LETTER

doi:10.1038/nature12496

Microbial colonization influences early B-lineage development in the gut lamina propria

Duane R. Wesemann[1,2,3,4], Andrew J. Portuguese[1,2,3], Robin M. Meyers[1,2,3], Michael P. Gallagher[1,2,3], Kendra Cluff-Jones[1,2,3], Jennifer M. Magee[1,2,3], Rohit A. Panchakshari[1,2,3], Scott J. Rodig[5], Thomas B. Kepler[6] & Frederick W. Alt[1,2,3]

The RAG1/RAG2 endonuclease (RAG) initiates the V(D)J recombination reaction that assembles immunoglobulin heavy (*IgH*) and light (*IgL*) chain variable region exons from germline gene segments to generate primary antibody repertoires[1]. *IgH* V(D)J assembly occurs in progenitor (pro-) B cells followed by that of *IgL* in precursor (pre-) B cells. Expression of IgH μ and IgL (Igκ or Igλ) chains generates IgM, which is expressed on immature B cells as the B-cell antigen-binding receptor (BCR). *Rag* expression can continue in immature B cells[2], allowing continued *Igκ* V(D)J recombination that replaces the initial VκJκ exon with one that generates a new specificity[3–5]. This 'receptor editing' process, which can also lead to *Igλ* V(D)J recombination and expression[3,6,7], provides a mechanism whereby antigen encounter at the *Rag*-expressing immature B-cell stage helps shape pre-immune BCR repertoires. As the major site of postnatal B-cell development, the bone marrow is the principal location of primary immunoglobulin repertoire diversification in mice. Here we report that early B-cell development also occurs within the mouse intestinal lamina propria (LP), where the associated V(D)J recombination/receptor editing processes modulate primary LP immunoglobulin repertoires. At weanling age in normally housed mice, the LP contains a population of *Rag*-expressing B-lineage cells that harbour intermediates indicative of ongoing V(D)J recombination and which contain cells with pro-B, pre-B and editing phenotypes. Consistent with LP-specific receptor editing, *Rag*-expressing LP B-lineage cells have similar V_H repertoires, but significantly different *Vκ* repertoires, compared to those of *Rag2*-expressing bone marrow counterparts. Moreover, colonization of germ-free mice leads to an increased ratio of *Igλ*-expressing versus *Igκ*-expressing B cells specifically in the LP. We conclude that B-cell development occurs in the intestinal mucosa, where it is regulated by extracellular signals from commensal microbes that influence gut immunoglobulin repertoires.

Pre-immune immunoglobulin diversification occurs within the gut or gut-associated structures in several vertebrate species, including sheep, rabbits, cattle, pigs and chicken[8–10]. In these species, cells harbouring a limited RAG-mediated V(D)J repertoire migrate to gut-associated structures to undergo further diversification through somatic mutation and/or gene conversion to generate a full pre-immune immunoglobulin repertoire[11]. These examples raise the notion that the gut environment may provide some benefit to the process of primary immunoglobulin diversification. In this regard, commensal bacteria are required for primary antibody repertoire diversification in pigs and rabbits[8,12], and may have an important role in stimulating this process in cattle, sheep and chickens shortly after birth[11]. In contrast to the above species, RAG-mediated V(D)J recombination is the major driver of pre-immune diversification in mice and humans. In this context, stimulated by our previous finding of mouse B-lineage tumours that arise in mesenteric lymph nodes from apparent receptor-editing B cells[13,14], we proposed that B-cell development might occur in the mouse gut and, thereby, allow *Rag* expression to diversify gut B cell pre-immune repertoires.

To test for RAG expression, we used our *Rag2* reporter mice, which contain a functional fusion with the green fluorescent protein gene (*Rag2-Gfp*) within the endogenous *Rag2* locus that provides *Rag2* expression functionally equivalent to that of the endogenous R*ag2* gene[15,16]. We used flow cytometry to test for RAG2–GFP in lymphocytes from mesenteric lymph nodes, small intestinal LP and intraepithelial lymphocytes of 3-week-old mice. Cells were gated on the CD19 pan-B-lineage marker, and GFP was plotted against the B220 pan-B-lineage marker. Staining with dual B-cell markers was done to optimize true GFP signal over background auto-fluorescence, which in wild-type controls was approximately 0.1% (Fig. 1a). With this method, we found essentially no *Rag2*-expressing B-lineage cells in the intraepithelial lymphocytes and mesenteric lymph nodes (Fig. 1a). However, we did find a population of *Rag2*-expressing, $CD19^+$ $B220^{low}$ cells within the LP that comprised approximately 3% of total $CD19^+$ cells (Fig. 1a). Quantitative PCR (qPCR) revealed *Rag1* and *Rag2* expression in wild-type small intestinal LP at a level of about 1–10% that of total bone marrow (BM), but little or no *Rag1* or *Rag2* expression in mesenteric lymph node or intraepithelial lymphocyte cells (Supplementary Fig. 1), confirming the flow cytometry results found with the *Rag2-Gfp* reporter mice. Large intestinal LP contained GFP^+ B-lineage cells as well, but at a lower level compared to that in the small intestinal LP (Supplementary Fig. 2). We did not find RAG2–GFP in Peyer's patch B cells (Supplementary Fig. 3) or mucosal T cells (Supplementary Fig. 4).

We examined various stages of early postnatal development to determine if levels of $RAG2^+$ B-lineage cells in the gut LP change over time. The proportion of LP RAG2–GFP^+ cells among total $CD19^+$ cells was low (<0.5%) in the first week of life; however, after that it gradually increased with levels peaking at approximately 4% at age 18–23 days before decreasing to undetectable levels by postnatal day 35 (Fig. 1b). In contrast, the $CD19^+$ B-cell population of peripheral blood contained 20–40% RAG2–GFP^+ cells during the first week of life, which then decreased over time to undetectable levels over the next 4 weeks (Fig. 1c). Similarly, RAG2–GFP^+ cell levels in the spleen appeared highest (10–15%) in the first week of life before decreasing to undetectable levels (Fig. 1d). The finding of low (<0.5%) levels of RAG2–GFP^+ cells in the gut LP in the first week of life, despite the presence of substantial proportions of RAG2–GFP^+ cells in the peripheral blood and spleen, suggests that the mechanism responsible for the later emergence of $RAG2^+$ cells in the gut may not be due to nonspecific dissemination driven by high levels of these cells in the blood. As $RAG2^+$ LP B-lineage cells do not express proteins known to promote gut lymphocyte tropism such as the α4β7 integrin or the CCR9 chemokine receptor (Supplementary Fig. 5), mechanisms underlying their appearance in the gut remain to be determined.

Sixteen days after intraperitoneal alum injection, RAG2–GFP^+ cells accumulate in the peripheral blood and spleens of adult mice owing to

[1]Program in Cellular and Molecular Medicine and Department of Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA. [4]Division of Rheumatology, Immunology and Allergy, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [5]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [6]Department of Microbiology, Boston University School of Medicine, Boston, Massachusetts 02215, USA.

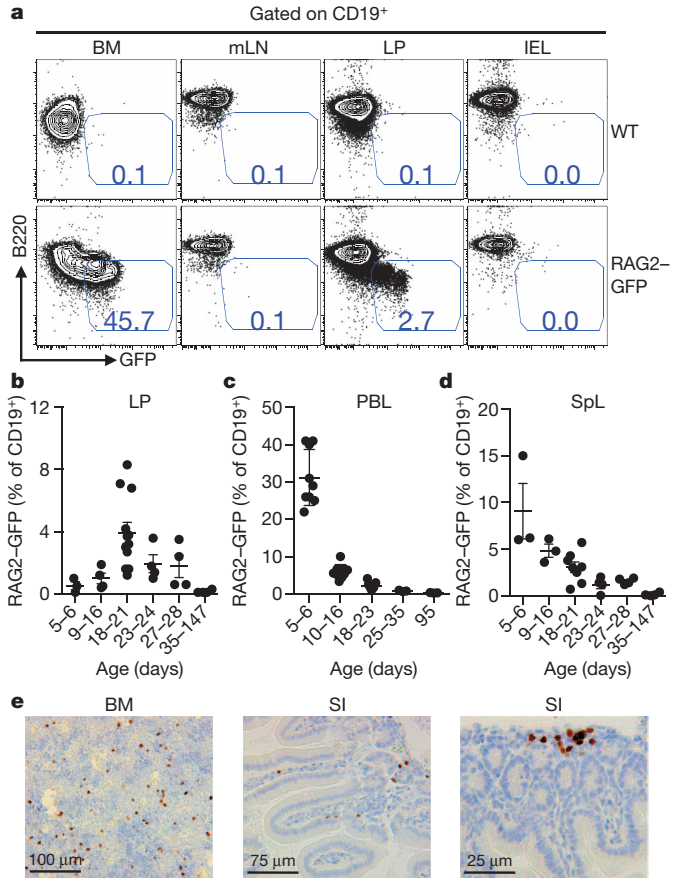


Figure 1 | Gut LP RAG2$^+$ B-lineage cells in weanling age mice. **a**, FACS plots of CD19$^+$ cells from the indicated tissues taken from wild-type (WT) (top) or homozygous *Rag2-Gfp* knock-in (bottom) mice. B220 expression is plotted against GFP fluorescence. Numbers denote percentage of CD19$^+$ B220^{low} RAG2–GFP$^+$ cells. IEL, intraepithelial lymphocyte; mLN, mesenteric lymph node. **b**–**d**, Dot plots showing percentage of RAG2–GFP$^+$ cells in LP (**b**), peripheral blood lymphocytes (PBL) (**c**) and spleen (SpL) (**d**) from indicated post-natal ages. Each point represents one mouse. Horizontal bars indicate mean values ± s.e.m. **e**, Immunohistochemistry of paraffin-embedded sections from bone marrow (BM) and small intestine (SI) stained with an anti-TdT antibody. Dark brown indicates TdT reactivity.

increased bone marrow output after initial alum-mediated bone marrow suppression[17,18]. To determine if the gut LP in adult mice maintains ability to support RAG2$^+$ B-lineage cells, we injected 4–6-month-old *Rag2-Gfp* mice with intraperitoneal alum and examined gut tissues on day 16. Following alum injection, low levels of RAG2–GFP$^+$ B-lineage cells appeared in intraepithelial lymphocytes, mesenteric lymph nodes and Peyer's patches; however, the most striking accumulation was in the LP, where RAG2–GFP$^+$ cells made up about 2.5% of total CD19$^+$ cells (Supplementary Fig. 6a, b). Appearance of RAG$^+$ B-lineage cells in the spleen and blood following alum injection is mediated by tumour necrosis factor alpha (TNF-α)[19]. To determine whether appearance of RAG$^+$ B-lineage cells in the small intestinal LP at weaning age is also TNF-α-dependent, we measured LP *Rag1* and *Rag2* expression by qPCR in 3-week-old *Tnfα* (also known as *Tnf*) knockout and wild-type control mice and found no differences (Supplementary Fig. 6c, d). Thus, the mechanism of gut LP RAG$^+$ B-lineage cell accumulation at weaning age seems distinct from that of peripheral RAG$^+$ cell accumulation that occurs after alum immunization.

RAG-expressing B-lineage cells in the BM comprise a heterogeneous population of early developmental subsets including pro-B, pre-B and immature B cells undergoing receptor editing[20]. The expression patterns of Igμ and Igκ (which accounts for ~95% of mouse IgL) can be used to distinguish these groups[5,21]. In this context, productive assembly of *Igh* V(D)J exons in pro-B cells leads to the cytoplasmic Igμ$^+$ pre-B cell stage[21]. Assembly of *Igκ* VJ exons in pre-B cells leads to formation of IgM and differentiation to the surface IgM$^+$ immature B cell stage. *Rag*-expressing cells with cytoplasmic Igκ and low or absent surface IgM have been defined as immature B cells undergoing receptor editing[2,5,17]. Staining of fixed/permeabilized CD19$^+$ B220^{low} RAG2–GFP$^+$ LP B cells for cytoplasmic Igμ and Igκ revealed similar relative levels of pro-B cells (Igμ$^-$, Igκ$^-$), pre-B cells (Igμ$^+$, Igκ$^-$) and editing B cells (Igμ$^+$, Igκ$^+$, RAG2$^+$), respectively, to those of the BM (Fig. 2a and Supplementary Fig. 7). Live CD19$^+$ B220^{low} RAG2–GFP$^+$ LP B cells that were surface IgM positive also had similarly low IgM levels to those of this putative editing B-lineage subset in BM (Fig. 2a and Supplementary Fig. 7). In addition, ligation-mediated qPCR further showed that sorted RAG2–GFP$^+$ LP B cells had a similar level of RAG-dependent DNA double-strand breaks at *Jκ* as RAG2–GFP$^+$ BM B-lineage cells (Fig. 2b), demonstrating that *Igκ* V(D)J recombination takes place in the RAG2$^+$ LP B cells at levels similar to those in RAG2$^+$ BM B-lineage cells. Finally, microarray analysis of RAG2$^+$ B-lineage cells in weanling age gut LP revealed no significant differences in general transcript expression profiles in RAG2$^+$ LP and RAG2$^+$ BM B-lineage cells (Supplementary Fig. 8), demonstrating a strong similarity between RAG2$^+$ cells in these two sites. Overall, these data indicate that the LP *Rag2*-expressing B-lineage cells contain early B-lineage developmental subsets representative of those found in the BM, supporting the occurrence of B-cell development in the LP.

We performed immunohistochemistry to confirm the LP localization of early B-lineage cells in mice at postnatal day 18–23. The terminal deoxynucleotidyl transferase (TdT) enzyme is present in pro-B cells and mediates addition of random nucleotides to *IgH* V(D)J DNA ends during V(D)J recombination to increase *IgH* V(D)J junctional diversity[22]. Staining for TdT, which is expressed similarly in BM and LP RAG2$^+$ B-lineage cells (Supplementary Fig. 8), revealed a dense nuclear expression pattern (brown) in small intestinal LP cells similar to TdT$^+$ cells in the BM (Fig. 1e). Small intestine sections were also subjected to dual staining for B220 plus TdT, which showed that the TdT$^+$ cells were also faintly B220-positive, analogous to TdT$^+$ cells in the BM (Supplementary Fig. 9). There were also cells in the small intestinal LP that stained strongly for B220 but were TdT$^-$, representing more mature B-cell populations (Supplementary Fig. 9, black arrows). TdT$^+$ cells were distributed throughout the LP, including within villi; however, they generally appeared to be more proximal to bases of villi, closer to the serosal

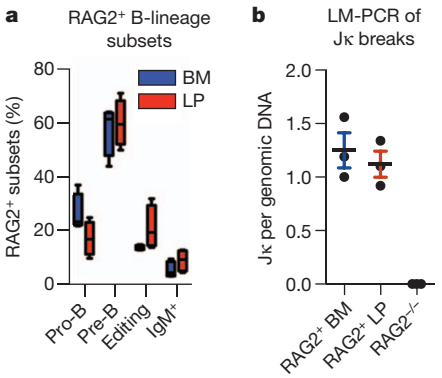


Figure 2 | RAG2–GFP$^+$ LP B-lineage developmental subsets. **a**, Plots show the relative percentage of *Rag2*-expressing pro-B cells, pre-B cells, editing B cells and surface IgM$^+$ B cells (see text for definition of each) in the bone marrow (blue bars) and lamina propria (red bars). Plotted are mean values ± s.e.m. and each are derived from experiments of 4 independent mice at postnatal day 17–24 (see Supplementary Fig. 7 for more details). **b**, Plots show quantitative ligation-mediated PCR (LM-PCR) of RAG2$^+$ BM B cells and RAG2$^+$ LP B cells normalized to genomic DNA. BM cells from *Rag2*$^{-/-}$ mice were a negative control. Values on the *y* axis are units relative to the signal obtained from RAG2$^+$ BM B-cell samples.

(antiluminal) intestinal surface compared to B220^{high} TdT^{-} B cells (Supplementary Fig. 10). These data indicate that gut-resident early B-lineage cells inhabit a generally distinct location within the LP compared to more mature B cells. Immunohistochemistry studies of human fetal intestine also identified serosally positioned pre-B cells in the intestinal LP, suggesting that similar early B-cell development may occur in the human gut[23].

Given our finding of primary B-cell development in the mouse intestinal LP, we asked whether this process contributes to differential diversification of pre-immune repertoires in developing RAG2^{+} LP B-lineage cells versus those from BM. To test this, we isolated RNA from sorted RAG2–GFP^{+} cells from the BM and LP (Supplementary Fig. 11) of 3-week-old *Rag2-Gfp* mice, and assessed productive V_H and $V\kappa$ use through 5′ rapid amplification of complementary DNA ends (5′RACE) generated from mature immunoglobulin gene transcripts, followed by 454 sequencing. To best visualize potential *V* segment usage differences between RAG2–GFP^{+} cells from LP versus BM, we plotted *V* segment use from the two sources against individual in-frame *V* segments in the order of highest to lowest usage in RAG2–GFP^{+} BM cells (Fig. 3 and Supplementary Fig. 12). These studies showed that V_H usage was very similar in RAG2–GFP^{+} LP cells and RAG2–GFP^{+} BM cells (χ^2 test, $P = 0.235$) (Fig. 3a and Supplementary Fig. 12a). As a positive control for the method, comparison of V_H usage between RAG2^{+} BM and total non-sorted splenic B cells showed an expected highly significant difference (χ^2 test, $P = 2.2 \times 10^{-16}$) (Supplementary Fig. 13a). The similar V_H repertoires of RAG2^{+} BM and LP B-lineage cells suggests that V_H use during primary IgH V(D)J recombination occurs similarly in the two locations. In contrast, we observed prominent and highly significant differences in $V\kappa$ usage in RAG2–GFP^{+} cells from LP versus BM (χ^2 test, $P = 0.00084$) (Fig. 3b and Supplementary Fig. 12b); as a negative control, we observed no significant overall differences in $V\kappa$ usage between RAG2^{+} BM samples compared with other RAG2^{+} BM samples; or RAG2^{+} LP samples compared with other RAG2^{+} LP samples from separate pools of mice (χ^2 test, $P = 0.560$ and 0.545, respectively) (Supplementary Fig. 13b, c). The finding of different $V\kappa$ repertoires within RAG2^{+} LP and BM B-lineage cells indicates that the LP versus BM location of B-cell development may influence $V\kappa$ usage in developing B cells. In this regard, the very similar V_H repertoires of LP and BM RAG2^{+} B-lineage cell indicates that a probable explanation for the marked difference in the LP $V\kappa$ repertoires from those of BM is that they are generated in the RAG2^{+} receptor-editing LP B-cell population.

Intestinal microflora have been shown to influence immune cell development in terms of lymphoid organization and T-cell subset accumulation and activity, both locally in the gut as well as systemically[24]. To assess the influence of microflora on B-cell development, we co-housed 3-week-old wild-type Swiss-Webster germ-free mice with regular specific pathogen free (SPF) mice for 7 days. Gram staining of small intestine contents was performed to confirm bacterial colonization of co-housed mice (Supplementary Fig. 14). We used qPCR to assay colonized mice and germ-free littermates for *Rag1* and *Rag2* expression, which was normalized to *Cd19* expression. In accord with ability of microflora to induce systemic effects on immune cell development[24], we observed colonization-dependent increases in *Rag1* and *Rag2* expression in the BM and spleen, as well as the gut LP (Supplementary Fig. 15). We also found pro-B cells (identified as CD19^{+} B220^{low} CD43^{+}) to represent an increased percentage of total CD19^{+} B cells in colonized mouse BM and LP, and potentially in the colonized spleen (Fig. 4a and Supplementary Fig. 16), correspondingly, the increased *Rag* expression in these tissues probably derives from the increased percentage of pro-B cells. We conclude that gut microflora induce increased levels of pro-B cells systemically, including in the gut LP, in previously un-colonized weaning-age mice.

Given our finding of different $V\kappa$ repertoires in LP versus BM RAG2^{+} B-lineage populations that might be generated via receptor editing, we examined the ratio of Igλ^{+} to Igκ^{+} B cells in the LP of colonized mice versus their germ-free mice littermates. In this regard, increased Igλ usage in the B-cell repertoire is another marker of receptor editing[3,6,7]. Notably, colonization led to a significant and reproducible increase of the ratio of Igλ^{+} to Igκ^{+} B cells in the LP but not the BM or spleen (Fig. 4b and Supplementary Fig. 17), consistent with a commensal-dependent process leading to increased editing specifically within LP B cells. However, as these analyses were not performed in *Rag2-Gfp* mice, only total B-cell populations could be analysed. Therefore, a non-mutually exclusive possibility would be selection for Igλ^{+} B cells in the gut after colonization with commensal microbes, a phenomenon not previously described. In this regard, we do find significant differences in both V_H and $V\kappa$ segment usage in the LP IgM^{+} B-cell population of colonized mice compared to germ-free littermates (Supplementary Fig. 18), indicating presence of commensals influences both IgH and IgL mature B-cell repertoires.

Consistent with growing evidence demonstrating that the microflora act as a regulators of T-lymphocyte subsets[24], we find that weanlings harbour an intestinal LP B-cell developmental process that is influenced in germ-free mice by microbial colonization. As weaning is concurrent with microbial expansion[25], occurrence of *Rag*-expressing B-lineage cell accumulation in weanlings may have evolved to allow B-cell primary repertoires to be modulated in response to colonization. In this regard, our findings also suggest that this LP B-cell developmental process includes BCR editing, which may contribute to significant differences in the primary $V\kappa$ repertoire of LP versus BM B-lineage populations. Past studies have implicated BCR editing in the BM as a negative selection process, largely based on studies of B cells engineered to make specific self-reactive, high-affinity BCRs. Given the natural repertoire in our studies, the degree to which the observed BCR-editing process in the gut represents a tolerance mechanism is unclear. However, given the potential special role of the gut in pre-immune diversification

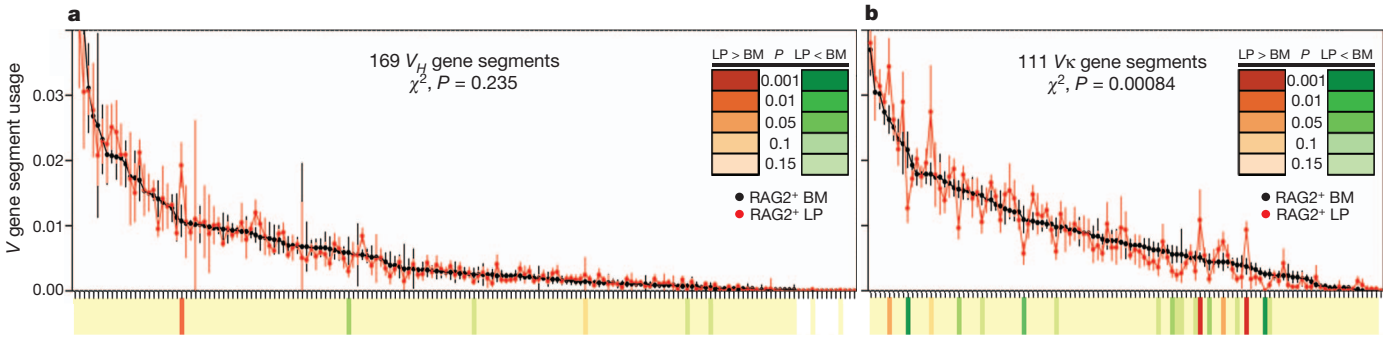


Figure 3 | Distinct $V\kappa$ segment usage in RAG2^{+} cells from BM versus LP. **a**, **b**, Dot plots show contributions (in order of highest to lowest BM use) of different V_Hs (**a**) and $V\kappa$s (**b**) to in-frame rearrangements in RAG2–GFP^{+} BM (black dots) and LP (red dots) cells. The two most highly used V_Hs are omitted to increase plot resolution. Each point shows mean ± s.e.m. of at least 4 experiments. The χ^2 calculated *P* values for overall differences between BM and LP are indicated. Significant *V* segment usage differences between BM and LP are indicated on heat map (*P* values scale indicated in inset). Full data set at increased resolution is in Supplementary Fig. 12.

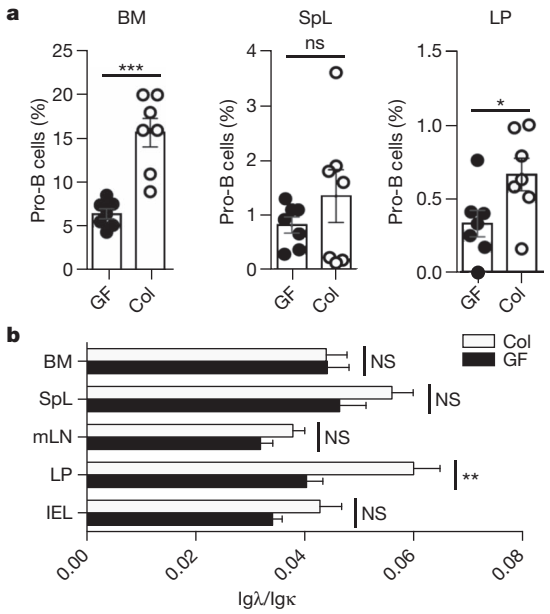


Figure 4 | Effects of gut colonization on development of LP B-lineage cells. **a**, Plots of percentage of pro-B cells versus total $CD19^+$ B-lineage cells from bone marrow (BM), spleen (SpL) and lamina propria (LP) of 4-week-old germ-free (GF) mice and littermates colonized (Col) by co-housing with specific pathogen-free mice for 7 days. **b**, Bar graphs show ratios of $Ig\lambda^+$ versus $Ig\kappa^+$ B cells within mesenteric lymph nodes (mLN), inter-epithelial lymphocyes (IEL) and tissues indicated in **a** from germ-free mice and littermates colonized by co-housing with specific pathogen-free mice. Mean values and s.e.m. are shown. $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$. NS, not significant. (Details in Supplementary Figs 16 and 17.)

in other vertebrates, RAG-dependent editing in the LP may also contribute diversification-related roles. In addition, immature gut-derived B cells also may have specialized roles, such as those suggested for immature splenic B cells[26]. Finally, primary B-cell development in the intestine, including mucosal B-cell receptor editing, might allow both luminal antigens and peripheral host mucosal components opportunities to shape the pre-immune repertoire. In this regard, the transient nature of LP B-cell development implies that there may be windows of opportunity for this influence to occur.

METHODS SUMMARY

All experiments with mice followed the protocols approved by the Animal Resources at Children's Hospital (ARCH) review committee and performed in accordance with NIH guidelines. Mice harbouring the *Rag2-Gfp* knock-in fusion gene at the endogenous locus were described previously[15] and were maintained on a 129/SvJ background. Wild-type BALB/c mice were purchased from Jackson Laboratories. Swiss-Webster germ-free mice were purchased from Taconic Farms. Peyer's patches, intraepithelial lymphocytes and LP lymphocytes were isolated essentially as described[27].

Full Methods and any associated references are available in the online version of the paper.

Received 22 February; accepted 23 July 2013.
Published online 21 August 2013.

1. Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F. W. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* **24,** 541–570 (2006).
2. Yu, W. *et al.* Continued RAG expression in late stages of B cell development and no apparent re-induction after immunization. *Nature* **400,** 682–687 (1999).
3. Tiegs, S. L., Russell, D. M. & Nemazee, D. Receptor editing in self-reactive bone marrow B cells. *J. Exp. Med.* **177,** 1009–1020 (1993).
4. Gay, D., Saunders, T., Camper, S. & Weigert, M. Receptor editing: an approach by autoreactive B cells to escape tolerance. *J. Exp. Med.* **177,** 999–1008 (1993).
5. Pelanda, R. *et al.* Receptor editing in a transgenic mouse model: site, efficiency, and role in B cell tolerance and antibody diversification. *Immunity* **7,** 765–775 (1997).
6. Retter, M. W. & Nemazee, D. Receptor editing occurs frequently during normal B cell development. *J. Exp. Med.* **188,** 1231–1238 (1998).
7. Hertz, M. & Nemazee, D. BCR ligation induces receptor editing in IgM^+IgD^- bone marrow B cells *in vitro*. *Immunity* **6,** 429–436 (1997).
8. Lanning, D., Zhu, X., Zhai, S. K. & Knight, K. L. Development of the antibody repertoire in rabbit: gut-associated lymphoid tissue, microbes, and selection. *Immunol. Rev.* **175,** 214–228 (2000).
9. Jenne, C. N., Kennedy, L. J. & Reynolds, J. D. Antibody repertoire development in the sheep. *Dev. Comp. Immunol.* **30,** 165–174 (2006).
10. Ratcliffe, M. J. Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. *Dev. Comp. Immunol.* **30,** 101–118 (2006).
11. Lanning, D. K., Rhee, K. J. & Knight, K. L. Intestinal bacteria and development of the B-lymphocyte repertoire. *Trends Immunol.* **26,** 419–425 (2005).
12. Butler, J. E. *et al.* Antibody repertoire development in fetal and neonatal piglets. VIII. Colonization is required for newborn piglets to make serum antibodies to T-dependent and type 2 T-independent antigens. *J. Immunol.* **169,** 6822–6830 (2002).
13. Wang, J. H. *et al.* Mechanisms promoting translocations in editing and switching peripheral B cells. *Nature* **460,** 231–236 (2009).
14. Wang, J. H. *et al.* Oncogenic transformation in the absence of Xrcc4 targets peripheral B cells that have undergone editing and switching. *J. Exp. Med.* **205,** 3079–3090 (2008).
15. Monroe, R. J. *et al.* RAG2:GFP knockin mice reveal novel aspects of RAG2 expression in primary and peripheral lymphoid tissues. *Immunity* **11,** 201–212 (1999).
16. Nagaoka, H., Yu, W. & Nussenzweig, M. C. Regulation of RAG expression in developing lymphocytes. *Curr. Opin. Immunol.* **12,** 187–190 (2000).
17. Nagaoka, H., Gonzalez-Aseguinolaza, G., Tsuji, M. & Nussenzweig, M. C. Immunization and infection change the number of recombination activating gene (RAG)-expressing B cells in the periphery by altering immature lymphocyte production. *J. Exp. Med.* **191,** 2113–2120 (2000).
18. Gärtner, F., Alt, F. W., Monroe, R. J. & Seidl, K. J. Antigen-independent appearance of recombination activating gene (Rag)-positive bone marrow B cells in the spleens of immunized mice. *J. Exp. Med.* **192,** 1745–1754 (2000).
19. Ueda, Y., Yang, K., Foster, S. J., Kondo, M. & Kelsoe, G. Inflammation controls B lymphopoiesis by regulating chemokine CXCL12 expression. *J. Exp. Med.* **199,** 47–58 (2004).
20. Jankovic, M., Casellas, R., Yannoutsos, N., Wardemann, H. & Nussenzweig, M. C. RAGs and regulation of autoantibodies. *Annu. Rev. Immunol.* **22,** 485–501 (2004).
21. Raff, M. C., Megson, M., Owen, J. J. & Cooper, M. D. Early production of intracellular IgM by B-lymphocyte precursors in mouse. *Nature* **259,** 224–226 (1976).
22. Desiderio, S. V. *et al.* Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311,** 752–755 (1984).
23. Golby, S. *et al.* B cell development and proliferation of mature B cells in human fetal intestine. *J. Leukoc. Biol.* **72,** 279–284 (2002).
24. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336,** 1268–1273 (2012).
25. Mackie, R. I., Sghir, A. & Gaskins, H. R. Developmental microbial ecology of the neonatal gastrointestinal tract. *Am. J. Clin. Nutr.* **69,** 1035S–1045S (1999).
26. Ueda, Y., Liao, D., Yang, K., Patel, A. & Kelsoe, G. T-independent activation-induced cytidine deaminase expression, class-switch recombination, and antibody production by immature/transitional 1 B cells. *J. Immunol.* **178,** 3593–3601 (2007).
27. Lefrancois, L. & Lycke, N. Isolation of mouse small intestinal intraepithelial lymphocytes, Peyer's patch, and lamina propria cells. *Curr. Protoc. Immunol.* Ch. 3, Unit 3.19 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health grants AI020047 (to F.W.A.) and AI89972 (to D.R.W.), Lymphoma and Leukemia SCOR 7009-12 (to F.W.A.), and National Institutes of Health research contract HHSN272201000053C (to T.B.K.). D.R.W. was also supported by an award from the American Academy of Allergy Asthma and Immunology and CSL-Behring and holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund. F.W.A. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions D.R.W. and F.W.A. designed the study; D.R.W., A.J.P., M.P.G., K.C.-J., J.M.M. and R.A.P. performed experiments; R.M.M. and T.B.K. performed computational analysis of sequencing data; S.J.R. performed immunohistochemistry experiments; D.R.W. and F.W.A. wrote the paper.

Author Information Microarray data have been deposited in MIAME format into the Gene Expression Omnibus (GEO) database under accession number GSE48870, and repertoire sequencing data have been deposited into the GEO database under accession number GSE48805. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.W.A. (alt@enders.tch.harvard.edu) or D.R.W. (dwesemann@research.bwh.harvard.edu).

METHODS

Mice, immunizations and colonization. Mice harbouring the *Rag2-Gfp* knock-in fusion gene at the endogenous locus were described previously[15] and were maintained on a 129/SvJ background. Wild-type BALB/c mice were purchased from Jackson Laboratories. Swiss-Webster germ-free mice were purchased from Taconic Farms. For each germ-free/colonization experiment, littermate germ-free mice were used as controls. Germ-free status and colonization status were confirmed by Gram staining of caecal and small intestinal contents as well as microbial culture. Immunization experiments were performed with alum as described[18]. All experiments with mice followed the protocols approved by the Boston Animal Care Facility of the Children's Hospital.

Cell isolation and flow cytometry. Peyer's patches, intraepithelial lymphocytes and LP lymphocytes were isolated essentially as described[27]. Peyer's patches were excised from the small intestine, and the remaining tissue was incubated with 1× Hank's balanced salt solution with 1 mM EDTA 10% FBS PBS for 30 min and room temperature three times for intraepithelial lymphocyte extraction. Residual intestinal tissue was digested in 20% FBS RPMI with 0.05% collagenase from *Clostridium histolyticum* (Sigma) for 1 h at 37 °C three times. Intraepithelial lymphocytes and LP cells were centrifuged over Lympholyte (Cedar Lane) per manufacturer's recommendations to minimize mucus contamination. Single-cell suspensions of mesenteric lymph nodes, Peyer's patches and spleen were prepared by mashing through a cell strainer (70 μm). Cells were stained with fluorophore-conjugated mouse antibodies, and flow cytometry was performed.

qPCR with reverse transcription and microarray analysis. For qPCR, total RNA was extracted using the TRIzol method (Invitrogen) and reverse transcribed into cDNA using qScript (Quanta Biosciences). *Rag1* and *Rag2* transcripts were then quantified using TaqMan qPCR assays Mm01270936_m1 and Mm00501300_m1, respectively (Applied Biosystems). The comparative C_t method was used to quantify transcripts that were normalized with respect to *Cd19* expression (Taqman assay Mm00515420_m1, Applied Biosystems). For comparative transcriptome analysis, B cells were isolated from BM and small intestinal LP of 3-week-old mice. Cells were sorted (BD FACSAria) into TRIzol on the basis of the following cell surface markers: $CD19^+$, $B220^{low}$, GFP^+. RNA was extracted and then amplified, labelled, and hybridized to Affymetrix GeneChip Mouse Gene 1.0 ST arrays (Expression Analysis). Raw data were normalized with the RMA algorithm implemented in the Expression File Creator module from the GenePattern suite[28]. Data were visualized with the Multiplot and Hierarchical Clustering Viewer modules. All cell populations analysed were generated in triplicate from independent experiments consisting of a pool of at least 8 mice for each experiment.

Immunohistochemistry. Immunohistochemistry was performed using 4-mm-thick formalin-fixed paraffin-embedded (FFPE) tissue sections. Slides were soaked in xylene, passed through graded alcohols, and put in distilled water. Slides were pretreated with EDTA (pH 8.0) retrieval solution (Zymed) in a steam pressure cooker (Biocare Decloaking Chamber CD2008US, Biocare Biomedical) at manufacturer's recommended settings. All further steps are performed at room temperature in a hydrate chamber. The slides were blocked for endogenous peroxidase activity with peroxidase block (DAKO), washed 5 min in buffer, and followed by 20 min incubation with serum free protein block (DAKO). For TdT single staining, a polyclonal rabbit antibody (DAKO catalogue no. A3524) was applied at 1:100 dilution for 1 h at room temperature followed by washing. The detection of antibody used DAKO Rabbit Envision and DAB according to the manufacturer's directions. For TdT/B220 double staining, rabbit anti-TDT was followed with Mach-2 Rabbit AP polymer (Biocare) and developed with Vulcan Fast Red (Biocare). Subsequently, rat anti-B220 (BD Pharmingen, catalogue no. 550286), was applied for 1 h at 1:200 dilution followed by Goat anti-Rat-HRP (Millipore) and developed with DAB. All slides were counterstained with Harris haematoxylin. Stained slides were scanned at ×200 magnification using an Aperio ScanScope XT workstation (Aperio Technology). Images were visualized, annotated and microscopic distances quantified using ImageScope software (version 10.0.35.1800, Aperio Technology).

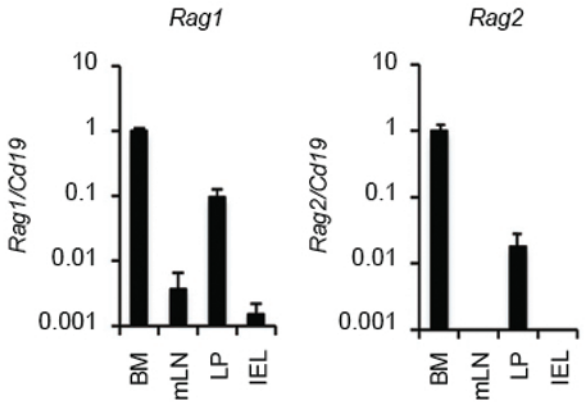
Ligation-mediated PCR. Sorted $CD19^+$ $B220^{low}$ RAG2–GFP^+ B-lineage cells from BM and LP were lysed in SDS lysis buffer (5 mM EDTA, 200 mM NaCl, 100 mM Tris-HCl pH 8.0, 0.2% SDS) with proteinase K (200 mg ml^{-1}) overnight at room temperature followed by incubation at 37 °C for 1 h. Following DNA isolation by phenol:chloroform separation and isopropanol precipitation, blunt-end ligation reactions were performed using an oligonucleotide duplex linker consisting of BW-1 (5′-GCGGTGACCCGGGAGATCTGAATTC-3′) and BW-2 (5′-GAATTCAGATC-3′). DNA was ligated overnight at 16 °C in ligation buffer (50 mM Tris pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP) and T4 DNA ligase (Promega). Ligase was inactivated by incubation for 10 min at 70 °C. Ligation reaction was diluted 1:3 in H_2O before being used for PCR. Nested PCR was used to detect ligation products resulting from both *Jκ1* and *Jκ2* double-stranded DNA breaks. The first round of amplification was performed using Qiagen Hot Star Taq (1.25 U per reaction) and primers Ko3 (5′-AGTGCCACTAACTGCTGAGAAACCT-3′) and BW-1H (5′-CCGGGAGATCTGAATTCCAC-3′). The PCR reaction was performed as follows: 95 °C for 15 min, followed by 26 cycles of 94 °C for 45 s, 57 °C for 45 s and 72 °C for 50 s, followed by 72 °C for 5 min. The second round of amplification was performed as qPCR using primers Ko (5′-CCACGCATGCTTGGAGAGGGGGTT-3′) and BW-1H primers with an internal Jκ probe (5′-56-FAM/ZEN-3-Iowa Black-TGAGGAGGGTTTTTGTACAGCCAGA-3′). Signals were normalized to actin amplified from genomic DNA (Mm00607939_s1). Each sample was calculated as a per cent of a BM standard run on each PCR plate to determine variability within BM and LP samples from three biologic replicates isolated from independent pools of 4–8 mice per experiment.

Repertoire sequencing. Total RNA was obtained from purified B cells using TRIzol reagents (Invitrogen). κ chain and μ chain cDNAs from each sample were synthesized using a SMARTer-RACE cDNA amplification kit (Clontech), according to the manufacturer's protocol. *Cμ*-specific (5′-CAGGTGAAGGAAATGGTGCT-3′) and *Cκ*-specific (5′-TTAACTGCTCACTGGATGGTG-3′) primers were used in lieu of oligo dT primers for cDNA synthesis. A total of 0.05 to 0.2 μg total RNA per sample was used. PCR was performed using Phusion DNA Polymerase (Thermo Scientific) and 12.5 μl of first-strand reactions with long and short universal primers (5′-CTAATACGACTCACTATAGGGCAAGCAGTGGTAACAACGCAGAGT-3′ and 5′-CTAATACGACTCACTATAGGGC-3′) together with either a biotinylated round-1 *Cμ* primer (5′-BIO-CTTATCAGACAGGGGGCTCTC-3′) or round-1 *Cκ* primer (5′-BIO-TCACTGGATGGTGGGAAGAT-3′) specific primers. First round PCR reaction conditions were then followed as described elsewhere[29]. PCR round-1 product sizes of 500–700 base pairs (bp) were extracted from agarose gels using a QIAquick gel extraction kit (Qiagen), enriched on streptavidin-coupled Dynabeads (Invitrogen), and purified with a QIAquick PCR purification columns (Qiagen) per manufacturer's instructions. Purified round-1 products were then subjected to a second round of PCR with nested primers containing 'A' and 'B' adaptor sequences as well as distinct 10-bp barcode sequences (to distinguish source material) using the nested universal primer (5′-CGTATCGCCTCCCTCGCGCCATCAG[unique 10-bp barcode sequence]ACGACTCACTATAGGGCAAGCAG-3′) together with either nested *Cμ* primer (5′-CTATGCGCCTTGCCAGCCCGCTCAG[unique 10-bp barcode sequence]GGGAAGACATTTGGGAAGGA-3′) or *Cκ* primer (5′-CTATGCGCCTTGCCAGCCCGCTCAG[unique 10-bp barcode sequence]TGGATGGTGGGAAGATGGAT-3′). PCR products (size 500–700 bp) were extracted from agarose gel, and 100 ng of each amplicon library was combined and used for 454 sequencing analysis. GS FLX Titanium sequencing kit XLR70 (Roche) was used for sample preparation. Data were collected at the sequencing core and at the University of Illinois. Data were analysed using the empirical Bayes procedure as described[30]. Clonotypes and clone assignments were determined using a recursive set of hypothesis tests on the equality of the V-gene segment mutation rate and that of CDR3. To control for false detection rate, comparisons were made between the same tissues of repeat experiments of BM and LP *Vκ* samples (Supplementary Fig. 13). V_H segment usage from $RAG2^+$ BM was also compared to total splenic B cell V_H segment usage to ensure that our analysis could identify the expected differences that occur due to selection between these populations.

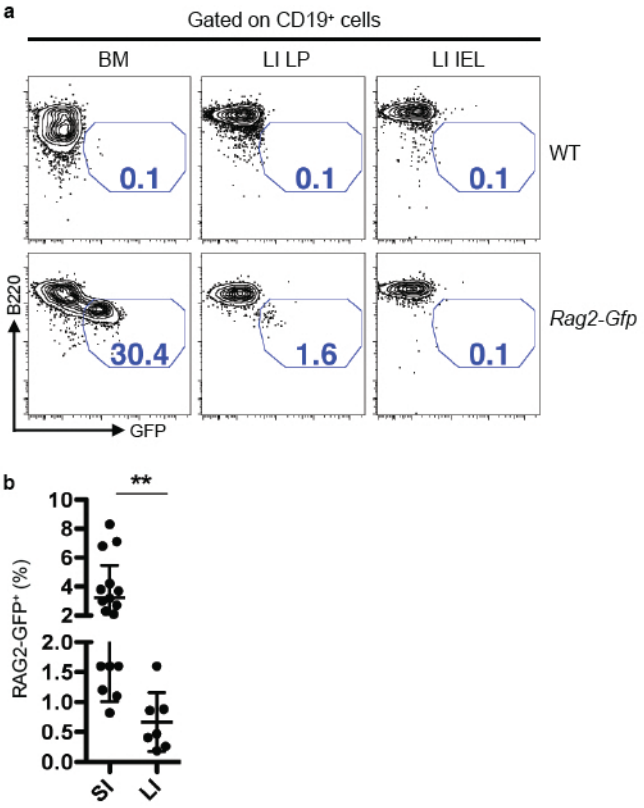
Statistics. If not otherwise stated, data were expressed as arithmetic means ± s.e.m., and statistical analyses were made by unpaired *t*-test, exact test, or χ^2 test where appropriate. $P < 0.05$ was considered statistically significant.

28. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38,** 500–501 (2006).
29. Warren, R. L. *et al.* Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res.* **21,** 790–797 (2011).
30. Kepler, T. B. Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors [v1; ref status: indexed, http://f1000r.es/z6], *F1000 Res.* **2,** 103 (2013).

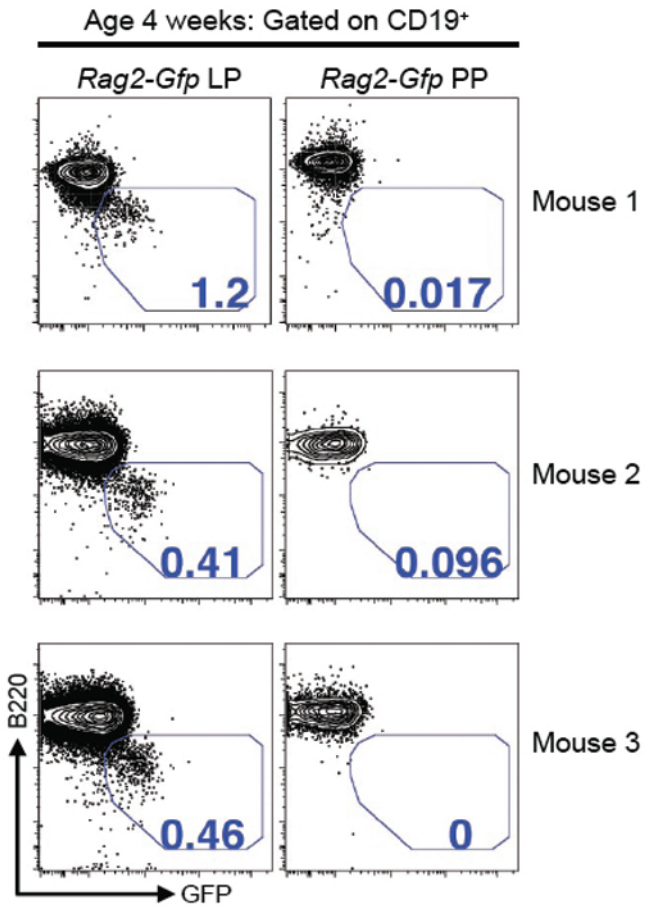
doi:10.1038/nature12496



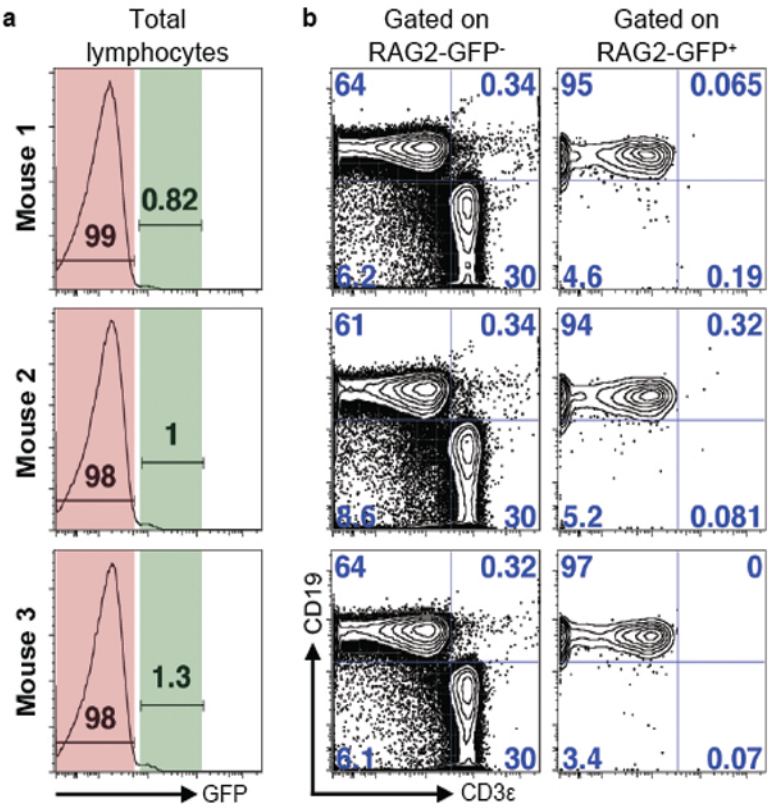
Supplementary Figure 1 | *Rag1* and *Rag2* are expressed in the small intestinal lamina propria of young wild type mice. Quantitative PCR analysis of *Rag1* and *Rag2* expression in the bone marrow (BM), mesenteric lymph nodes (mLN), small intestinal lamina propria (LP), and intraepithelial lymphocytes (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.



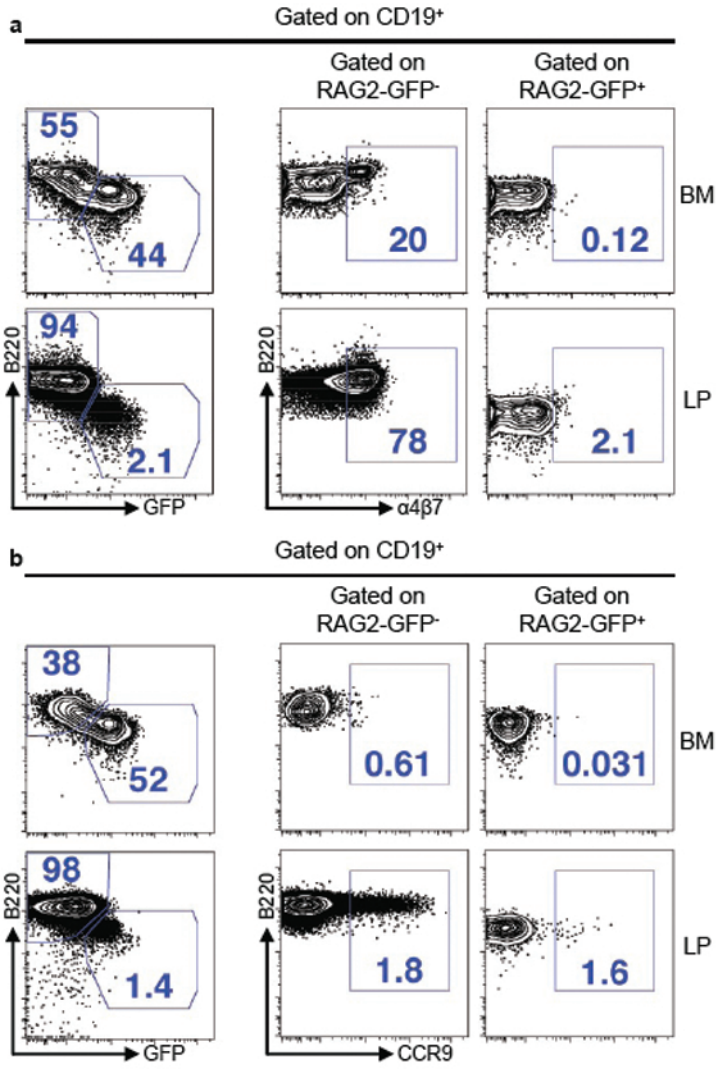
Supplementary Figure 2 | RAG2-GFP$^+$ B cells are found in the large intestine at lower levels compared with the small intestine. **a**, FACS plots of CD19$^+$ gated cells from bone marrow (BM), large intestinal intraepithelial lymphocytes (LI IEL) and large intestinal lamina propria (LI LP) from wild 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).



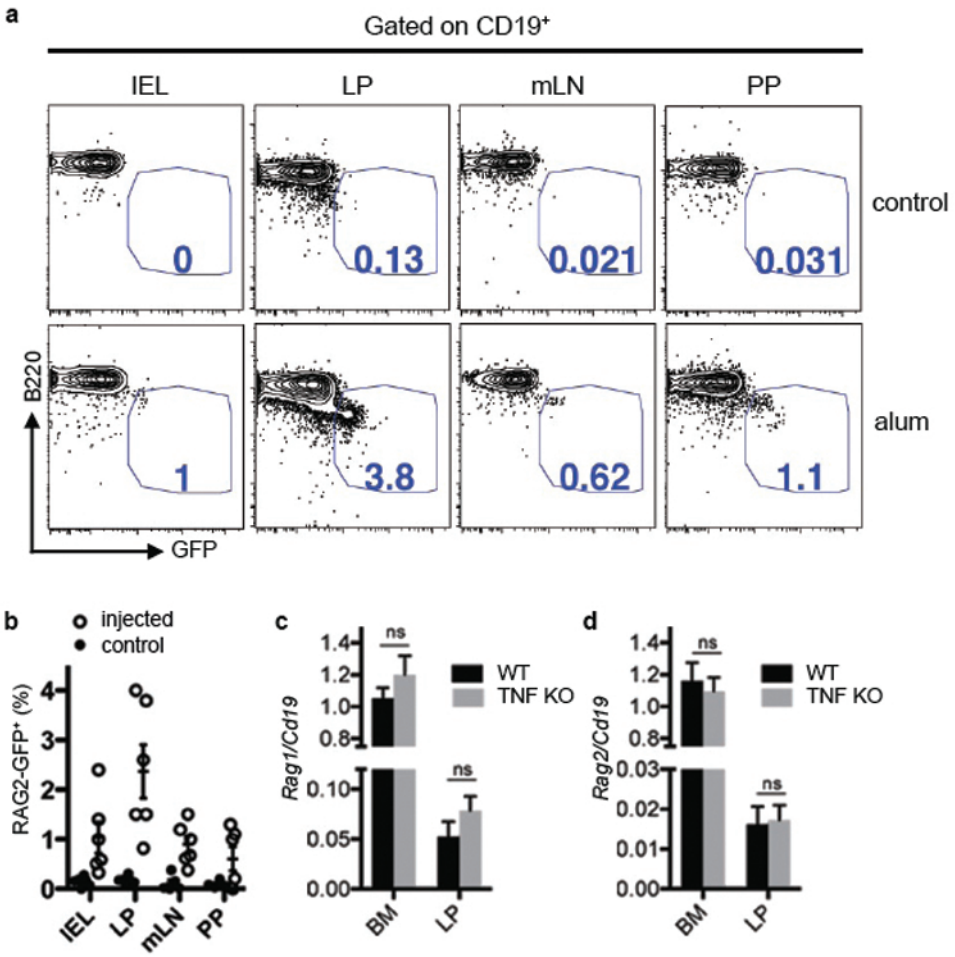
Supplementary Figure 3 | Peyer's patches do not harbor RAG2-GFP$^+$ cells. FACS plots of CD19$^+$ gated cells from the lamina propria (LP) and Peyer's patches (PP) isolated from three independent *Rag2-Gfp* mice on post-natal day 28. The numbers in the gates shown indicate percentage of B220^{low} RAG2-GFP$^+$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



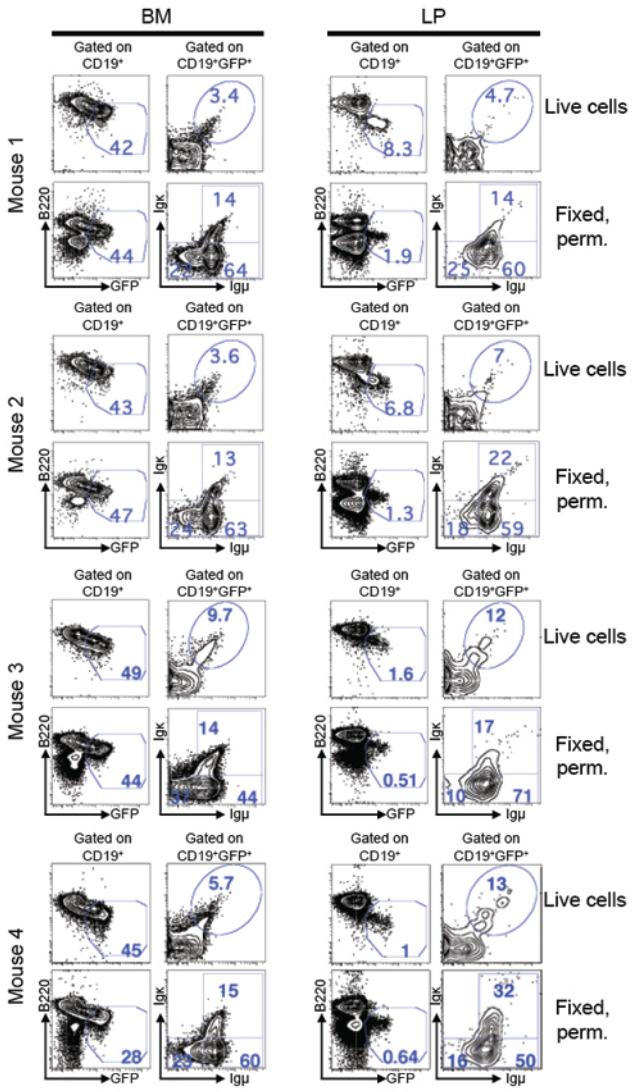
Supplementary Figure 4 | RAG2-GFP is not detected in LP T lineage cells. **a**, Histogram plots showing flow cytometric measurement of RAG2-GFP$^-$ (red) and RAG2-GFP$^+$ (green) cell populations from total small intestinal LP lymphocytes at post-natal day 18-21. Cells were stained for CD19 and the pan 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.



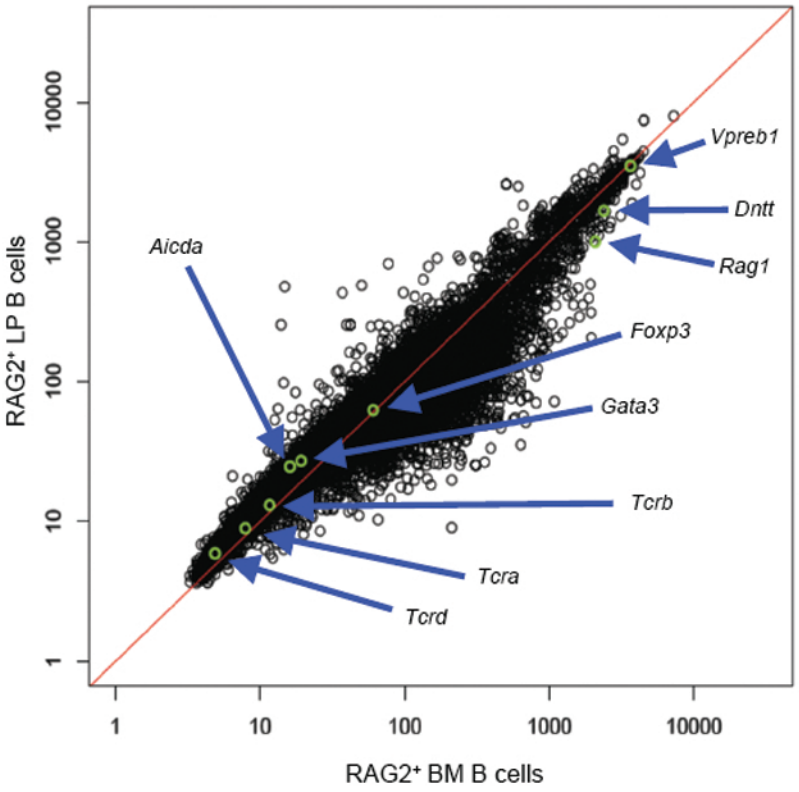
Supplementary Figure 5 | Small intestinal LP-resident RAG2-GFP$^+$ B lineage cells do not express α4β7 or CCR9. **a,b**, Both α4β7 integrin and CCR9 chemokine receptor have been implicated in lymphocyte homing to the gut[31]. Lymphocytes from *Rag2-Gfp* mice were stained for surface expression of CD19, 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.



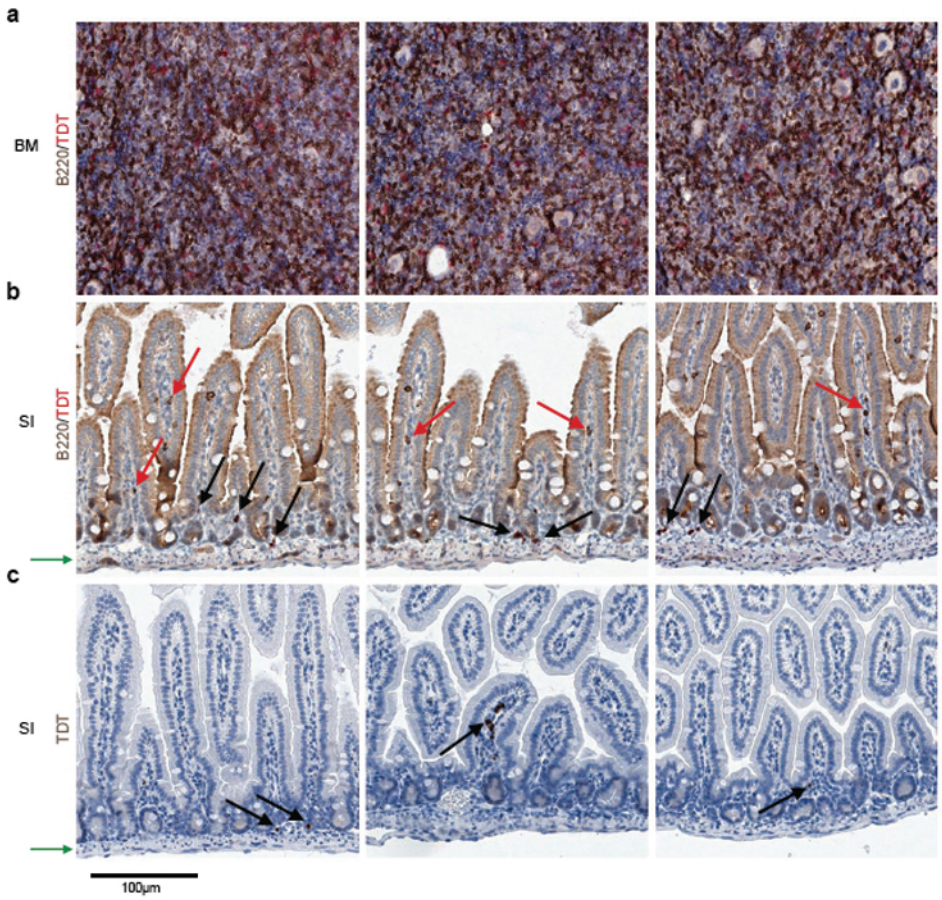
Supplemental Figure 6 | Intraperitoneal alum injection leads to the appearance of RAG2^{+} cells in the gut mucosa. **a**, FACS plots of CD19^{+} gated cells showing the percentages of RAG2-GFP^{+} B220^{low} cells from uninjected (control) or mice injected with intraperitoneal alum. **b**, Scatter dot 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).



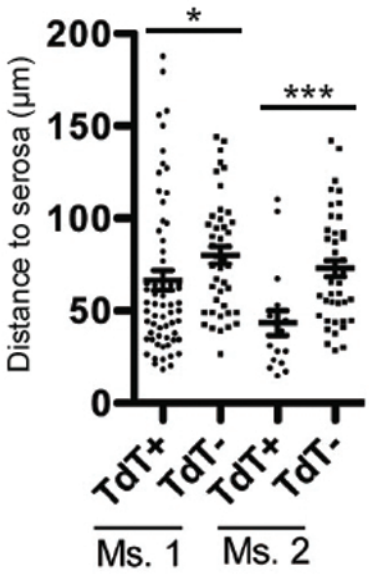
Supplementary Figure 7 | Analysis of RAG2-GFP$^+$ B lineage developmental subsets. FACS plots showing bone marrow (BM) and small intestinal lamina propria (LP) cells from post-natal day 17-24 *Rag2-Gfp* mice, which were stained live (top plots), or after fixation and permeabilization (bottom plots) for 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.



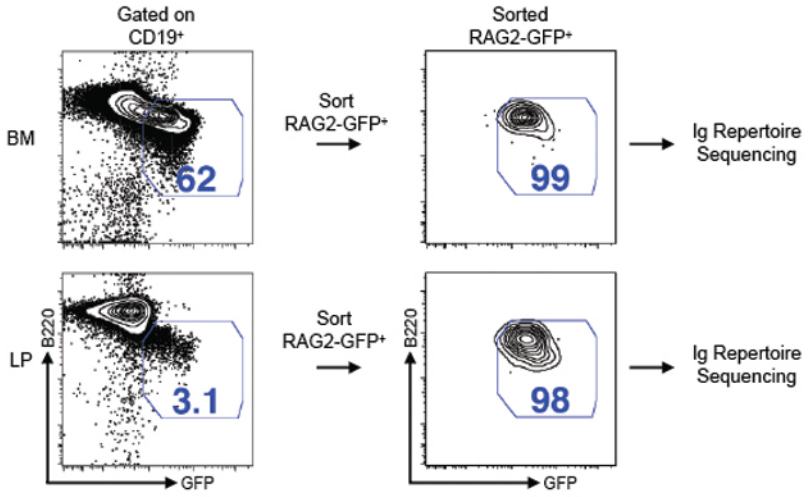
Supplementary Figure 8 | Comparative transcriptome analysis indicates a high degree of similarity between bone marrow (BM) and small intestinal lamina propria (LP) RAG2-GFP⁺ cells. Scatter plot of affymetrix 1.0 ST microarray data comparing RAG2-GFP⁺ cells from BM versus RAG2-GFP⁺ cells sorted from LP. 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.



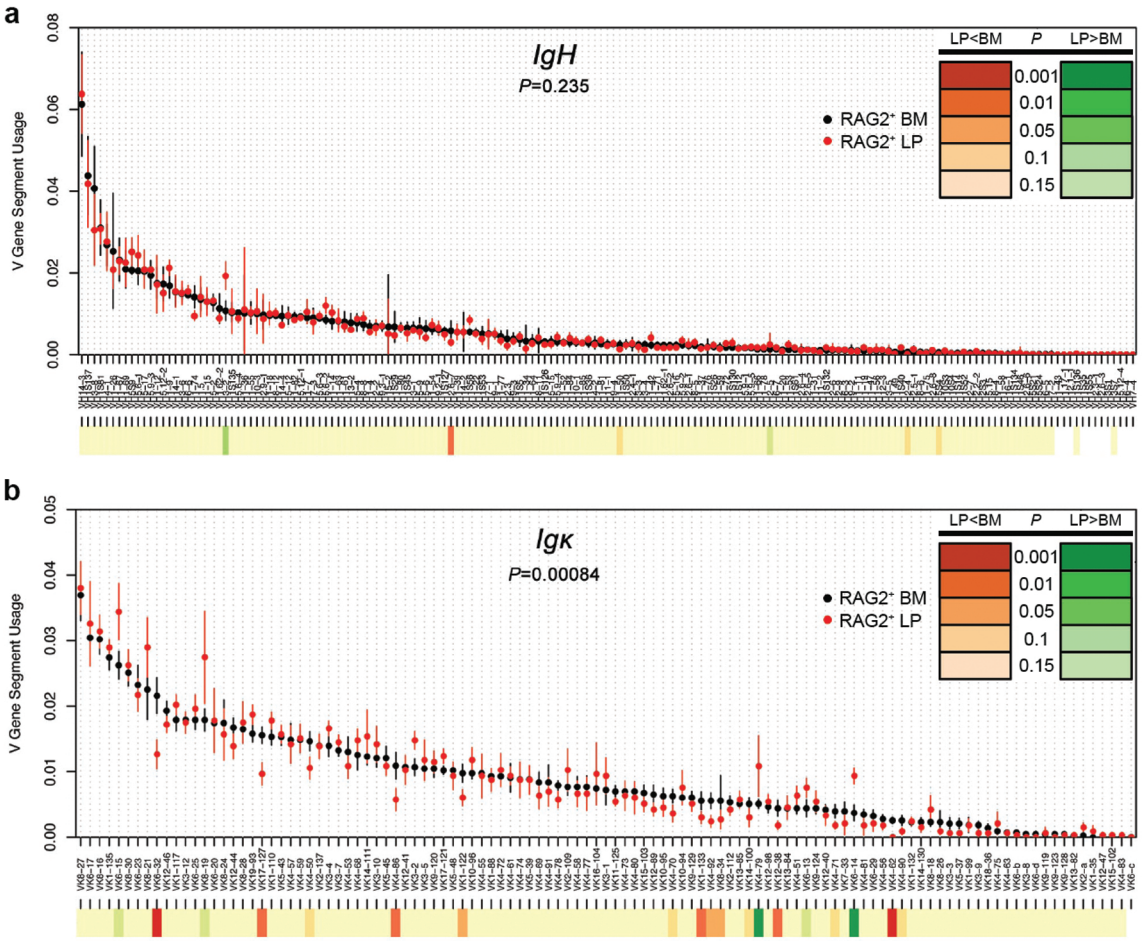
Supplementary Figure 9 | Immunohistochemistry (IHC) identifies TdT$^+$ B lineage cells in the gut lamina propria. **a, b**, IHC of paraffin-embedded section from BM (**a**) and small intestines (SI, **b**) stained with an anti-TdT antibody (red stain) plus anti-B220 antibody (brown stain). **b**, Red 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.



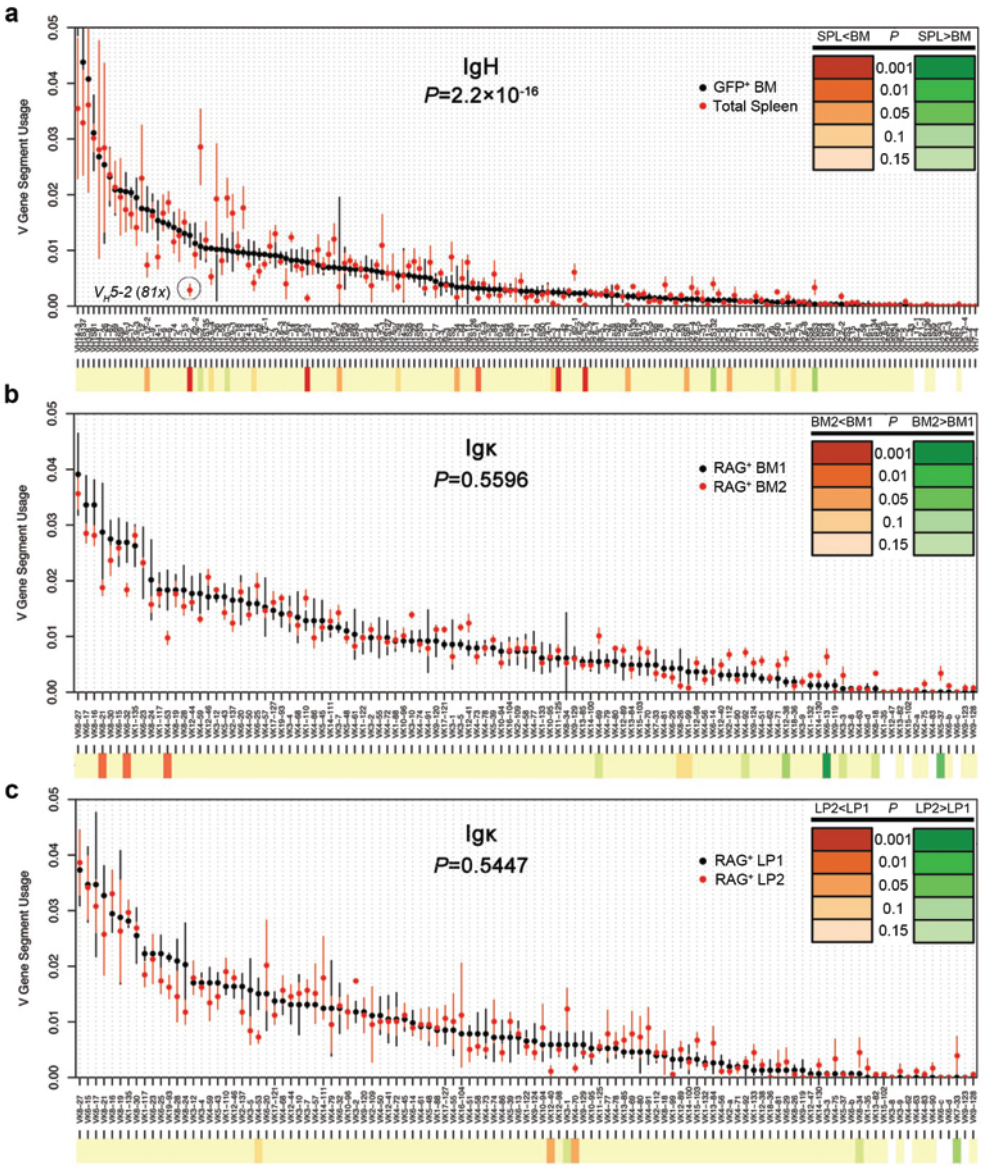
Supplementary Figure 10 | Intestinal TdT$^+$ B lineage cells occupy distinct location compared to TdT$^-$ B cells. Quantitative measurements of the distance between individual small intestinal lamina propria resident TdT$^+$ or TdT$^-$ B220$^+$ cells and the serosal (antiluminal) surface from two independent 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).



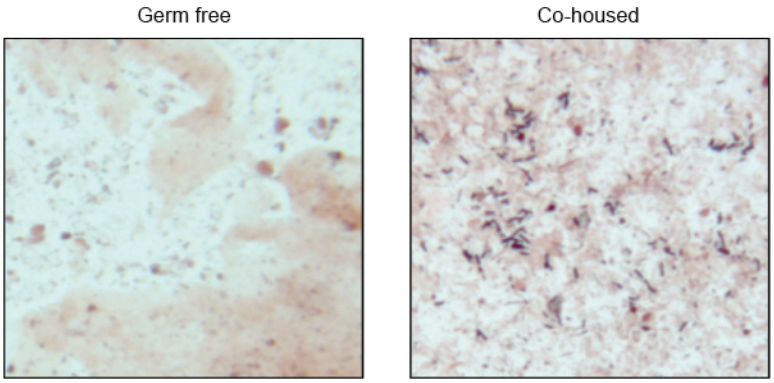
Supplementary Figure 11 | Sorting strategy for repertoire sequencing. This is a schematic diagram showing the sorting strategy to prepare samples for repertoire sequencing. The FACS plots on the left show $CD19^+$ gated cells from bone marrow (BM) and small intestinal lamina propria (LP) with additional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



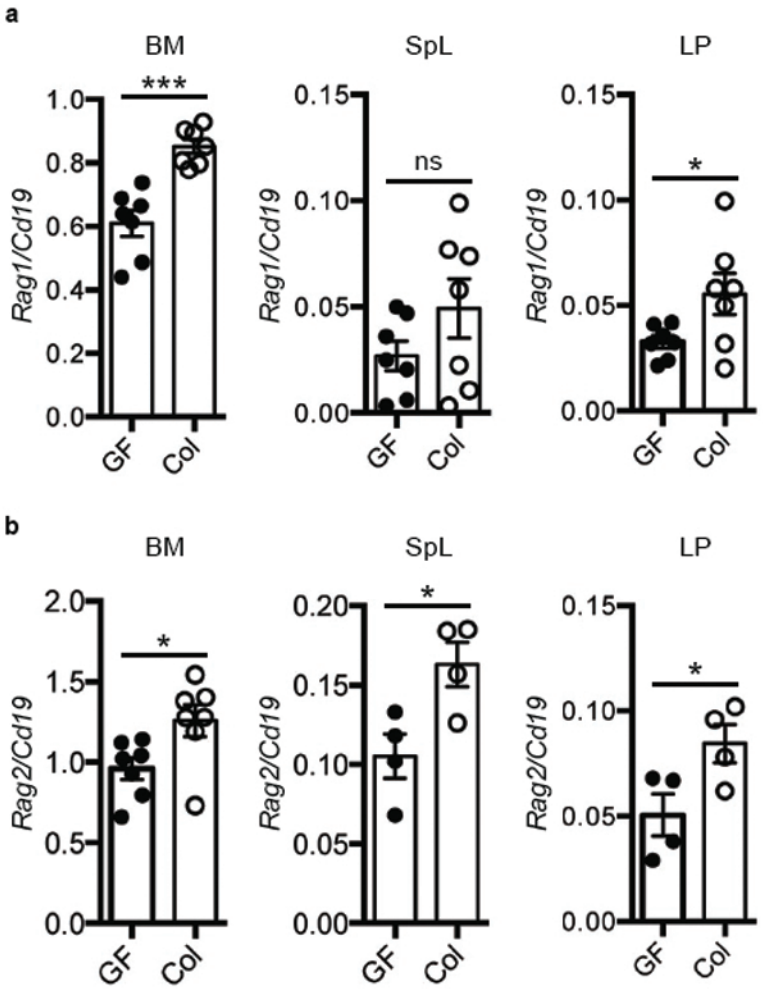
Supplementary Figure 12 | Distinct *Vκ* segment usage in RAG2$^+$ cells from BM and LP. **a,b**, Dot plots showing distribution of *IgH V* segment (V_H) (**a**) and *Igκ V* segment (*Vκ*) (**b**) usage in RAG2-GFP$^+$ cells sorted from bone marrow (BM, black dots) and small intestinal lamina propria (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.

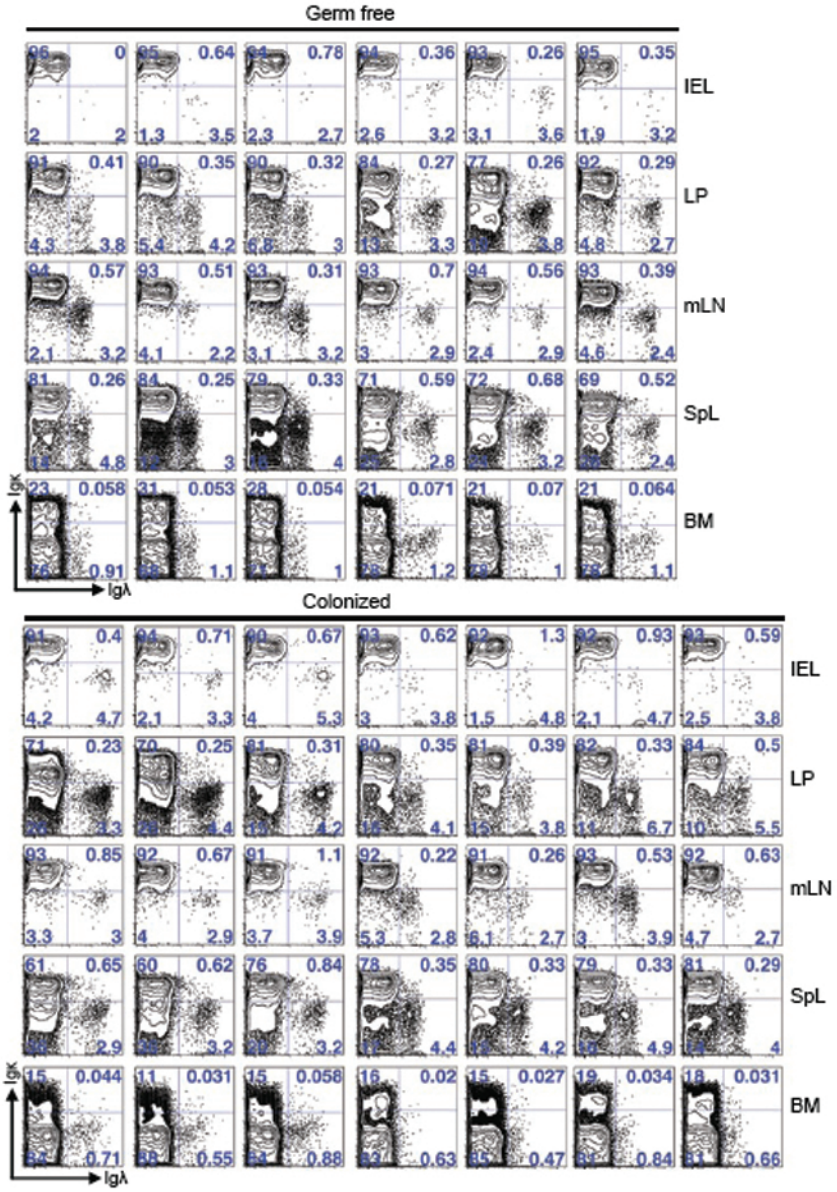


Supplementary Figure 14 | Gram stain of small intestinal contents from germ-free and co-housed mice. Photographs of gram-stained small intestinal (distal) contents from 4 wk-old germ-free, littermates that were colonized by cohousing with specific pathogen free mice for 7 days.

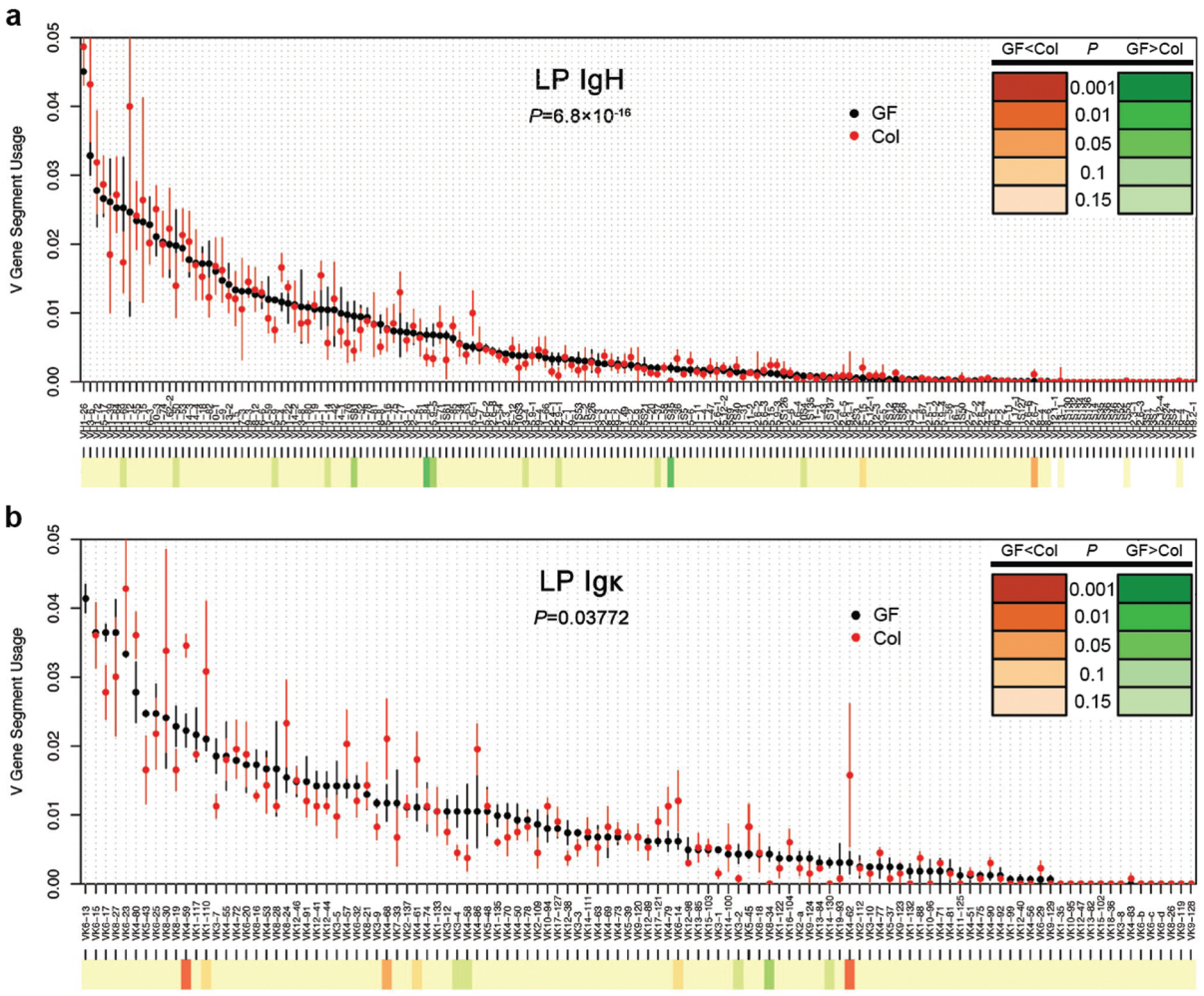


Supplementary Figure 15 | Gut colonization leads to increased gut lamina propria *Rag* expression levels. **a, b**, Bar graphs showing quantitative PCR for *Rag1* (**a**) and *Rag2* (**b**) expression in bone marrow (BM), spleen (SpL) and small intestinal lamina propria lymphocytes (LP) of 4 wk-old germ 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 $V\kappa$ 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 ($V\kappa$) 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 Vκ gene segments by the total number of in-frame V_H or Vκ gene segments represented in our sequencing data set, respectively. Individual V_H and $V\kappa$ 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).