Mucosal or systemic microbiota exposures shape the B cell repertoire

Hai Li [1*], Julien P. Limenitakis [1*], Victor Greiff [2], Bahtiyar Yilmaz [1], Olivier Schären [3], Camilla Urbaniak [4], Mirjam Zünd [1], Melissa A.E. Lawson[1], Ian D. Young [1], Sandra Rupp [1], Mathias Heikenwälder [5], Kathy D. McCoy [1], Siegfried Hapfelmeier [3], Stephanie C. Ganal-Vonarburg [1*†], Andrew J. Macpherson [1*†]

[1] Maurice Müller Laboratories (DBMR), Universitätsklinik für Viszerale Chirurgie und Medizin Inselspital, Murtenstrasse 35, University of Bern, 3008 Bern, Switzerland.

[2] University of Oslo, Dept. of Immunology, 0372 Oslo, Norway.

[3] Institute for Infectious Diseases, University of Bern, 3001 Bern, Switzerland.

[4] McMaster University Medical Centre, Hamilton, Ontario, Canada.

[5] Division of Chronic Inflammation and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany.

* These authors have made an equal contribution.

† Correspondence to: andrew.macpherson@dbmr.unibe.ch; stephanie.ganal@dbmr.unibe.ch

University Hospital Bern (Inselspital), D117, BHH, Bern 3010, Switzerland, Tel. +41 31 632 8025

FIRST PARAGRAPH

Microbiota colonization causes profound B cell stimulation and immunoglobulin induction, yet mammals colonized with many taxa have highly complex individualized immunoglobulin repertoires [1,2]. To deconstruct how the microbiota shapes the B cell pool and its functional responsiveness we have used a simplified model of defined transient microbial exposures by different taxa in germ-free mice [3]. B cell immunoglobulin repertoire development was followed by deep sequencing and in single cells. Intestinal mucosal exposure generated oligoclonal responses which differed from germ-free controls or from the diverse repertoire generated after intravenous systemic exposure. The IgA repertoire, predominantly to cell-surface antigens, did not expand following dose escalation, whereas increased systemic exposure broadened the IgG repertoire to both microbial cytoplasmic and cell-surface antigens. These microbial exposures induced characteristic immunoglobulin heavy chain B cell repertoires mainly at memory and plasma cell stages. Whereas sequential systemic exposure to different taxa diversified the IgG repertoire and facilitated alternate specific responses, sequential mucosal exposure produced limited overlapping repertoires and attrition of initial IgA binding specificities. This shows a contrast between a flexible response to systemic exposure with the need to avoid fatal sepsis, and a restricted response to mucosal exposure reflecting the generic nature of mucosal microbial mutualism.

TEXT

Mammalian immunoglobulins (Ig) are composed of heavy (H) and light (L) chains, each assembled from one of many $V_{H/L}$, (D_H) and $J_{H/L}$ gene segments during B cell development. The resulting genetic and structural immunoglobulin repertoire diversity and antigen recognition possibilities are further increased by additional nucleotide insertion during recombination, later somatic mutation, or class switch recombination from IgM and/or IgD to IgG, IgE or IgA. Selection and clonal expansion of particular B cells with an appropriate immunoglobulin specificity endows the system to respond to a huge variety of antigens. Comparisons of colonized and germ-free animals show that B cell numbers and immunoglobulin levels (especially IgG and IgA) are considerably amplified by colonization with a microbiota, some of which penetrate mucous membranes to prime systemic secondary lymphoid structures even in healthy hosts. [4,5] B cell repertoires have so far been shown to be largely unique to each individual animal [1,2], although whether this is caused by differences in microbiota composition or the sequence of colonization in each animal is unknown.

Using localized time-limited exposures of defined doses of single benign microbial taxa in germ-free animals [3], we have addressed how the B cell repertoire is shaped by microbiota exposure at mucous membranes or in systemic lymphoid tissues, and how interactions between different exposure sites or subsequent exposures to different taxa affects the outcome.

Distinct B cell Ig repertoires to intestinal microbes depending on mucosal or systemic exposure site

We used standardized doses of the live non-replicating *E. coli* strain HA107 to transitorily expose germ-free mice to live microbes, either confined within the intestine or via the blood stream. After exposure, all mice rapidly return to germ-free status, owing to HA107 auxotrophy for the essential bacterial amino acids D-alanine and diaminopimelic acid. [3] We found distinct antibody repertoires computed on the basis of CDR3 heavy chain amino acid identity [6] from entire IgA, IgG and IgM repertoires, depending on whether the exposure had been mucosal or systemic (Figure 1a-c, Extended data Figure 1a). Clonotype overlap was significantly greater within each exposure condition compared with different conditions (Figures 1a-c right

panels) and different exposure conditions segregated in unsupervised tree clustering (Figures 1a-c lower panels). A clear separation according to microbial exposure condition was also found with shared clonotypes computed as sequences encoded by the same V_H and J_H segments, with an identical nucleotide sequence in the CDR3 region [7] (Extended Data Figure 1b-g).

We used *Clostridium orbiscindens* (DMSZ 8061) as a second transitory colonizing taxon to confirm that the repertoires depend on route of microbial exposure (Figure 1d and Extended Data Figure 1h). Flow cytometric sorting prior to analysis with standardized B cell subset numbers showed that the distinct effects of different microbial exposure routes occurred in the memory and plasma cells (Figure 1e). Although single cell experiments currently only capture a fraction of the clonotype repertoire compared with bulk analysis (in our case 14±6% $\bar{x}$±SD, n=8), an MDS plot combining the heavy and light chain information also segregated according to exposure route (Figure 1f). Subsequent analysis considering single cell heavy chain and light chain clonotypes individually showed that the selective effects of microbial exposure route segregate with the Ig heavy chain (Figure 1g).

Using bacterial-specific flow cytometry, we verified a preferential isotype class switch to IgG2b after systemic exposure or IgA after mucosal exposure (Extended Data Figure 1i and j) [3]. In line with recent reports [8-12], we also found that mucosal exposure could generate a small serum IgG response (red symbols Extended Data Figure 1k). Conversely, exclusive systemic exposure generated a specific secretory IgA response in the intestine (blue symbols Extended Data Figure 1l), confirming that both the mucosal and systemic immune systems can be primed by microbial exposure from either route.

Alternative approaches are being used by different laboratories to avoid clonal frequency over-estimation due to sequencing errors. We carried out a series of control experiments to confirm that equivalent biological clustering results were obtained with either computational consolidation or use of unique molecular identifier barcodes (Extended Data Figure 2a-c). Although each set of bulk sequencing results in this paper are from samples processed together in a batch, we also verified the technical reproducibility of the same sample processed on different occasions (Extended Data Figure 2d).

Different thresholds of Ig priming according to microbial exposure route

In our non-replicating transitory exposure system in germ-free mice, test doses can be calibrated without further expansion in vivo. We showed that mucosal doses of 10^6-10^8 colony forming units (c.f.u) were needed in order to reshape the mucosal IgH repertoire (Extended Data Figure 3a, e and f) in line with the high doses known to induce specific IgA binding (Extended Data Figure 3b and [3]). In contrast, the systemic IgH repertoire was reshaped by as little as 10,000 c.f.u. (Extended Data Figure 3c, g and h). The doses required for detectable humoral responses were generally at least an order of magnitude greater than for repertoire programming (Extended Data Figure 3b and d).

Differential B cell targeting of microbial subfractions according to in vivo exposure route

To ask why Ig heavy chain clonotype repertoires differed depending on mucosal or systemic microbial exposure, we investigated the targets for bacterial antigen binding according to exposure route. Intestinal IgA bound most effectively to the purified bacterial membrane fraction of *E. coli*, whereas serum IgG bound to both membrane and cytoplasmic fractions (Figure 2a). Ribosomal proteins of *E. coli* dominated the cytoplasmic binding of serum IgG antibodies and were confirmed in a second context by proteomic analysis

in SPF mice systemically exposed to *Enterobacter cloacae* (Extended Data Figure 4a-c). Increased exposure of bacterial membranes relative to the cytoplasmic contents in the mesenteric lymph nodes after intestinal treatment was shown by higher lipopolysaccharide/16S ribosome ratios compared with the spleen after systemic treatment (Extended Data Figure 4d). These differences may be caused by the intestinal efficiency of bacterial lipid uptake compared with bacterial cytoplasmic protein contents, whereas systemic exposure would prime equivalent immune responses to membrane and cytoplasmic fractions after opsonization and phagocytosis [13,14]. Supporting this, B cell clonotype repertoires after systemic (but not mucosal) challenge with ultrasound lysed bacteria overlapped with the repertoires after exposure with intact bacteria (Extended Data Figure 4e and f). Conversely, in the absence of local anatomic processing, the intrinsic antigen-specific and -nonspecific responses to membrane or cytoplasmic bacterial antigens presented directly *in vitro* did not differ depending on mesenteric or splenic B cell origin (Extended Data Figure 4g). We concluded that the different repertoires from mucosal or systemic B cell priming are a consequence of the way in which microbial membrane and cytoplasmic antigens are processed and presented in the different anatomic compartments.

IgA repertoire restriction and IgG repertoire expansion with increased microbial exposure doses

IgA has previously been found to be oligoclonal [15,16]. We next followed how repertoire expansion or restriction developed according to isotype, exposure route and dose. Whereas IgG diversity generally increased considerably after systemic exposure, no IgA diversity increase in any tissue was found after maximal mucosal exposure (Figure 2b), and both were considerably less than IgM diversity (Extended Data Figure 5a). As the mucosal exposure dose increased, CDR3 clonotype diversity in the IgA heavy chain became more restricted, particularly in the mesenteric lymph nodes (Figure 2c). By comparison, IgG heavy chain diversity in both the spleen and MLN was maintained or further increased with higher levels of systemic exposure (Figure 2c). We also confirmed directly in sorted cells that repertoire diversity is increased after systemic microbial exposure in IgG memory and plasma cells but not after either route of exposure in IgA plasma cells (Figure 2d). Transitory exposure by both the systemic and mucosal routes only generated small numbers of somatic mutations which were not significantly increased by increasing the transitory exposure doses (Extended Data Figure 5b and c). The converse of the breadth of the isotype repertoires is the degree of clonal relatedness, which was higher for IgA than for IgG2b (Extended Data Figure 5d and e). Mucosal and systemic exposure clearly have different physical constraints in terms of the epithelial barrier and scalability of local antigen processing and presentation to B cells. Nevertheless, our experiments show that increased mucosal microbial doses only narrow the IgA repertoire whereas the systemic IgG repertoire is progressively diversified as the exposure dose is increased.

To compare directly how immunoglobulin heavy and light chain sequences evolve in different animals during microbial exposure at the different sites, we analysed single cell networks of paired heavy and light chains in naïve and class-switched B cells. As shown in Figures 1g, 2e and f, and Extended Data Figure 6a and b, light chain diversity made only a minor contribution to the networks. IgA immunoglobulins in the mesenteric lymph nodes after intestinal exposure predominantly showed heavy chain connectivity including between germ-free and mucosally-stimulated animals, whereas IgG networks from the spleen of systemically exposed animals had very few germ-free sequences (Figure 2e and f). By contrast, germ-free immunoglobulin sequences were present in our analysis of naïve B cells and intercalated into the networks of both mucosally and systemically exposed animals (Extended Data Figure 6a and b). There was spontaneous activation and class-switch

recombination to IgA, even in the mesenteric lymph nodes of germ-free animals, whereas splenic B cells became significantly activated after systemic microbial exposure (Extended Data Figure 6c-f). We concluded that there are clear relationships of class-switched immunoglobulin sequences between different microbe exposed animals in both the mesenteric lymph nodes and the spleen. In diversely colonized animals polyspecific antibodies have been shown to arise from specificities in the germ-free repertoire [16]. We now show that the developing IgA repertoire characteristic of mucosal exposure is related to the class-switched Ig sequences present in germ-free animals.

Oral sensitization of systemic immunoglobulin responses

Since mucosal or systemic microbial exposures generate different B cell repertoire responses with different thresholds, we asked whether exposure at one site could determine the characteristics of a secondary response at the other. The real-world situation is that there are variable degrees of sequential exposure in both mucosal and systemic compartments.

Oral feeding of a soluble protein antigen tolerises subsequent B- and T cell responses to systemic exposure with the same antigen [17]. In contrast, a sequence of microbial mucosal exposure followed by systemic exposure reduced the threshold for the systemic response by several orders of magnitude compared with intestinal exposure alone (Figure 3a, Extended Data Figure 7a). This effect is T cell dependent as the enhanced systemic response could be abrogated through CD4 T cell depletion at the time of the initial intestinal exposure (Extended Data Figure 7c-f). Although systemic responses were sensitized by prior mucosal exposure to the same microbe, this did not work the other way around, as systemic exposure did not reduce the subsequent high threshold for mucosal IgA responses primed by the mucosal route (Figure 3b, Extended Data Figure 7b).

Given that individual mucosal and systemic exposures lead to different B cell repertoires, we asked whether the order of sequential exposure (systemic→mucosal or *vice versa*) makes a difference. This showed that for splenic IgM plasma cells the site of first exposure determined the resulting repertoire even after subsequent exposure at a different site (Figure 3c), although the priming may be dominated by the site of first exposure.

Functional effects on the resultant repertoire after sequential exposures to different microbes

Our final question was how IgA and IgG repertoires build up with successive transitory exposures to different microbes. Mucosal exposure to non-replicating *Salmonella typhimurium* HA218 (engineered from a non-invasive non-virulent parent strain on the same principles as *E. coli* HA107 [18]) attenuated an established mucosal IgA response to *E. coli* HA107, or *vice versa* with the opposite order (Figure 3d, e and Extended Data Figure 8a). In contrast, the broader systemic IgG response to either microbe given systemically was unrestricted irrespective of exposure order (Figure 3h,i and Extended Data Figure 8b). The CDR3 clonotype repertoires were also biased by the latest mucosal exposure, whereas after sequential systemic exposures the clustering was distinct regardless of exposure order (Figure 3g and k). The clonal relatedness and diversity of IgA showed no significant increase after successive treatments, contrasting with increased IgG clonal diversity and relatedness in the spleen as successive systemic treatments were administered (Figure 3f and j). These results are consistent with greater flexibility of new systemic IgG specificities to accommodate a response to another microbe to avoid systemic sepsis, whereas IgA mucosal protection requires more generic and likely lower affinity responses [2,16] which adapt sequentially to many different varieties of overlapping antigen exposure [2,19-21].

In this paper we have shown that the route and order of microbial antigen exposures determine the repertoire and functional outcome for B cell immune responses. In general, the systemic exposure thresholds are lower, yet trigger responses that expand and diversify the B cell repertoire in contrast with mucosal exposure. This is consistent with the IgA repertoire building on an evolutionarily-determined baseline, as some of the clonotypes overlap with immunoglobulin sequences from defined natural antimicrobial specificities (Extended Data Table 1). These different functional demands from B cell immunity offer an explanation for the different characteristics of responses of the B cell system against challenges from the microbiota in different host compartments, the expansion of the same B cell clone in multiple Peyer's patches [22] and poor responses to the mucosal route for vaccination where limited hygiene results in an increased burden of environmental microbes and recurrent mucosal infections [23,24].

ACKNOWLEDGEMENTS

We would like to thank Mercedes Gomez de Agüero, Tosso Leeb, Pamela Nicholson, Markus Geuking and Daniel Candinas. The Clean Mouse Facility is supported by the Genaxen Foundation, Inselspital and the University of Bern. This work was funded by the Swiss National Science Foundation (SNSF 310030B_160262, SNSF 310030_179479) and the European Research Council (H2020 ERC-2016-ADG HHMM_Neonates, Grant Agreement: 742195) to A.J.M. S.C.G.V. was funded by a Marie Curie Intra-European Fellowship (FP7-PEOPLE-2013-IEF Project No. 627206) and a long-term fellowship from the European Molecular Biology Organization. V.G. acknowledges the support of UiO:LifeSciences Convergence Environment Immunolingo and EU H2020 iReceptorplus (#825821). S.H. was funded by SNSF grant 169791 and ERC StG 281904. K.D.M. was funded by the ERC StG 281785. M. H. was funded by an ERC CoG (HepatoMetaboPath). J.P.L. was supported by a SystemsX Transition Postdoc Fellowship (TPdF2013/139).

AUTHOR CONTRIBUTIONS

H.L., J.P.L., S.C.G.V and A.M conceived the study, interpreted data and wrote the manuscript. H.L. and S.C.G.V. performed most experiments. J.P.L. and V.G carried out computational analysis. B.Y., M.A.E.L., S.R., M.H. and K.D.M. helped with repertoire experiments. O.S. and S.H. contributed bacterial strains and culture preparation. C.U., M.Z. and I.D.Y. helped with ex-vivo analyses of bacterial fractions and cellular responses.

COMPETING INTERESTS

The authors have no competing interests.

SUPPLEMENTARY INFORMATION

Summary table containing a) summary of sequenced data in the paper; b) top 100 clonotypes in Extended Data Figure 1e-h; c) top 100 clonotypes in Extended Data Figure 3 a and c; and d) SI Figure 1: gating strategy.

REFERENCES

1 Lindner, C. *et al.* Age, microbiota, and T cells shape diverse individual IgA repertoires in the intestine. *J Exp Med* **209**, 365-377, (2012).

2 Lindner, C. *et al.* Diversification of memory B cells drives the continuous adaptation of secretory antibodies to gut microbiota. *Nat Immunol* **16**, 880-888, (2015).

3 Hapfelmeier, S. *et al.* Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science* **328**, 1705-1709, (2010).

4 Berg, R. D. Bacterial translocation from the gastrointestinal tract. *Adv Exp Med Biol* **473**, 11-30, (1999).

5 Lockhart, P. B. *et al.* Bacteremia associated with toothbrushing and dental extraction. *Circulation* **117**, 3118-3125, (2008).

6 Xu, J. L. & Davis, M. M. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* **13**, 37-45, (2000).

7 Soto, C. *et al.* High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* **566**, 398-402, (2019).

8 Koch, M. A. *et al.* Maternal IgG and IgA Antibodies Dampen Mucosal T Helper Cell Responses in Early Life. *Cell* **165**, 827-841, (2016).

9 Gomez de Aguero, M. *et al.* The maternal microbiota drives early postnatal innate immune development. *Science* **351**, 1296-1302, (2016).

10 Zeng, M. Y. *et al.* Gut Microbiota-Induced Immunoglobulin G Controls Systemic Infection by Symbiotic Bacteria and Pathogens. *Immunity* **44**, 647-658, (2016).

11 Chen, Y. *et al.* Microbial symbionts regulate the primary Ig repertoire. *J Exp Med* **215**, 1397-1415, (2018).

12 Wilmore, J. R. *et al.* Commensal Microbes Induce Serum IgA Responses that Protect against Polymicrobial Sepsis. *Cell Host Microbe* **23**, 302-311 e303, (2018).

13 Pepys, M. B. Role of complement in induction of antibody production in vivo. Effect of cobra factor and other C3-reactive agents on thymus-dependent and thymus-independent antibody responses. *J Exp Med* **140**, 126-145, (1974).

14 Sorman, A., Zhang, L., Ding, Z. & Heyman, B. How antibodies use complement to regulate antibody responses. *Mol Immunol* **61**, 79-88, (2014).

15 Stoel, M. *et al.* Restricted IgA Repertoire in Both B-1 and B-2 Cell-Derived Gut Plasmablasts. *J Immunol* **174**, 1046-1054, (2005).

16 Bunker, J. J. *et al.* Natural polyreactive IgA antibodies coat the intestinal microbiota. *Science* **358**, eaan6619, (2017).

17 Mowat, A. M., Faria, A. M. & Weiner, H. L. in *Mucosal Immunology* Vol. 1 (eds J. Mestecky *et al.*) 487-537 (Elsevier, 2005).

18 Pfister, S. P. *et al.* Uncoupling of invasive bacterial mucosal immunogenicity from pathogenicity. *Nat Commun* **11**, 1978, (2020).

19 Boursier, L., Dunn-Walters, D. K. & Spencer, J. Characteristics of IgVH genes used by human intestinal plasma cells from childhood. *Immunology* **97**, 558-564, (1999).

20 Casola, S. *et al.* B cell receptor signal strength determines B cell fate. *Nat Immunol* **5**, 317-327, (2004).

21 Dunn-Walters, D. K., Boursier, L. & Spencer, J. Hypermutation, diversity and dissemination of human intestinal lamina propria plasma cells. *Eur J Immunol* **27**, 2959-2964, (1997).

22 Bergqvist, P. *et al.* Re-utilization of germinal centers in multiple Peyer's patches results in highly synchronized, oligoclonal, and affinity-matured gut IgA responses. *Mucosal Immunol* **6**, 122-135, (2013).

23 Levine, M. M. Immunogenicity and efficacy of oral vaccines in developing countries: lessons from a live cholera vaccine. *BMC Biol* **8**, 129, (2010).

24 Valdez, Y., Brown, E. M. & Finlay, B. B. Influence of the microbiota on vaccine effectiveness. *Trends Immunol* **35**, 526-537, (2014).

Figure 1: Antibody repertoires in mucosal and systemic tissues following transitory oral or systemic exposure with a commensal microbe

Germ-free mice were either orally or systemically primed three times every other day by intragastric (10^{10} c.f.u.) or intravenous (10^{8} c.f.u.) *E. coli* HA107 (a-c, e-g), or *Clostridium orbiscindens* (d), and compared with germ-free controls. Immunoglobulin (Ig) repertoire sequencing at 21 days for IgA (a, d), IgG2b (b), and IgM (c) in ileum, mesenteric lymph nodes (MLN), bone marrow (BM) and spleen (SPL) was performed. Left: MDS plot based on Ig heavy chain CDR3 amino acid sequences of the entire repertoire. Right: Tukey plots of clonal overlap between samples of different priming conditions or the same priming condition (sample comparisons from the same mouse were excluded) are shown with arithmetic mean and whiskers extending to 1.5x interquartile range. Two-sided Wilcoxon rank-sum tests were performed with P_{adj}-values as shown. Lower: Hierarchical clustering. Dendrogram branch lengths show distance between repertoires based on CDR3 amino acid sequence similarity with vertical dotted lines designating principal unbiased separations. (e) IgM, IgA and IgG2b of indicated sorted B cell populations at 21 days. Hierarchical clustering dendrograms from the indicated sorted cell populations, branch length shows the distance between CDR3 amino acid repertoires. (f, g) Single cell VDJ repertoire sequencing analysis. (f) MDS plot of distance between repertoires based on combined Ig heavy and light chain CDR3 amino acid sequences. Grey shading visually indicates GF repertoires. (g) Hierarchical clustering of single cell heavy or light chain Ig repertoires from mesenteric lymph nodes and spleen. Data are representative of nine (a-c), two (e) or three (f, g) independent experiments. All datapoints represent organs from individual mice.

Figure 2: Differences between B cell repertoires following systemic or mucosal exposure

Germ-free mice treated with *E. coli* HA107 at the indicated doses either orally or systemically as in Figure 1. (a) Binding of systemic or intestinal antibodies to *E. coli* membrane versus cytoplasmic fractions evaluated using ELISA ($\bar{x}$±SD, n=6, two-sided unpaired t-test). (b) Rarefaction plots of Ig heavy chain sequencing at 21 days for IgA and IgG2b in MLN and spleen. (c) Ig heavy chain sequencing at 21 days for IgA in MLN and IgG2b in spleen (n=3 mice for each bacterial exposure group and each defined dose, n = 15 germ-free controls, * p=0.036 MLN IgA both cases, p=0.046 spleen and p=0.036 MLN IgG2b, two-sided Wilcoxon rank-sum test). (d) IgM, IgG2b and IgA heavy chain sequencing of splenic sorted B cell populations at 21 days, n=6 for each condition. Tukey plots indicate calculated number of clonotypes per 1000 cells in B cell populations. P=0.003 two-sided Wilcoxon rank-sum test. (e, f) Single cell sequencing analysis. Networks built on combined heavy and light chain CDR3 amino acid sequences each from single spleen IgG- (e) or MLN IgA-expressing (f) B cells (singlets excluded: n=1601 and 670 for E and F respectively), edges show Levenshtein distance =1 or 2. Networks show immunoglobulin sequence relationships within and between different mice in the same experiment, colour-coded according to mouse and treatment. Pie charts (left) indicate the percentage of edge connections that are based on Levenshtein distance 1 on the light chain compared to the heavy chain, or (right) indicate the distribution of edge connections between individual B cells within or between exposure conditions. Tukey plots in c and d are shown with arithmetic mean and whiskers of 1.5x interquartile range. Data are representative of two independent experiments. All datapoints are from individual mice.

Figure 3: Antimicrobial antibody responses following combined mucosal and systemic exposure or exposure to two different taxa.

(a, b) Germ-free mice were intestinally (a) or systemically (b) exposed with *E. coli* HA107 on alternate days or remained germ-free throughout. On day 21, all mice were intravenously (a) or intestinally (b) exposed with indicated doses of HA107 (schema Extended Data Figure 7a, b). Flow cytometric analysis of specific bacterial binding of serum IgG2b (a) or intestinal IgA (b) on day 42. (c) Germ-free mice were primed as in a and b. Ig-Seq for IgM of sorted splenic plasma cells. MDS plot of repertoire separation. (d-g) Germ-free mice were either mucosally primed with three doses of *E. coli* HA107 or *Salmonella typhimurium* HA218 or remained germ free. At day 21, half of each group received a second schedule of priming with the opposite taxon, n=3 for each condition. (schema Extended Data Figure 8a). (d, e) Bacterial flow cytometry of *E. coli* HA107-specific (d) or *S. typhimurium* HA218-specific (e) intestinal IgA at d42. (f) Comparisons of mean repertoire diversity and median percentage of expanded clonotypes within the IgA repertoires in the spleen and mesenteric lymph nodes. (g) Hierarchical clustering of full-length IgA heavy chain sequences. (h-k) Experimental design as (d-g) except both reversible taxa given systemically, n=3 for each condition (schema Extended Data Figure 8b). (h, i) Bacterial flow cytometry for serum IgG bacterial binding. (j) Comparisons of mean repertoire diversity and median percentage of expanded clonotypes within the IgG2b repertoires in the spleen and mesenteric lymph nodes. (k) Hierarchical clustering of full length IgG2b heavy chain sequences. (f,j) Unpaired two-tailed t-test was performed on the number of expanded clonotypes: Bars show ranges for each dimension. Data are representative of three (a) or two (b-j) independent experiments. All datapoints are from individual mice.

Extended Data Figure 1: Mucosal and systemic exposure differentially shape the repertoires of the various Ig isotypes

Germ-free mice were either orally or systemically primed three times every other day by intragastric (10^{10} c.f.u.) or intravenous (10^{8} c.f.u.) delivery of *E. coli* HA107 (a-g) or *Clostridium orbicindens* (h) and compared with germ-free controls. Immunoglobulin heavy chain repertoire clonotypes at day 21 are defined by V and J segment usage in combination with CDR3 nucleotide sequences, except for panel A where clonotypes are based on CDR3 amino acid sequences as in Figure 1. (a) MDS plot of IgA, IgG2b and IgM showing distinct isotype repertoires (in each case all isotypes were sequenced from each individual sample from at least 3 mice). (b-d) MDS plot from ileum, mesenteric lymph nodes (MLN), bone marrow (BM) or spleen (SPL) of immunoglobulin repertoire sequencing for IgA (b), IgG2b (c) and IgM (d). (e-h) Heat maps showing the 100 most abundant non-unique clonotypes for IgA (e, h), IgG2b (f), IgM (g). Clonotype specifics for each panel are shown in the supplementary information. Samples of ileum, BM, MLN and SPL from each mouse are included. Individual mice are colour-coded on the x-axis of heatmaps. (i, j) Immunoglobulins from intestinal wash of intestinally exposed mice (I) or from serum of systemically exposed mice (j) on day 21 were assessed in bacterial flow cytometry for specific IgM, IgG1, IgG2b, IgG2c, IgG3 or IgA binding to *E. coli* HA107. (k, l) *E. coli* HA107 was incubated with day 21 serum or intestinal wash Ig of the three groups of mice followed by detection of bound murine IgG2b (k) or IgA (l) by flow cytometry. All datapoints are from organs of individual mice.

Extended Data Figure 2: Comparison of computational correction method with unique molecular identifiers (UMIs) for PCR artifacts and sequencing reproducibility on different occasions

Germ-free mice were either orally or systemically primed three times every other day by intragastric (10^{10} c.f.u.) or intravenous (10^{8} c.f.u.) exposure to *E. coli* HA107 and compared with germ-free controls. Immunoglobulin repertoire sequencing at 21 days for IgA in MLN and IgG2b in spleen was performed in parallel from the same RNA samples with 2 different primers containing different UMIs or primers without UMIs. (a, b) MDS plot of MLN IgA (a) or spleen IgG2b (b) repertoires showing Euclidean distance between points representing the similarity between results whether computational correction (filled symbols) or UMI correction (open and cross symbols for the different UMIs) was used. (c) Heatmap based on CDR3 sequence identity reflecting similarity between UMI-corrected and computational-corrected IgG2b repertoires. (d) Identity matrix based on CDR3 sequences of technical replicates of the same biological sample that had been used on two separate occasions for the entire pipeline of cDNA preparation, IgA amplicon PCR, library preparation and MiSeq sequencing. The distinctions between replicate mice and technical repeats are shown in the diagram.

Extended Data Figure 3: Threshold differences for shaping systemic or mucosal B cell repertoires and induction of antibody responses

Reversible *E. coli* HA107 were given to germ-free mice at the indicated doses either orally (a, b, e, f) or systemically (c, d, g, h) as in the legend to Figure 1. (a, c) Ig repertoire analysis at 21 days showing heatmaps of the top 100 non-unique clonotypes for IgA and IgG2b in MLN and spleen based on V, J segment usage combined with CDR3 nucleotide sequences. (b, d) Flow cytometric analysis of *E. coli*-specific Ig binding from intestinal or serum samples from the corresponding mice in A and C respectively. (e-h) Hierarchical clustering of the indicated immunoglobulin repertoires. Dendrogram of samples, branch length shows the distance between repertoires based on CDR3 amino acid sequence similarity of entire B cell repertoires. The dotted lines indicate the principal separations in unsupervised analyses as an independent assessment of thresholds of exposure required for repertoire shaping. Each column within the heatmaps or each dilution series from antibody-binding bacterial flow cytometry are from individual mice, in every case 3 mice were used for every experimental condition studied. The supplementary information contains a table specifying the top 100 clonotypes in each case for panels a and c. Data are representative of two independent experiments

Extended Data Figure 4: Differences in microbial antigen processing and presentation in the mucosal and systemic compartments

(a) Reversible *E. coli* HA107 were given to germ-free mice at the indicated doses either orally or systemically as in the legend to Figure 1. Binding of systemic or intestinal antibodies to *E. coli* non-ribosomal versus ribosomal proteins of the cytoplasmic fraction evaluated using ELISA ($\bar{x}$±SD, n=6, two-sided unpaired t-test). (b, c) *Enterobacter cloacae* proteins were either separated by FPLC or subject to ribosomal protein purification, prior to 12% SDS-polyacrylamide gel electrophoresis. Silver stain for total protein (B). Western blot for immunoreactivity against serum IgG raised from specific pathogen-free C57BL/6 mice injected with 10^8 c.f.u. *E. cloacae* 20 days previously (c). Control experiments verified earlier published data[25] that unmanipulated control mice showed no immunoreactivity against this dominant intestinal aerobe. Proteomic analysis of extracted bacterial ribosomal proteins confirmed identities as follows: 50SL24, 50SL11, 50SL4, 50SL3, 50SL1, 30SS4, 30SS1. (d) Germ-free mice were either orally (n = 6) or systemically (n = 5) exposed once by intragastric (10^{10} c.f.u.) or intravenous (10^8 c.f.u.) delivery of *E. coli* HA107. Mesenteric lymph nodes and spleens were analysed 18 hours later for LPS levels and 16S rRNA ($\bar{x}$±SD, two-sided unpaired t-test). (e, f) Indicated doses of intact or ultrasound-lysed *E. coli* HA107 were administered either orally (e) or systemically (f) three times to germ-free mice every other day and compared to germ-free controls. Ig heavy chain repertoire analysis at 21 days for IgA in mesenteric lymph nodes (e) or IgG2b in the spleen (f). (g) *In vitro* culture of leukocytes from the mesenteric lymph nodes or spleen of germ-free mice stimulated with either cytoplasmic or membrane fractions of *E. coli* HA107. Activated B cells were sorted on day 5 after stimulation and IgM heavy chain sequencing was carried out. (e-g) In all cases Euclidean distance in MDS plots reflects the distance between indicated repertoires based on CDR3 amino acid sequences of the entire B cell repertoire. (a-d) Data are representative of two independent experiments. (e,f and g) Data are from two single experiments. Each datapoint is from the organ of an individual mouse.

Extended Data Figure 5: Network formation of different isotypes depending on transitory microbial treatment and comparison with strong cholera toxin immunogen

Germ-free mice were either orally or systemically primed three times every other day by a range of intragastric (10^2-10^{10} c.f.u.) or intravenous (10^2-10^8 c.f.u.) doses of *E. coli* HA107, compared with priming by cholera toxin or to germ-free controls. (a) Rarefaction plots of immunoglobulin repertoire sequencing at 21 days for IgM or IgG2b (inset: same data as in Figure 2b but different y-axis scale) in MLN and spleen. Colour-coding indicates the route of exposure. (b, c) Median number of mutations per clonotype for IgA in the MLN (b) or IgG2b in splenic B cells (c). Tukey plots in each case are shown with whiskers at 1.5x interquartile range (n=3 for each condition). (d, e) Germ-free mice were either orally or systemically primed three times every other day by intragastric (10^{10} c.f.u.) or intravenous (10^8 c.f.u.) delivery of *E. coli* HA107, or by intragastric (15 μg) or intravenous (15 μg) delivery of cholera toxin B (CTB) and compared with germ-free controls. Ig repertoire sequencing was carried out for IgA, IgM and IgG2b at 21 days. (d) Tukey plots with whiskers at 1.5x interquartile range indicate proportions of expanded clonotypes (excluding singletons) within the entire CDR3 amino acid IgA repertoire in ileum, MLN and spleen or within the IgG2b repertoire in MLN and spleen after the indicated exposures (n=3 mice in each bacterial/cholera toxin priming group, n=12 germ-free controls, two-sided Wilcoxon rank-sum test). P_{adj}-values as shown. (e) Radial plot showing median values for mutational levels in MLN, ileum and spleen in individual mice. Peripheral displays of representative network structures of clonotypes showing relatedness with edges representing Levenshtein distance 1 (blue). Singletons are shown in orange. (d,e) All conditions were repeated over 10 times, except for cholera toxin i.v. priming which was carried out once and cholera toxin intestinal priming which was carried out twice. All datapoints are from individual mice.

Extended Data Figure 6: Characteristics of naïve B cell repertoires following systemic or mucosal exposure and sites of B cell activation

(a, b) Characteristics of naïve single cell B cell repertoires corresponding to class-switched single cell repertoires shown in Figure 2e and f. Germ-free mice were either orally or systemically primed three times every other day by intragastric (10^{10} c.f.u.) or intravenous (10^{8} c.f.u.) delivery of *E. coli* HA107 and compared with germ-free controls. Single cell VDJ sequencing analysis on day 21. Network built on combined heavy and light chain CDR3 amino acid sequences each from a single splenic (a) or MLN (b) naïve IgD-expressing B cells, excluding singlets (n=4272 or 5672 for A and B respectively). Edges show Levenshtein distance 1 or 2. Networks show the immunoglobulin sequence relationships within and between different mice in the same experiment, colour-coded according to mouse and treatment. Pie charts (left) indicate the percentage of edge connections that are based on a Levenshtein distance of 1 on the light chain compared to the heavy chain, or (right) indicate the distribution of edge connections between individual B cells within or between exposure conditions.

(c-d) Germ-free mice were systemically (i.v.) primed with two doses of 10^{8} c.f.u. *E. coli* HA107, 7 days apart, and compared with germ-free controls. Lymphocytes from spleen were isolated 18 h after the i.v. injection, stained with fluorescent-labelled antibodies and analysed by flow cytometry. (c) Overall quantifications ($\bar{x}$±SD, n=11 germ-free and 9 systemic exposure) of live GL-7^{+} CD19^{+} lymphocytes as a proportion of all CD19^{+} lymphocytes from two independent experiments. (d) Lymphocytes from spleen were isolated 3d after the last injection for culture. IgA and IgG2b antibody levels were determined in culture supernatants at 5d by ELISA. Plots show pooled data from two experiments (geometric mean, n=7 germ-free and 8 systemic exposure for both isotypes). (e, f) Germ-free mice were reversibly exposed with 3 gavage doses of 10^{10} c.f.u. *E. coli* HA107 and compared with germ-free controls. (e) Overall quantifications ($\bar{x}$±SD, n=4 for each condition) of live GL-7^{+} Bcl-6^{+} CD19^{+} lymphocytes as a proportion of all CD19^{+} lymphocytes from MLN 24 h after the last gavage ($\bar{x}$±SD). (f) Lymphocytes from MLN were isolated 3d after the last injection for culture. IgA and IgG2b antibody levels were determined at 5d in culture supernatants by ELISA. (geometric means, n=4 for each condition except IgG2b intestinal exposure =3). (c-f) P-values with unpaired t-test are indicated as shown in the figure. All datapoints are from individual mice.

Extended Data Figure 7: CD4 T cells are required for systemic immune memory following intestinal exposure to reversible *E. coli* HA107

(a, b) Schematic experimental designs to Figures 3a-c: (a) Germ-free mice were either intestinally exposed with 10^{10} c.f.u. reversible *E. coli* HA107 on alternate days or remained germ-free throughout. On day 21, all mice were reversibly intravenously exposed with 10^{3}-10^{7} c.f.u. HA107 as shown. Mice were analysed on day 42. (b) As in a, although the initial reversible exposure was given systemically with subsequent reversible intestinal exposures at different doses. (c) Schematic experimental design. Germ-free mice were either mucosally exposed with 3 doses of 10^{10} c.f.u. *E. coli* HA107 on days 0, 2 and 4 or left germ-free. In both groups, half the mice were treated with anti-CD4 depleting antibody i.p. on day -3, the other half received a control isotype. CD4+ T cells were absent (<0.1 % of blood leukocytes) from day 0 until at least day 10, but were shown to have recovered by day 21. On day 21, mice were intravenously primed with 10^{7} c.f.u. *E. coli* HA107. Control groups included intestinally only exposed mice and untreated germ-free mice. (d) Representative dot plots of flow cytometric analysis of the blood on days 0 and 21 in both the isotype and anti-CD4 antibody treated groups. (e) Bacterial flow cytometry at day 42 analysing specific bacterial surface binding against *E. coli* HA107 from serum IgG2b of the indicated groups. (f) Immunoglobulin repertoire sequencing for IgG2b in the spleen on day 42. Euclidean distance in the MDS plot reflects the distance based on CDR3 amino acid sequences of the entire repertoires. Data (c-f) are representative of two independent experiments. Each dilution series (e) or datapoints (f) are from individual mice.

Extended Data Figure 8: Experimental schemes to Figure 3

(a) Schematic experimental design to Figure 3d-g: Germ-free mice were either mucosally primed with three doses of 10^{10} c.f.u. *E. coli* HA107 or reversible *Salmonella typhimurium* HA218 on alternate days or remained germ free. At day 21, half of each group received a second schedule of priming, but with the opposite taxon. Recovery of germ-free status was verified after each stage. (b) Schematic experimental design to Figure 3h-k as in panel (a) although both reversible taxa were given systemically at doses of 10^{8} c.f.u. as shown.

METHODS

Mice

Germ-free C57BL/6 mice (either males or females, 8-16 weeks of age) were bred and maintained in flexible-film isolators at the Clean Mouse Facility, University of Bern, Switzerland. Gender balance was ensured in all experimental groups and germ-free status was routinely monitored by culture-dependent and -independent methods. All mouse experiments were performed in accordance with Swiss Federal regulations approved by the Commission for Animal Experimentation of Kanton Bern. Group sample sizes of ≥3 mice were determined from preliminary experiments that indicated sufficient power to discriminate repertoire effects. Computational analysis but not experimental processing was blinded to the treatment of each animal.

Bacterial culture

Escherichia coli HA107 and non-invasive non-virulent *Salmonella typhimurium* HA218 (UK-1 Δ*asd*, Δ*alr*, Δ*dadX*, Δ*metC::tetRA*, Δ*ssaV::camR*, Δ*invC::aphT*) were cultured overnight in LB medium containing 100 μg/ml meso-DAP and 400 μg/ml D-alanine at 37°C, shaking at 200 rpm [3]. A full description of in vivo characteristics of the reversible Salmonella strain will be published separately [18]. *Enterobacter cloacae* was cultured in LB medium at 37°C, 200 rpm. *Clostridium orbiscindens* (DSMZ 8061) was cultured in BHI medium with additional hemin and vitamin K at 37°C in a Whitley MG500 anaerobic incubator gassed with 10% (v/v) H_2, 10% CO_2, and 80% N_2. Reversible colonisations with *Clostridium orbiscindens* (DSMZ 8061) were confirmed through culture-dependent and -independent methods to return to germ free status after live exposure. To prepare gavage solutions or intravenous injections, bacteria were centrifuged for 10 min at 4000 **g** and washed twice with sterile PBS. The required dose was resuspended in 500 μl of sterile PBS and administered to germ-free mice by oral gavage or injection into the tail vein.

Preparation of *E. coli* HA107 lysates and subfractions

Overnight culture of *E. coli* HA107 was harvested, washed with PBS and resuspended at 10^{10} c.f.u. /ml in PBS. Bacterial cells were disrupted by ultrasound using the Bioruptor system for 15 minutes at 4°C (Diagenode, 30 seconds ON/30 seconds off mode). The lysate was centrifuged at 10,000 **g** for 10 minutes, diluted and sterilized by 0.2 μm filtration before in vivo treatment or further fractionation. The lysate was fractionated into membrane and cytoplasmic fractions by ultracentrifugation (Optima MAX-TL benchtop ultracentrifuge, Beckman) for 10 min 100,000 **g**. The cytoplasmic supernatant was collected and the pellet containing membranes was resuspended in PBS. An 18-hour ultracentrifugation at 100,000 **g** was performed on the cytoplasmic fraction to separate the ribosomal fraction from the non-ribosomal fraction [26]. Protein concentrations were quantified using the Pierce BCA assay kit (Thermo Fisher Scientific).

Systemic and intestinal exposure to live reversible *E. coli* HA107

For systemic exposure, germ-free mice were aseptically transferred to a laminar flow hood. Doses (10^2 -10^8 c.f.u. as shown) of *E. coli* HA107 were administered into the tail vein. For intestinal exposure, mice received

10^2 -10^{10} c.f.u. of *E. coli* HA107 by oral gavage within germ-free isolators [3]. When multiple doses were administered, mice received 3 doses on alternate days. Mice were analysed 2 weeks after administration of the last exposure dose.

Systemic and intestinal priming of immune responses with *E. coli* HA107 lysate

The preparation of bacterial lysates was performed as described above. Germ-free mice were treated with bacterial lysates (dose equivalent to 10^8 intact bacteria) either systemically or orally on alternate days. Mice were analysed 2 weeks after administration of the last priming dose.

Combined intestinal and systemic priming with *E. coli* HA107

Germ-free C57BL/6 mice were divided into two groups. In the first scenario, a group of mice was treated with three doses of 10^{10} c.f.u. of HA107 orally on days 0, 2 and 4, whilst the other group were unprimed controls. Both groups of mice (intestinally primed with 3 doses of 10^{10} c.f.u. of *E. coli* HA107 and control mice) received an i.v. injection of 10^3 - 10^7 c.f.u. (as shown) of *E. coli* HA107 on day 21.

In the second converse scenario, a group of mice was treated three times with 10^8 c.f.u. of HA107 systemically on days 0, 2 and 4 while the other group were unprimed controls. Both groups of mice subsequently received 3 intragastric doses of 10^7 - 10^9 c.f.u. (as shown) of *E. coli* HA107 on days 21, 23 and 25. All mice from both scenarios were analysed on day 42.

Combined priming with *E. coli* HA107 and *S. typhimurium* HA218

For mucosal priming, germ-free mice received 10^{10} c.f.u. of *E. coli* HA107 by gavage on days 0, 2 and 4 or were unprimed controls. Subsequently half the mice in both groups received 3 doses of 10^{10} c.f.u. *S. typhimurium* HA218 on days 21, 23 and 25 by gavage, while the other half received no further priming. Within the same experiment further groups received the opposite treatment sequence of *S. typhimurium* HA218 → *E. coli* HA107. Systemic priming was carried out using groups of germ-free mice with the above schedule in both treatment orders, except that 3 doses of 10^8 c.f.u. of *E. coli* HA107 or *S. typhimurium* HA218 by intravenous injection were used instead of the intestinal treatments for mucosal priming. Recovery of germ-free status was verified at each stage. All mice were analysed on day 42.

Priming with cholera toxin subunit B

Germ-free mice each received 3 doses of intragastric (15 μg) or intravenous (15 μg) cholera toxin B (CTB) on alternate days and were analyzed 3 weeks after the last priming dose. For intragastric delivery, CTB was dissolved in 0.1M sodium bicarbonate (pH 9), systemic delivery was in PBS.

Bacterial FACS

Bacterial FACS was performed as described previously [3], with the following refinements. *E. coli* HA107 was incubated with serum or intestinal washes with starting dilutions at 1/10 for serum, or undiluted for intestinal washes: in each sample the concentrations of different immunoglobulin isotypes had been measured separately by ELISA as detailed below. After incubation and washing to remove unbound antibodies, the samples were incubated further at 4°C for 2 hours either with secondary anti-murine IgA-FITC (BD Biosciences C10-3) or anti-murine IgG2b-FITC (BD Biosciences R12-3). Bacteria were acquired on a BD FACSArray using FSC

and SSC parameters in logarithmic mode. The FACS data were analysed using FlowJo software (TreeStar, Inc.), and the levels of bacteria-specific antibodies present in the samples were expressed as the geometric mean against the absolute Ig concentration in the sample.

ELISA

Immunoglobulin-specific ELISAs were performed as described [27] using goat anti-mouse IgA, IgM, IgG1, IgG2b, IgG2c, IgG3 coating antibodies from SouthernBiotech. Antibody standards were purified mouse myeloma Ig of the required isotype with peroxidase-coupled secondary antibodies as previously detailed. The same protocol was used for antigen-specific ELISA except that ELISA plates were first coated overnight with 10 μg/ml of HA107 membrane, cytoplasmic, ribosomal or non-ribosomal fractions.

Immunoblotting

Bacterial (*E. cloacae*) cytoplasmic proteins (10 μg/ml) were loaded to 12% SDS-PAGE and electrophoresed for 45 minutes at 90V; subsequently they were either stained directly for proteins or transferred to a nitrocellulose membrane. After blocking with 3% BSA, the membrane was incubated with i.v. primed mouse serum diluted at 1:30 in PBS for 1hr at room temperature. Peroxidase-conjugated goat anti–mouse IgG catalysis of 3, 3'-diaminobenzidine was used to detect bound IgG [25].

LPS versus 16S rRNA measurements

Germ-free mice were treated either systemically with 10^8 c.f.u. of *E. coli* HA107 or mucosally with 10^{10} c.f.u. of *E. coli* HA107. Mice were dissected 18 hours later. Spleen and MLN were collected and disaggregated on baked 40 μm-metal cell strainers using PBS. Samples were then split into two portions for either LPS or 16S rRNA estimation.

LPS measurements

Samples were centrifuged (350 **g** 5 mins) to remove intact cells. Supernatants were heated at 70°C for 30 mins and cleared by centrifugation at 10,000 **g**. LPS was quantified using the EndoZyme® II Recombinant Factor C Endotoxin Detection Assay (Hyglos GmbH, BioMérieux #890030) according to the manufacturer's instructions, using 10-fold dilutions to circumvent interference. After incubation at 37°C for 90 minutes fluorescence was measured at an excitation of 380 nm and an emission of 445 nm (Infinite Pro 200, Tecan). End point fluorescent measurements with blank corrections and standards from 0.005 - 5 EU/ml were used. Standard curve generation and regression analysis performed in Graphpad prism were used to estimate LPS concentrations in samples, with extrapolation of the standard curve based on the regression parameters. Samples and standards were measured in duplicate.

16S measurements

Total RNA was extracted from samples using the DNA-RNA all prep kit (Qiagen), and real-time PCR was performed on synthesized cDNA using forward (5'-ATG CGT AGA GAT CTG GAG G-3') and reverse primers (5'-CAA CCT CCA AGT CGA CAT C-3') for 16S ribosomal gene amplification.

Lymphocyte isolation

To isolate lymphocytes from spleens, lymph nodes and Peyer's patches, tissues were cut into small pieces and digested in IMDM (2 % FCS) containing collagenase type IA (1 mg/ml, Sigma) and DNase I (10 U/ml, Roche)

at 37°C for 30 min. Cellular suspensions were passed through a cell strainer (40 µm) and washed with IMDM (2 % FCS, 2mM EDTA). Erythrocytes in splenic suspensions were lysed using 0.84 % NH_4Cl for 2 min at room temperature. Cells were finally resuspended in FACS buffer and counted in a Neubauer chamber.

Flow cytometry

After isolation, cells were washed once with PBS before staining with fixable viability dye diluted in PBS (eBisocience) for 30 min on ice. Single cell suspensions were sequentially incubated with primary/biotin- and fluorescence-coupled antibodies diluted in FACS buffer. The following mouse-specific conjugated antibodies were used: B220-FITC (RA3-6B2, Biolegend), Bcl-6-AlexaFluor647 (K112-91, BD Biosciences), CD3-BV785 (17A2, Biolegend), CD4-PE (RM4-5, Biolegend), CD11b-PEcy7 (M1/70, Biolegend), CD19-AF700 (6D5, Biolegend), CD21-FITC (7E9, Biolegend), CD22-FITC (OX-97, Biolegend), CD23-PerCPcy5.5 (B3B4, Biolegend), CD25-PerCP-cy5.5 (PC61, Biolegend), CD43-FITC (1B11, Biolegend), CD80-PerCPcy5.5 (16-10A1, Biolegend), CD86-BV421 (GL-1, biolegend), CD93-PEcy7 (AA4.1, Biolegend), CD138-PE (281-2, Biolegend), CD273-APC (TY25, Biolegend), CXCR5-biotin (2G8, BD Biosciences), IgA-FITC (C10-3, BD Biosciences), IgG2b-FITC (R12-3, BD Biosciences), IgM-PEcy7 (RMM-1, Biolegend), Ly-77-BV421 (GL-7, BD Biosciences), PD-1-PE (RMP1-30, Biolegend). Streptavidin-APC (BD Biosciences) was used together with biotin-coupled antibodies. Data were acquired on an LSRFortessa (BD Biosciences) and analysed using FlowJo software (Tree Star Inc.). In all experiments, FSC-H versus FSC-A was used to gate singlets, and to exclude dead cells using the fluorescence-coupled fixable viability dye (eBisocience). B-lineage subsets were stained for flow cytometry cell sorting as: pre B cells from bone marrow (live/dead$^-$ CD3$^-$ CD19$^+$ CD138$^-$ IgM$^-$ CD25$^+$), plasma cells from bone marrow (live/dead$^-$ CD3$^-$ CD138$^+$ CD19$^-$ CD22$^-$), transitional B cells from spleen (live/dead$^-$ CD3$^-$ CD19$^+$ CD138$^-$ GL7$^-$ CD93$^+$), splenic IgM$^+$ plasma cells (live/dead$^-$ CD3$^-$ CD19$^-$ CD138$^+$ IgM$^+$), splenic IgG2b$^+$ plasma cells (live/dead$^-$ CD3$^-$ CD19$^-$ GL7$^-$ CD138$^+$ IgG2b$^+$), follicular splenic B cells (live/dead$^-$ CD3$^-$ CD19$^+$ CD93$^-$ CD138$^-$ CD23^{hi} CD21^{lo}), splenic marginal zone B cells (live/dead$^-$ CD3$^-$ CD19$^+$ CD93$^-$ CD138$^-$ CD23^{lo} CD21^{hi}), splenic IgM$^+$ memory cells (live/dead$^-$ CD3$^-$ CD19$^+$ CD138$^-$ CD80$^+$ CD273$^+$ IgM$^+$) [28], splenic IgG2b$^+$ memory cells (live/dead$^-$ CD3$^-$ CD19$^+$ GL7$^-$ CD138$^-$ IgG2b$^+$), peritoneal B2 cells (live/dead$^-$ CD3$^-$ CD19$^+$ CD23$^+$ CD43$^-$), peritoneal B1a cells (live/dead$^-$ CD3$^-$ CD19$^+$ CD23$^-$ CD43$^+$ B220^{lo} IgM^{hi}), peritoneal B1b cells (live/dead$^-$ CD3$^-$ CD19$^+$ CD23$^-$ CD11b$^+$ B220^{hi} IgM^{lo}), splenic/MLN IgA$^+$ plasma cells (live/dead$^-$ CD3$^-$ CD19$^-$ GL7$^-$ CD138$^+$ IgA$^+$).

***In vitro* primary cell culture**

Primary cells were isolated from spleen, PP and MLN from germ-free or *E. coli* HA107 conditioned mice. Tissues were disaggregated through 40 µm cell strainers and flushed with 10 ml IMDM+GlutaMax (GIBCO) supplemented with 10% FCS (Invitrogen). Cell suspensions were centrifuged (350 **g**, 5 min, 4°C) and erythrocyte lysis from splenic samples was performed as described above.

For the estimation of IgA and IgG2b production by B cells, cells were centrifuged again and resuspended in complete IMDM medium IMDM+GlutaMAX (GIBCO), supplemented with 10% FCS (Invitrogen), 10,000 µg/ml streptomycin (GIBCO), 10,000 U/ml penicillin (GIBCO), 50 µM 2-mercaptoethanol (Merck). The cell suspensions were seeded at a density of 10^6 cells/ml in 200 µl of complete IMDM medium per well in 96-well round-bottom tissue culture plates. The cell cultures were incubated at 37ºC, 5% CO_2. Viability of the cultured cells was controlled daily microscopically. Supernatants were collected after 5 days and antibody

concentrations were determined using ELISA as described above.

For the assessment of repertoire reprogramming to antigen exposure, an aliquot of cells was collected and $CD86^+$ $CD138^+$ antigen-experienced plasma cells were sorted and used as pre-treatment controls. Remaining cells were seeded at a density of 10^6 cells/ml in RPMI 1640 medium supplemented with FCS, streptomycin, penicillin, and β-mercaptoethanol as above with the addition of 10 ng/ml recombinant IL4 (Peprotech) in a 48-well flat-bottomed plate. *E. coli* cytoplasmic and membrane fractions were added at 5 µg/ml (protein equivalent) to culture wells. The culture conditions were as detailed above. On day 6, cells were harvested, and $CD86^+$ $CD138^+$ plasma cells were FACS-sorted for repertoire analysis.

CD4 T cell depletion

Germ-free mice were injected intraperitoneally either with 250 µg of anti-CD4 antibody (clone GK5.1, Bioexcel) or an isotype control (clone LTF-2, Bioexcel) diluted in 250 µl of PBS on day -3 before intestinal exposure with 3 x 10^{10} c.f.u. *E. coli* HA107 on days 0, 2 and 4. T cell depletion and recovery were confirmed by flow cytometry on blood samples taken on days 0 and 21 respectively. All mice received a single intravenous dose of 10^7 c.f.u. *E. coli* HA107 on day 21. Mice were analysed on day 42.

Immunoglobulin repertoire sequencing bulk

Tissues were removed and snap-frozen in Trizol reagent using liquid nitrogen. Thawed tissues were immediately homogenized (Retsch bead-beater) in 1 ml of Trizol reagent (Life Technologies). Chloroform (200 µl) was added, samples were mixed, and centrifuged (12,000 **g**, 15 min, 4°C). The upper phase was collected, and RNA was precipitated with ice-cold isopropanol by centrifugation (12,000 **g**, 10 min, 4°C). The RNA pellet was washed with 75% (v/v) ethanol, dried and resuspended in RNase-free water. RNA concentrations and purity were analysed using a Nanodrop2000 (Thermo Scientific).

For IgA, cDNA synthesis was performed by mixing 700 ng of RNA, 1 µl of 2 µM gene specific primer mix [1], 1 µl of 10 mM of dATP, dCTP, dGTP and dTTP (Invitrogen), with dH_2O to a total volume of 13 µl. For IgM and IgG2b heavy chains and light chains, oligo(dT) primers were used for cDNA synthesis. Samples were then heated to 65°C for 5 min, after which they were put on ice for a minimum of 1min. Tubes were shortly centrifuged to collect contents, and in a second reaction step, 4 µl 5X first strand buffer (Invitrogen), 1 µl 0.1 M DTT (Invitrogen), 1 µl RNaseOUT (Invitrogen) and 1µl Superscript III RT enzyme (Invitrogen) were added to the reaction, mixed and incubated at 55°C for 50 min. The reaction was terminated by heat inactivation at 70°C for 15 min followed by cooling at 4°C. Samples were stored at -20°C until further processing.

Amplicon PCR for IgA, IgM or IgG2b heavy chains and kappa or lambda light chains was performed by mixing 5 µl of 10X PlatinumTaq PCR buffer (Qiagen), 1 µl of 10 mM mixed dNTPs, 0.2 µl of PlatinumTaq DNA Polymerase (Qiagen), 1.5 µl of 50 mM $MgCl_2$, 1 µl of each 100 µM forward primer mix and reverse primers (sequences see below according to chain and isotype), 5 µl template cDNA, in a total volume of 50 µl. The forward primers consisted of a degenerative mix of 19 different primers binding into FR1 as previously described [29,30]. Reverse primers were complementary to the constant region of each isotype (sequences see below).

PCR products were loaded onto a 1.5% (w/v) agarose gel and purified with the QIAquick Gel Extraction kit (QIAGEN) according to the manufacturer's instructions. The purified DNA was tested for quality and quantity

using a 2100 Bioanalyzer (Agilent) and Qbit (Thermofisher). A second PCR was performed to attach sample indices and sequencing adapters using the Nextera® XT Index Kit (Illumina). Subsequent sample sizes and quality were tested with a Fragment Analyzer™ (Advanced Analytical), normalized and libraries were pooled for sequencing on the MiSeq Illumina sequencer in the paired 250bp mode, or the 300bp mode for experiments containing UMIs.

Primer sequences

IgH Forward Mix	**Illumina Adapter sequence read 1 + Diversity region + VH 5' specific Region**
IgH-UAd-fw1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAKGTRMAGCTTCAGGAGTC
IgH-UAd-fw2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAGGTBCAGCTBCAGCAGTC
IgH-UAd-fw3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCAGGTGCAGCTGAAGSASTC
IgH-UAd-fw4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAGGTCCARCTGCAACARTC
IgH-UAd-fw5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCAGGTYCAGCTBCAGCARTC
IgH-UAd-fw6	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCAGGTYCARCTGCAGCAGTC
IgH-UAd-fw7	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCAGGTCCACGTGAAGCAGTC
IgH-UAd-fw8	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAGGTGAASSTGGTGGAATC
IgH-UAd-fw9	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAVGTGAWGYTGGTGGAGTC
IgH-UAd-fw10	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAGGTGCAGSKGGTGGAGTC
IgH-UAd-fw11	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAKGTGCAMCTGGTGGAGTC
IgH-UAd-fw12	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAGGTGAAGCTGATGGARTC
IgH-UAd-fw13	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAGGTGCARCTTGTTGAGTC
IgH-UAd-fw14	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGARGTRAAGCTTCTCGAGTC
IgH-UAd-fw15	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAAGTGAARSTTGAGGAGTC
IgH-UAd-fw16	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCAGGTTACTCTRAAAGWGTSTG
IgH-UAd-fw17	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCAGGTCCAACTVCAGCARCC
IgH-UAd-fw18	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGATGTGAACTTGGAAGTGTC
IgH-UAd-fw19	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGAGGTGAAGGTCATCGAGTC
IgH Reverse (isotype specific)	**Illumina Adapter sequence read 2 + diversity region + Ig constant region specific**
IgA-const-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGCTCGTGGGAGTGTCAGTG
IgM-const-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGACGAGGGGGAAGACATT
IgG2b-const-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNTTGTATCTCCACACCCAGGG

IgH Forward UMI	**Illumina Adapter sequence read 2 + Diversity region + VH 5' specific Region**
IgH-UAd-fw1_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAKGTRMAGCTTCAGGAGTC
IgH-UAd-fw2_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGGTBCAGCTBCAGCAGTC
IgH-UAd-fw3_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCAGGTGCAGCTGAAGSASTC
IgH-UAd-fw4_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGGTCCARCTGCAACARTC
IgH-UAd-fw5_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCAGGTYCAGCTBCAGCARTC
IgH-UAd-fw6_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCAGGTYCARCTGCAGCAGTC
IgH-UAd-fw7_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCAGGTCCACGTGAAGCAGTC
IgH-UAd-fw8_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGGTGAASSTGGTGGAATC
IgH-UAd-fw9_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAVGTGAWGYTGGTGGAGTC
IgH-UAd-fw10_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGGTGCAGSKGGTGGAGTC
IgH-UAd-fw11_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAKGTGCAMCTGGTGGAGTC
IgH-UAd-fw12_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGGTGAAGCTGATGGARTC
IgH-UAd-fw13_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGGTGCARCTTGTTGAGTC
IgH-UAd-fw14_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGARGTRAAGCTTCTCGAGTC
IgH-UAd-fw15_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAAGTGAARSTTGAGGAGTC
IgH-UAd-fw16_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCAGGTTACTCTRAAAGWGTSTG
IgH-UAd-fw17_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCAGGTCCAACTVCAGCARCC
IgH-UAd-fw18_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGATGTGAACTTGGAAGTGTC
IgH-UAd-fw19_UMI	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGAGGTGAAGGTCATCGAGTC
Reverse UMI_1	**Illumina Adapter sequence read 1 + UMI_1 + IgA constant region specific**
IgG2b-const-rev_UMI_1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNNNNNNNTTGTATCTCCACACCCAGGG
IgA-const-rev_UMI_1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNNNNNNNGAGCTCGTGGG AGTGTCAGTG
Reverse UMI_2	**Illumina Adapter sequence read 1 + UMI_2 + IgA constant region specific**
IgG2b-const-rev_UMI_2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGHHHHHACAHHHHHACAHHHHHHNTTGTATCTCCACACCCAGGG
IgA-const-rev_UMI_2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGHHHHHACAHHHHHACAHHHHHHNGAGCTCGTGGGAGTGTCAGTG

IgL Forward Mix	**Illumina Adapter sequence read 1 + Diversity region + VH 5' specific Region**
IgL_nonfw1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATCCAGCTGACTCAGCC

IgL_nonfw2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTTCTCWCCCAGTC
IgL_nonfw3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTTCTCWCCCAGTC
IgL_nonfw4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTGYTRACACAGTC
IgL_nonfw5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTRATGACMCAGTC
IgL_nonfw6	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTMAGATRAMCCAGTC
IgL_nonfw7	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTCAGATGAYDCAGTC
IgL_nonfw8	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATYCAGATGACACAGAC
IgL_nonfw9	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTTCTCAWCCAGTC
IgL_nonfw10	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGWGCTSACCCAATC
IgL_nonfw11	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTSTRATGACCCARTC
IgL_nonfw12	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYRTTKTGATGACCCARAC
IgL_nonfw13	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTGATGCBCAGKC
IgL_nonfw14	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTGATAACYCAGGA
IgL_nonfw15	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTGATGACCCAGWT
IgL_nonfw16	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTGTGATGACACAACC
IgL_nonfw17	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNAYATTTTG CTGACTCAGTC
IgL_nonfw18	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNARGCTGTTGTGACTCAGGAATC
IgL Reverse	**Illumina Adapter sequence read 2 + diversity region + IgL constant region specific**
IgL_rev1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNACGTTTGATTTCCAGCTTGG
IgL_rev2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNACGTTTTATTTCCAGCTTGG
IgL_rev3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNACGTTTTATTTCCAACTTTG
IgL_rev4	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNACGTTTCAGCTCCAGCTT GG
IgL_rev5	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNACCTAGGACAGTCAGTTTGG
IgL_rev6	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNACCTAGGACAGTGACCTTGG

Repertoire sequencing and pre-processing

Antibody variable heavy chain (VH) libraries were prepared as previously described [31] and sequenced on the Illumina MiSeq platform (2x250 cycles, paired-end) with 10% PhiX control library. Mean base call quality of all samples was in the range of Phred score 30. The resulting FASTQ files were pre-processed (VDJ alignment, clonotyping) using the MiXCR software package v2.1.12 (clonotype formation by CDR3 amino acid region annotated using C57BL/6J germline gene data[32] (available at http://mixcr.milaboratory.com/ and https://github.com/milaboratory/mixcr/).

MIXCR considers sequence quality and corrects PCR and sequencing errors in the alignment step to the reference V, D, J germline segments. The error correction is done by assembling the clonotypes with a heuristic multi-layer clustering which briefly can be summarized as follows : the algorithm consists of two steps: (A) a K-mer chaining algorithm is used, then in order to find the optimal reference sequence a heuristic search is performed to maximize a scoring function which takes into account variations (deletions, insertions, mutations) in the sequence reads. (B) Subsequently, multiple candidate alignments are built and a classical alignment score (scoring matrix and penalties for indels) is calculated. Reads are filtered based on these values. Then clonotypes are assembled and those presenting errors are further removed by clustering similar sequences based on a defined similarity threshold (clonotyping). Finally, MIXCR rescues low-quality reads in a later step by mapping them to previously assembled high-quality clonotypes to preserve maximal quantitative information. Singletons (supported by only a single read) were excluded from clonotype output [29,32] as previously described [33].

For all analyses, except where nucleotide analysis combined with V and J segment information is specified, clones were defined by 100% amino acid sequence identity of CDR3 regions. CDR3 regions were defined by MiXCR according to the nomenclature of the Immunogenetics database (IMGT) [34].

Clonal analysis and quantification of repertoire convergence

MIXCR output files were further processed in VDJtools post-analysis framework [35] (available at https://github.com/mikessh/vdjtools) to further analyse clonotype diversity, V gene usage and sample clustering based on CDR3 abundance. Further filtering was applied in order to keep only in-frame, productive sequences. Basic statistic segment usage was calculated, and weighted variable usage profiles were hierarchically clustered and visualized as heatmaps. Repertoire overlap was measured by calculating the geometric mean of relative overlap frequencies between variable segment usage profiles or CDR3 amino acid sequence usage. The relative overlap similarity was represented as hierarchical clustering or multi-dimensional scaling (MDS) plots. Heatmaps following the top100 clonotypes (ranked by abundance) present in at least 2 samples were generated based on an all-vs-all intersection between chosen samples with settings = "strict", meaning intersections had to contain the identical nucleotide sequence (CD3nt sequence + V gene + J gene). The corresponding CDR3 amino acid sequences are provided in the supplementary information table respective to each heatmap found in the figures.

Technical replicate sequencing and UMI pre-processing

Technical reproducibility as previously described [29,31] was tested on cDNA derived from 9 C57BL/6 mice. cDNA was split into three portions and library preparation was performed in parallel with different Illumina indices as described above. Replicate samples were sequenced with the same sequencing depth. cDNA was amplified using primers with the alternative primer design approaches of 15 nucleotide unique molecular identifiers (UMI_1) [36] or with UMI with fixed positions preventing secondary structures (UMI_2) [37] by PCR in a multiplex reaction using primer sets as mentioned above. Raw reads were demultiplexed and exported without the sample barcodes or Illumina clustering adapters. Pre-processing of the raw demultiplexed reads was carried out using pRESTO [38]. Reads with a mean Phred quality score below 20 were removed, then those without valid forward (V-region) and reverse (isotype/constant region) primer sequence matches were removed (match error rate of 0.2). The forward primer regions were masked (with Ns) and the reverse primer regions were deleted from the sequence. Reads with identical UMIs were collapsed into a single consensus sequence for each UMI. UMI read groups having a nucleotide diversity score exceeding 0.1 were discarded. Finally, corrected consensus sequence read pairs were assembled into full length sequences with a maximum allowed error rate of 0.2 and p-value threshold of 0.1. All unique sequences that were not represented by at least two raw reads were removed from further analysis and duplicate full-length sequences were discarded. Following pre-processing, sequences were further processed using MIXCR as described above with the computational correction of PCR errors option turned off (-OcloneClusteringParameters=null).

Mutation and N-joints

To resolve junctional and N-joint regions, the V, N1, D, N2, and J subregions were defined on the sequences exported using MIXCR. Normalized CDR3 subregion (V, N1, D, N2, J) lengths (median) of clonotypes by B cell subpopulation were calculated. Mutation analysis was performed on the annotated sequences exported using MIXCR excluding N-joint regions. Each mutation was reported at the amino acid level for V and J segments counted per sequence. The median number of mutations per sequence over a given repertoire is reported.

Rarefaction plots

In order to explore the relationship between sample size and diversity coverage, rarefaction and extrapolation sampling curves were generated using the iNEXT R package [39].

Determination of sequence similarity among clones within a repertoire

The Levenshtein distance between all pairwise CDR3 amino acid sequence combinations of identical CDR3 amino acids was calculated, and each Levenshtein distance was subsequently normalized by the sequence length of the respective sequence combination. Levenshtein distances were computed using the stringdist package in R [40].

Network construction and analysis

The networks were constructed by representing each unique amino acid CDR3 sequence (clone) as nodes. Clones that were exactly one amino acid substitution, insertion or deletion (Levenshtein distance = 1) apart were connected by edges as previously described [41]. An antibody repertoire network is an undirected graph G = (V, E) described as a set of nodes (CDR3 vertices, V) together with a set of connections (similarity edges, E), representing the adjacency matrix A of pairwise Levenshtein distances (LD) between CDR3 amino acids. Degrees (number of similar CDR3 sequences to a specific CDR3 sequence) were calculated for each similarity layer LD1 for each CDR3 sequence in each sample. CDR3 with zero degrees that were not similar to any other CDR3 in the network were excluded in order to fit degree distributions. The contribution of expanded clonotypes within a given repertoire consisted of the sum of clonotypes with a degree ≥1 on a Levenshtein distance 1 network over the size of the repertoire. The networks were visualized employing Cytoscape or the Fruchterman–Reingold layout and Kamada–Kawai layout algorithms using the igraph R package [42]. For networks based on single cell data concatenated heavy and light chain amino acid CDR3 sequences from the same cell were analysed.

Single cell transcriptomics and repertoire sequencing

Germ-free mice were either mucosally exposed to 3 x 10^{10} c.f.u. *E. coli* HA107, systemically exposed to 3 x 10^{8} c.f.u. *E. coli* HA107 or maintained as germ-free controls. Spleens and mesenteric lymph nodes were removed on day 21 and disaggregated to single cell suspensions. Cells were enriched for $CD138^{+}$ plasma cells by magnetic cell sorting. Cells were washed once and resuspended in PBS/0.05% (w/v) BSA. The 10x Genomics platform was used to perform concurrent B cell VDJ repertoire sequencing and single cell 5' gene expression analysis. Prepared libraries were sequenced on an Illumina NovaSeq 6000 (2x150 cycles, paired-end mode).

V(D)J immunoglobulin single cell sequencing analysis

Processing raw sequencing data

The CellRanger software version 3.0.1 was used to process the raw sequencing data produced by the 10x Chromium direct Ig enrichment. The 10x genomics mouse genome GRCm38.93 3.0.0 release was used as reference for the alignments. The "filtered contigs" files were subsequently used for downstream analyses.

VDJ annotation of BCR contigs

Following pre-processing, sequences were further analysed using MIXCR v2.1.12 as described above with the computational correction of PCR errors option turned off (-OcloneClusteringParameters=null). VDJ alignment and annotation was performed against C57BL/6J germline gene data. CDR3 regions were defined by MiXCR according to the nomenclature of the Immunogenetics database (IMGT). Clonotypes were defined by 100% amino acid sequence identity of CDR3 regions.

Filtering and display of relationships between exposure conditions in BCR contigs and cells

As with bulk Repseq analyses, only productively rearranged clonotypes were used. Cells with one heavy chain contig paired with one light chain contig were kept for downstream analysis. Cells remaining had been checked to have a transcriptomic profile consistent with B-cells (cells with any one of the following markers: CD19, CD138 or the isotype-encoding genes with a positive log-normalized expression level). Given the limited output at single-cell level, repertoire overlap was relaxed from 100% to 90% amino acid identity incorporating a length correction [2]. The relative overlap was normalized for the resulting repertoire size and similarity was represented as hierarchical clustering or multi-dimensional scaling (MDS) plots.

Statistical significance

Statistical significance was tested either using Student t-test or the Wilcoxon rank-sum test followed by a Benjamini-Hochberg correction for false discovery rate. Results were considered significant for $p_{adj}<0.05$.

Data availability

The raw files for the datasets generated during this study is available on the SRA bioproject PRJNA625440 access number. The pre-processed files (clonotype list and properties) are available on the github repository https://github.com/Mucosal-Immunology-Bern/Manuscript-microbiota-and-B-cell-repertoire. The associated codes for the analysis using R packages are to be found in the same github repository.

Availability of biological materials

The bacterial strains for reversible germ-free colonisation can be obtained without restriction from A.J.M and S.H. following a material transfer agreement with the University of Bern.

SUPPLEMENTARY REFERENCES

25 Macpherson, A. J. *et al.* A primitive T cell-independent mechanism of intestinal mucosal IgA responses to commensal bacteria. *Science* **288**, 2222-2226, (2000).

26 Rivera, M. C., Maguire, B. & Lake, J. A. Isolation of ribosomes and polysomes. *Cold Spring Harb Protoc* **2015**, 293-299, (2015).

27 Slack, E. *et al.* Innate and adaptive immunity cooperate flexibly to maintain host-microbiota mutualism. *Science* **325**, 617-620, (2009).

28 Tomayko, M. M., Steinel, N. C., Anderson, S. M. & Shlomchik, M. J. Cutting edge: Hierarchy of maturity of murine memory B cell subsets. *J Immunol* **185**, 7146-7150, (2010).

29 Greiff, V. *et al.* Quantitative assessment of the robustness of next-generation sequencing of antibody variable gene repertoires from immunized mice. *BMC Immunol* **15**, 40, (2014).

30 Krebber, A. *et al.* Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J Immunol Methods* **201**, 35-55, (1997).

31 Menzel, U. *et al.* Comprehensive evaluation and optimization of amplicon library preparation methods for high-throughput antibody sequencing. *PLoS One* **9**, e96727, (2014).

32 Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* **12**, 380-381, (2015).

33 Greiff, V. *et al.* Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development. *Cell Rep* **19**, 1467-1478, (2017).
34 Lefranc, M. P. *et al.* IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res* **27**, 209-212, (1999).
35 Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput Biol* **11**, e1004503, (2015).
36 Stern, J. N. *et al.* B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci Transl Med* **6**, 248ra107, (2014).
37 Khan, T. A. *et al.* Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Sci Adv* **2**, e1501371, (2016).
38 Vander Heiden, J. A. *et al.* pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930-1932, (2014).
39 Hsieh, T., Ma, K. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution* **7**, 1451-1456, (2016).
40 van der Loo, M. The stringdist package for approximate string matching. *R Journal* **6**, 111-122, (2014).
41 Miho, E., Roskar, R., Greiff, V. & Reddy, S. T. Large-scale network analysis reveals the sequence space architecture of antibody repertoires. *Nat Commun* **10**, 1321, (2019).
42 Csardi, G. & Nepusz, T. The igraph software package for complex network research. *Interjournal, Complex Systems*, 1695, (2006).

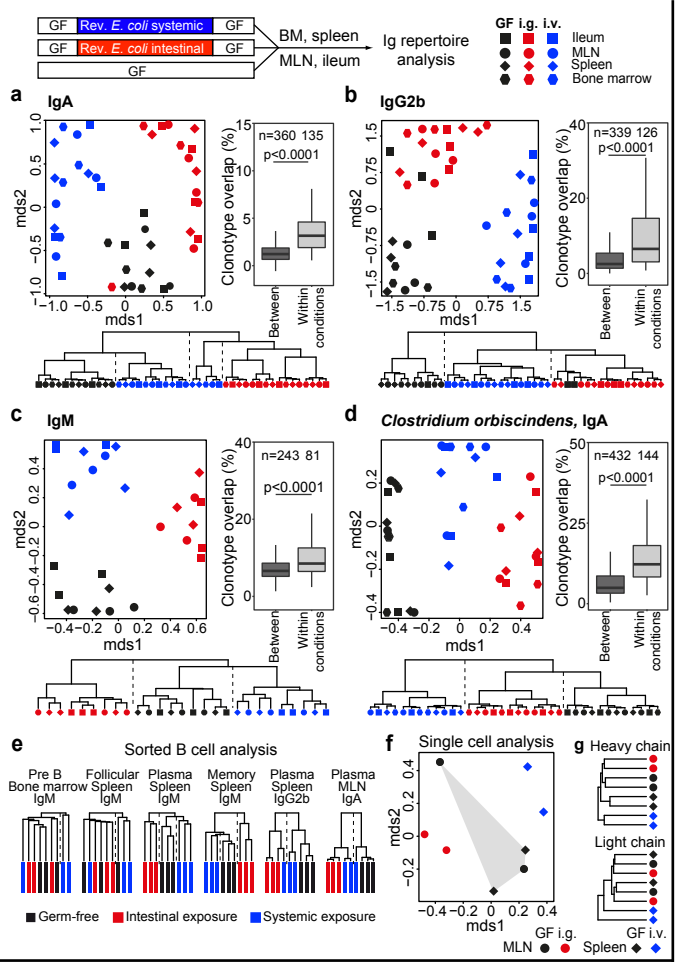

GF
Rev. E. coli systemic
GF
GF
Rev. E. coli intestinal
GF
GF
BM, spleen
MLN, ileum
Ig repertoire analysis
GF i.g. i.v.
Ileum
MLN
Spleen
Bone marrow
a
IgA
b
IgG2b
c
IgM
d
Clostridium orbiscindens, IgA
mds1
mds2
Clonotype overlap (%)
n=360 135
p<0.0001
n=339 126
p<0.0001
n=243 81
p<0.0001
n=432 144
p<0.0001
Between
Within conditions
e
Sorted B cell analysis
Pre B Bone marrow IgM
Follicular Spleen IgM
Plasma Spleen IgM
Memory Spleen IgM
Plasma Spleen IgG2b
Plasma MLN IgA
Germ-free
Intestinal exposure
Systemic exposure
f
Single cell analysis
GF i.g.
MLN
Spleen
g
Heavy chain
Light chain
GF i.v.

Figure 1

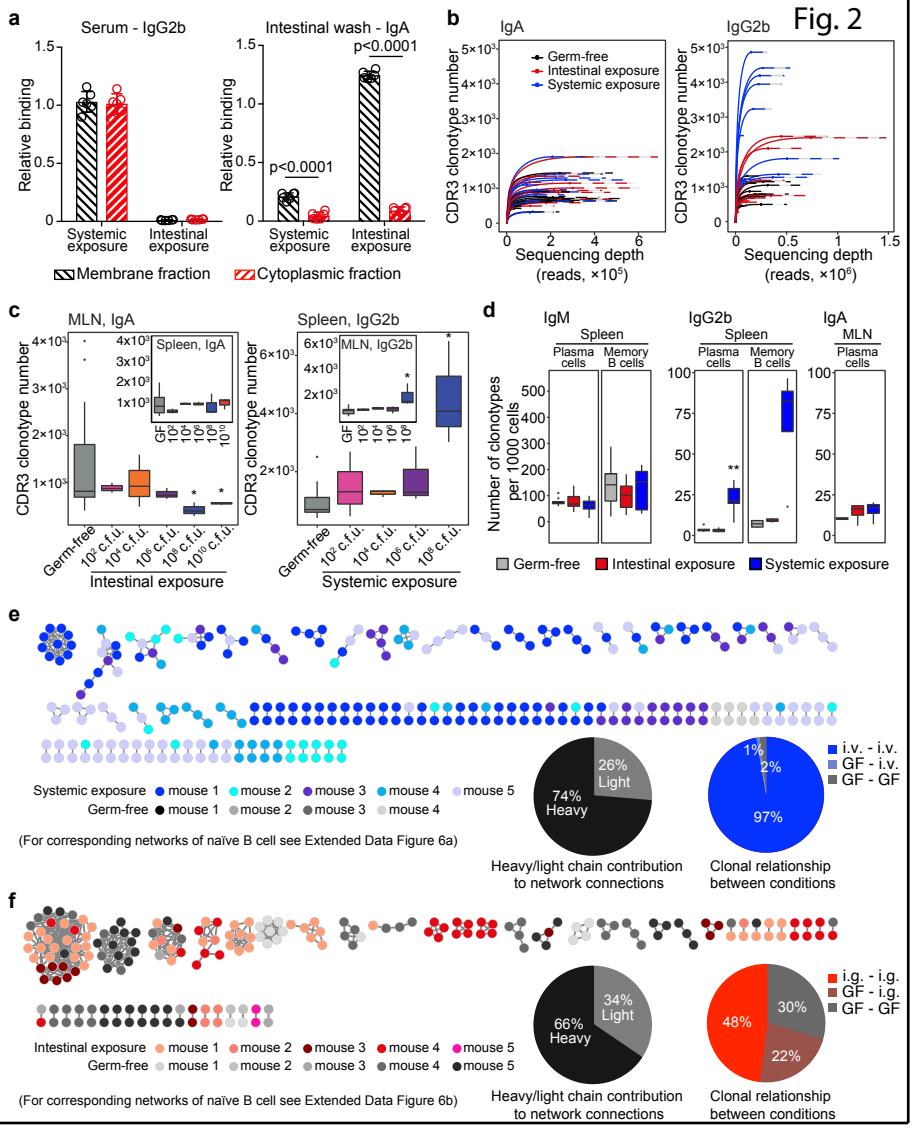


Fig. 2
a
Serum - IgG2b
Intestinal wash - IgA
Relative binding
p<0.0001
Systemic exposure
Intestinal exposure
Membrane fraction
Cytoplasmic fraction
b
IgA
IgG2b
Germ-free
Intestinal exposure
Systemic exposure
CDR3 clonotype number
Sequencing depth (reads, ×10^5)
Sequencing depth (reads, ×10^6)
c
MLN, IgA
Spleen, IgG2b
Spleen, IgA
MLN, IgG2b
Germ-free
Intestinal exposure
Systemic exposure
d
IgM
IgG2b
IgA
Spleen
MLN
Plasma cells
Memory B cells
Number of clonotypes per 1000 cells
Germ-free
Intestinal exposure
Systemic exposure
e
Systemic exposure mouse 1 mouse 2 mouse 3 mouse 4 mouse 5
Germ-free mouse 1 mouse 2 mouse 3 mouse 4
(For corresponding networks of naïve B cell see Extended Data Figure 6a)
26% Light
74% Heavy
Heavy/light chain contribution to network connections
1%
2%
97%
i.v. - i.v.
GF - i.v.
GF - GF
Clonal relationship between conditions
f
Intestinal exposure mouse 1 mouse 2 mouse 3 mouse 4 mouse 5
Germ-free mouse 1 mouse 2 mouse 3 mouse 4 mouse 5
(For corresponding networks of naïve B cell see Extended Data Figure 6b)
34% Light
66% Heavy
Heavy/light chain contribution to network connections
48%
30%
22%
i.g. - i.g.
GF - i.g.
GF - GF
Clonal relationship between conditions

Figure 2

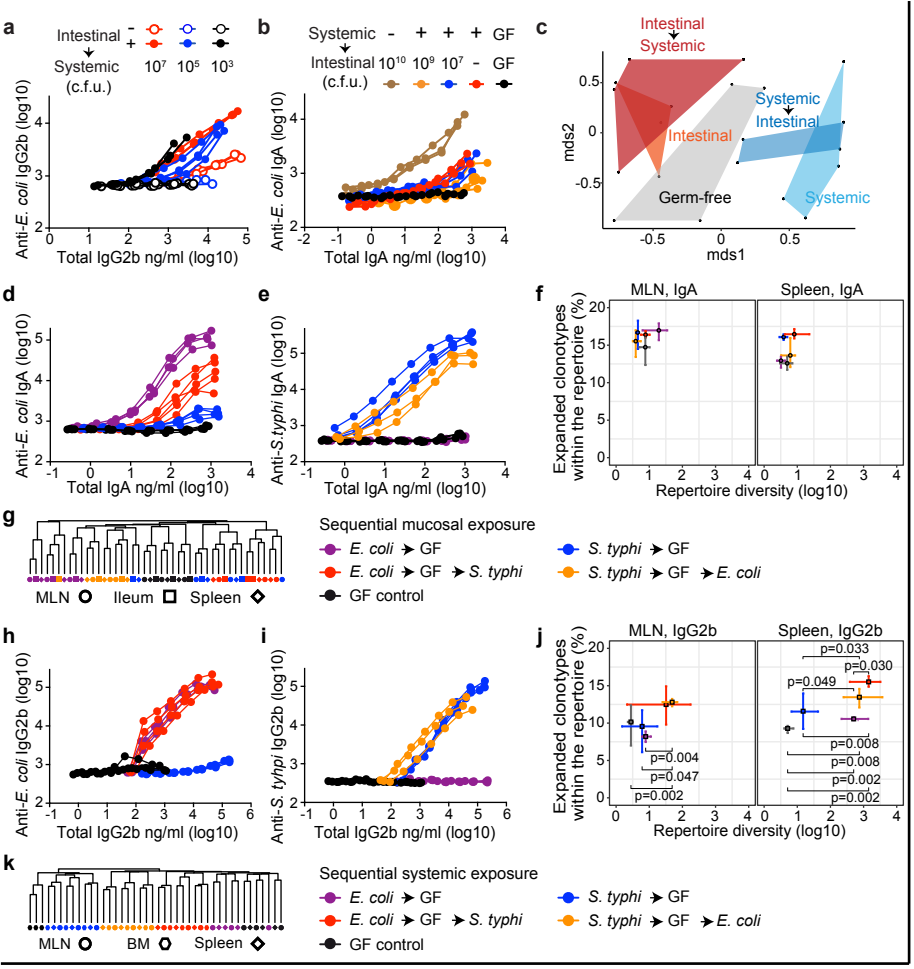
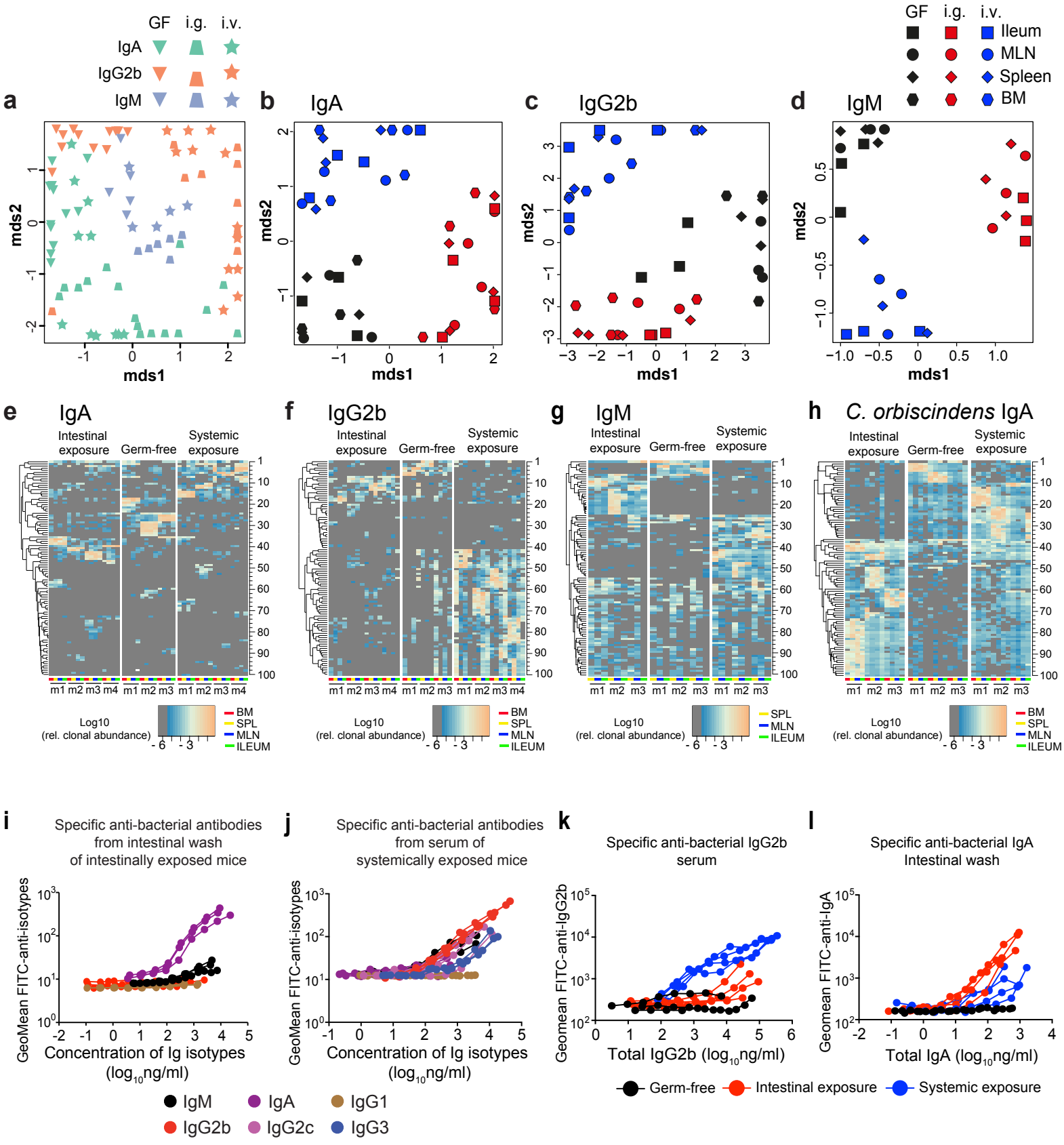
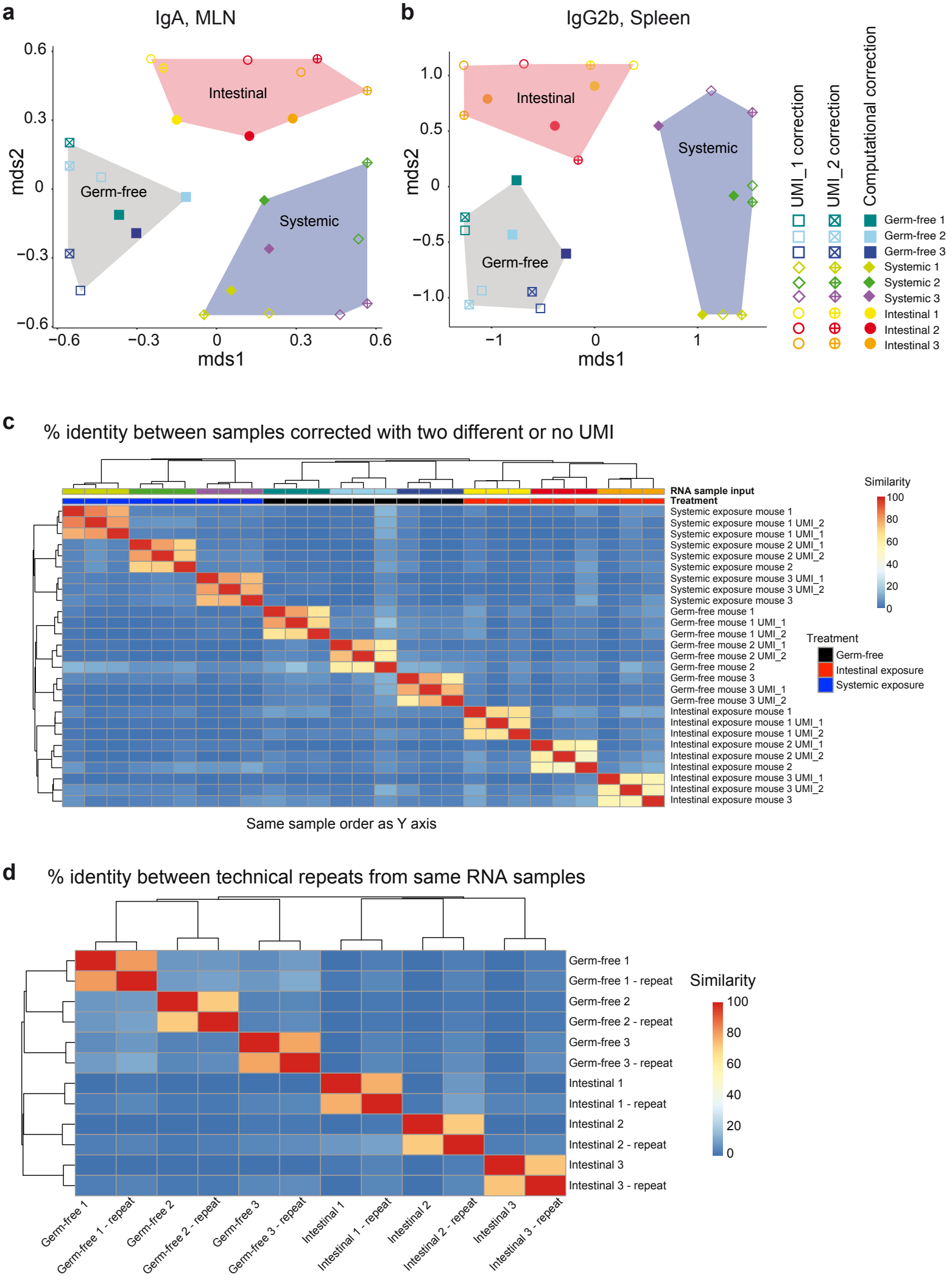

a
Intestinal
Systemic
(c.f.u.)
Anti-E. coli IgG2b (log10)
Total IgG2b ng/ml (log10)
b
Systemic
Intestinal
(c.f.u.)
GF
Anti-E. coli IgA (log10)
Total IgA ng/ml (log10)
c
Intestinal
Systemic
Systemic
Intestinal
Germ-free
Systemic
mds1
mds2
d
Anti-E. coli IgA (log10)
Total IgA ng/ml (log10)
e
Anti-S.typhi IgA (log10)
Total IgA ng/ml (log10)
f
MLN, IgA
Spleen, IgA
Expanded clonotypes within the repertoire (%)
Repertoire diversity (log10)
g
MLN
Ileum
Spleen
Sequential mucosal exposure
E. coli ➤ GF
E. coli ➤ GF ➤ S. typhi
GF control
S. typhi ➤ GF
S. typhi ➤ GF ➤ E. coli
h
Anti-E. coli IgG2b (log10)
Total IgG2b ng/ml (log10)
i
Anti-S. typhi IgG2b (log10)
Total IgG2b ng/ml (log10)
j
MLN, IgG2b
Spleen, IgG2b
Expanded clonotypes within the repertoire (%)
Repertoire diversity (log10)
p=0.033
p=0.030
p=0.049
p=0.008
p=0.008
p=0.002
p=0.002
p=0.004
p=0.047
p=0.002
k
MLN
BM
Spleen
Sequential systemic exposure
E. coli ➤ GF
E. coli ➤ GF ➤ S. typhi
GF control
S. typhi ➤ GF
S. typhi ➤ GF ➤ E. coli

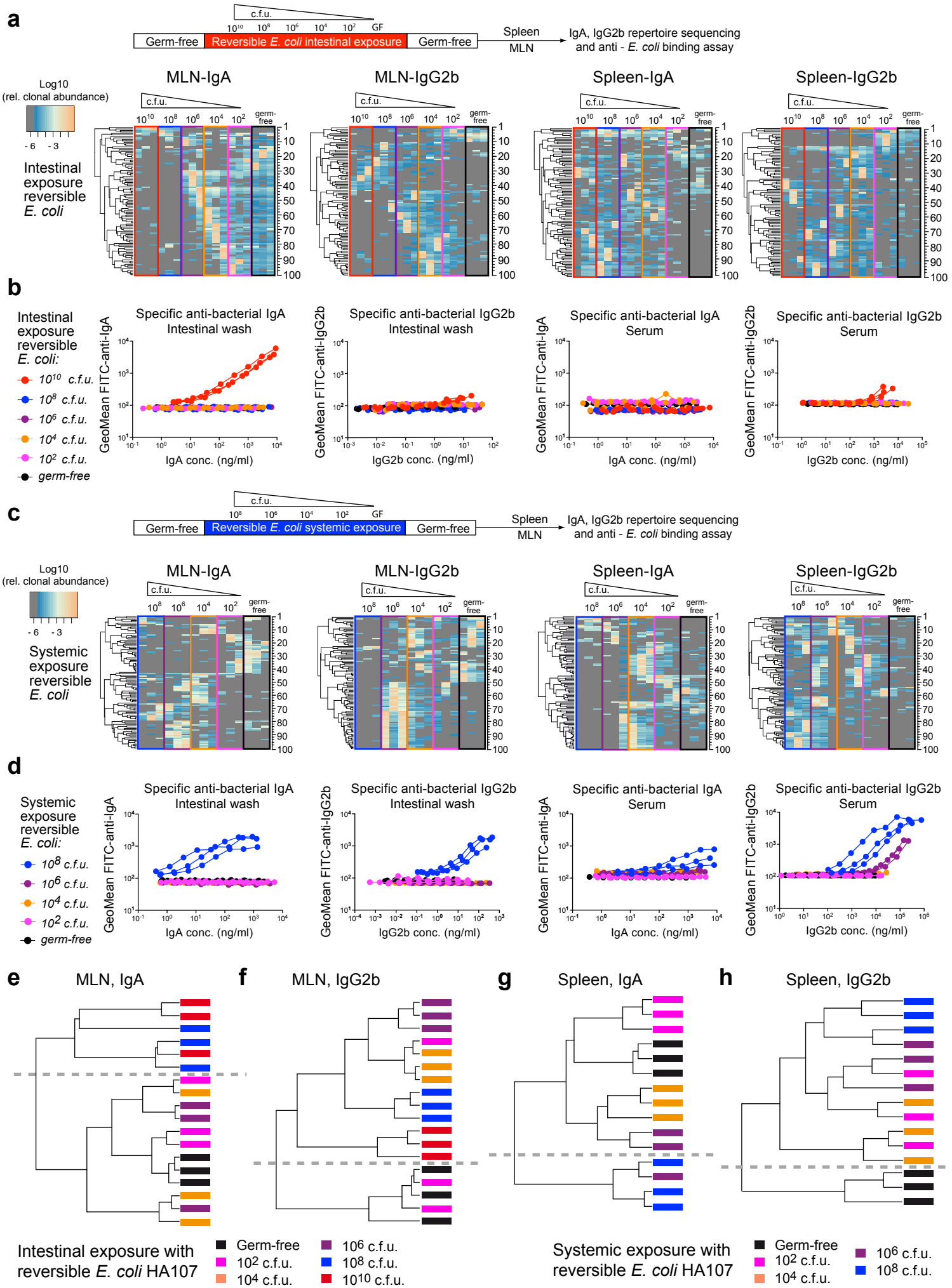
Figure 3



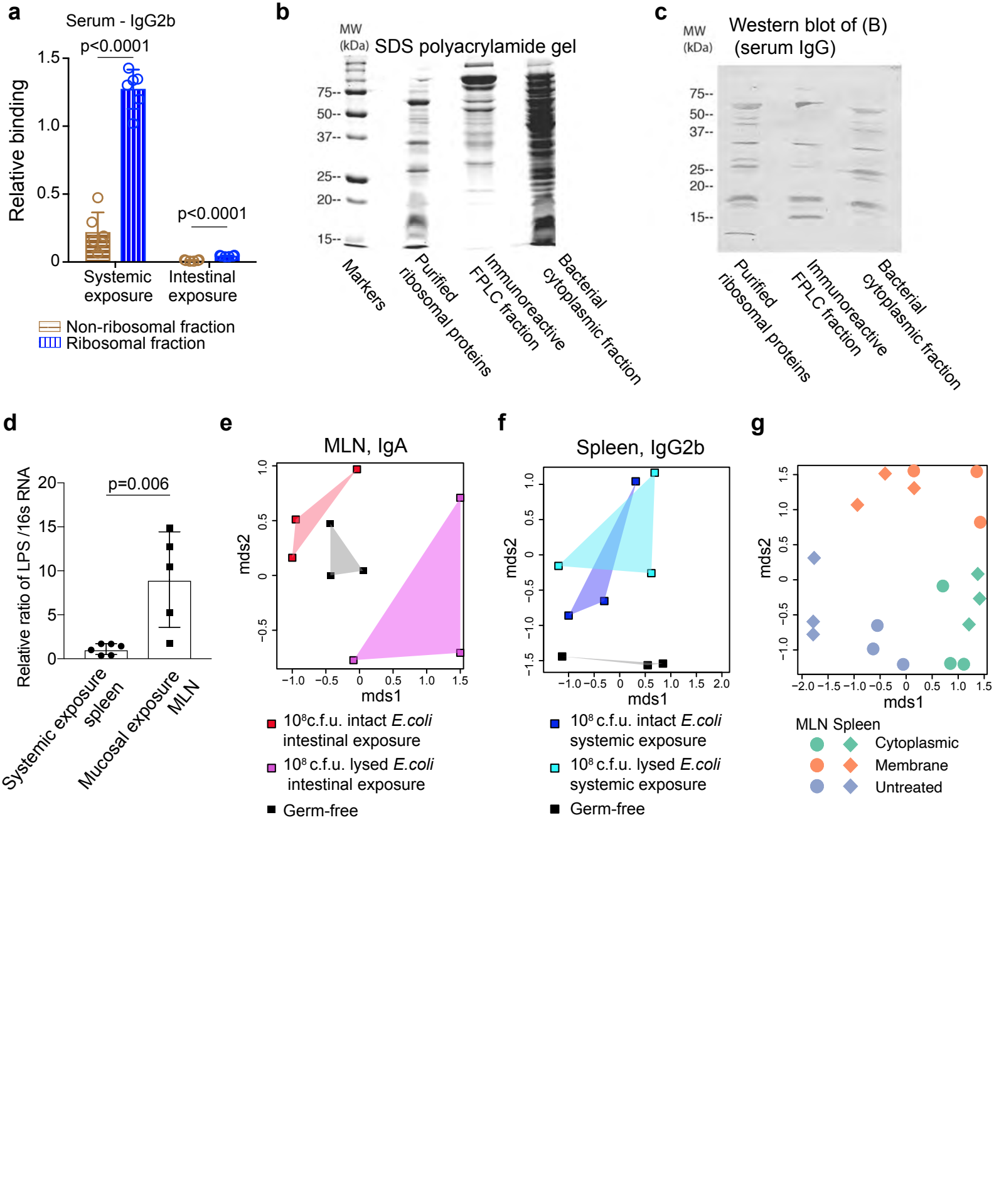
Extended Data Fig. 1



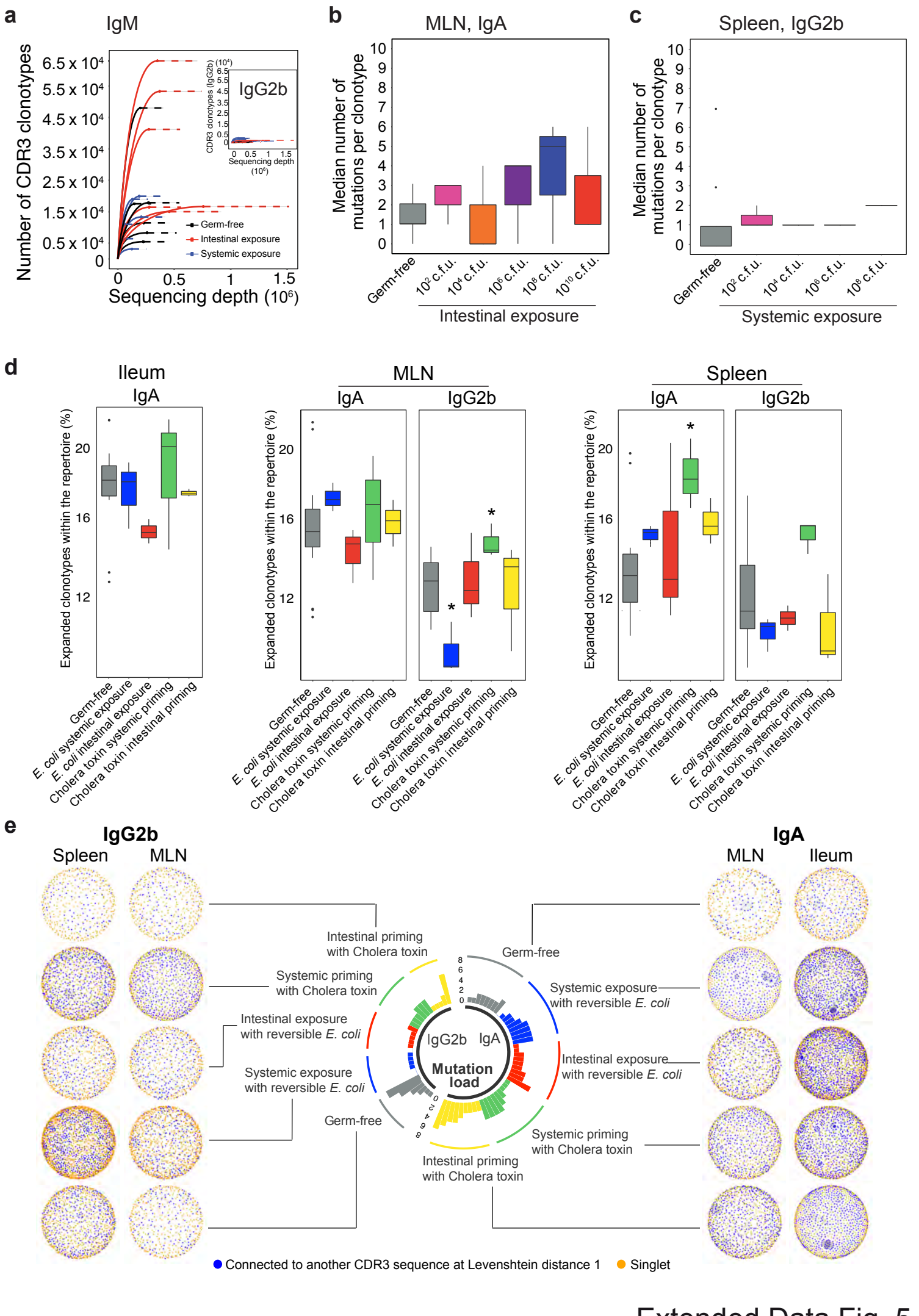
Extended Data Fig. 2



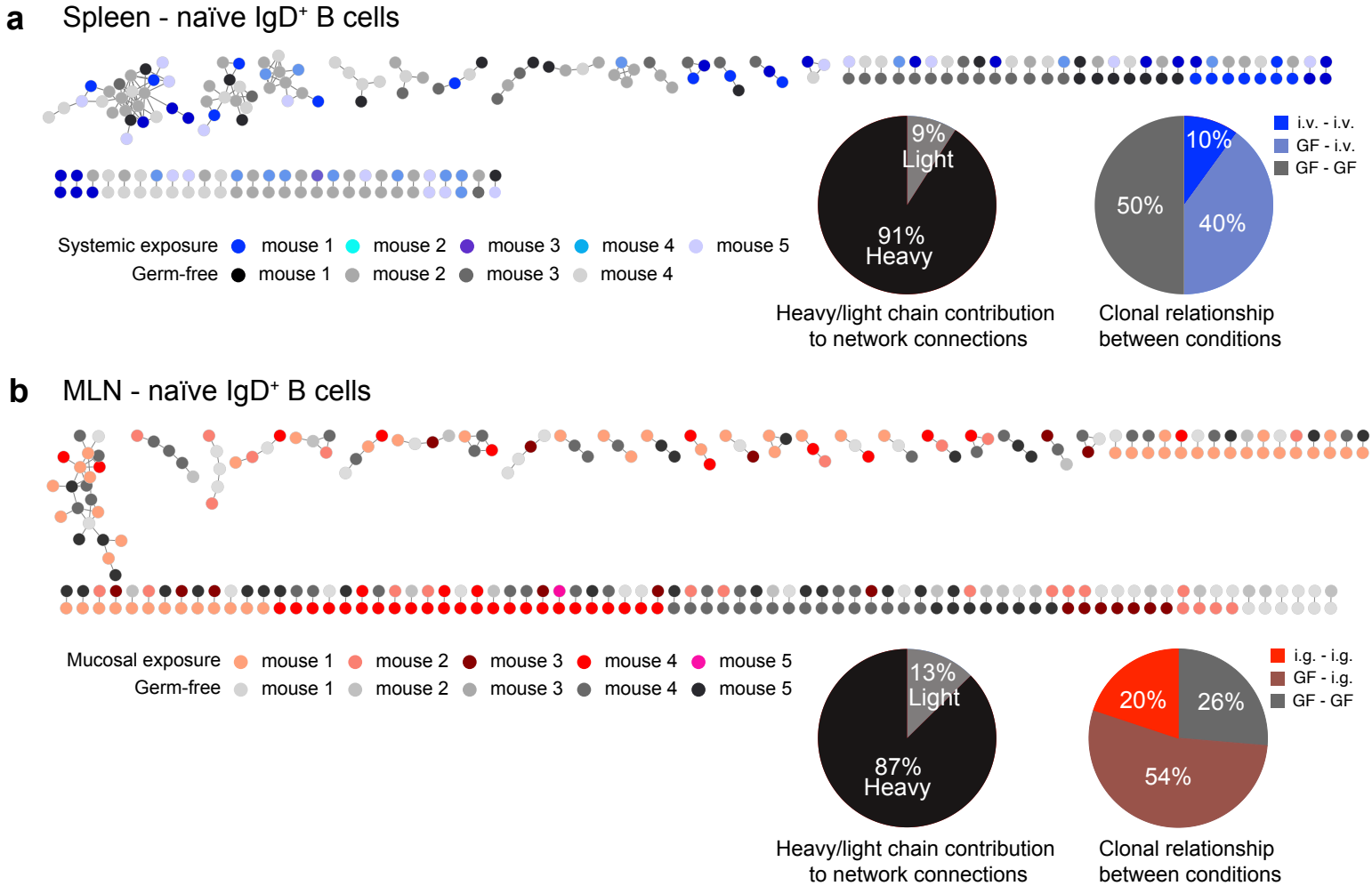
Extended Data Fig. 3

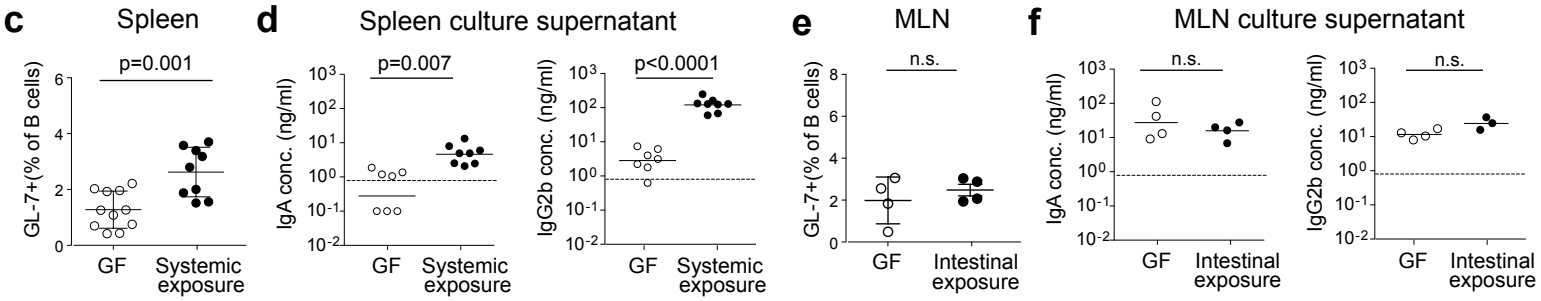


Extended Data Fig. 4

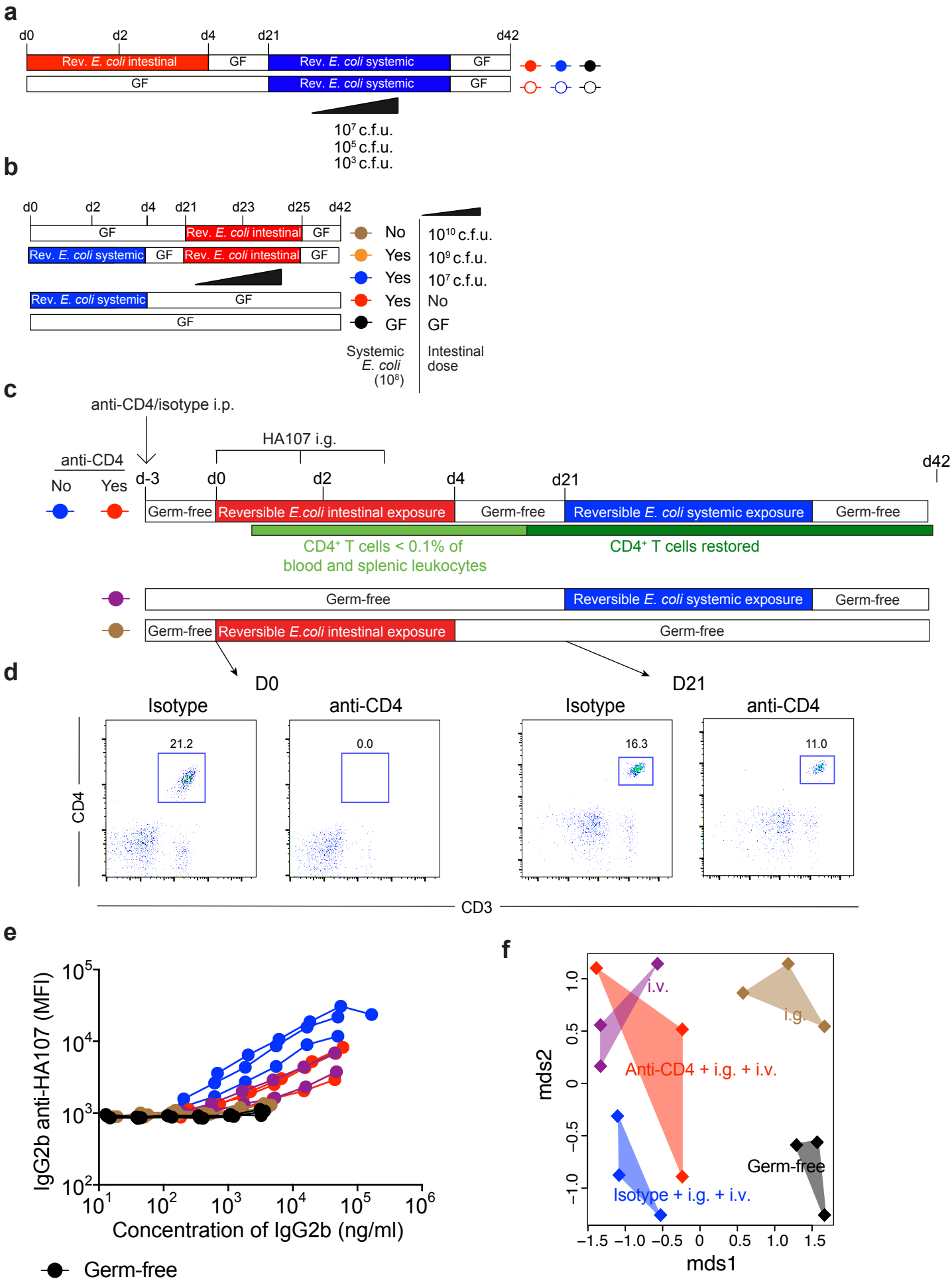


Extended Data Fig. 5



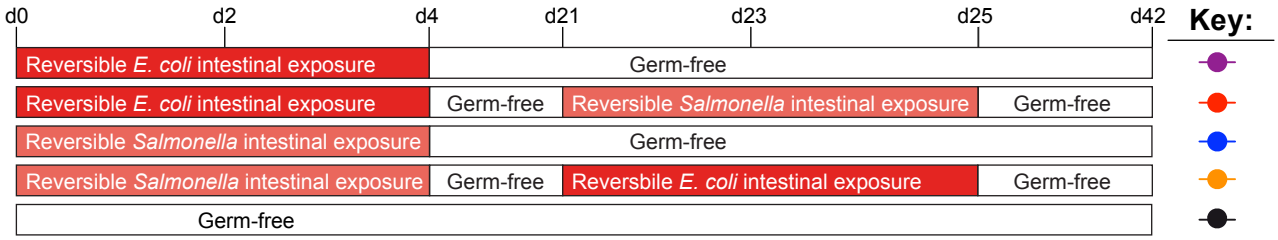


Extended Data Fig. 6

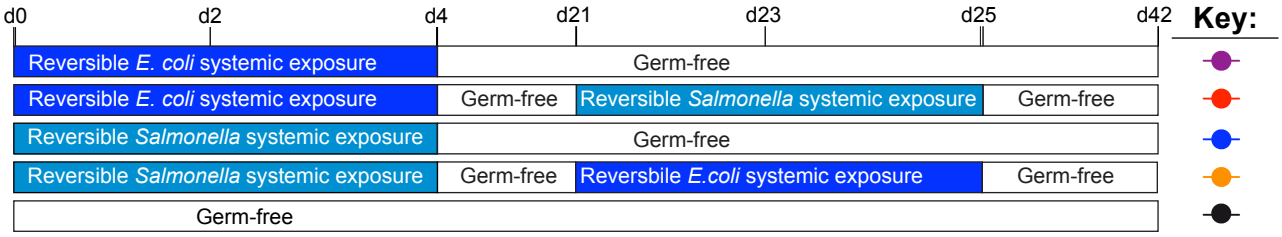


Extended Data Fig. 7

a



b



Extended Data Fig. 8