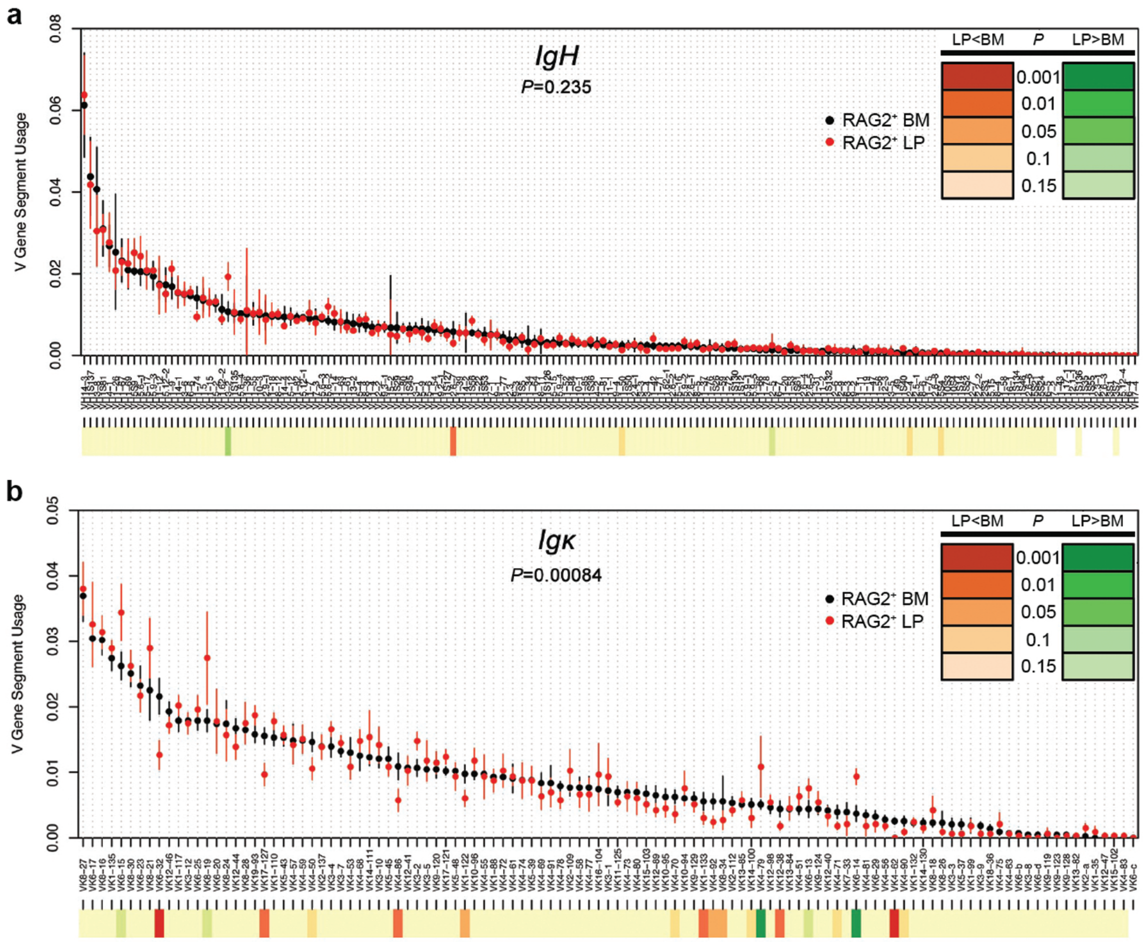
Supplementary Figure 9 | Immunohistochemistry (IHC) identifies TdT$^+$ B lineage cells in the gut lamina propria. a, b, IHC of paraffin-embedded section from BM (**a**) and small intestines (SI, **b**) stained with an anti-TdT antibody (red stain) plus anti-B220 antibody (brown stain). **b**, Red arrows point to LP-resident B220^{high} TdT negative cells that represent mature B cells. Black arrows point to LP-resident TdT$^+$ B220^{low} cells that resemble BM pro-B cells (**a**). **c**, IHC sections of small intestines (SI) stained with an anti-TdT antibody alone. Dark brown indicates TdT-reactivity. Sections were counterstained with alcian green to identify nuclei. These SI images are representative of what we used to calculate distance of TdT$^+$ and TdT$^-$ B lineage cells to the serosal (antiluminal) surface shown in Supplementary Figure 10. The serosal surfaces are indicated by the green arrows.



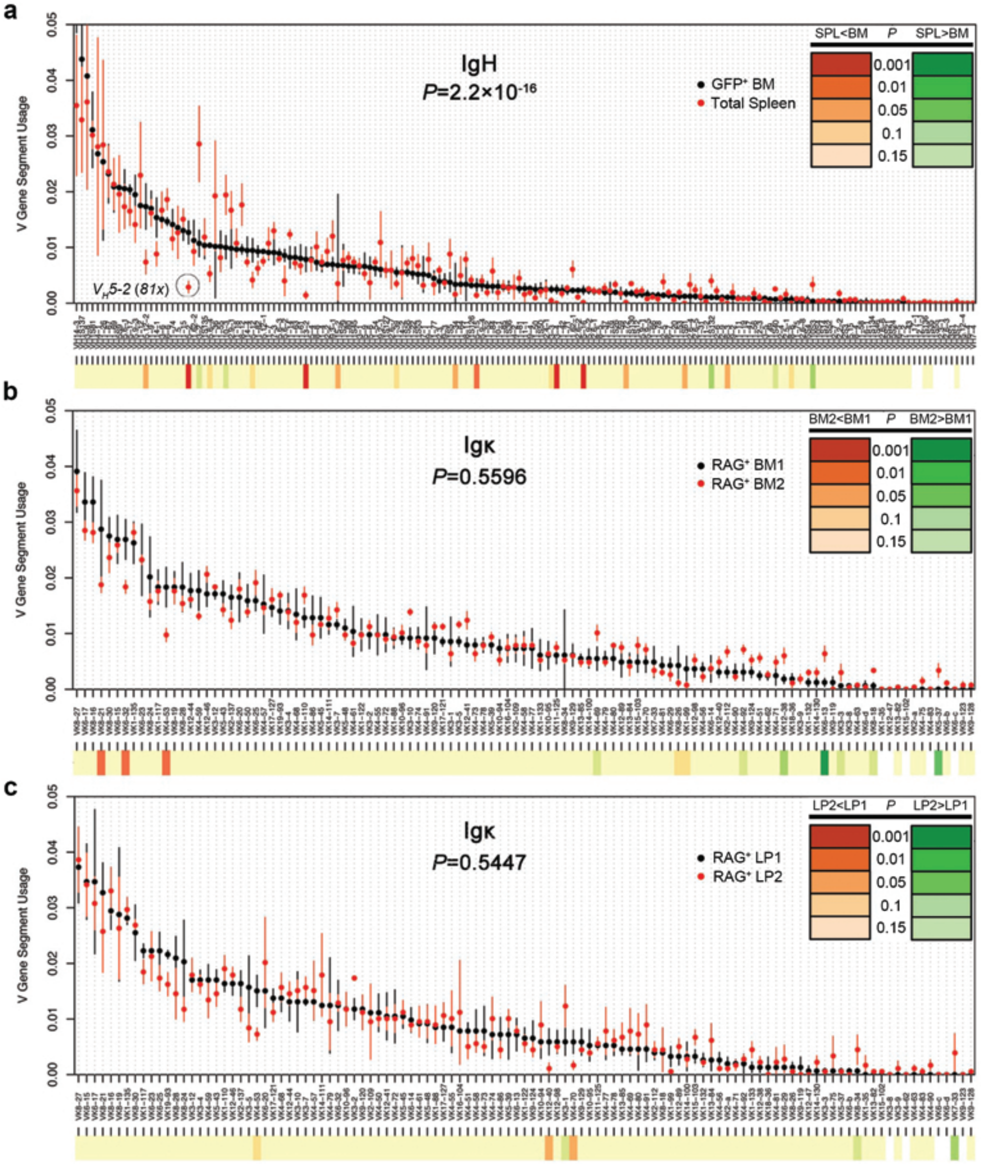
Supplementary Figure 10 | Intestinal TdT+ B lineage cells occupy distinct location compared to TdT- B cells. Quantitative measurements of the distance between individual small intestinal lamina propria resident TdT+ or TdT- B220+ cells and the serosal (antiluminal) surface from two independent mice (Ms.1 and Ms. 2). Mean levels ± s.e.m. are also shown. *P<0.05, ***P<0.001 (two-tailed Student's t-test).



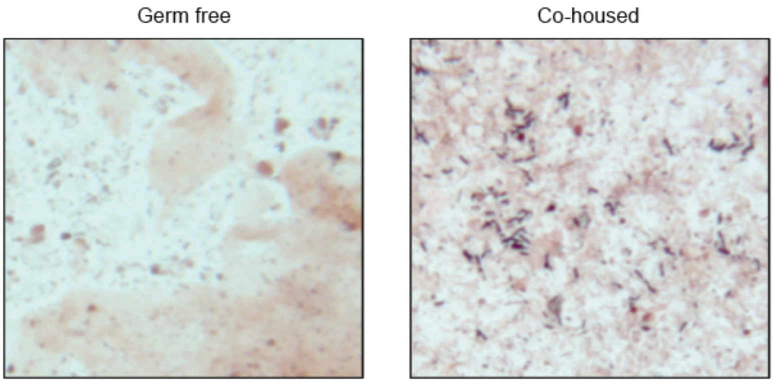
Supplementary Figure 11 | Sorting strategy for repertoire sequencing. This is a schematic diagram showing the sorting strategy to prepare samples for repertoire sequencing. The FACS plots on the left show CD19$^+$ gated cells from bone marrow (BM) and small intestinal lamina propria (LP) with additional polygonal gates showing RAG2-GFP$^+$ cells before (left) and after (right) sorting. Sorted cells were then subjected to repertoire sequencing.



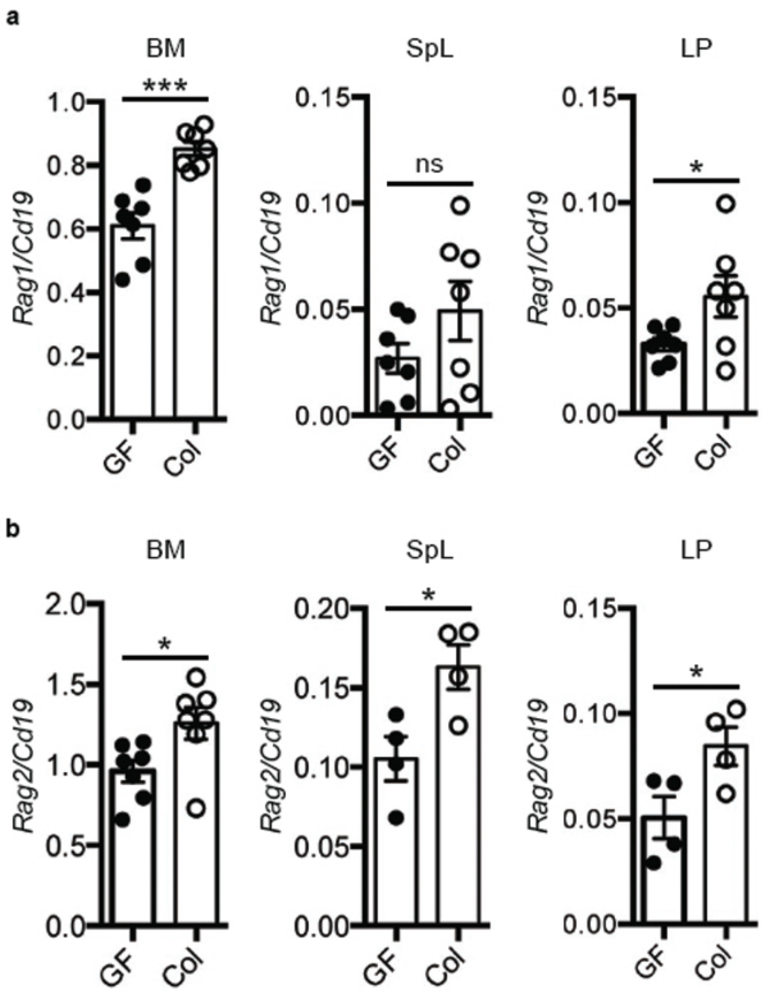
Supplementary Figure 12 | Distinct *Vκ* segment usage in RAG2$^+$ cells from BM and LP. a,b, Dot plots showing distribution of *IgH V* segment (*V_H*) (**a**) and *Igκ V* segment (*Vκ*) (**b**) usage in RAG2-GFP$^+$ cells sorted from bone marrow (BM, black dots) and small intestinal lamina propria (LP, red dots) as determined by 454 pyrosequencing. Aligned sequences with unique (determined by *V(D)J* junction anlysis), in-frame V(D)J junctions were analyzed. Individual *V* gene segment usage (*y*-axis) was calculated by dividing the number of individual *V_H* or *V*κ gene segments by the total number of in-frame *V_H* or *V*κ gene segments represented in our sequencing data set, respectivley. Individual *V_H* and *Vκ* gene segments are arranged on the *x*-axis in order of highest to lowest utilization found in the bone marrow samples. Individual *V* gene segment names are shown. Plotted are the means ± s.e.m. of at least 4 experiments each consisting of a pool of 8-12 mice for each independent experiment. The *P* value for overall difference between BM and LP *V* segment utilization was calculated with the χ^2 test. Heat map under dot plots shows sequence utilization differences between RAG2$^+$ BM vs. RAG2$^+$ LP with intensity of color specifying increased significance of the indicated *P* values as calculated by the exact test for differential expression (*P* values corresponding to color shown in inset).



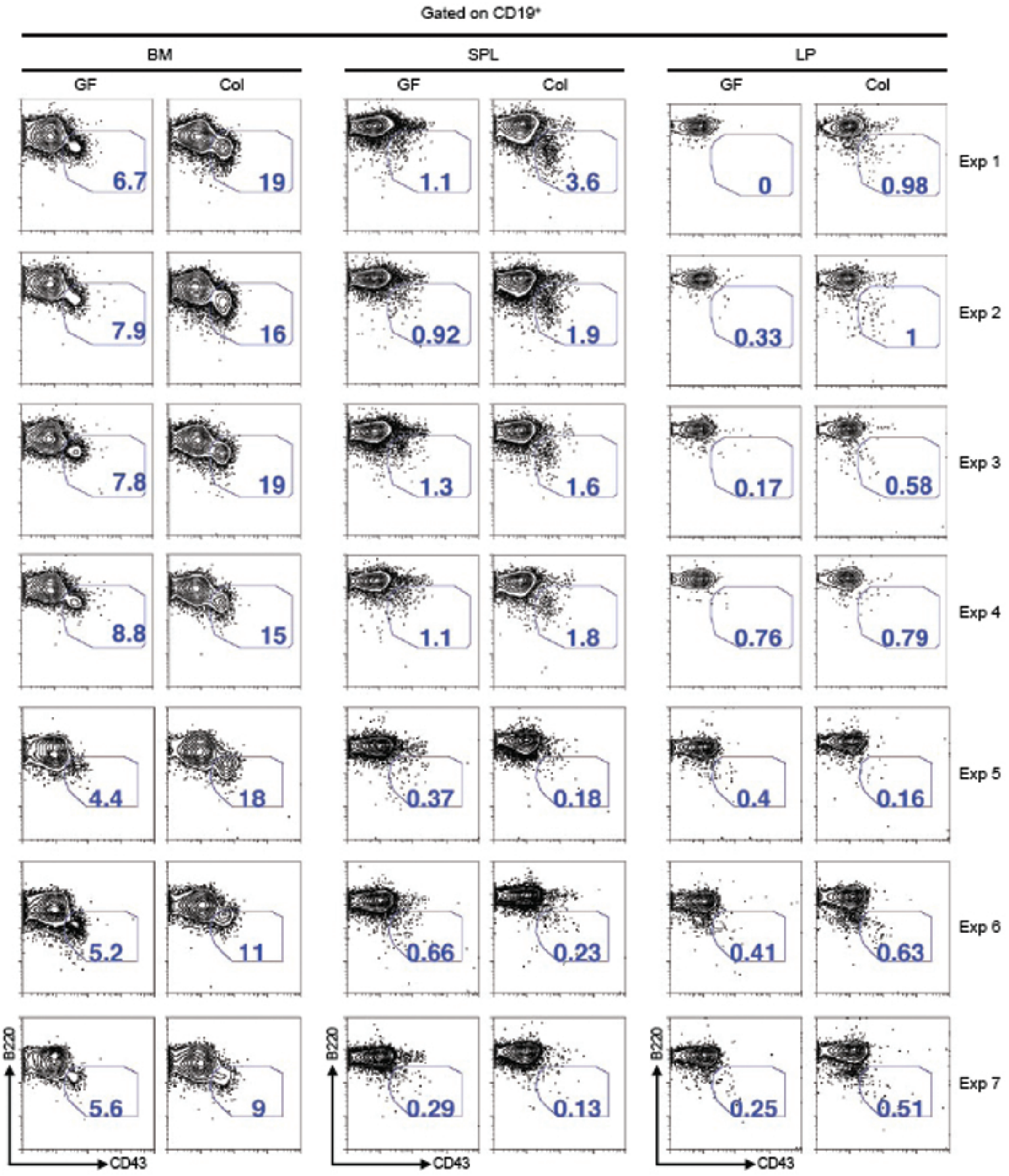
Supplementary Figure 13 | Control comparisons of the repertoire sequencing data. **a**, Dot plot and heat map comparing *IgH V* segment (*V_H*) usage between RAG2$^+$ BM B lineage cells and total splenic B cells. As expected, several prominent and significant differences are observed between these subsets including the *V_H* segment *5-2* (also known as *81x*), which is known to be utilized less in the splenic B cell repertoire as compared to the early lineage BM B cell repertoire[32]. **b, c,** Dot plot and heat map comparing *Igκ V* segment (*Vκ*) usage from RAG2-GFP$^+$ BM (**b**) and LP (**c**). To further validate our repertoire sequencing approach, the 8 independent experiments where the *Igκ* repertoire was sequenced were divided randomly into two groups of 4 experiments each and compared against each other to reveal false detection rates due to multiple comparisons within the same tissue from independent pools of mice. The indicated *P* values were calculated from the χ^2 test and indicate no significant difference between these samples. The indicated *P* value was calculated from the χ^2 as above.



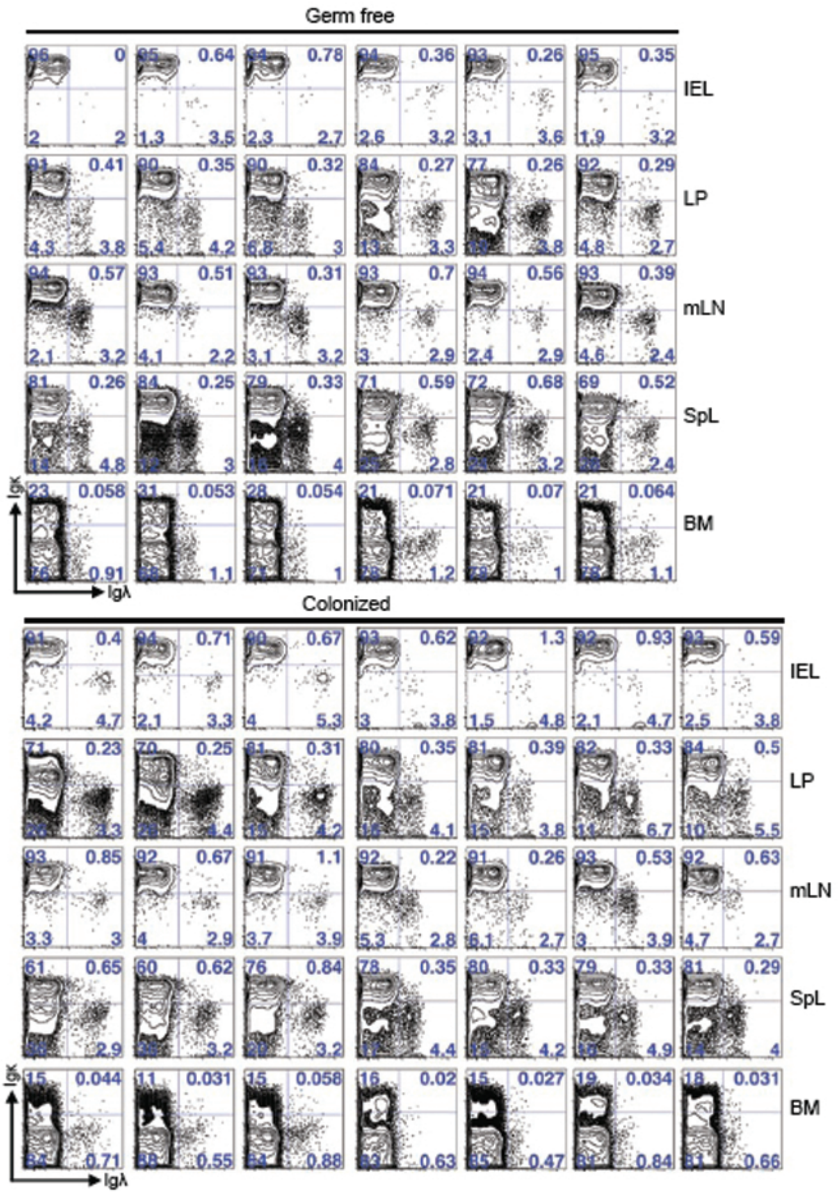
Supplementary Figure 14 | Gram stain of small intestinal contents from germ-free and co-housed mice. Photographs of gram-stained small intestinal (distal) contents from 4 wk-old germ-free, littermates that were colonized by cohousing with specific pathogen free mice for 7 days.



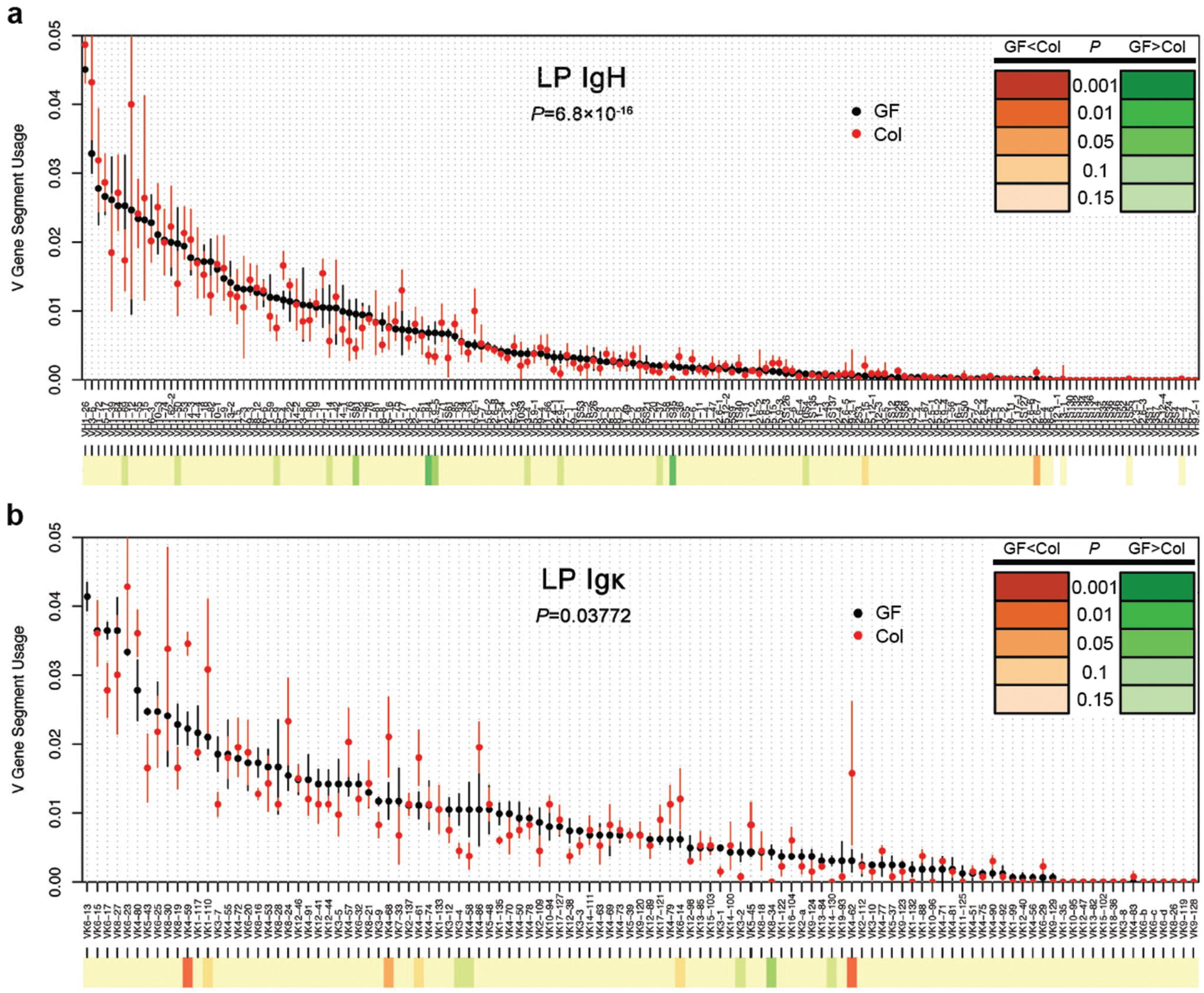
Supplementary Figure 15 | Gut colonization leads to increased gut lamina propria *Rag* expression levels. a, b, Bar graphs showing quantitative PCR for *Rag1* (**a**) and *Rag2* (**b**) expression in bone marrow (BM), spleen (SpL) and small intestinal lamina propria lymphocytes (LP) of 4 wk-old germ free (GF) mice and littermates that were colonized (Col) by co-housing with regular serum pathogen free (SPF) mice for 7 days prior to analysis. *Rag1* and *Rag2* levels were normalized to *Cd19* expression. *y*-axis values signify levels relative to wild type Balb/c BM. Shown are mean values ± s.e.m. of at least 3 independent experiments. **P*<0.05, ****P*<0.001 (two-tailed Student's *t*-test).



Supplementary Figure 16 | Colonization of germ-free mice leads to increased proportion of pro-B cells. FACS plots of CD19$^+$ gated cells from bone marrow (BM), spleen (SpL) and small intestinal lamina propria (LP). Total BM, SpL and LP lymphocytes were isolated from 4 wk-old germ-free (GF) mice or littermates that were colonized (Col) with microflora by co-housing with regular SPF mice for 7 days prior to analysis. The polygonal gates shown encompass cells with the pro-B phenotype (B220^{low} CD43$^+$). Numbers shown in the gates indicate percentages. Seven experiments are shown. Statistical analysis of these data is shown in Figure 4a.



Supplementary Figure 17 | Colonization of germ-free mice leads to increased Igλ/Igκ ratio specifically in lamina propria B cells. FACS plots of CD19$^+$ gated cells from intraepithelial lymphocytes (IEL), mesenteric lymph nodes (mLN), bone marrow (BM), spleen (SpL) and small intestinal lamina propria (LP) showing Igκ and Igλ expression. Total BM, SpL and mLN as well IEL and LP lymphocytes were isolated from 4 wk-old germ-free (GF) mice or littermates that were colonized (Col) with microflora by co-housing with regular SPF mice for 7 days prior to analysis. The plots are gated in quadrants. The upper left quadrant encompasses CD19$^+$ Igκ$^+$ B cells and the lower right quadrant encompasses CD19$^+$ Igλ$^+$ cells. The numbers shown in the gates indicate percentages of CD19$^+$ cells. Plots from 6 germ-free mice are shown above and plots from 7 colonized mice are shown. Statistical analysis of these data is shown in Figure 4b.



Supplementary Figure 18 | Distinct V_H and *Vκ* and segment usage in germ-free versus colonized mice. a,b, Dot plot and heat map comparing *IgH V* segment (V_H) usage (**a**) and *Igκ V* segment (*Vκ*) usage (**b**) from germ-free Swiss Webster mice (GF, black) and littermates cohoused with SPF mice for 7 days (Col, red) as determined by 454 pyrosequencing. Aligned sequences with unique (determined by *V(D)J* junction anlysis), in-frame V(D)J junctions were analyzed. Individual *V* gene segment usage (*y*-axis) was calculated by dividing the number of individual V_H or *V*κ gene segments by the total number of in-frame V_H or *V*κ gene segments represented in our sequencing data set, respectively. Individual V_H and *Vκ* gene segments are arranged on the *x*-axis in order of highest to lowest utilization found in the bone marrow samples. Individual *V* gene segment names are shown. Plotted are the means ± s.e.m. of at least 3 experiments. The *P* value for overall difference between GF and Col *V* segment utilization was calculated with χ^2 test. Heat map under dot plots shows sequence utilization differences between GF vs. Col with intensity of color specifying increased significance of the indicated *P* values (shown in insets) as calculated by the exact test for differential expression.

Supplementary References

31. Mora, J.R. & Von Andrian, U.H. Specificity and plasticity of memory lymphocyte migration. *Curr Top Microbiol Immunol* **308**, 83-116 (2006).

32. Yancopoulos, G.D., et al. Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. *Nature* **311**, 727-733 (1984).