Tunable dynamics of B cell selection in gut germinal centres

Carla R. Nowosad1,4, Luka Mesin1,4, Tiago B. R. Castro1,2, Christopher Wichmann2,3, Gregory P. Donaldson1, Tatsuya Araki1, Ariën Schiepers1, Ainsley A. K. Lockhart2, Angelina M. Bilate2, Daniel Mucida2✉ & Gabriel D. Victora1✉

Germinal centres, the structures in which B cells evolve to produce antibodies with high affinity for various antigens, usually form transiently in lymphoid organs in response to infection or immunization. In lymphoid organs associated with the gut, however, germinal centres are chronically present. These gut-associated germinal centres can support targeted antibody responses to gut infections and immunization. But whether B cell selection and antibody affinity maturation take place in the face of the chronic and diverse antigenic stimulation characteristic of these structures under steady state is less clear. Here, by combining multicolour ‘Brainbow’ cell-fate mapping and sequencing of immunoglobulin genes from single cells, we find that 5–10% of gut-associated germinal centres from specific-pathogen-free (SPF) mice contain highly dominant ‘winner’ B cell clones at steady state, despite rapid turnover of germinal-centre B cells. Monoclonal antibodies derived from these clones show increased binding, compared with their unmutated precursors, to commensal bacteria, consistent with antigen-driven selection. The frequency of highly selected gut-associated germinal centres is markedly higher in germ-free than in SPF mice, and winner B cells in germ-free germinal centres are enriched in ‘public’ clonotypes found in multiple individuals, indicating strong selection of B cell antigen receptors even in the absence of microbiota. Colonization of germ-free mice with a defined microbial consortium (Oligo-MM12) does not eliminate germ-free-associated clonotypes, yet does induce a concomitant commensal-specific B cell response with the hallmarks of antigen-driven selection. Thus, positive selection of B cells can take place in steady-state gut-associated germinal centres, at a rate that is tunable over a wide range by the presence and composition of the microbiota.

Our intestines are constantly exposed to large amounts of antigens derived from diet and commensal microbes. The interaction of these antigens with the immune system takes place primarily in gut-associated secondary lymphoid structures, including gut-draining mesenteric lymph nodes (mLNs) and Peyer’s patches, where gut-associated germinal centres (gaGCs) provide a site for the hypermutation of immunoglobulin genes even under steady state. B cell antigen receptor (BCR)-driven selection and affinity maturation of antibodies occur efficiently in gaGCs upon oral immunization. However, given previous reports that steady-state gaGCs can form in a BCR-independent fashion, show little evidence of BCR-driven selection of specific antibodies at the sequence level, and are associated instead with the selection of polyreactive immunoglobulins, it has been postulated that gaGCs may act predominantly as diversifiers of the immunoglobulin repertoire, rather than fostering affinity maturation towards commensal microbes (reviewed in refs. ). We thus sought to determine the extent to which germinal-centre selection and antibody affinity maturation occur in the midst of chronic antigenic stimulation, and to define the impact of the microbiota on these processes.

To estimate the rate of B cell selection in steady-state gaGCs, we first used in situ photoactivation of mice engineered to express photoactivatable green fluorescent protein (PA-GFP) (Fig. 1a and Extended Data Fig. 1a) to sequence B cell immunoglobulin heavy chain genes (Igh) from 20 individual gaGCs from various mLNs of 5 mice housed under SPF conditions. Clonal diversity in SPF gaGCs spanned a wide range, with a median of 33 clones per germinal centre (using the Chao1 estimator function), a D50 value (the fraction of clones accounting for 50% of sequenced cells) of 0.20, and 30% of B cells belonging to the largest clone in the germinal centre (Fig. 1b, c). One of the 20 samples sequenced contained a highly dominant clone that accounted for 64% of cells in that germinal centre (Fig. 1b, c). Analysis of somatic mutations within this clone (Fig. 1d) showed the nested expansion of nodes with increasing numbers of mutations (indicated by arrows) typical of sequential positive selection.
This was confirmed in a much larger number of germinal centres using Brainbow\textsuperscript{12} multicolour fate-mapping\textsuperscript{13} (see Supplementary Information). We fate-mapped steady-state gaGC B cells in \textit{Aicda}^{CreERT2/+}, \textit{Rosa26}\textsuperscript{Confetti} mice held under SPF conditions by administering two doses of tamoxifen, two days apart (Fig. 1e). This labelled roughly 50% of B cells in both mLN s and Peyer’s patches, as estimated by the density of coloured cells in the germinal centres and by flow-cytometry experiments (Fig. If, g and Extended Data Fig. 1b). This fraction decreased progressively after labelling, probably as a result of germinal-centre B cells being replaced by incoming unlabelled clones as the response evolved (Fig. If, g). Because density estimations using Brainbow are sensitive to dropout of low-fluorescence germinal centres, we measured germinal-centre turnover by flow cytomtery using the \textit{Sippr}^{CreERT2} \times \textit{Rosa26}\textsuperscript{Stop-tdTomato} reporter. Pulse-labelling using this model allowed us to place the half-life of gaGC B cells at roughly two weeks under SPF conditions (Fig. 1h and Extended Data Fig. 1c).
Despite this rapid turnover, the normalized dominance score (NDS, an estimate of the frequency of B cells in a germinal centre that carry the dominant colour11,15; see Supplementary Information) in gaGCs increased progressively to day 23 after tamoxifen, when 11% of germinal centres (6 of 57) scored 0.75 or higher (Fig. 1f, g). The strongest clonal expansions that occur in mLN germinal centres are therefore large and rapid enough to generate dominant lineages, despite the replacement of labelled clones with incoming unlabelled B cells. In germinal centres with NDS values of more than 0.5 could occasionally be detected as early as day 14 after tamoxifen, peaking at day 23 after tamoxifen, when 15% of Peyer’s patches (3 of 20) had reached NDS values of more than 50% (Extended Data Fig. 1e). We conclude that clonal selection is detectable in gaGCs, despite chronic exposure to a high burden and diversity of foreign antigens and the rapid turnover of B cell clones.

To understand the relationship between clonal selection and affinity maturation in gaGCs, we used vibratome slicing of agarose-embedded AID-Confetti lymph nodes44 (Extended Data Fig. 2a) to isolate gaGCs containing ‘winner’ clones, where antigen-driven selection is most likely to have occurred44. Sequencing of IgH from B cells sorted from...
such germinal centres showed evidence of ‘clonal bursts’—jackpot-type positive selection events in which multiple B cells descending from a single somatic hypermutation (SHM) variant account for a large fraction of cells in a germinal centre (Fig. 2a and Extended Data Fig. 2a–c). Because clonal bursts are regularly associated with the acquisition of affinity-enhancing mutations, we produced recombinant monoclonal antibodies using burst-point sequences to probe for binding to commensal bacteria (Supplementary Table 1). Despite the variation inherent to bacterial flow cytometry, two of seven antibodies produced from burst-associated immunoglobulin sequences reproducibly bound faecal bacteria (Fig. 2b and Extended Data Fig. 2d, e). Binding followed different patterns: whereas monoclonal antibody S078 bound strongly to a small population of bacteria, SI20 bound with moderate intensity to a much larger cohort (Fig. 2b). These two antibodies—as well as two other clones (S210 and S212) —reacted with bacteria-rich centrifugation fractions, as measured by enzyme-linked immunosorbent assay
To investigate the influence of commensal diversity on gaGC selection dynamics, we rederived AID-Confetti mice into germ-free conditions, in which germinal centres still form, as well as into stable vertical colonization with a consortium of 12 bacterial strains representing major phyla present in the mouse gut (Oligo-MM12; Extended Data Fig. 3f). When compared with SPF mice, germ-free mice had higher frequencies of germinal-centre B cells in jejunal and ileal mLNs and lower frequencies in duodenal and caecal-colonic mLNs and Peyer’s patches (Extended Data Fig. 4a, b). Germ-free germinal centres were strongly skewed away from IgG2a and IgA towards IgG1 (Extended Data Fig. 4c).

Colonization with Oligo-MM12 microbiota partly restored the phenotype of SPF mice, increasing germinal-centre B cell frequency in distal Peyer’s patches and IgG2a in proximal Peyer’s patches (Extended Data Fig. 4d).

Multicolour fate-mapping showed that strongly selected germinal centres accumulated at a markedly faster rate in germ-free mice than they did under SPF conditions (Fig. 3a–c). By day 23 after tamoxifen, roughly 56% of mLN germinal centres had reached an NDS of 0.75 or higher, compared with around 11% in SPF mice (Fig. 3c). This was confirmed at the Igh sequence level: 11 of 14 mLN germinal centres picked at random from vibratome slices had dominant clones that accounted for more than 50% of all B cells in the germinal centre, compared with 1 of 20 germinal centres in SPF conditions (Fig. 1a–c and Extended Data Fig. 5a–d). Faster selection of Brainbow colours was also observed in germ-free Peyer’s patches, where 6 of 9 germinal centres exceeded an NDS of 0.5 by day 20–23 after tamoxifen, compared with 3 of 20 under SPF conditions (Fig. 3a–c). Germinal-centre selection in Oligo-MM12-colonized mice fell between the SPF and germ-free rates (Extended Data Fig. 5e–g). Therefore, selection in gaGCs is not dependent on a fully diverse microbiota, and in fact becomes accelerated in the absence of commensal bacteria.

Clonal phylogenies of single-coloured germinal centres from germ-free and Oligo-MM12-colonized mice revealed strong clonal bursting, as shown by the presence of large expansions of B cells with identical variable heavy chain (VH) immunoglobulin sequences and multiple inferred descendants (Fig. 3d, e and Extended Data Fig. 6a, b, arrows). We produced monoclonal antibodies from 16 immunoglobulin sequences that were strongly selected under each condition (including those indicated by named arrows in Fig. 3d, e and Extended Data Fig. 6a, b; Supplementary Table 1). Of seven monoclonal antibodies cloned from Oligo-MM12, three (M216, M218, and M220) bound faecal bacteria fractions from Oligo-MM12-colonized mice (Fig. 3f, i, j), and one (M218) bound to cultured Enterococcus faecalis, as assessed by both flow cytometry and dot blotting (Fig. 3g, h). Reversion of somatic mutations resulted in larger decreases in binding for M216 and M220 and a modest decrease for M218, as shown by flow cytometry and ELISA (Fig. 3h–j). Thus, as with SPF microbiota, vertical colonization with Oligo-MM12 triggers efficient antigen-driven maturation towards commensals in steady-state gaGCs.

We subjected the nine monoclonal antibodies obtained from germ-free winner clones to an array of assays that covered major potential sources of antigen, including food, autoantigens (anti-nuclear antibody and intestinal tissue antigens), faecal bacteria, and a standard polyclonality panel. None of the germ-free antibodies reacted above background levels in any of these assays (Extended Data Fig. 6c–f). To determine whether germ-free germinal centres were indeed populated in a BCR-dependent manner, or simply stochastically owing to a lack of antigenic stimulation, we searched for commonalities in the Igh sequences of winner clones, along with additional sequences obtained from single germinal-centre B cells from mLN vibratome slices and whole mLNs and Peyer’s patches of wild-type germ-free mice. This revealed substantial overlap of clonotype ’themes’ across individuals, which we regarded as unlikely to be random given the small pool of germinal centres sampled. Two themes were particularly prevalent (Fig. 4a, b). One used the relatively rare VH1-47–47 segment, coupled to joining segments JH2 or JH3 via an 11–12-amino-acid heavy-chain complementarity-determining region 3 (CDRH3) sequence that begins with the consensus sequence ARGSRNY (Fig. 4a). No commonalities in light-chain usage were detected at this sampling depth. After allowing a one-amino-acid substitution in the ARGSNY motif, this theme was present in 5 of 7 germ-free mice, accounting for 16.6% of all cells sequenced (Fig. 4b and Extended Data Fig. 7a). Clones with these characteristics represented only 0.00006% of all reads (1 in roughly 17,000) in a previously published database of naïve B cell Igh sequences from C57BL6 mice, containing more than 30 million reads representing 2.5 million unique rearrangements from 5 mice (P < 2.2 × 10^-16 compared with germ-free gaGCs).

A second public clonotype was encoded by the rare V1-12 segment, with the stricter seven-amino-acid consensus CDRH3 sequence, AREGFAY, followed by JH3 (Fig. 4a). Again, no patterns of light-chain usage were identified. Allowing a one-amino-acid substitution in CDRH3, this...
clonotype was present in 2 of 7 germ-free mice, representing roughly 3% of all cells sequenced. This clonotype was also heavily dominant in all 3 single-coloured germinal centres sorted from different organs of one of three Oligo-MM1–colonized mice, accounting for 171 of 191 cells sequenced from this mouse (Fig. 4b and Extended Data Fig. 8a). Despite its short length, the Vι1–12/AREGFAY/Jκ3 combination was seen only 7 times (in 1 of 5 mice) in the more than 30 million reads of the naïve B cell database, and only twice more in one other mouse if a single amino-acid substitution in the AREGFAY motif was allowed (P < 2.2 × 10−10 compare with germ-free gaGCs). Both clonotypes accumulated somatic mutations that converged across mice and between SPF and Oligo-MM1 conditions (Extended Data Fig. 7b). In agreement with their failure to bind food protein extracts (Vι1–47 and Vι1–12 clonotypes are represented by monoclonal antibodies G082/G226 and M228/M232, respectively) (Fig. 3d–j and Extended Data Fig. 6), both clonotypes were also detected in mice fed a custom-made protein-free chow formulated from purified ingredients (Extended Data Fig. 8b, c and Supplementary Table 2).

To assess whether reliance on public clonotypes is broadly characteristic of germ-free gaGCs when sampled in an unbiased manner, we developed a multiwell incidence-based approach to measure clonal overlaps between mice with high confidence. In total, we sequenced roughly 80 thousand cells from gaGCs in the mLNs and Peyer’s patches of 6 germ-free, 6 SPF and 7 Oligo-MM1–colonized mice (Extended Data Fig. 9a–d). Confirming our initial findings, the Vι1–12/AREGFAY/Jκ3 theme (regardless of Jκ3 usage) was present in 6 of 6 germ-free and 5 of 7 Oligo-MM1–colonized mice, corresponding to 4.4% (173 of 3,929) and 2.4% (98 of 4,043) of clone wells (a number obtained by multiplying each clone by the number of wells it was found in) (Extended Data Fig. 9a–c) in each condition (Fig. 4c). One such clone was also detected in a single SPF mouse (5 of 3,629 clone wells) (Fig. 4c), and 3 others (one using the Jκ4 segment) were observed in our photoactivation data (see Fig. 1 and Supplementary Spreadsheet 1), indicating that full bacterial colonization is not sufficient to completely exclude such clonotypes from gaGCs. Clonotype Vι1–12/AREGFAY was present, again at a lower frequency, in 5 of 6 germ-free mice (44 of 3,929 clone wells) (Fig. 4c).

Analysis of Levenshtein distances showed that, in germ-free and Oligo-MM1–colonized but not SPF mice, CDR3 sequences of Vι1–47 and Vι1–12 clones were closer in sequence to the ARGNSY and AREGFAY motifs, respectively, than were CDR3 sequences of clones using other V segments (Extended Data Fig. 9e). Of note, both Vι1–12/AREGFAY/Jκ3 and Vι1–12/AREGFAY/Jκ3 rearrangements were found recurrently in Peyer’s patches of germ-free mice in independent work published while this paper was under review, further underscoring the public nature of these two clonotypes. Finally, public clonotypes in general (defined as any recurrent Vι/Jκ combination with exactly matching CDR3 amino acid sequences) were markedly more frequent across germ-free and Oligo-MM1 gaGCs than across SPF mice (Fig. 4d and Extended Data Fig. 9f). Although almost all germ-free-associated clonotypes were either eliminated or reduced to below detection levels by SPF colonization, the Oligo-MM1 gaGC repertoire overlapped more substantially with that of germ-free mice, indicating that Oligo-MM1 colonization is insufficient to completely replace the germ-free response (Fig. 4d and Extended Data Fig. 9g). Thus, rather than being stochastically populated, gaGCs display stringent selection driven by BCR specificity under conditions of low antigenic diversity, resulting in rapid focusing of germinal centres on dominant lineages and pronounced reliance on clonotypes found repeatedly across different mice.

We have shown that gut-associated germinal centres undergo clonal selection and antigen-driven maturation in the absence of infection or immunization, and that the rate of selection varies markedly depending on the presence and complexity of the gut microbiota. Under microbial-replete conditions, selection of highly dominant clones is relatively rare and is associated with improved binding of commensal-derived antigens. At the other extreme, gaGC selection accelerates precipitously in the absence of microbes, leading to strong convergent selection of IgH clonotypes across mice. Thus, clonal selection in steady-state gaGCs is a tunable process (see Supplementary Information for further discussion). The ability to generate specific, affinity-matured responses to commensals would allow targeted control of individual bacterial species and may thus play a part in maintaining the composition of the gut microbiota.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2865-9.
Methods

Mice and treatments

Mice were housed at a temperature of 72°F and humidity of 30–70% in a 12-h light/dark cycle with ad libitum access to food and water. Male and female mice aged 8–12 weeks at the start of the experiment were used. PA-GFP transgenic mice were generated in our laboratory2. Rosa26Confetti (B6.129P2-Gt(Rosa)26Sor/J) and Rosa26Stop-tomato (B6. Cg-Gt(Rosa)26Sor/J) mice were obtained from Jackson Laboratories. Rosa26Confetti mice were backcrossed to C57BL/6 mice for several generations in our laboratory and restricted to the Igk<sup>–/–</sup> haplotype. Aicda<sup>−/−</sup> (Aicdajtm1Jdd/J) mice were a gift from C.-A. Reynaud and J.-C. Weill (Institut Necker, Paris). Sipr2<sup>CreERT2</sup> (Tg(Sipr2-cre/ERT2)#Kuro) BAC-transgenic mice<sup>4</sup> were a gift from T. Okada (Riken, Tokyo). We colonized germ-free AID-Confetti breeders with a single gavage of Oligo-MM<sup>22</sup> and monitored colonization (including the presence of the entire consortium in successive generations) by specific amplification of individual bacterial members by quantitative polymerase chain reaction (qPCR; see below). Mice were bred and maintained in isolators. Vertically colonized AID-Confetti Oligo-MM<sup>22</sup> mice were used for all experiments. To deplete protein antigen from the diet, we used a custom solid diet containing free amino acids (Modified TestDiet 9GCV with 5% cellulose; composition details are in Supplementary Table 1). Diets were irradiated at more than 45 kGy to ensure sterility for germ-free conditions and were provided to mice from one week of age until the time of analysis.

Recombination of floxed alleles in both AID-Confetti and Sipr2-Tomato mice was induced by two gavages of 10 mg tamoxifen (Sigma, catalogue number T5648) dissolved in corn oil at 50 mg ml<sup>–1</sup>, 2 days apart. To ensure that selection was not a function of the route of administration of either tamoxifen or corn oil, we also injected two AID-Confetti mice intra-peritoneally once with 10 mg tamoxifen for analysis at the SPF 21–23 day time point. In germ-free AID-Confetti mice, tamoxifen was prepared and administered under sterile conditions to mice housed in individually ventilated isocages (Techniplast). Sample sizes were not calculated a priori. Given the nature of the comparisons (mice born under differing microbial colonization status), mice were not randomized into each experimental group and investigators were not blinded to group allocation.

All animal procedures were approved by the Institutional Animal Care and Use Committee of the Rockefeller University.

Oligo-MM<sup>22</sup> qPCR

Colonization of mice by the Oligo-MM<sup>22</sup> consortium was confirmed and monitored over generations by qPCR, using primer pairs specific to each species as previously validated (individual strain primer sequences are in Supplementary Table 5, adapted from ref. 20; universal bacterial qPCR primers are as follows: UNIF340-ACTCCTACGGGAGGCAGCAGT and UNIR44R-ATTACCGCGGCTGCTGGC). DNA was extracted from faecal samples using the ZR Fecal DNA kit (Zymo Research) according to the manufacturer’s instructions. Quantitative PCR was performed with the Power SYBR Green master mix (Applied Biosystems). The average cycle threshold (Ct) value of two technical replicates was used to quantify the relative abundance of each species’ 16S ribosomal RNA using the ΔΔCt method, with the universal 16S rRNA primers serving as controls between samples. Relative abundance was corrected according to the genome copy number of 16S rRNA for each species.

Multiphoton imaging and photoactivation

mLNs and Peyer’s patches were collected and imaged as previously described<sup>23</sup>. In brief, adipose tissue and excess epithelium were removed under a dissecting microscope, and mLNs and Peyer’s patches were mounted in phosphate-buffered saline (PBS) between two coverslips held together with vacuum grease, as previously described<sup>23</sup>. Mounted mLNs and Peyer’s patches were imaged on an Olympus FV1000 upright microscope with a ×25 Plan water-immersion objective (numerical aperture (NA) 1.05) and a Mai-Tai DeepSee titanium-sapphire laser (Spectraphysics). Confetti alleles were imaged at an excitation wavelength of λ = 930 nm. Fluorescence emission was collected in three channels, using a pair of CFP (480/40 nm) and YFP (525/50 nm) filters separated by a 505 nm dichroic mirror to detect CFP, GFP and YFP, and a dedicated RFP filter (605/70 nm). In situ photoactivation was performed as previously described<sup>12,23</sup>. PA-GFP transgenic mice were injected intravenously with 5 μg of a non-blocking antibody to CD35 (clone SC12, produced in house) conjugated to Cy3 to label follicular dendritic cells (FDCs). Clusters of CD35<sup>–</sup> cells were identified by imaging at λ = 950 nm, where photoactivation does not take place, and three-dimensional (3D) regions of interest were photoactivated by higher-power scanning at λ = 830 nm. Fragments of lymph nodes were then processed for flow cytometry as described below.

Image analysis

Colour dominance in AID-Confetti germinal centres was determined in 3D data sets reconstructed using ImageJ software and the Bio-formats plugin. Cells of each colour or colour combination were counted manually in two or more 2D slices, at least 20 μm apart, using the Cell Counter plugin. For overviews of Peyer’s patches, colours were counted in the same way, but from a single imaging plane. The normalized dominance score (NDS) was calculated as before<sup>23</sup>. In brief, we first estimated the fraction of recombined cells in each germinal centre by calculating the density of fluorescent cells per area unit (100 μm<sup>2</sup>) in anatomically defined dz areas in which cell distribution was homogeneous, then multiplied the density by the fraction of coloured cells accounted for by the dominant colour in the germinal centre. Fully recombined germinal centres have a cell density of very close to 1 (ref. 1), making an NDS of 1 a good approximation of 100% occupancy by the dominant colour. The sizes of germinal centres in mLNs and Peyer’s patches were calculated as the cross-sectional area of the largest available z-section. All image analysis was carried out in ImageJ.

Lymphocyte flow cytometry and sorting

To evaluate the dynamics of germinal centres and the distribution of isotypes across space and time, we isolated individual mLNs that drain the duodenum, jejunum, ileum or caecum/colon as previously described<sup>10</sup>. Pairs of Peyer’s patches were isolated from the most proximal part of the duodenum or the most distal part of the ileum. Cells were isolated by maceration using disposable micropesites (Axygen) in 100 μl of PBS supplemented with 0.5% bovine serum albumin (BSA) and 1 mM EDTA (constituting PBE buffer), and single-cell suspensions were obtained by two passes through a 70-μm mesh. Cells were stained with antibodies against B220 protein, T cell antigen receptor (TCR) α/β chains (or a ‘dump’ mixture containing antibodies against CD4, CD3, CD8 and NK1.1), CD38, Fas, IgM, IgG<sub>1</sub>, IgG<sub>2a</sub> and IgA, supplemented with Fc block (see Supplementary Table 3) for 30 min on ice. Samples were run on a FACS LSRII or FACS Symphony (BD Biosciences).

To sort B cells from single mLN germinal centres, we first determined the localization of single-coloured germinal centres (see ‘Multiphoton imaging and photoactivation’ section above). As previously described<sup>10</sup>, we embedded selected mLNs in 4% low-melt NuSiue GT agarose in PBS that had been heated to boiling then cooled to 37 °C before embedding. We then cut lymph nodes into 300-μm slices using a Leica VT1000A vibratome. Slices were further dissected under a Leica M165FC fluorescence stereomicroscope using a double-edged razor blade to isolate single germinal centres from slices in which several germinal centres were present. Slice fragments were placed in micro-centrifuge tubes containing 100 μl PBE, macerated using disposable micropesites and dissociated into single-cell suspensions by gentle
vortexing. We then added 100 µl of 2× antibody stain (comprising antibodies against CD38, Fas, B220 and TCR-β) supplemented with Fc block; see Supplementary Table 3) to the cell suspension, which was incubated on ice for 30 min. Single cells were sorted as described below using FACS Aria II or III cell sorters. Cells positive for any fluorescent Confetti colour, detected as previously described, were index-sorted into 96-well plates containing 5 µl TCL buffer (Qiagen) supplemented with 1% β-mercaptoethanol. The precise assignment of colours from index-sorted cells was carried out post-acquisition using Diva software, version 8.0.2, using all four channels.

For the isolation of single follicles from non-fluorescent germ-free mice, mice were injected with anti-CD35 (8C12) Alexa Fluor 594, individually dissected, imaged, and sliced as above. Slices of 250 µm thickness were examined under a stereomicroscope and manually microdissected into slice fragments comprising roughly one follicle each, as defined by staining of FDCs (see the section ‘Multiphoton imaging and photoactivation’ above). Two slice fragments were prepared per mLN, isolated from slices that were at least 1,000 µm (four 250-µm slices) apart in the intact node. Slice fragments were prepared for single-cell sorting as above. For the isolation of single germinal centres from fluorescent germ-free mice, we combined Confetti fate-mapping with anti-CD35 follicle staining. Germ-free AID-Confetti mice were treated with tamoxifen as described above, and, 24 h before injection, were injected with 5 µg anti-CD35-Cy3 antibody. Mesenteric lymph nodes were sliced and imaged and fluorescent clusters were manually excised, avoiding any other follicles (marked with anti-CD35 antibody). In this case, both fluorescent and non-fluorescent germinal-centre B cells were single-sorted for sequencing.

For single-cell sequencing from whole mLNs and Peyer’s patches, organs from non-fluorescent germ-free mice were isolated and processed into single-cell suspensions by maceration using disposable micropipettes. After staining through a 70-µm mesh, cells were stained and sorted as above. For incidence-based sequencing experiments, samples from mLNs and Peyer’s patches were stained as above, and sorted at 100 cells per well into 16–32 individual wells in a 96-well plate containing 10 µl TCL buffer per well (Qiagen) supplemented with 1% β-mercaptoethanol. Analysis of data from flow-cytometry experiments was carried out using FlowJo software, version 10.5.3 (Tree Star Inc.).

**Immunoglobulin sequencing**

RNA from single cells or 100-cell pools was reverse-transcribed using oligo-dT primers as in refs. 11,16. PCR primers were used as previously described with the addition of IgA-specific amplification primers Cα outer (ATCAGGCAGCCGATTATCAC) and Cα inner (GAGCTCGTGGGAGTGTAGTG)26. Pooled PCR products were then purified using SPRI beads (0.7× volume ratio), gel purified and sequenced with a 500-cycle Reagent Nano kit v2 for single-cell libraries and with a 600-cycle Reagent kit v3 for 100-cell pool libraries on the Illumina MiSeq platform.

**Sequence analysis**

For single-cell Igk analyses, raw paired-end sequences were merged across the overlapping regions using PANDAs (version 2.11) for full amplicon reconstruction29, then processed with the FASTX toolkit. Only those sequences with high counts for each single cell/well were analysed. For annotation of V(D)J gene rearrangement, the seqs sequences obtained were submitted to both HighV-QUEST (version 1.6.9)30 and Vbase2 (ref. 31) databases, choosing the assignment that yielded the lowest number of somatic mutations in case of discrepancy. Sequences with common Vh/Jh genes and the same CDR3 length were grouped into clonal lineages when CDR3 nucleotide identity was 75% or more. Igk sequences were determined using the same method when needed for the cloning of monoclonal antibodies. Clonal lineage trees were plotted using GCTree32, with the unmaturated V gene sequence of the V(D)J rearrangement used as an outgroup. Logo plots for public clonotypes were created by first using the T-Coffee algorithm33 to align all CDR3 sequences bearing Vh, Jh and CDR3 connected by a 11-amino-acid or 12-amino-acid CDR3. The results of this alignment were then processed using the WebLogo3 web server34. Public clones were searched in the Greiff et al.31 database using R. Matching required exact Vh,Jh and CDR3 length matches and up to one-amino-acid mismatch in the ARGSNY or AREGFAY motifs. Dendrograms were generated using ClustalX (version 2.1) and FigTree (version 1.4.4) and branches were coloured in Adobe Illustrator according to the sequence annotations.

For incidence-based sequencing, raw paired-end sequences were merged as above and submitted to HighV-QUEST (version 1.6.9) for annotation30. The output database was then processed in the R environment to remove non-functional and out-of-frame sequences. Any sequences with less than six reads were discarded. Expanded clonal populations were defined using Change-O35. Briefly, sequences within the same mouse that shared Vh,Jh genes and with a maximum CDR3 hamming distance of four nucleotides were grouped into a single clone*well of size equivalent to the number of wells this sequence was found in and retaining the full list of CDR3 sequences in the cluster (generating a CDR3list for that clone*well). To remove the possibility of errors due to sequence misassignment and interwell contamination, sequences found in a single well within their mouse of origin were eliminated. CDR3 sequences were translated to amino-acid sequences and non-CDR3 sequences were discarded before assignment of public clones. Our definition of a public clone required that clones have the same Vh and Jh segments, and an exact match between any of the CDR3 sequences in the clone*well’s CDR3lists. Public clonotypes, Vh,1–47/ARGSNY and Vh,1–12/AREGYF were identified in this database using the same criteria as above but without restriction of Jh segment usage. Circular ideogram plots were created using Circos (version 0.69-9), with each individual mouse represented by a single ideogram bar. The full analysis pipeline is available at https://github.com/victoraLab/MIBS. Levenshtein distances were calculated in R using the stringdist package. Wells assigned to the same clone within the same mouse were counted only once to avoid overrepresentation due to clonal expansion. All sequences with 15 or more reads were used in this analysis, regardless of the number of wells in which a clone was present. For the Vh,1–47/ARGSNY clonotype, distances were calculated starting from the string ARGSNYYXXXDY and plotted as the resulting difference minus 4 to correct for the 4 ‘X’ characters.

**Production of monoclonal antibodies**

Heavy- and light-chain sequences obtained from the same expanded nodes in somatic hypermutation (SHM) phylogenies of germinal-centre B cells, as well as their deduced unmutated ancestors, were produced and assembled into custom mammalian expression vectors (modified from ref. 36) encoding the human IgG, and IgK constant regions (Twist Biosciences). Plasmids were transfected into Freestyle 293-F suspension cells (obtained from Life Technologies and tested for mycoplasma contamination in our laboratory), and monoclonal antibodies were purified using protein-G-affinity chromatography, as previously described11,37. Integrity of all monoclonal antibodies was assayed/quantified by SDS-PAGE and biolayer interferometry on an Octet Red96 instrument using protein-G-coated sensors (FortéBio).

**Assays for monoclonal antibody binding**

Bacterial flow cytometry was carried out using protocols adapted from refs. 4,38. Freshly collected faeces from SPF or Oligo-MM3 mice from our own facility were macerated with micropestles in 100 µl of ice-cold PBS per 10 mg of faeces and vortexed for 5 min. Large debris was removed by spinning at 400g for 5 min at 4 °C. Supernatant containing bacteria was removed and pelleted by centrifugation at 8,000g before staining with SYTO BC bacterial DNA stain (Thermo Fisher; 1:5,000). For staining of cultured bacteria, overnight cultures were pelleted at 8,000g and resuspended in SYTO BC at approximately 10^7 colony-forming units.
(CFU) per millilitre (a density similar to that of faecal bacteria). Stained bacteria were incubated with monoclonal antibodies at 10 μg ml⁻¹ for faeces, or 1 μg ml⁻¹ for cultures in PBS supplemented with 0.25% BSA, for 1 h on ice before washing and incubating with Alexa Fluor 647 conjugated goat anti-human IgG, secondary antibody (Thermo Fisher, catalogue number A-21445; 1:2,000) with 5% v/v normal goat serum in PBS plus 0.25% BSA for 30 min. Bacteria were washed in PBS plus 0.25% BSA for 15 min, and 0.25 μg ml⁻¹ 4′,6-diamidino-2-phenylindole (DAPI) was added immediately before sample acquisition. Samples were run on a FACs Symphony (BD Biosciences), with forward scatter and side scatter set to logarithmic mode. Within each experiment, the same number of events was acquired for all samples. Binding was evaluated in SYTO BC² DAPI live bacteria. Data were analysed using FlowJo software version 8.7 or 10.3 (Tree Star Inc.).

Faecal bacterial ELISAs were performed with bacteria isolated from freshly collected faeces of either Oligo-MM²³ or SPF mice. To regulate bacterial diversity sampled in SPF mice, faeces from multiple cages of C57BL/6 mice were pooled and prepared as a single sample. Faeces were macerated with disposable micropestles and centrifuged as described in the paragraph above for bacterial flow cytometry. Bacteria were fixed in 0.5% paraformaldehyde for 20 min with continual rotation to avoid clumping, then washed three times in PBS. Poly-L-lysine-coated high-binding ELISA plates were incubated overnight with bacterial preparations at an optical density at 600 nm (OD₆₀₀) of 0.2 for Oligo-MM²³ or an OD₆₀₀ of 0.35 for SPF. Plates were blocked with PBS plus 1% BSA for 2 h, washed with PBS plus 0.05% Tween, and incubated for 1 h with serial dilutions of monoclonal antibodies, starting at a concentration of 900 nM. Plates were washed and incubated with a horseradish peroxidase (HRP)-conjugated goat anti-human IgG (Fcγ-specific) secondary antibody (Jackson Immuno Research, catalogue number 109-035-098) at 1 μg ml⁻¹ for 1 h. Assays were developed using a chromogenic substrate (Sigma) according to the manufacturer’s instructions. Absorbance was read at 450 nm using an AccuScan plate reader (Fisher Scientific).

Polyreactivity ELISAs were carried out as previously described²⁴, with the following modifications. Briefly, high-binding ELISA plates (Costar) were coated with single-stranded DNA, double-stranded DNA, lipopolysaccharide (LPS) and keyhole limpet haemocyanin (KLH) at 10 μg ml⁻¹ and insulin at 5 μg ml⁻¹ in PBS overnight at 4°C. Plates were washed with PBS plus 0.05% Tween (PBST) and incubated in PBST for 1 h at room temperature. Plates were incubated with serial dilutions of monoclonal antibodies in PBST for 2 h at room temperature, starting at a concentration of 10 μg ml⁻¹. Plates were incubated with HRP-conjugated secondary antibodies, developed and absorbances read as above. Self-reactivity was tested using a QUANTA Lite ANA ELISA kit (Inova Diagnostics) according to the manufacturer’s instructions.

For testing of reactivity to food proteins, crude protein extracts were prepared by grinding 10 g of germ-free autoclaved chow to a powder in a pestle and mortar, then shaking at 120 r.p.m. overnight in 40 ml PBS at 37°C. Extracts were clarified by spinning at 4,000 g for 10 min and then 21,130 g for 1 h at 4°C before filtering through a 0.22-μm filter. Final protein concentrations were determined using a bicinchoninic acid (BCA) assay (Pierce). ELISA plates were coated overnight with 100 μg ml⁻¹ food extract, blocked with PBS plus 2% BSA, and incubated with serial dilutions of monoclonal antibody in PBST plus 2% BSA starting at 10 μg ml⁻¹ for 2 h. Assays were incubated with HRP-conjugated secondary antibodies, developed, and absorbances read as above.

For dot blots of bacterial lysates, overnight cultures of bacteria were centrifuged at 8,000g for 8 min to pellet bacteria and washed once with brain–heart infusion (BHI) medium, then resuspended at a 1× concentration in BHI. These concentrations were normalized to an OD₆₀₀ of 4.0, then mixed 1:1 with Laemmli sample buffer and boiled for 10 min to lyse cells. Next, 2 μl of prepared bacterial lysates were spotted on nitrocellulose and dried at room temperature overnight. Blots were blocked in PBST (1× PBS and 0.1% Tween) with 1% BSA for 2 h, washed briefly with PBST, then stained with 4 μg ml⁻¹ primary antibody (monoclonal human IgG, s) in PBST with 1% BSA for 2 h. Blots were washed for 5 min in PBST three times, stained with 1:1,000 peroxidase-conjugated goat anti-human IgG (Jackson ImmunoResearch, catalogue number 109-035-098) in PBST with 1% BSA for 1 h, then washed for 5 min in PBST three times. Blots were dried on filter paper before incubation with the Clarity enhanced chemiluminescence (ECL) substrate (Bio-Rad); chemiluminescence images were acquired with an ImageQuant LAS4000 (GE) using the same exposure time for all blots.

Self-reactivity was assessed in western blots that probed whole-tissue lysates. The small intestine was dissected from wild-type SPF mice and a 1-cm segment of the distal ileum was removed, cut open longitudinally, and washed thoroughly with PBS to remove luminal content. The tissue was snap-frozen and then bead-beat directly in Læmmli buffer to pulverize the tissue, lyse the cells, and reduce/denature proteins in one step. Samples were boiled for 10 min and then centrifuged at 16,000g for 5 min. Supernatants were run on 4–20% Tris-glycine gels with SDS, and transferred onto polyvinylidifluoride (PVDF) membranes. Blots were blocked and stained using the procedure described above for dot blots.

Bacterial culture

A list of bacterial strains is provided in Supplementary Table 4. Bacteria were grown in an anaerobic atmosphere of 10% carbon dioxide, 5% hydrogen, and 85% nitrogen. Members of the Oligo-MM²³ consortium were cultured in BHI (Becton Dickinson) supplemented with 5 μg ml⁻¹ hemin (Sigma), 5 μg ml⁻¹ vitamin K1 (Sigma), 250 μg ml⁻¹ cysteine, 250 μg ml⁻¹ sodium sulfide, and 4 μg ml⁻¹ porcine mucin type 3 (Sigma) (the latter only for Akkermansia muciniphila YL44), with the exception of Lactobacillus reuteri 20791. These concentrates were normalized to an OD₆₀₀ of 4.0, 0.25% BSA for 15 min, and 0.25 μg ml⁻¹ 4′,6-diamidino-2-phenylindole (DAPI) was added immediately before sample acquisition. Samples were run on a FACs Symphony (BD Biosciences), with forward scatter and side scatter set to logarithmic mode. Within each experiment, the same number of events was acquired for all samples. Binding was evaluated in SYTO BC² DAPI live bacteria. Data were analysed using FlowJo software version 8.7 or 10.3 (Tree Star Inc.).

Statistical analysis

No statistical methods were used to predetermine sample size. Unless otherwise noted, statistical calculations were performed using the tests described in the figure legends in GraphPad Prism version 8.3.0. The Chao1 formula was used to estimate total clonal richness in photo-activated germinal centres, calculated using the EstimateS package. Proportions of clones bearing public-clonotype motifs in our sample versus the Greiff et al. database were compared using Fisher’s exact test calculated in R. For the ARG5NYXXXXDY CDRH3, Levenshelt distances were compared between conditions; for the AREFGAY CDRH3, distances were compared between V₅₁-12 and all clones using a Mann–Whitney test. Exact binomial confidence intervals were calculated using the JavaStat online tool at https://stattages.info/confint.html.

Reporting summary

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

Data availability

Incidence-based sequencing raw and processed data are available through BioProject (https://www.ncbi.nlm.nih.gov/bioproject; identification code PRJNA647715); the analysis pipeline is available at https://github.com/victorialab/MBIS.
29. Lefranc, M. P. et al. IMGT, the international ImMunoGeneTics information system. Nucleic Acids Res. 37, D1006–D1012 (2009).
30. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 13, 31 (2012).
31. Retter, I., Althaus, H. H., Münch, R. & Müller, W. VBASE2, an integrative V gene database. Nucleic Acids Res. 33, D671–D674 (2005).
32. DeWitt, W. S., III, Mesin, L., Victora, G. D., Minin, V. N. & Matsen, F. A., IV. Using genotype abundance to improve phylogenetic inference. Mol. Biol. Evol. 35, 1253–1265 (2018).
33. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217 (2000).
34. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190 (2004).
35. Gupta, N. T. et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics 31, 3356–3358 (2015).
36. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1635–1645 (2009).
37. Pasqual, G., Angelini, A. & Victora, G. D. Triggering positive selection of germinal center B cells by antigen targeting to DEC-205. Methods Mol. Biol. 1291, 125–134 (2015).
38. Palm, N. W. et al. Immunoglobulin A coating identifies colitogenic bacteria in cultured bacteria and carried out and dot and western blots. A.M.B. optimized and performed flow cytometry of faecal bacteria. A.A.K.L. established the protein-free-diet protocol. L.M. and T.B.R.C. designed and performed all bioinformatics analyses. C.R.N., D.M. and G.D.V. conceptualized the study, designed all experiments, and wrote the manuscript with input from all authors.

Acknowledgements We thank all members of the Victora and Mucida laboratories, past and present, for assistance with experiments, fruitful discussions and critical reading of the manuscript. In particular we thank: A. Rogoz and G. Fayzikhodjaeva for maintaining gnotobiotic mice; S. Gonzalez for maintaining SPF mice; K. Gordon and K. Chhosphel for FACS; the Rockefeller University Genomics Center for RNA sequencing; and employees of the Rockefeller University for continuous assistance. We thank A. Vale (Universidade Federal do Rio de Janeiro, Brazil) for help with analysis of public clonotypes, K. McCoy (University of Calgary, Canada), S. Y. Wong and K. Cadwell (New York University, USA) for providing Oligo-MH+ strains. J. Faith (Mount Sinai School of Medicine, USA) for other bacterial strains, and J. Dibritz and E. Wirthgen (University of Rostock, Germany) for contributing to the training and supervision of C.W. This work was supported by National Institutes of Health (NIH)/National Institute of Allergy and Infectious Diseases (NIAID) grants R01AI119006 and R01AI139117 (to G.D.V.), and NIH/NIAID/National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grants R01DK093674, R01DK113375 and R21AI144827 (to D.M.), with additional support from NIH grant DPA144248 (Pioneer Award) and from the Robertson Foundation to G.D.V. and NIH grant R01DK16646 (Transformative Award) to D.M. C.R.N. is a Human Frontier of Science Program postdoctoral fellow. G.F.D. is a Robert Black Fellow of the Damon Runyon Cancer Research Foundation. A.S. is a Boehringer-Ingelheim Fonds PhD fellow. G.D.V. and D.M. are Burroughs-Wellcome Investigators in the Pathogenesis of Infectious Disease. G.D.V. is a Searle Scholar, a Pew-Stewart Scholar, and a MacArthur Fellow.

Author contributions C.R.N. and L.M. performed all mouse and antibody-sequencing experiments, with help from T.A. C.W. and C.R.N. produced and assayed the reactivity of monoclonal antibodies by ELISA, with help from A.S. G.P.D. stained monoclonal antibodies in cultured bacteria and carried out and dot and western blots. A.M.B. optimized and performed flow cytometry of faecal bacteria. A.A.K.L. established the protein-free-diet protocol. L.M. and T.B.R.C. designed and performed all bioinformatics analyses. C.R.N., D.M. and G.D.V. conceptualized the study, designed all experiments, and wrote the manuscript with input from all authors.

Competing interests The authors declare no competing financial interests.

Additional information Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2865-9.

Correspondence and requests for materials should be addressed to D.M. or G.D.V.

Peer review information Nature thanks Ramy Arnaout, Rachael Bashford-Rogers and Hai Qi for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Clonal replacement in steady-state gaGCs. a, Gating strategy for PA-GFP mice used in Fig. 1a–d. GC, germinal centre. b, Gating strategy and efficiency of labelling in germinal centres of AID-Confetti mice seven days after the administration of tamoxifen, as in Fig. 1e. The labelling efficiency is calculated as 100% minus the product of the percentage of unlabelled cells in the GFP/YFP, RFP and CFP channels. c, Gating strategy for the S1pr2CreERT2/tdTomato fate-mapping experiments shown in g and Fig. 1h. All flow plots are representative of multiple experiments. d, Multiphoton images of Peyer’s patches from SPF mice at different times after tamoxifen treatment (see Fig. 1f). Values in parentheses in images are NDS values. e, Quantification of multiple images as exemplified in d (see also Fig. 1g). Data are from three to five mice per time point at days 14–35, and one to two mice per time point for day 7 and later times. f, Size of germinal centres in Peyer’s patches (PPs) versus mLNs, calculated from samples obtained seven days after tamoxifen treatment as in d and Fig. 1f, plotted as the cross-sectional area of the largest available z-section. Each symbol represents one germinal centre. Lines show medians; P values are from two-tailed Mann–Whitney U tests. Data are pooled from multiple mLNs and Peyer’s patches of two mice from two independent experiments. g, Turnover of B cell clones in germinal centres of Peyer’s patches from S1pr2CreERT2 × Rosa26Stop-tdTomato mice (see Fig. 1h).
Extended Data Fig. 2 | Binding characteristics of ’winner’ gaGC clones from SPF mice. a, Gating strategy for isolating AID-Confetti single germinal centres shown in b, c, Figs. 2a, 3d, e and Extended Data Figs. 6a, b, 8a. CR, CFP and/or RFP; non CR, non-CFP, non-RFP, GY, GFP and/or YFP. b, c, Additional Igh sequence relationships among B cells from high-NDS germinal centres (b) and one low-NDS germinal centre (c) (see Fig. 2a). Scale bars, 50 μm. In c, each tree is for a separate clone (defined as a unique V(D)J rearrangement). Only clones with more than five cells are shown (grey slices in pie charts). d, Gating strategy for bacterial flow cytometry, performed in e, Figs. 2b, 3f, h, i and Extended Data Fig. 6d. e, Flow-cytometry analysis of the binding of recombinant monoclonal antibodies to faecal bacteria isolated from SPF mice. Plots gated as in d. All plots are representative of data obtained from at least two separate experiments. f, Summary of the reactivity of SPF monoclonal antibodies, assayed by ELISA against food protein extracts, autoantigens (anti-nuclear antibody, ANA), and a five-antigen polyreactivity panel comprising single-stranded DNA, double-stranded DNA, keyhole limpet haemocyanin (KLH), insulin and LPS. Shown are background-subtracted OD₄₅₀ values. Data representative of assays repeated in at least three separate experiments.
Extended Data Fig. 3 | Stable vertical transmission of the Oligo-MM\textsuperscript{12} consortium. a–c, qPCR of total (a) and strain-specific (b, c) 16S DNA from faecal samples of mice stably colonized with the Oligo-MM\textsuperscript{12} consortium. In a, \(\Delta C_t\) values were calculated in respect to a reference SPF sample, marked by the black filled symbols, with which all other values were compared. In c, \(C_t\) values were used to quantify the relative abundance of each species (see Methods). LOD, limit of detection. F\textsubscript{1} refers to the first generation after the parental strain (P, colonized by gavage). Note that *Bifidobacterium animalis* (YL2) is usually undetectable in faeces\textsuperscript{19}.
Extended Data Fig. 4 | Frequency and isotype distribution of gáGCs in germ-free and Oligo-MM12-colonized mice. 

**a**, Gating strategy for analysing the frequency of germinal centres and distribution of isotypes (results shown in **b**–**d**). 

**b**, Frequency of cells with the phenotype of germinal centres (CD38- FAS hi) among total B220+ B cells in the indicated organs of mice raised under the indicated conditions. Each symbol represents one mouse. SPF, n = 25; germ-free (GF), n = 16; Oligo-MM12, n = 11. 

**c**, Frequency of germinal-centre B cells positive for the indicated surface BCR isotype in different organs of mice raised under the indicated conditions. Data are from at least three mice per group, as in **d**. Data are presented as means ± s.e.m. 

**d**, Statistical analysis of selected isotypes and anatomical locations, using data from **c**. Each symbol represents one mouse. Lines indicate medians; P values are obtained from two-tailed Kruskall–Wallis tests carried out on each trio, with Dunn’s multiple comparisons post-test. All P values below 0.05 are reported.
Extended Data Fig. 5 | Clonal selection in germ-free and Oligo-MM12-colonized mice. a. Gating strategy for germ-free AID-Confetti single germinal centres used in b–d. b–d. Sequencing of IgH genes from B cells obtained from individual mLNs of germinal centres. Germinal-centre B cells were single-cell-sorted from fragments of vibratome slices containing single germinal centres. To avoid biased selection of germinal centres based on NDS or loss of germinal centres with a low density of coloured cells, mLNs were harvested at five to seven days after treatment with tamoxifen, before extensive selection or clonal turnover; both fluorescent and non-fluorescent cells were included in the sample. This unbiased selection ensures that data are comparable to those obtained using in situ photoactivation (Fig. 1a–d), which we could not perform because the photoactivatable GFP-transgenic strain is not available under germ-free status. b. Clonal composition of individual germinal centres from five mice (GF1–GF5). c. caecal-colonic mLNs; j, jejunal mLNs. c. Quantification of data from b. Each symbol represents one germinal centre. d. Proportion of germinal centres in which the largest clone accounts for more than 50% of all B cells in mLNs of SPF mice (data from Fig. 1b) and germ-free mice (data from b). P values are from two-tailed Fisher’s exact tests. Centre bars represent the proportion in the sample; error bars show the exact binomial 95% confidence interval. e. Multiphoton images of Oligo-MM12 mLNs and Peyer’s patches at different times after treatment with tamoxifen. Blue represents collagen (second harmonic); white shows autofluorescence; other colours are from the Confetti allele. Scale bars, 200 μm (overviews), 50 μm (close-ups). N/D, NDS not determined owing to a low density of coloured cells. f. Quantification of images as in e for mLNs (top) and Peyer’s patches (bottom). Each symbol represents one germinal centre. Medians are indicated. Only germinal centres with a density of more than 0.4 fluorescent cells per 100 μm² are included in the NDS calculations. g. Proportion of germinal centres with NDS values of more than 0.75 in mLNs (top) and more than 0.5 in Peyer’s patches (bottom) under SPF, germ-free and Oligo-MM12 conditions at 20–23 days after tamoxifen; SPF and germ-free data are as in Fig. 3c. For SPF, Oligo-MM12 and germ-free mLNs, n = 57, 16 and 27, respectively; for Peyer’s patches, n = 20, 10 and 9, respectively. P values obtained by two-tailed Fisher’s exact tests. Error bars represent exact binomial 95% confidence intervals. All data are from three to five mice per time point.
Extended Data Fig. 6 | Characteristics of ’winner’ gaGC clones from germ-free and Oligo-MM12-colonized mice. a, b, Additional Igh sequence relationships among B cells from high-NDS germinal centres of germ-free (a) and Oligo-MM12-colonized (b) mice. Details are as in Fig. 2a. Scale bars, 50 μm. c, Reactivity summary of germ-free monoclonal antibodies assayed by ELISA against food protein extracts, autoantigens (anti-nuclear antibody, ANA), and a five-antigen polyreactivity panel. Shown are background subtracted OD450 values. d, Flow-cytometry analysis of the binding of monoclonal antibodies from germ-free mice to faecal bacteria from SPF mice. Details are as in Fig. 2b. e, ELISA analysis of the binding of monoclonal antibodies from germ-free mice to faecal bacterial fractions from SPF mice. MG053 was assayed at three dilutions only. Other monoclonal antibodies were assayed at dilutions indicated on the x-axis. Lines show the means of two assays. f, Western blot (WB) analysis of the binding of monoclonal antibodies from germ-free mice to a protein extract from mouse ileum tissue, run on a single-well 4–15% gel and blotted using a multiwell mask. Monoclonal antibody 3H9 is a DNA-specific negative control. Data in c–f are representative of two or more independent experiments.
Extended Data Fig. 7 | Mutational patterns in germ-free/Oligo-MM12 public clonotypes. a. Dendrograms showing the sequence relationships between VH1-47 and VH1-12 clones in different mice. All clones with up to two-amino-acid differences from the public-clonotype CDRH3 motifs are included. b. Heat maps showing the frequency of amino-acid replacements along the VH1-47 and VH1-12 families in germ-free (blue) and Oligo-MM12 (green) mice, using the same data as in Fig. 4b. Only mice with more than two cells within the specified clone were included in the analysis. The number of cells analysed per mouse is indicated at the top of each column. Only those amino acids mutated in at least three (VH1-47) or two (VH1-12) mice are listed on the left, using Immunogenetics (IMGT; http://www.imgt.org) numbering; to the right, the most frequent amino-acid replacement in each mouse is given. Arrows indicate recurrent amino-acid mutations found in five of six mice (VH1-47) or three of three mice (VH1-12).
Extended Data Fig. 8 | Stereotypical germ-free IgH clonotypes are present in Oligo-MM<sup>12</sup> and germ-free/dietary-protein-free conditions. a, Massive expansion of a public VH<sub>1</sub>–12 clonotype across different secondary lymphoid organs of mouse MM<sup>12</sup> (from Fig. 4b), at 21 days after tamoxifen treatment. Multiphoton images show all three germinal centres sequenced from this mouse (yellow dotted boxes), magnified in the side panels. Scale bars, 200 μm (overviews) and 50 μm (close-ups). mLN close-ups are from different image acquisitions of the same germinal centre. A clonal tree of all cells from this clone is shown at the bottom right. Arrowheads indicate clonal bursts and the organ of origin of cells with that particular sequence. b, Frequency of cells with a germinal-centre phenotype (CD38<sup>dim</sup> FAS<sup>hi</sup>) among total B220<sup>+</sup> B cells in the indicated organs of mice raised on protein-free chow (PFC). Data for SPF and germ-free mice are reproduced from Extended Data Fig. 4b. Each symbol represents one mouse. For PFC, n = 8 mice. c, Clonal distribution of germinal-centre B cells sequenced from the indicated tissues of three separate mice (PFC1–3), with public clonotypes colour-coded. See also Fig. 4b, c, caecal colonic mLN; D, duodenal mLN; I, ileal mLN; PP, Peyer’s patch.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | Multiwell incidence-based \textit{ig}h sequencing reveals clonal overlap among individual mice and between microbial colonization conditions. a, Overview of the incidence-based \textit{ig}h sequencing method used for c–g and Fig. 4c, d. To identify expanded public clonotypes among gaGC samples from multiple mice with high confidence, we developed an incidence-based sequencing strategy based on repeated sampling of the same germinal-centre B cell population. We sorted multiple samples of 100 germinal-centre B cells (usually 32 for mLN and 16 for Peyer’s patches) from 6 germ-free, 6 SPF, and 7 Oligo-MM\textsuperscript{12}-colonized mice, and sequenced all BCRs in each sample, for a total of roughly 80 thousand input B cells, plus 32 wells each of non-germinal-centre B cells from the mLN of 3 germ-free and 3 SPF mice as controls. To avoid counting as ‘public’ sequences that were spuriously present in different mice owing to barcode misassignment or DNA contamination, we included in our analysis only those clones that were represented by more than five reads in any single well and found in at least two wells from the same sample. Key bioinformatics steps are described in the figure; see Methods for a full description of the bioinformatic pipeline. b, Gating strategy used for data in c–g and Fig. 4c, d, described in a. c, Number of distinct clones per well, after collapsing sequences with matching V\textsubscript{H}, J\textsubscript{H}, and CDR\textsubscript{H3} nucleotide sequences. Each symbol represents one well. Boxes represent medians and interquartile ranges. As expected, non-germinal-centre B cell samples had many more total clones per well than did germinal-centre B cells. d, Proportion of expanded clones (present in more than one well per sample) in germinal-centre and non-germinal-centre samples from mLNs and Peyer’s patches of mice held under the specified conditions. e, Histograms showing Levenshtein distances between the indicated consensus CDR\textsubscript{H3} sequence and the CDR\textsubscript{H3} sequence of all clones in the indicated category. For ARG\textsubscript{S}NYXXXDY, distances are plotted for clones carrying the ‘correct’ V\textsubscript{H1}–47 gene or two ‘control’ V\textsubscript{H} regions with similar usage frequency in our sample. P values were obtained by Kruskal–Wallis test comparing all three conditions. Owing to the very low number of total V\textsubscript{H1}–12 clones outside of the germ-free condition, distances to the AREGFAY CDR\textsubscript{H3} are compared between V\textsubscript{H1}–12 clones and all clones. P values obtained by two-tailed Mann–Whitney U test. f, Fraction of clone*wells containing public clonotypes in each condition, pooled from all mice. P values were obtained by Fisher’s exact test. g, Venn diagram showing the number of clones per condition (pooled from all mice) and overlap between conditions. The clone in the centre of the graph (SPF/Oligo-MM\textsuperscript{12}/germ-free overlap) corresponds to the V\textsubscript{H1}–47 public clonotype. In f, g, data are as in Fig. 4d.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  
  *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

BD FACSDiva software v8.0.2 was used for flow cytometry data acquisition. Sequencing data was acquired using the Illumina MiSeq platform. Microscopy was acquired on the Olympus FluoView v4.2 acquisition software.

Data analysis

Data was analysed using FlowJo (TreeStar) v8.7 and v10.5.3, Prism (GraphPad) v8.3.0, Imagei v1.51 and R v3.6.3. Sequencing analysis was carried out using PANDASeq v.2.11, HighVQUEST v. 1.6.9, VBASE2, FASTX Toolkit v0.0.13, Change-O v0.4.6, the T-coffee algorithm (Notredame et. al. 2000) and GCTree v1 (deWitt et. al. 2018). Circular ideogram plots were created using Circos v. 0.69-9. Dendograms were generated using clustalx v2.1 and FigTree v1.4.4. Data was presented using Adobe Illustrator v23.0.4 and 15.1.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Single cell sequences are available as a supplementary spreadsheet. These refer to Figures 2, S2, 3, 4, 5 and S6 as labeled in the spreadsheet. Incidence-based sequencing data is available at https://github.com/victoraLab/MIBS.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical method were used to determine sample size. Each experiment was repeated across multiple animals.

Data exclusions

Data only excluded for technical reasons. In bacterial flow cytometry data was excluded if the background secondary antibody binding was too high and no mAb binding could be observed (experiments appear in figures 2, S2 and 4.)

Replication

All experiments were reproducible and were repeatable as detailed in figure legends. Sequencing in Figure 5c-g was carried out once, but on a large sample size (6 or 7 animals per group in 3 groups).

Randomization

Mice of either sex were used for most studies. For sequencing experiments in Figure 5c-g age and sex matched mice were used. Mice were allocated into groups based on genotype and colonization status (not randomized)

Blinding

Investigators were not blinded to group allocation; all of the measurements reported objectively quantifiable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☑️  | Antibodies            |
| ☑️  | Eukaryotic cell lines |
| ☑️  | Palaeontology         |
| ☑️  | Animals and other organisms |
| ☑️  | Human research participants |
| ☑️  | Clinical data         |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☑️  | ChIP-seq              |
| ☑️  | Flow cytometry        |
| ☑️  | MRI-based neuroimaging |

Antibodies

| Antibodies used | |
|-----------------|------------------|
| B220- BV421, BioLegend, #103240, RA3-6B2, Lot: B288312, 1:200 dilution |
| B220- BV605, BioLegend, #563708, RA3-6B2, Lot: B201934, 1:200 dilution |
| B220- BV711, BioLegend, #103255, RA3-6B2, Lot: B267109, 1:200 dilution |
| CD38- APC, BioLegend, #17-0381-82, 90, Lot: 2068810, 1:200 dilution |
| CD38- PerCP-Cy5.5, BD, #562770, 90, Lot: 9112983, 1:200 dilution |
| FAS- PE-Cy7, BD, #557653, Jo2, Lot: 9039631, 1:400 dilution |
| FAS- BV421, BD, #562633, Jo2, Lot: 9029848, 1:400 dilution |
| CD16/32 (Fc block)- Bio-X-Cell, BE0307, 2.4G2, 1:200 dilution |
| TCR β- APC-e780, invitrogen, #47-5961-82, H57-597, Lot: 2114497, 1:200 dilution |
| CD3- BV785, BioLegend, #100232, 17A2, Lot: B277518, 1:200 dilution |
| CD4- BV785, BioLegend, #100552, RM4-5, Lot: B264992, 1:200 dilution |
| CD8- BV785, BioLegend, #100749, 53-6.7, Lot: B258589, 1:200 dilution |
| NK-1- BV785, BioLegend, #108749, PK136, Lot: B279624, 1:200 dilution |
| IgM- PE-Cy7, eBioscience, #25-5790-81, 114/1, Lot: 2039912, 1:200 dilution |
| IgG2b- AF488, Southern Biotech, #1090-30, Lot: A2513-X65G, 1:500 dilution |
| IgA- Biotin, Southern Biotech, #1040-08, Lot: I5613-P366E, 1:500 dilution |
| IgA- PE, eBioscience, #12-4204-83, mA-6E1, Lot: E01650-1634, 1:200 dilution |
| IgG1- APC, BioLegend, #406610, RMG-1, Lot: B247242, 1:200 dilution |
| Streptavidin- APC-e780, Invitrogen, #47-4317-82, Lot: 2005846, 1:200 dilution |
| goat anti-human IgG1- Alexa Fluor 647, Thermo Fisher, #A-21445, Lot: 1962791, 1:2,000 dilution |
| goat anti-human IgG (Fcy specific)- HRP, Jackson Immuno Research, #109-035-098, Lot: 13741, 1:1,000 dilution |

MAbs produced in-house:
Validation

All fluorescent antibodies validated as described on the manufacturers website. HRP-conjugated antibodies validated in-house by ELISA measuring full length IgG1 antibody concentration of commercially purchased standards. mAbs produced by us in this study were validated by SDS-PAGE, ELISA, spectrophotometry (nanodrop) and bio-layer interferometry (Octet Red 96) to ensure proper expression, folding and concentration.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)
ATCC, HEK 293F cells

Authentication
Cell lines were not authenticated; validation of functionality was established measuring the quantity and quality of the produced antibody.

Mycoplasma contamination
All cell lines tested negative for mycoplasma contamination.

Commonly misidentified lines
None used.

Animals and other organisms

Policy information about studies involving animals. ARRIVE guidelines recommended for reporting animal research

Laboratory animals
Mice both sexes were used in all studies. 8-12 week old mice were used in all studies and strains. Rosa26.Confetti (013731) and Rosa26.Stop.TdTomato (007914) mice were from The Jackson Laboratory. AicdaCreERT2 mice were provided by Claude-Agnès Reynaud and Jean-Claude Weill (Institut Necker). S1pr2CreERT2 BAC transgenic mice provided by T. Okada (RIKEN Yokohama) and T. Kurosaki (U. Osaka). PA-GFP mice were generated by G. Victora and M. Nussenzweig (Rockefeller University).

Wild animals
The study did not involve wild animals.

Field-collected samples
This study did not involve samples collected from the field.

Ethics oversight
All animal procedures were approved by the Institutional Animal Care and Use Committee of the Rockefeller University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.
Flow Cytometry

Plots

Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cells were isolated by maceration with disposable micropestles (Axygen) in 100 μl of PBS supplemented with 0.5% BSA and 1 mM EDTA (PBE), and single cell suspensions obtained by two passes through a 70 μm mesh. Cells were stained with fluorescently labeled antibodies on ice for 30 minutes.

Instrument

Samples were run on a FACS LSRII or FACS Symphony (BD). For cell sorted samples were run on a FACS ARIA (BD).

Software

BD FACSDiva software v8.0.2 was used for flow cytometry data acquisition. Analyzed using FlowJo software package (Tri-Star, USA) v10.5.3 and v8.7.

Cell population abundance

Most cells were single cell index sorted into 96-well PCR plates, with single-cell precision. For bulk sorting, 100 cells per well were sorted with single-cell precision.

Gating strategy

All positive and negative populations were determined by compensation with single color controls. For sorting and analysis, all lymphocytes were first gated based on SSC-A vs FSC-A, followed by 2 singlet gates (FSC-H vs FSC-A and SSC-H vs SSC-A). For GC gating, cells were gated on either TCRbeta or Dump-, B220+, CD38-, Fas+ and interrogated for IgM, IgG1, IgG2b or IgA. For AID-confetti sorting experiments cells were gated on SSC-A vs FSC-A in the same way as GC cells. TCRbeta-B220+CD38-Fas+ cells (GC B cells) were then plotted as follows: CFP vs RFP, GFP vs YFP and all colored GC cells were single-cell index sorted. For PA-GFP sorting experiments cells were gated as above for GC cells, then GFP+ (photoactivated) GC cells were single-cell index sorted. For GF single GC experiments (AID-Confetti), cells were gated as described, but fluorescent and non-fluorescent TCRbeta-B220+CD38-Fas+ cells were index sorted. For non-fluorescent sequencing analysis, cells were gated as above for GC cells and single-cell index sorted. For bacterial flow cytometry, cells were gated on SSC-A vs FSC-A with only far outliers removed. Then, SYTO +DAPI- live bacteria were assayed for mAb binding.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.