Article

Mapping the T cell repertoire to a complex gut bacterial community

https://doi.org/10.1038/s41586-023-06431-8

Received: 25 April 2022

Accepted: 13 July 2023

Published online: 16 August 2023

Check for updates

Kazuki Nagashima[1,2,3], Aishan Zhao[1,2,3], Katayoon Atabakhsh[1,2,3], Minwoo Bae[4], Jamie E. Blum[5,6], Allison Weakley[3,7], Sunit Jain[3,7], Xiandong Meng[3,7], Alice G. Cheng[8], Min Wang[1,2,3], Steven Higginbottom[2,3], Alex Dimas[2,3], Pallavi Murugkar[3], Elizabeth S. Sattely[5,6], James J. Moon[9,10], Emily P. Balskus[4,11] & Michael A. Fischbach[1,2,3,7]✉

Certain bacterial strains from the microbiome induce a potent, antigen-specific T cell response[1–5]. However, the specificity of microbiome-induced T cells has not been explored at the strain level across the gut community. Here, we colonize germ-free mice with complex defined communities (roughly 100 bacterial strains) and profile T cell responses to each strain. The pattern of responses suggests that many T cells in the gut repertoire recognize several bacterial strains from the community. We constructed T cell hybridomas from 92 T cell receptor (TCR) clonotypes; by screening every strain in the community against each hybridoma, we find that nearly all the bacteria-specific TCRs show a one-to-many TCR-to-strain relationship, including 13 abundant TCR clonotypes that each recognize 18 Firmicutes. By screening three pooled bacterial genomic libraries, we discover that these 13 clonotypes share a single target: a conserved substrate-binding protein from an ATP-binding cassette transport system. Peripheral regulatory T cells and T helper 17 cells specific for an epitope from this protein are abundant in community-colonized and specific pathogen-free mice. Our work reveals that T cell recognition of commensals is focused on widely conserved, highly expressed cell-surface antigens, opening the door to new therapeutic strategies in which colonist-specific immune responses are rationally altered or redirected.

Immune modulation by the gut microbiome plays an important role in a variety of diseases and therapeutic indications[5–7]. As a result, strategies for controlling or redirecting colonist-specific immune responses have considerable therapeutic promise. However, progress towards this goal is impeded by an incomplete understanding of the logic underlying immune recognition of this microbial community.

In pioneering efforts to date, microbiome immunology has been characterized at two levels: the community and the strain. Transplanting human faecal communities into germ-free mice has led to important insights into a variety of immune-linked disease phenotypes including inflammatory bowel disease, obesity, autism, malnutrition and the response to cancer immunotherapy[8–14], but identifying the causative strains has been challenging.

At the strain level, two themes have emerged. First, the nature of the strain determines what type of immune cell is induced; for example, the gut colonists segmented filamentous bacterium (SFB)[15] and *Helicobacter*[16,17] induce T helper 17 (T_H17) cells and peripheral regulatory T (pT_{reg}) cells, respectively. Second, in many cases, the immune cells elicited have a T or B cell receptor specific for an epitope from the strain that elicited them[1–4,18,19]. These studies have led to the hypothesis that a few 'keystone' species dominantly influence the repertoire and function of gut T cells[5,20].

Efforts have been made to bridge community- and strain-level analyses. In a series of pioneering studies, strains responsible for immune modulation have been identified from undefined communities[21–24]; in another effort, a large set of isolates were profiled under conditions of mono-colonization[25]. Both approaches identify strains that are capable of modulating immune function, but neither one shows how a strain behaves in the context of a complex community. Notably, the phenotype of T cells induced by *Akkermansia muciniphila* changes depending on the complexity of the community in which it resides[4]; it is unclear whether the same is true for other common species. Moreover, it is unknown what each strain in the community contributes to the 'sum total' phenotype of immune modulation by the microbiome.

A complex community models the microbiome

As a starting point for our work, we set out to determine whether germ-free mice colonized by a complex defined community have a similar profile of T cell subtypes to that of conventionally colonized mice. To address this question, we colonized germ-free C57BL/6 mice with a 97- or 112-member gut bacterial community (hereafter, hCom1d and hCom2d) (Supplementary Table 1), derivatives of gut bacterial communities recently developed as a defined model system for the

[1]Department of Bioengineering, Stanford University, Stanford, CA, USA. [2]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford University, Stanford, CA, USA. [3]ChEM-H Institute, Stanford University, Stanford, CA, USA. [4]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [5]Department of Chemical Engineering, Stanford University, Stanford, CA, USA. [6]Howard Hughes Medical Institute, Stanford, CA, USA. [7]Chan Zuckerberg Biohub, San Francisco, CA, USA. [8]Department of Gastroenterology, Stanford School of Medicine, Stanford, CA, USA. [9]Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Boston, MA, USA. [10]Harvard Medical School, Boston, MA, USA. [11]Howard Hughes Medical Institute, Cambridge, MA, USA. ✉e-mail: fischbach@fischbachgroup.org

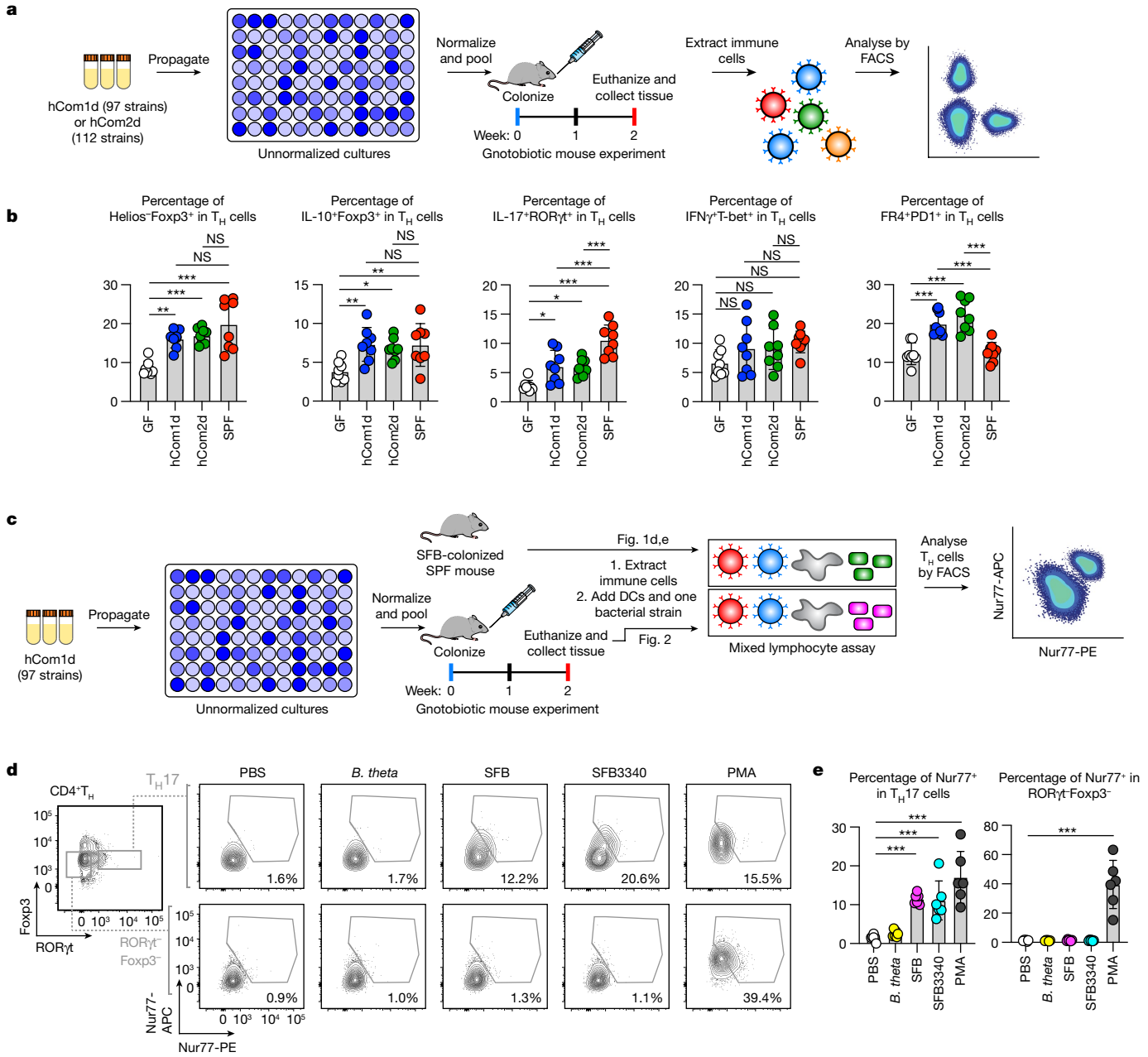


Fig. 1 | A model system for studying immune modulation by the gut microbiome. a, Schematic of the experiment. Frozen stocks of 97 strains (hCom1d) or 112 strains (hCom2d) were used to inoculate cultures that were grown for 48 h, diluted to similar optical densities and pooled. The mixed culture was used to colonize germ-free C57BL/6 mice by oral gavage. Mice were housed for 2 weeks before euthanasia. Immune cells from the large intestine were extracted, stimulated by PMA–ionomycin and analysed by flow cytometry. **b**, T_H cell subtypes, as a percentage of the total T_H cell pool, were broadly similar among hCom1d-colonized, hCom2d-colonized and SPF mice and distinct from germ-free mice. $n = 8$ mice per group (see Supplementary Fig. 1 for the gating strategy). GF, germ free; NS, not significant. **c**, Schematic of the mixed lymphocyte assay. SFB-colonized SPF mice (data in Fig. 1d,e) or hCom1d-colonized mice (data in Fig. 2) were euthanized; immune cells from the intestine were extracted and cocultured with a heat-treated bacterial strain and dendritic cells. After 4 h, cells were fixed, stained with two antibodies specific for Nur77 and analysed by flow cytometry. DCs, dendritic cells; PE, phycoerythrin. **d**, The gating strategy for the mixed lymphocyte assay. Expression of Nur77 by cells from the small intestine was analysed in T_H17 cells and RORγt⁻Foxp3⁻ T_H cells to evaluate TCR stimulation. T_H17 cells were stimulated by SFB (faecal pellets from SFB-mono-colonized mice), purified SFB3340 peptide (an antigen from SFB) and PMA–ionomycin, a positive control. *B. theta*, *Bacteroides thetaiotaomicron*. **e**, Statistical analysis for the mixed lymphocyte assay. $n = 6$ mice per group. Statistical significance was assessed using a one-way analysis of variance in **b** and **e** (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS > 0.05). Data shown are mean ± standard deviations from two independent experiments in **b** and **e**.

human gut microbiome[26]. Strains in hCom1d and hCom2d were chosen on the basis of a rank-ordered list of species prevalence in samples from the National Institutes of Health (NIH) Human Microbiome Project. After two weeks of colonization, we isolated intestinal T cells and profiled them by flow cytometry (Fig. 1a and Supplementary Fig. 1). Although there are certain differences, we find that the levels of T cell subsets in hCom1d- and hCom2d-colonized mice are more similar to those of conventionally colonized (specific pathogen-free, SPF) mice than germ-free mice (Fig. 1b and Extended Data Fig. 1), and within a range of physiological variation that includes large differences in T_H17 cells in SPF mice from different vendors[15]. The pattern of T cell differentiation was consistent when germ-free mice were

cohoused with an hCom2d-colonized cage-mate, suggesting that the immune modulatory effects are not an artefact of gavage (Extended Data Fig. 2). We conclude that hCom1d- and hCom2d-colonized mice are a reasonable setting in which to study immune modulation by the gut microbiome.

Profiling T cell responses to each strain

Next, we sought to understand how each strain modulates T cell immunity in the physiologic setting of a native-scale community. We designed a screen in which T cells isolated from hCom1d-colonized mice could be incubated in vitro with each bacterial strain in the community—one at a time—plus dendritic cells for antigen presentation. As primary T cells can change state while being cultured ex vivo for an extended period of time, a requirement for this experiment is an assay that would give us a rapid, sensitive readout of primary T cell stimulation. To this end, we established a new assay in which we measure the amount of Nur77, an early marker of TCR stimulation[27,28]. We found that monitoring Nur77 expression with two distinct antibodies constituted a sensitive and specific assay for primary T cell stimulation (Fig. 1c and Supplementary Fig. 2a). Although Nur77-enhanced green fluorescent protein (EGFP) mice have been useful for monitoring TCR-stimulated cells in vivo[27,28], germ-free Nur77-EGFP mice were not suitable for restimulation due to a high background level of Nur77 expression (Supplementary Fig. 2b). We calibrated the assay using a well-established commensal-specific T cell response, T_H17 cell induction by SFB[1]. T helper (T_H) cells were extracted from SFB-positive or negative mice and cocultured with antigens and antigen-presenting cells (APCs) (Fig. 1d,e and Supplementary Fig. 2c). We observed the upregulation of Nur77 only in T_H17 cells cocultured with SFB or the SFB antigen SFB3340, whereas phorbol 12-myristate 13-acetate (PMA) stimulated Nur77 expression to a similar extent in T_H17 cells and RORγt$^-$ T_H cells, demonstrating that the coculture assay is sufficiently sensitive and specific to identify interactions between commensal bacterial strains and antigen-specific T cells.

To determine which T cell subtypes are restimulated by strains in the community, we colonized germ-free C57BL/6 mice with hCom1d, waited two weeks to give naïve T cells time to differentiate and then isolated T cells from the large intestine. We incubated these T cells with each of the 97 strains in hCom1d individually, using murine dendritic cells for antigen presentation (Fig. 1c); as a control, we used germ-free mice from which T cells were collected and profiled in the same manner. We measured the proportion of pT_{reg} cells, T_H17 cells and FR4$^+$ T_H cells restimulated by each strain using flow cytometry (Supplementary Fig. 3), because pT_{reg}, T_H17 and FR4$^+$ T_H cells are induced by colonization with hCom1d or hCom2d (Fig. 1b). Although follow-up validation is warranted, we draw the following provisional conclusions: most of the strains in the community did not restimulate T cells from germ-free mice (Fig. 2). By contrast, most of the strains in the community (70 of 97 strains, $P < 0.05$) restimulated at least one type of T cell from hCom1d-colonized mice, and more than a third of the strains restimulated several T cell subtypes (31 of 97) (Fig. 2). More strains activate pT_{reg} cells (51) than T_H17 (36) or FR4$^+$ T_H (21) cells. The restimulation of T cells required MHCII (major histocompatibility complex II) on APCs, and non-community member strains failed to stimulate T cells (Extended Data Fig. 3). *A. muciniphila* is known to induce a context-dependent T cell response: predominantly T follicular helper cells in the presence of a simple defined community, and other effector T_H cells in the presence of a more complex SPF microbiome[4]. We observed that the intestinal T cell pool in hCom1d-colonized mice contains *A. muciniphila*-specific pT_{reg} and T_H17 cells, providing further evidence that hCom1d mimics the function of a complex gut microbiome.

If each strain induced its own 'private' T cell subpopulation, the sum over all the T cells stimulated by each strain would be 100% or less of the gut T cell population. The sum over all the stimulated T cells is far greater than 100% (Fig. 2): for Nur77$^+$ pT_{reg}, T_H17 and FR4$^+$ T_H cells, a total of 863%, 811% and 536%, respectively. We reasoned that this pattern of restimulation arises from a scenario in which T cells are specific to many strains, and the proportion of cells exceeding the 100% threshold is a consequence of many strains restimulating the same T cell clonotypes.

Identifying microbiome-responsive TCRs

To test this hypothesis, we sought to identify T cell clonotypes—individual cells or groups of cells that express the same TCR—that are responsive to the gut microbiome so we could generate a higher-resolution map of strains to clonotypes. We colonized germ-free mice with hCom1d or hCom2d. We pooled three mice per condition, isolated immune cells from the small and large intestine, enriched the CD3$^+$ fraction by fluorescence-activated cell sorting (FACS) and analysed these cells using a combination of single-cell RNA sequencing (scRNA-seq) and single-cell TCR sequencing (scTCR-seq) (Fig. 3a). The dataset of transcriptomes from 35,237 cells (4,908–7,029 cells per condition) were filtered, normalized and analysed by uniform manifold approximation and projection (UMAP). Unbiased clustering defined 25 clusters (Extended Data Fig. 4a). We visualized the expression of canonical subset markers (Extended Data Fig. 4b–d) and matched the 25 clusters to 16 cell types (Extended Data Fig. 5a). Effector T cell subsets showed a continuous distribution, consistent with a recent report[29]. The FR4$^+$ T_H cell type, which corresponds to FR4$^+$PD1$^+$ T_H cells in Fig. 1b, expresses Il21 and Cxcr5 and may affect B cell class switching (Extended Data Fig. 4b–d). The 'other effector T_H' cell type does not express T_H-subset-defining markers but showed high expression of Il21r, Bcl2a1 and Tgfb1, suggesting that this uncharacterized population may play a distinct role in intestinal immunity (Extended Data Fig. 4b–d). We found that colonization by hCom1d or hCom2d changed the composition of immune cell subsets more profoundly in the large intestine than the small intestine; colonic pT_{reg}, T_H17 and FR4$^+$ T_H were induced (Extended Data Fig. 5b,c). These data support our initial findings from the flow cytometry analysis in Fig. 1b.

To explore the pattern of TCR specificities across T cell subsets, we combined the scRNA-seq and scTCR-seq data. The pattern of TCR clonotypes was consistent with a well-established model: after a naïve T cell differentiates into an effector T cell or a pT_{reg}, it clonally expands (Extended Data Fig. 5a). Notably, the frequency of clonotypes in which the same TCR is found on a pT_{reg} and an effector T cell increases when the mice are colonized (Extended Data Fig. 5d).

With the combined scRNA-seq and scTCR-seq data in hand, we sought to determine how microbiome-specific TCRs map to strains from the community. We reasoned that if we could identify TCR clonotypes that were specific for the microbiome, the corresponding TCR genes could be used to construct T cell hybridomas, enabling us to assay TCR specificity against each strain from the community (Fig. 3a). We used two orthogonal criteria to select a group of TCR clonotypes that were likely to be microbiome specific (Supplementary Table 2).

First, we chose 55 clonotypes that were 'expanded' in that they occurred more than twice in our combined pool of 35,237 cells (range 2–84). We reasoned that expanded TCR clonotypes may have derived from T cell clones that divided in response to bacterial stimulation. Moreover, they may be functionally important given their prevalence in the repertoire. Clonotypes that represent a TCR shared by pT_{reg} and effector T cells were of particular interest as they increase after colonization.

Second, we analysed scRNA-seq data to identify genes that were differentially expressed in community-colonized versus germ-free mice (Extended Data Fig. 5e). We used these colonization-induced or repressed genes together with T cell subset markers to choose 37 clonotypes that harboured an expression signature consistent

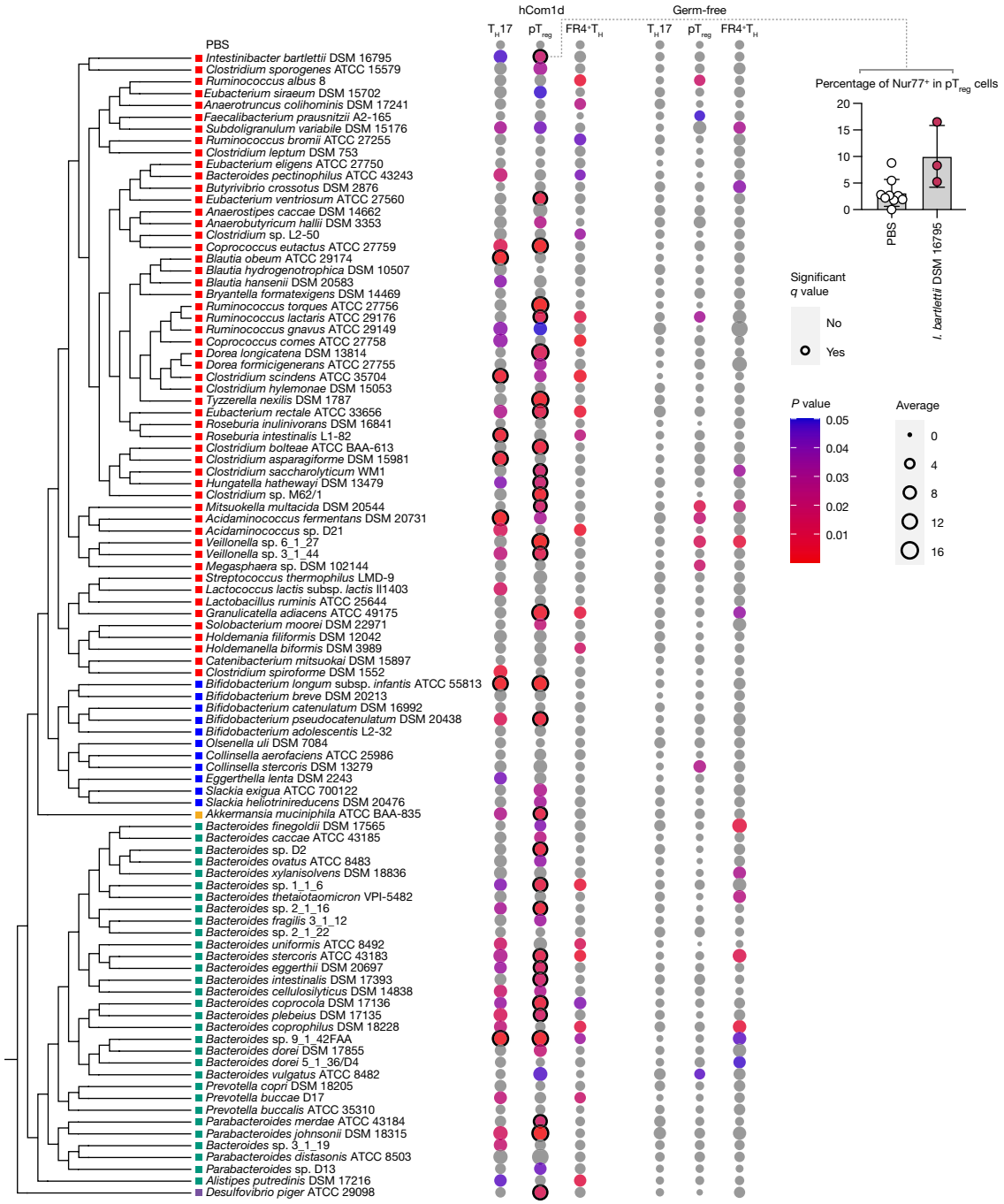


Fig. 2 | Strain-by-strain profiling of T cell responses to a complex defined community. Profiling T cell reactivity against each strain in hCom1d (see Fig. 1c for schematic and Supplementary Fig. 3 for the gating strategy). Immune cells were isolated from the large intestine of 10–15 hCom1d-colonized mice, pooled and cocultured with each of the 97 heat-treated strains in hCom1d, along with dendritic cells for antigen presentation. After 4 h, immune cells were fixed, stained and analysed. As a negative control, immune cells from germ-free mice were cocultured and profiled in the same manner. Dot size shows the average percentage of Nur77$^+$ cells after coculture. Dot colour represents *P* value in comparison with the negative control, treatment with PBS. Data showing the restimulation of pT_{reg} cells by *Intestinibacter bartlettii* DSM 16795 are shown as an example (upper right). A phylogenetic tree of the strains in hCom1d was generated on the basis of a multiple sequence alignment generated from conserved single-copy genes. The coloured square to the left of each strain name indicates its phylum: Firmicutes, red; Actinobacteria, blue; Verrucomicrobia, orange; Bacteroidetes, green; and Proteobacteria, purple. $n = 9$ replicates were used for PBS and $n = 3$ replicates were used for the remaining samples. The data represent three independent experiments per colonization condition. For statistical analysis, multiple comparison testing was performed by GraphPad Prism using the two-stage linear step-up method (Benjamini, Krieger and Yekutieli) for controlling the false discovery rate. *q* value (adjusted *P* value, two-sided) <0.05: significant.

with being microbiome specific (Extended Data Fig. 5f and Supplementary Table 2). For both categories, clonotypes were selected to represent a mixture of tissue origins (small versus large intestine) and T cell subtypes (pT_{reg}, $FR4^+$ T_H, T_H17, T_H1, T_H2 or a combination thereof).

Mapping strains to TCR clonotypes

To determine how strains map to TCR clonotypes, we constructed 92 T cell hybridomas, each one expressing a single TCR. To construct hybridomas en masse (Fig. 3a), we ordered synthetic expression constructs

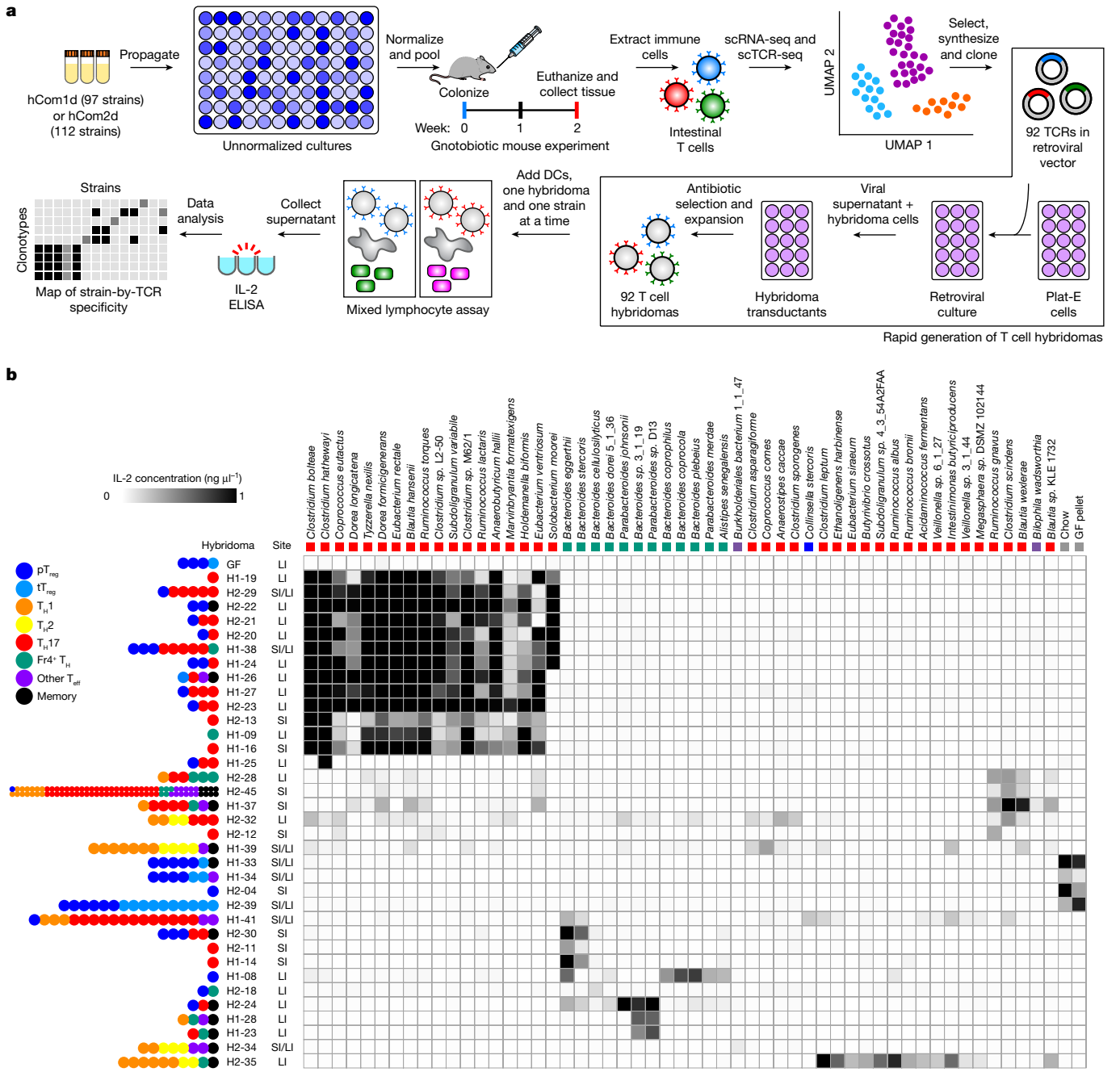


Fig. 3 | scRNA-seq and scTCR-seq to identify microbiome-responsive T cell clonotypes. a, Schematic of the experiment. After 2 weeks of colonization, intestinal T cells were isolated, purified and analysed by scRNA-seq and scTCR-seq. For generating TCR hybridomas, 92 TCR clonotypes were selected on the basis of one of two criteria: (1) 55 were 'expanded' in that they occurred more than twice in our combined pool of 35,237 cells (range 2–84) and (2) 37 harboured an expression signature consistent with being microbiome specific. Synthetic expression constructs that consist of the TCR α and β chains separated by a self-cleaving P2A peptide were cloned into a retroviral vector and used to transfect Plat-E cells. Retrovirus was collected and used to transduce NFAT-GFP hybridoma cells, which were enriched for transductants using a selectable marker. Each T cell hybridoma was cocultured with every strain in hCom1d and hCom2d, one at a time. We measured IL-2 production by the T cell hybridomas to detect TCR stimulation. Data from this experiment were analysed to create a map of strain-TCR specificity. **b**, A map of TCR-strain specificity. Each T cell hybridoma was cocultured with every strain individually. Shown are the 35 TCRs reactive to at least one bacterial strain and the 55 antigens (53 strains, chow and a GF faecal pellet) that stimulated at least one TCR. Thirteen T cell hybridomas were reactive to a subset of 15–18 Firmicutes. GF, a T cell hybridoma derived from germ-free mice, is shown as a negative control. Coloured dots at left represent the primary T cells in which the corresponding TCR was expressed. H1-## indicates hybridomas expressing a TCR selected from hCom1d-colonized mice and H2-## are TCR hybridomas from hCom2d-colonized mice. The coloured square below each strain name indicates its phylum: Firmicutes, red; Bacteroidetes, green; and Proteobacteria, purple. SI, small intestine; LI, large intestine; tT_{reg}, thymic T_{reg}. The data are an average of two independent experiments.

that consist of the TCR α and β chains separated by a self-cleaving P2A peptide. These constructs were cloned into a retroviral vector and propagated in Plat-E cells in a 96-well format; the virus-containing supernatant was used to transduce the NFAT-GFP (green fluorescent protein) hybridoma cell line[30]. We used puromycin to enrich for transductants and used the enriched pools for the experiments described below.

A variety of T cell epitopes from other gut commensals have been discovered and characterized[1,2,4,31–35]. We started by testing 20 such antigens against the 92 hybridomas using dendritic cells for antigen presentation (Extended Data Fig. 6). After overnight coculture, we measured the concentration of IL-2 in the culture supernatant as a reporter of TCR stimulation and signal transduction. None of the hybridomas reacted to any of the antigens, indicating that the hybridomas are not specific for previously discovered epitopes.

Next, we incubated each hybridoma with every strain individually (120 antigens × 92 hybridomas = 11,040 assays) (Fig. 3b). Roughly 30% of the T cell hybridomas (31 out of 92) were responsive to at least one strain, and roughly 45% of the bacterial strains (53 out of 118) stimulated at least one TCR. Four more hybridomas were stimulated by homogenized mouse chow, indicating that they are food responsive; the remainder are not responsive to the microbiome or food. We confirmed this result by an independent repeat of the coculture assays using the microbiota- and food-reactive hybridomas (55 antigens × 36 hybridomas = 1,980 assays). TCR sequences selected from the 'expanded' clonotypes were more responsive to bacterial strains or food (26 out of 55 hybridomas) than clonotypes selected for an apparent microbiome-induced gene signature (nine out of 37 hybridomas).

The hybridoma-strain specificity data reveal clusters of 3–5 TCRs that are specific for 2–4 strains in the community (or for chow). There are apparent similarities among the cell type(s) of origin of the TCRs, and phylogenetic similarities among the strains for which they are specific. These data represent only a miniscule portion of the T cell repertoire, so no broad conclusions about frequency or absolute numbers can be drawn from this experiment. Nonetheless, it is striking that even in a portion of the repertoire this small, many TCRs seem to be specific to several bacterial strains.

In addition to these smaller clusters, we were intrigued by the observation that 13 of the microbiome-responsive hybridomas were stimulated by a subset of 15–18 Firmicutes. These are not the most abundant strains in hCom1d-colonized mice (Extended Data Fig. 7), so this is not a trivial consequence of bacterial cell numbers in vivo. Many of these clonotypes were found simultaneously in pT_{reg} and effector T cells in the population analysed by scRNA-seq, indicating that simultaneous restimulation of pT_{reg} and effector T cells observed in the mixed lymphocyte screen occurs not just at the population level but within individual T cell clonotypes. These data are consistent with the possibility that there exist abundant T cells that target many strains of Firmicutes.

Identifying the antigen for the TCR

To determine the molecular basis for the specificity of these TCR clonotypes, we developed a library-on-library scheme that would enable us to map T cell epitope libraries against a T cell hybridoma library efficiently (Fig. 4a). We generated genomic libraries in *Escherichia coli* from three of the strains that stimulated all the Firmicute-reactive hybridomas: *Clostridium bolteae, Tyzzerella nexilis* and *Subdoligranulum variabile*. We organized each genomic library into 480 pools of 30 clones each and screened the pools against 13 cocultured hybridomas. We found a rare positive clone pool from each of the three libraries; a subsequent deconvolution step yielded individual stimulatory clones from *C. bolteae, T. nexilis* and *S. variabile*. These three clones are related in primary amino acid sequence (Fig. 4b): each of them contains a region from the C-terminal domain of a substrate-binding protein (SBP) that is predicted to function as part of an ATP-binding cassette (ABC) transport system for monosaccharide use (Fig. 4c). SBPs in Gram-positive bacteria are extracellular lipoproteins that are anchored in the outer leaflet of the plasma membrane[36,37]. Two computational tools predicted that the SBP from *T. nexilis* has a lipoprotein signal peptide (Extended Data Fig. 8a,b). The predicted extracellular localization of the SBP may contribute to its strong antigenicity by making the antigen more accessible to immune cells. Of note, among the strains in hCom1d and hCom2d, there is a near-perfect correspondence between the presence of the SBP in the genome and stimulation of the hybridomas in question; the only exception is the actinobacterium *Collinsella stercoris* DSM 13279 (Fig. 4d and Extended Data Fig. 8c). Although they are all Firmicutes (except *C. stercoris* DSM 13279), they do not derive from a monophyletic clade (Fig. 4e), so the trait(s) they share seem to be shaped by gene acquisition or loss.

To narrow down the region within the C terminus that harbours the MHCII epitope, we synthesized 42 peptides from the *T. nexilis* SBP that tile the 53 amino acid region of overlap among the three clones from the screen. We cocultured each candidate peptide with a pool of 13 mixed TCR hybridomas (Extended Data Fig. 8d). Truncated peptides containing the 9-mer YDAFAINMV fully stimulated the mixed hybridomas, whereas peptides that lack any of these residues are inactive or only weakly stimulatory.

This epitope is identical among the Firmicutes that were found to stimulate this subset of hybridomas but the surrounding sequence is not. We tested 12 22–23-mer peptides surrounding this epitope, representing the SBP subsequences from all the stimulatory Firmicutes, against each of the 13 hybridomas (Fig. 4f). This experiment yielded two findings: first, we confirm that the epitope is YDAFAINMV (hereafter, $SBP_{405–413}$). In support of this assignment, *Solobacterium moorei* has an F to Y mutation in this region; as a result, the *S. moorei* peptide only weakly stimulates the corresponding T cell hybridomas.

Second, $SBP_{405–413}$ is the target of only six of the 13 hybridomas. Notably, an *E. coli* clone harbouring the genomic region from *T. nexilis*—which covers most of the amino acid sequence of the SBP—stimulates all 13 hybridomas. Using synthetic peptides that tile the remainder of the SBP, we discovered that the remaining seven TCRs recognize an N-terminal epitope, $SBP_{76–84}$, that is also highly conserved in the Firmicutes strains (Extended Data Fig. 9a). An AlphaFold2-based prediction of the SBP structure suggests that both epitopes lie within the central beta sheet of the membrane-proximal domain (Fig. 4g), where amino acid sequences are more conserved than in variable loop regions, providing one possible explanation for the degree of its conservation.

We sought to determine the prevalence of SBP-specific T cells in the gut repertoire. We isolated T cells from hCom1d- and hCom2d-colonized mice, SPF mice and germ-free mice as a control (Fig. 4h). As expected, SBP restimulates pT_{reg} and T_H17 cells from hCom1d and hCom2d-colonized mice, albeit at a lower proportion in the total repertoire than in the subset of expanded clonotypes profiled in the panel in Fig. 3b. Notably, SBP also stimulates pT_{reg} cells from SPF mice. These mice are colonized by an entirely murine microbiota; they have little overlap in microbial composition with hCom1d or hCom2d (composed of human isolates). The broad conservation of SBP among a subset of Firmicutes makes SBP-specific T cells a substantial portion of the gut repertoire, even in mice who do not harbour the bacterial strains from which SBP was discovered. SBP-specific T cells are also observed when we colonize with a human faecal community or vary the method of colonization, diet, or mouse genetic background, indicating that their presence is robust to experimental variation (Extended Data Fig. 9b,c).

We wondered whether the characteristics of SBP are specific to this antigen or more broadly applicable to antigens from other commensals. From a distinct feature in the heat map in Fig. 3b, we cloned an antigen from *Bacteroides eggerthii* that stimulates TCRs from three small intestine-derived T cell clonotypes: H2-11, H2-30 and H1-14. The antigen was a tetratricopeptide repeat lipoprotein (TPRL) that is highly conserved among *Bacteroides* species and predicted to be localized to the bacterial cell surface (Extended Data Fig. 10). The epitope $TPRL_{29–53}$ restimulates pT_{reg} and T_H17 cells from hCom2d-colonized mice.

We reasoned that SBP and TPRL are selected for immune recognition because the proteins from which they derive are highly expressed. To investigate this hypothesis, we analysed transcriptomic data from individual strains grown in vitro and from faecal communities obtained

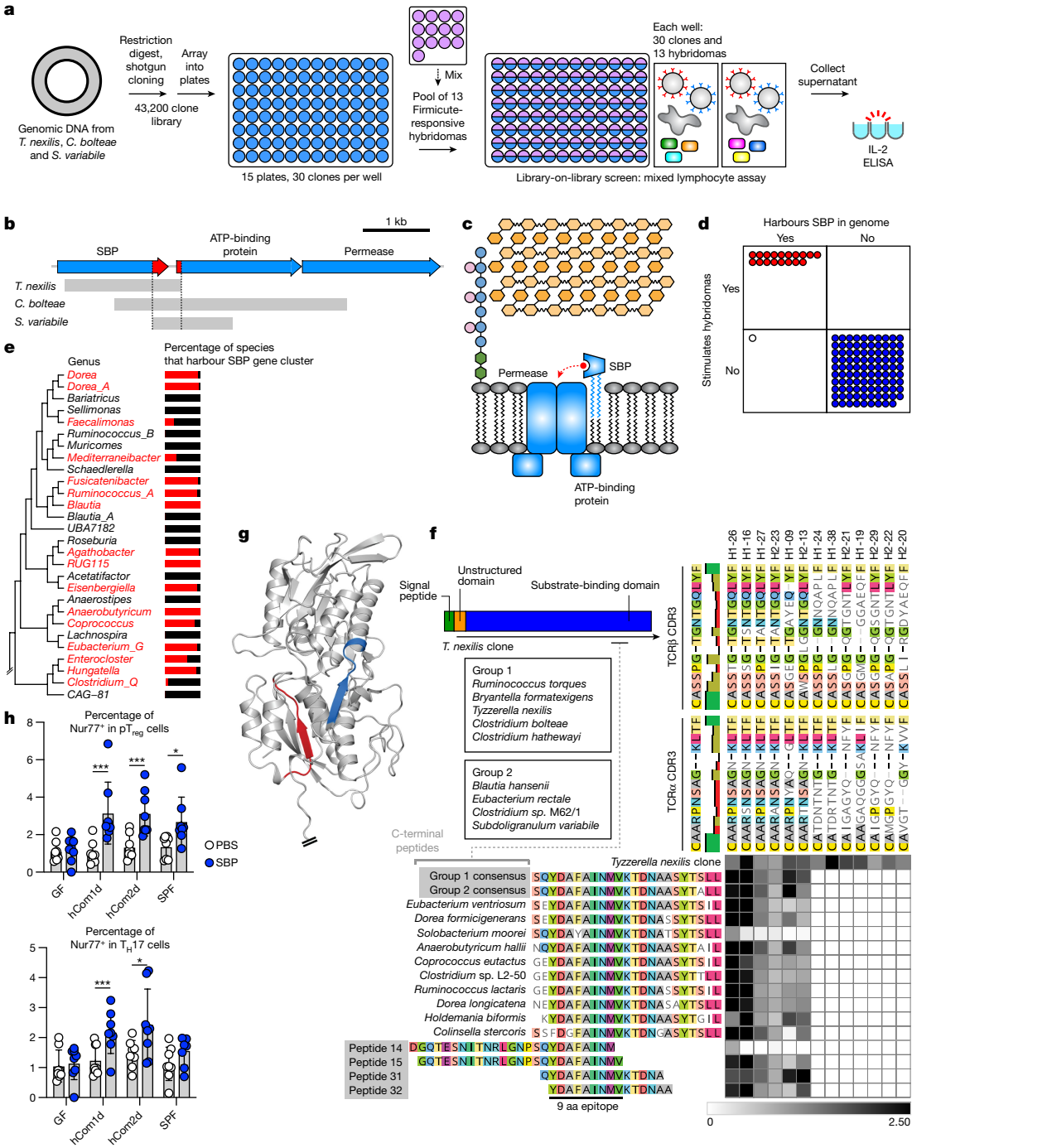


Fig. 4 | Discovery of a conserved Firmicutes antigen. a, Schematic of the antigen identification experiment. Genomic DNA from *T. nexilis*, *C. bolteae* and *S. variabile* was used to create libraries in *E. coli*. Each of three libraries was arrayed in plates as 480 pools of 30 clones. Each library was screened against a mixture of 13 Firmicutes-reactive hybridomas (Fig. 3b). TCR stimulation was monitored by measuring IL-2 production. **b**, One positive clone was identified from each of the genomic libraries. The genomic fragment in each clone overlaps around the C-terminal domain of a SBP, a component of a predicted ABC transport system for monosaccharide use. **c**, Schematic of the ABC transport system and SBP. The SBP is a predicted lipoprotein that is anchored in the the plasma membrane. **d**, BLAST-based genomic analysis of SBP distribution in hCom strains. The presence of an SBP homologue is almost perfectly correlated to the strain's ability to stimulate the 13 hybridomas from Fig. 3b. **e**, A phylogenetic tree showing the distribution of the SBP gene cluster among host-associated Firmicutes. Red, species with the SBP cluster. **f**, Identification of the antigen epitope $SBP_{405–413}$. C-terminal SBP peptides from 19 stimulatory strains from hCom1d and hCom2d were cocultured with the T cell hybridomas. The truncated peptides 14, 15, 31 and 32 were tested to define the minimal epitope. Six of the 13 TCRs were responsive to the synthetic peptides. The seven remaining TCRs are reactive to the N-terminal epitope $SBP_{76–84}$ in Extended Data Fig. 9a. There is a strong correlation between the reactivity of TCRs and the TCR CDR3 sequences. **g**, Predicted structure of the SBP from AlphaFold2. The $SBP_{405–413}$ and $SBP_{76–84}$ epitopes, shown in red and blue, respectively, lie within the central beta sheet of the membrane-proximal domain. **h**, Induction of SBP-specific T cells in vivo. Colonic T cells from hCom1d- or hCom2d-colonized mice were isolated and cocultured with $SBP_{405–413}$ and DCs and monitored for Nur77 expression by FACS. $n = 8$ mice per group from two independent experiments. Statistical analysis was a two-sided *t*-test. Error bars are standard deviations. $^*P < 0.05$; $^{***}P < 0.001$; NS > 0.05. In hCom1d and hCom2d-colonized mice, T_H17 and pT_{reg} cells showed an antigen-specific response to SBP. Notably, SBP-specific T cells were also found in SPF mice, suggesting that the SBP is broadly conserved among Firmicutes that colonize the intestine.

from a cohort of 307 healthy human subjects[38] (Extended Data Fig. 11). SBP was among the ten most highly expressed genes in *C. bolteae* and *Clostridium hathewayi* cultured in vitro, and the SBP and TPRL genes were among the most highly expressed genes from select SBP and TPRL-encoding species in vivo, as assessed by the ratio between RNA and DNA level from paired metagenomic and metatranscriptomic samples.

To confirm the presence of SBP and TPRL proteins in the human microbiome, we performed an experiment in which the T cell hybridomas that respond to $SBP_{405–413}$, $SBP_{76–84}$ and $TPRL_{29–53}$ are restimulated by human faecal communities (Extended Data Fig. 12). We found that six out of six human faecal communities stimulated the $SBP_{405–413}$ and $SBP_{76–84}$-specific hybridomas and five out of six stimulated the $TPRL_{29–53}$-specific hybridomas. These data indicate that the SBP and TPRL are highly expressed by human gut isolates under native conditions of colonization and are a prominent target for immune surveillance by intestinal T cells.

Discussion

The system we introduce here has three features that make it powerful for studying immune modulation by the gut microbiome. First, the gut community is defined, enabling the interrogation of individual strains to quantify their contribution to a community-level phenotype. In this way, we identified a variety of strains capable of restimulating pT_{reg} and T_H17 cells (Fig. 2). Together with another study using a similar approach[39], these data are a promising starting point for identifying new antigen-specific T cell inducers that function robustly in the setting of a complex community.

Second, the use of scRNA-seq and scTCR-seq, in conjunction with T cell hybridoma construction, enabled us to construct a map of strains to TCR clonotypes. Two factors were important: the expansion of TCR clonotypes in colonized mice was a useful predictor of microbiome specificity, and a protocol for the rapid construction of more than 90 hybridomas in parallel—by enrichment rather than cell sorting—enabled us to test TCRs at scale. The combination of scRNA-seq and scTCR-seq has been used in previous work[29,40], but studies have been limited to the comparison of TCR sequences[41] and have not yet identified the antigens they recognize. Our approach provides a method for mapping dozens of TCRs to antigens in parallel.

Among 31 TCRs reactive to bacterial strains, 30 reacted to many strains. This pattern deviates from the current model, in which T cells are specific for an individual colonist[1–4]. Instead, it suggests that in the physiological setting of a complex microbial community, expanded TCR clonotypes mediate the response to a group of colonists that share a conserved epitope, although the generality of this phenomenon is not yet clear. Most TCRs targeted several strains of Firmicutes or Bacteroidetes, as observed recently in different settings[32,42] and a distinction from previous work reporting polyreactive immunoglobulin A (IgA) that recognizes strains of Proteobacteria[43]. Future experiments are needed to find dietary antigens for the four TCRs that were specific to chow and faecal pellets from germ-free mice.

Third, in cloning the epitope recognized by the Firmicutes-specific T cell hybridomas, it was essential that our pool of possible TCR epitopes was completely defined. Each strain in the community is sequenced and arrayed as a pure culture, enabling the construction of genomic libraries from strains of interest. Screening pools of genomic clones against pooled hybridomas made it possible to test 43,200 clones against 13 hybridomas (561,600 combinations) in just 1,440 ELISA assays (arrayed across 15 plates), leading to the identification of three clones—one from each genomic library—which encoded overlapping fragments of orthologous SBPs from an ABC transport system. The SBP and TPRL are expressed on the cell surface, widely conserved and highly expressed, features that are likely to be found in other epitopes recognized by T cells that patrol the gut microbiome[32].

We propose that a T cell clone specific for one of the conserved SBP epitopes, $SBP_{405–413}$ or $SBP_{76–84}$, may expand quickly because it encounters SBP-expressing Firmicutes frequently at barrier surfaces along the intestine. Notably, the Firmicutes that express SBP are not the most abundant strains in the community. We speculate that they might have privileged access to a barrier surface, or they produce a metabolite that facilitates immune modulation.

If the ability of microbiome-directed TCRs to recognize many bacterial species is borne out by subsequent studies, it will have important implications for the logic of immune surveillance against the gut microbiota. A typical gut community consists of very many strains and epitopes; mounting a prophylactic defence against all of them, using a limited set of lymphocyte clonotypes, is an imposing challenge for the host. Our data indicate that T cell clonotypes that recognize broadly conserved antigens undergo positive selection, providing single-clonotype defence against a swath of strains. This model echoes some of the characteristics of B cell response to the microbiome[44,45] as well as broadly neutralizing antibodies against HIV and SARS-CoV-2 (refs. 46,47), which are specific at the level of molecular recognition but provide broad protection against pathogens because their target is widely conserved across variants.

Our work has four important limitations. First, we analysed gnotobiotic mice after two weeks of colonization. To characterize physiological immune development, a different time point (for example, colonization at weaning and sampling 2–3 months later) would be useful to assess in the future.

Second, we only profiled a small portion of the intestinal T cell repertoire by scRNA-seq and constructed hybridomas from an even smaller pool of T cells. Future efforts should focus on generalizing and streamlining this approach so that a larger portion of the repertoire can be profiled in settings in which the host genotype, the microbial colonization state and environmental parameters such as diet are altered. Moreover, it will be interesting to explore whether a similar approach could be used to identify the targets of T cell surveillance in the human gut microbiome.

Third, although $SBP_{405–413}$ and $SBP_{76–84}$ are conserved across 18 Firmicutes species from hCom1d and hCom2d, it is unclear whether all 18 strains contribute to SBP-specific T cell responses in vivo or a subset are dominant. This question could be addressed in future work by constructing dropout communities in which a subset of the 18 strains are missing and assessing the number and subtype of SBP-specific T cells. A similar approach could be used to investigate whether the mixture of pT_{reg} cells and effector T cells observed among SBP-specific T cells arises from individual strain(s) that induce both phenotypes versus many strains that induce a single T cell subtype.

Fourth, the function of SBP or TPRL-specific T cells is not yet known. Given their abundance, they have the potential to play an important role in immune response to commensals. Moreover, they are a mixture of pT_{reg} cells and T_H17 cells, raising the possibility that the balance within the SBP or TPRL-specific pool might be relevant to inflammatory and autoimmune disease[5]. Further work is needed to understand the physiological relevance of these T cells.

An epitope that is the focus of a concerted T cell response against the microbiome creates new therapeutic opportunities. By rationally altering the composition of the community, it may be possible to influence the balance of pT_{reg} cells and effector T cells in the SBP or TPRL-specific pool. Alternatively, key SBP or TPRL-harbouring strains could be engineered to express non-native antigens, redirecting this pool against a target of therapeutic interest[48]. Finally, it might be possible to engineer SBP or TPRL-specific T cells directly using new approaches to target and transduce T cells on the basis of their TCR specificity[49,50]. These approaches could open the door to new therapeutic communities for inflammatory bowel disease, cancer, infectious disease and autoimmune disease.

Online content

Any methods, additional references, Nature Portfolio 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/s41586-023-06431-8.

1. Yang, Y. et al. Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature* **510**, 152–156 (2014).
2. Xu, M. et al. c-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature* **554**, 373–377 (2018).
3. Linehan, J. L. et al. Non-classical immunity controls microbiota impact on skin immunity and tissue repair. *Cell* **172**, 784–796.e18 (2018).
4. Ansaldo, E. et al. *Akkermansia muciniphila* induces intestinal adaptive immune responses during homeostasis. *Science* **364**, 1179–1184 (2019).
5. Ivanov, I. I., Tuganbaev, T., Skelly, A. N. & Honda, K. T cell responses to the microbiota. *Annu. Rev. Immunol.* **40**, 559–587 (2022).
6. Gilbert, J. A. et al. Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).
7. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).
8. Smith, M. I. et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* **339**, 548–554 (2013).
9. Sharon, G. et al. Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice. *Cell* **177**, 1600–1618.e17 (2019).
10. Ridaura, V. K. et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013).
11. Schaubeck, M. et al. Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut* **65**, 225–237 (2016).
12. Gopalakrishnan, V. et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359**, 97–103 (2018).
13. Routy, B. et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359**, 91–97 (2018).
14. Matson, V. et al. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* **359**, 104–108 (2018).
15. Ivanov, I. I. et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
16. Kullberg, M. C. et al. Bacteria-triggered CD4(+) T regulatory cells suppress *Helicobacter hepaticus*-induced colitis. *J. Exp. Med.* **196**, 505–515 (2002).
17. Chai, J. N. et al. *Helicobacter* species are potent drivers of colonic T cell responses in homeostasis and inflammation. *Sci. Immunol.* **2**, eaal5068 (2017).
18. Palm, N. W. et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000–1010 (2014).
19. Fagarasan, S., Kawamoto, S., Kanagawa, O. & Suzuki, K. Adaptive immune regulation in the gut: T cell-dependent and T cell-independent IgA synthesis. *Annu. Rev. Immunol.* **28**, 243–273 (2010).
20. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).
21. Surana, N. K. & Kasper, D. L. Moving beyond microbiome-wide associations to causal microbe identification. *Nature* **552**, 244–247 (2017).
22. Atarashi, K. et al. T_{reg} induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232–236 (2013).
23. Tanoue, T. et al. A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. *Nature* **565**, 600–605 (2019).
24. Atarashi, K. et al. Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2011).
25. Geva-Zatorsky, N. et al. Mining the human gut microbiota for immunomodulatory organisms. *Cell* **168**, 928–943.e11 (2017).
26. Cheng, A. G. et al. Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell* **185**, 3617–3636.e19 (2022).
27. Moran, A. E. et al. T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. *J. Exp. Med.* **208**, 1279–1289 (2011).
28. Ashouri, J. F. & Weiss, A. Endogenous nur77 is a specific indicator of antigen receptor signaling in human T and B cells. *J. Immunol.* **198**, 657–668 (2017).
29. Kiner, E. et al. Gut $CD4^+$ T cell phenotypes are a continuum molded by microbes, not by T_H archetypes. *Nat. Immunol.* **22**, 216–228 (2021).
30. Ise, W. et al. CTLA-4 suppresses the pathogenicity of self antigen-specific T cells by cell-intrinsic and cell-extrinsic mechanisms. *Nat. Immunol.* **11**, 129–135 (2010).
31. Wegorzewska, M. M. et al. Diet modulates colonic T cell responses by regulating the expression of a Bacteroides thetaiotaomicron antigen. *Sci. Immunol.* **4**, (2019).
32. Bousbaine, D. et al. A conserved Bacteroidetes antigen induces anti-inflammatory intestinal T lymphocytes. *Science* **377**, 660–666 (2022).
33. Kuczma, M. P. et al. Commensal epitopes drive differentiation of colonic Tregs. *Sci. Adv.* **6**, eaaz3186 (2020).
34. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T. R. & Elson, C. O. A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **106**, 19256–19261 (2009).
35. Lee, S.-J. et al. Temporal expression of bacterial proteins instructs host CD4 T cell expansion and Th17 development. *PLoS Pathog.* **8**, e1002499 (2012).
36. van der Heide, T. & Poolman, B. ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep.* **3**, 938–943 (2002).
37. Brautigam, C. A., Deka, R. K., Liu, W. Z. & Norgard, M. V. The Tp0684 (MglB-2) lipoprotein of *Treponema pallidum*: a glucose-binding protein with divergent topology. *PLoS ONE* **11**, e0161022 (2016).
38. Mehta, R. S. et al. Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.* **3**, 347–355 (2018).
39. Spindler, M. P. et al. Human gut microbiota stimulate defined innate immune responses that vary from phylum to strain. *Cell Host Microbe* **30**, 1481–1498.e5 (2022).
40. Miragaia, R. J. et al. Single-cell transcriptomics of regulatory T cells reveals trajectories of tissue adaptation. *Immunity* **50**, 493–504.e7 (2019).
41. Muschaweck, M. et al. Cognate recognition of microbial antigens defines constricted $CD4^+$ T cell receptor repertoires in the inflamed colon. *Immunity* **54**, 2565–2577.e6 (2021).
42. Perez-Muñoz, M. E., Joglekar, P., Shen, Y.-J., Chang, K. Y. & Peterson, D. A. Identification and phylogeny of the first T cell epitope identified from a human gut bacteroides species. *PLoS ONE* **10**, e0144382 (2015).
43. Bunker, J. J. et al. Natural polyreactive IgA antibodies coat the intestinal microbiota. *Science* **358**, eaan6619 (2017).
44. 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).
45. Rollenske, T. et al. Parallelism of intestinal secretory IgA shapes functional microbial fitness. *Nature* **598**, 657–661 (2021).
46. Cameroni, E. et al. Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* **602**, 664–670 (2022).
47. Burton, D. R. & Hangartner, L. Broadly neutralizing antibodies to HIV and their role in vaccine design. *Annu. Rev. Immunol.* **34**, 635–659 (2016).
48. Chen, Y. E. et al. Engineered skin bacteria induce antitumor T cell responses against melanoma. *Science* **380**, 203–210 (2023).
49. Dobson, C. S. et al. Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat. Methods* **9**, 449–460 (2022).
50. Yu, B. et al. Engineered cell entry links receptor biology with single-cell genomics. *Cell* **185**, 4904–4920.e22 (2021).

Methods

Mice

Gnotobiotic mouse experiments were performed on C57BL/6 or Swiss Webster germ-free mice originally obtained from Taconic Biosciences maintained in aseptic isolators. SFB$^-$ C57BL/6 mice (Stock no. 000664), CD45.1$^+$ C57BL/6 mice (Stock no. 002014), Nur77-EGFP C57BL/6 mice (Stock no. 016617), MHCII-deficient C57BL/6 mice (Stock no. 003584) and OTII-TCRtg C57BL/6 mice (Stock no. 004194) were purchased from the Jackson Laboratory. All mice were 8–15 weeks old on the day of euthanasia. Both male and female mice were used, and sex-matched littermates were used as a control throughout the study. The animal facility was operated at Stanford University with the following housing conditions: 12/12 light/dark cycle, room temperature between 20 and 26 °C and humidity level 40–65%. All mouse experiments were conducted under a protocol approved by the Stanford University Institutional Animal Care and Use Committee.

Bacterial strains and synthetic community construction

hCom1d and hCom2d were prepared from individual strain stocks as described previously[26]. For germ-free mouse experiments, strains were revived from frozen stocks and cultured for 48 h anaerobically in an atmosphere consisting of 10% CO_2, 5% H_2 and 85% N_2. Strains were propagated in sterile 2.2-ml 96-well deep well plates in their respective growth medium (Supplementary Table 1): Mega Medium[26] supplemented with 400 μM vitamin K2, or Chopped Meat Medium supplemented with Mega Medium carbohydrate mix[26] and 400 μM vitamin K2. The optical density at 600 nm (OD_{600}) of each revived strain was measured. We pooled appropriate volumes of each culture corresponding to 2 ml at OD_{600} = 1.3, centrifuged for 5 min at 5,000*g*, and resuspended the pellet in 2 m of 20% glycerol. For each inoculum preparation cycle, a small number of strains typically did not reach OD_{600} of roughly 1.3. For these strains, the entire 4 ml culture volume was added to the pooled strain mixture. Following pooling and preparation, 1.2 ml of the synthetic community was aliquoted into 2 ml cryovials and stored at −80 °C. To colonize C57BL/6 germ-free mice, a single cryovial was thawed and 200 μl was administered by oral gavage into 8–12-week-old mice. Two weeks after colonization, mice were euthanized for analysis as detailed below.

To prepare heat-killed bacterial strains for in vitro coculture experiments, individual strains were revived from frozen stocks under the anaerobic conditions described above. The growth medium was filtered to avoid contamination by precipitates in the later coculture experiment. After measuring the OD_{600}, each culture was centrifuged for 5 min at 5,000*g* and resuspended into a volume of PBS required to normalize OD_{600} to 1.0. Normalized cultures were heat treated in a water bath at 70 °C for 30 min and then stored at −30 °C.

Metagenomic sequencing and sequence analysis were performed as described previously[26]. In brief, faecal pellets were collected from mice into sterile tubes. Genomic DNA was extracted using the DNeasy PowerSoil HTP kit (Qiagen) and quantified in 384-well format using the Quant-iT PicoGreen double-stranded DNA Assay Kit (ThermoFisher). Sequencing libraries were generated in 384-well format using a custom low-volume protocol based on the Nextera XT process (Illumina). Sequencing reads were generated using a NovaSeq S4 flow cell or a NextSeq High Output kit, in 2 × 150 basepair (bp) configuration. Between 5 and 10 million paired-end reads were targeted for isolates and 20–30 million paired-end reads for communities. NinjaMap[26] was used to calculate relative abundance of each strain.

Isolating immune cells from the intestinal lamina propria

Mice were euthanized by cervical dislocation. The small and large intestine were collected by dissection and the mesenteric lymph nodes, Peyer's patches and caecum were removed. Intestinal tissue was shaken at 225 rpm for 40 min at 37 °C in DMEM (Gibco) containing 5 mM EDTA and 1 mM dithiothreitol. Tissues were washed with DMEM, manually shaken to detach the epithelial layer and cut into pieces. Tissue fragments were then digested using a mouse Lamina Propria Dissociation Kit (Miltenyi Biotec, no. 130-097-410) and a gentleMACS dissociator (Miltenyi Biotec). After digestion, debris was removed using a 40 and 80% Percol gradient purification (GE Healthcare).

Induction of splenic DCs by injecting B16-FLT3L

Here, 8–12-week-old CD45.1$^+$ C57BL/6 mice were purchased from the Jackson Laboratory (Stock no. 002014). Here, 5.0×10^6 B16-FLT3L cells[51] were administered to mice by intraperitoneal injection to expand FLT3L-dependent DCs. Mice were euthanized 12–14 d after injection. Spleens were excised and digested using a spleen dissociation kit (mouse) (Miltenyi Biotec, no. 130-095-926) and a gentleMACS dissociator (Miltenyi Biotec). Red blood cells were lysed and CD11c$^+$ dendritic cells were enriched using CD11c MicroBeads UltraPure (mouse) (Miltenyi Biotec, no. 130-125-835). In the coculture assay, contamination of splenic T cells from B16-FLT3L-injected CD45.1$^+$ C57BL/6 mice was gated out in the FACS analysis using the congenic marker CD45.1.

Coculture of primary intestinal immune cells with splenic DCs and antigens

Here, 5.0×10^5 CD45.1$^+$ FLT3L-induced DCs were cocultured in DMEM + 10% FBS with one of the following stimulants for 30 min at 37 °C in a 5% CO_2 incubator: 5 μl of PBS, 5 μl of of a synthetic antigen peptide (Genscript, 0.2 mg ml^{-1} in PBS), eBioscience Cell Stimulation Cocktail (Fisher Scientific 00-4975-93) or 5 μl of heat-killed bacteria in PBS, prepared as described above. After 30 min, 5.0×10^4 freshly isolated CD45.2$^+$ cells from the intestinal lamina propria were added. To avoid loss of cells, we did not purify T cells from freshly isolated lamina propria (LP) cells. As Nur77 reacts specifically to TCR stimulation, other immune cell subsets in the coculture are unlikely to affect Nur77 expression[27,28]. We cocultured cells from CD45.1$^+$ and CD45.2$^+$ mice for 4 h before staining and fixation. To screen T cell responses to bacterial strains from hCom1d, we pooled T cells from 10–15 mice in each experiment.

Cell fixation, staining and flow cytometry

Single cell suspensions were stained with Fixable Viability Dye eFluor 780 (eBioscience, 65-0865-18) before fixation to detect dead cells. Cells were then fixed for 30 min at room temperature using the fixation/permeabilization buffer supplied with the eBioscience Foxp3/Transcription Factor Staining buffer set (ThermoFisher Scientific, catalogue no. 00-5523-00).

After fixation, cells were stained with combinations of the following primary antibodies in permeabilization buffer for 30 min at room temperature: Helios (22F6, BioLegend, catalogue no. 137214), T-bet (4B10, BioLegend, catalogue no. 644806), Nur77-PE (12.14, eBioscience, catalogue no. 12-5965-82), Foxp3 (FJK-16s, eBioscience, catalogue no. 25-5773-82), Nur77-purified (11C1052, LSBio, catalogue no. LS-C183978-100), CD45.1 (A20, eBioscience, catalogue no. 47-0453-82), CD4 (RM4-5, BioLegend, catalogue no. 100559), CD3e (145-2C11, BioLegend, catalogue no. 100351), FR4 (12A5, BD, catalogue no. 744122), IL17A (eBio17B7, eBioscience, catalogue no. 12-7177-81), RORγt (B2D, eBioscience, catalogue no. 17-6981-82 or Q31-378, BD Bioscience, catalogue no. 562894), IFNγ (XMG1.2, BioLegend, catalogue no. 505830), PD1 (29F.1A12, BioLegend, catalogue no. 135208) and IL10 (JES3-9D7, BioLegend, catalogue no. 505010). Cells stained with the Nur77-purified antibody (11C1052, LSBio, catalogue no. LS-C183978-100) were further stained with a secondary APC Goat antirabbit IgG (Polycolonal, eBioscience, catalogue no. A-10931). All staining was performed at the dilution of 1:200 in the presence of purified antimouse CD16/32 (clone 93, catalogue no. 553141) and 10% fetal bovine serum to block non-specific binding. Cells were analysed using an LSR II flow cytometer (BD Biosciences) and data were processed using FlowJo (v.10.8.1, TreeStar).

scRNA-seq and scTCR-seq analysis

We colonized 8–10-week-old C57BL/6 germ-free mice with hCom1d or hCom2d. We pooled three gender-matched mice per condition, isolated lymphocytes from the lamina propria of the small and large intestine and enriched the live $CD3^+$ T cell fraction by cell sorting (FACS Aria II, BD Biosciences). Sorted cells were resuspended in PBS with 0.05% BSA and roughly 1×10^4 cells were loaded onto the Chromium controller (10X Genomics). Chromium Single Cell 5′ reagents were used for library preparation according to the manufacturer's protocol. The libraries were sequenced on an Illumina HiSeq 4000. Sequencing data were aligned to the reference mouse genome mm10 with Cell Ranger (10X Genomics). The data were processed using the R packages Seurat v.3.1.5 (ref. 52) and scRepertoire v.1.0.0 (ref. 53). R v.3.6.2 (2019-12-12) on the platform x86_64-apple-darwin15.6.0 (64-bit) and running under macOS v.10.16 operating system.

For scRNA-seq analysis, we excluded cells with fewer than 500 and more than 4,500 detected genes. We further eliminated cells with more than 25% mitochondrial genes. Data were clustered using the FindClusters function of Seurat. Visualization of the clusters on a two-dimensional (2D) map was performed with UMAP (RunUMAP function of Seurat, dims = 1:20). To assign cell clusters to T cell subsets, gene expression levels in clusters were visualized by dot plots, violin plots and feature plots using Seurat. Differentially expressed genes of each T cell subset between colonization conditions were identified using the FindMarkers function of Seurat.

scRepertoire was used to merge scTCR-seq data with the Seurat object of scRNA-seq data. The expansion of TCR clonotypes was visualized on UMAP by the DimPlot function of Seurat (group.by = cloneType). To characterize T cell phenotypes of TCR clonotypes, a metadata file was generated from the Seurat object and analysed and quantified using Microsoft Excel (v.16.73).

Generating TCR hybridomas

We generated the retroviral vector pMSCV-mCD4-PIG TCR-OTII to transduce the NFAT-GFP 58α$^-$β$^-$ hybridoma cells with synthetic TCR constructs. pMSCV-PIG (catalogue no. 21654) was purchased from Addgene, subjected to restriction digest by BglII, and purified. To generate pMSCV-PIG TCR-OTII, a gBlock encoding the OTII TCR was ordered from IDT and ligated into BglII-digested pMSCV-PIG using HiFi DNA assembly master mix (E2621S, New England Biolabs). pMSCV-PIG TCR-OTII was digested with BstXI and SalI and then ligated with a gBlock encoding the mammalian CD4 gene to make pMSCV-mCD4-PIG TCR-OTII. Plasmid and gBlock sequences are listed in Supplementary Table 3.

pMSCV-mCD4-PIG TCR-OTII, our backbone vector for TCR expression, was provided to Twist Bioscience. Sequences of the TCR α and β genes we selected (92 TCRs in total) were obtained from the JSON files generated by scTCR-seq. A cassette that consists of the TCR α and β genes separated by a self-cleaving P2A peptide was synthesized (Twist Bioscience) and used to replace the OTII TCR gene in pMSCV-mCD4-PIG.

The Platinum-E (Plat-E) retroviral packaging cell line (RV-101, Cell Biolabs) was used for viral packaging of pMSCV-mCD4-PIG TCR vectors. Plat-E cells were cultured in a 96-well plate to 60–80% confluence and transfected with pMSCV-mCD4-PIG TCR vectors as a DNA–lipid complex using Lipofectamine 3000 (L3000001, ThermoFisher Scientific) following the manufacturer's instructions. Then 24 h after transfection, culture supernatant containing viral particles was harvested.

Here, 2.5×10^4 NFAT-GFP cells were centrifuged, resuspended in 200 µl of the virus-containing supernatant supplemented with 10 µg ml^{-1} protamine sulfate (MP Biomedicals) and plated in U-bottom 96-well plates. NFAT-GFP cells were spin-transduced by centrifugation at 1,000*g* at 32 °C for 120 min, then pipetted up and down and cultured overnight at 37 °C in a 5% CO_2 incubator. Cells were then transferred into a flask and incubated for 1–2 weeks with 10 ml of culture medium (DMEM with 10% FCS, penicillin-streptomycin and 2 mM L-glutamine) under puromycin selection (1 µg ml^{-1}). After selection and propagation, TCR expression by NFAT-GFP cells was confirmed by flow cytometry.

TCR hybridoma coculture experiment

Here, 2.0×10^4 TCR-transduced NFAT-GFP cells were cocultured in a 96-well plate with 1.0×10^5 FLT3L-induced DCs and one of the following stimuli at 37 °C in a 5% CO_2 incubator: 5 µl of PBS, 5 µl of synthetic antigen peptide (Genscript, 0.2 mg ml^{-1} in PBS) or 5 µl of heat-killed bacteria in PBS. After 24 h of coculture, cells were centrifuged and culture supernatant was collected. TCR stimulation was evaluated by measuring IL-2 concentration in culture supernatant using an IL2 ELISA kit (catalogue nos. 423001, 423501, 431001 and 421101, BioLegend).

Antigen discovery using *E. coli* shotgun genomic libraries

E. coli shotgun genomic libraries were generated as described[1,33,54] with minor modifications. Bacterial genomic DNA was purified from liquid cultures of *T. nexilis*, *C. bolteae* and *S. variabile* using a QIAamp PowerFecal DNA Kit (Qiagen). 0.1 mg of genomic DNA per strain was partially digested by Sau3AI. For partial digestion, Sau3AI was serially diluted and aliquots of genomic DNA were added. After 10 min of digestion, Sau3AI was heat inactivated. All the reactions from the serially dilutions of Sau3AI were mixed after digestion and subjected to gel electrophoresis. DNA fragments between 500 and 5,000 bp were gel purified. The pGEX-4T1 expression vector (GE28-9545-49, Sigma-Aldrich) was digested with BamHI and dephosphorylated using the Quick Calf Intestinal Alkaline Phosphatase kit (M0525L, New England Biolabs). CIP was heat inactivated and the linearized pGEX-4T1 vector was gel purified.

Sau3AI-digested bacterial genomic DNA and the linearized, dephosphorylated pGEX-4T1 backbone were ligated using the T4 DNA Ligase kit (M0202M, New England Biolabs). Ligation products were transformed into ElectroMAX DH10B competent cells (18290015, ThermoFisher Scientific) by electroporation. DH10B cells were plated on Luria-Bertani (LB) agar + 100 µg ml^{-1} carbenicillin. Colony-forming units were counted and several colonies were picked for Sanger sequencing to verify the presence of a bacterial genomic fragment. DH10B cells were then adjusted to 30 colony-forming units per 100 µl of LB medium containing 10% glycerol (v/v) and 100 µg ml^{-1} carbenicillin and cultured in 96-well plates overnight. Five plates were prepared per bacterial strain to generate shotgun genomic libraries. Then 7 µl of DH10B culture per well was subcultured in 200 µl of fresh LB-carbenicillin medium in 96-well plates to an OD_{600} of 0.5–0.6; the remainder of the DH10B culture was stored at −20 °C. IPTG (isopropyl-β-D-thiogalactoside) (Sigma-Aldrich) was added to a final concentration of 0.5 mM and DH10B cells were incubate for 16 h at 18 °C to induce the expression of inserted bacterial genes. DH10B cells were centrifuged, washed with PBS, heat-killed by incubating at 75 °C for 1 h and stored at −20 °C until use.

For coculture experiments, 13 Firmicutes-reactive NFAT-GFP hybridomas were mixed: 2.0×10^5 mixed NFAT-GFP cells were cocultured in 150 µl medium (DMEM with 10% FCS, Pen/Strep) in a 96-well plate with 2.0×10^5 FLT3L-induced DCs and 5 µl of heat-killed DH10B cells in PBS. After 24 h of coculture, cells were centrifuged and culture supernatant was collected to measure IL-2 concentration as a readout of TCR stimulation. After identifying a pool of 30 *E. coli* clones that stimulates TCR hybridomas, we cultured the pool on LB agar. Single colonies were picked from the LB agar plate and cultured in LB medium in a 96-well plate. Expression was induced with IPTG and, following 16 h of incubation, cells were heat treated as described above. Each heat-treated *E. coli* clone was cocultured with TCR hybridomas to identify stimulatory *E. coli* clones. One stimulatory clone was identified from *C. bolteae, T. nexilis* and *S. variabile*.

Protein structure prediction by AlphaFold2 with MMseqs2

To predict the three-dimensional structure of the SBP, we used ColabFold: AlphaFold2 protein structure[55] and complex[56] prediction using

multiple sequence alignments generated through MMseqs2 (ref. 57). Predicted structures were visualized by PyMOL v.2.5 (Schrödinger, Inc.).

Phylogenetic distribution of the SBP gene cluster

Metadata were obtained for all genomes in the Unified Human Gastrointestinal Genome (UHGG) v.2 database as of 25 March 2022 was obtained from their FTP site (https://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/human-gut/v2.0/genomes-all_metadata.tsv). This list was filtered such that only the isolate genomes that belonged to the phylum Firmicutes remained—2,651 genomes across 237 genera remained after filtering. The list was filtered further to keep only those genera that contained at least ten isolate genomes. Our final dataset contained 2,174 genomes from 60 Firmicutes genera. These 2,174 genomes and their genomic feature files were downloaded from the FTP site. All genomic features were extracted from the genomes using the 'getfasta' module from the bedtools toolkit. A representative genome from each genus was picked to generate a tree using GTDBtk classify workflow using the release202 version of the database.

The three proteins from the SBP operon were extracted from the genome sequence of *Clostridium nexile* DSM 1787 and a protein blast database was created using these sequences. BlastX was run with all the genes from the 2,174 genomes as query and the proteins from the operon as the database (non-default settings -dbsize 1,000,000 -num_alignments 5 -outfmt '6 std qlen slen qcovs ppos'). Another BlastX search was run with the genes from the Firmicutes in hCom1d and hCom2d using the same database and blast settings. The blastx results were filtered such that only those hits were considered that had a bit score more than or equal to 560 and percentage identity more than or equal to 60%. Finally, of the 386 UHGG genomes that contained a hit that passed our thresholds, 366 genomes across 18 genera contained proteins homologous to the proteins in the SBP operon. A custom R script (github, decorate_tree.Rmd) was used to integrate the UHGG blast results and the genera tree created from GTDBtk. Figure 4e shows a portion of the tree enriched in SBP$^+$ genera, and the whole tree is included in Source Data for Fig. 4.

In vitro transcriptional analysis of SBP homologues

A turbid 48-h starter culture of an SBP-encoding strain, *Enterocloster bolteae* (formerly known as *C. bolteae*) DSM 15670, in BHI medium was inoculated 1:100 into 5 ml of BHI medium, and cultures were grown at 37 °C anaerobically in triplicate. The cultures were harvested when they reached OD_{600} = 0.900. They were centrifuged, and cell pellets were resuspended in 500 µl of Trizol reagent (ThermoFisher, catalogue no. 15596026).

RNA extractions, QC, library preparations and sequencing reactions were conducted at GENEWIZ, LLC/Azenta. (South Plainfield) as follows: total RNA was extracted using Trizol following manufacturer's instructions (ThermoFisher Scientific). RNA samples were quantified using Qubit 2.0 Fluorometer (ThermoFisher Scientific) and RNA integrity was checked with 4200 TapeStation (Agilent Technologies).

Ribosomal RNA depletion sequencing library was prepared by using three probes from QIAGEN FastSelect rRNA 5S/16S/23S Kit (Qiagen), respectively. RNA sequencing library preparation uses NEBNext Ultra II RNA Library Preparation Kit for Illumina by following the manufacturer's recommendations (NEB). Briefly, enriched RNAs were fragmented for 15 min at 94 °C. First strand and second strand complementary DNA were subsequently synthesized. cDNA fragments were end repaired and adenylated at 3′ ends, and universal adapters were ligated to cDNA fragments, followed by index addition and library enrichment with limited cycle PCR. Sequencing libraries were validated using the Agilent Tapestation 4200 (Agilent Technologies), and quantified using Qubit v.2.0 Fluorometer (ThermoFisher Scientific) as well as by quantitative PCR (KAPA Biosystems).

The sequencing libraries were multiplexed and clustered onto one lane of a flowcell. After clustering, the flowcell was loaded onto the Illumina HiSeq instrument (4,000 or equivalent) according to the manufacturer's instructions. The samples were sequenced using a 2 × 150 bp paired-end configuration. Image analysis and base calling were conducted by the Illumina Control Software. Raw sequence data (.bcl files) generated were converted into fastq files and demultiplexed using Illumina bcl2fastq v.2.17 software. One mismatch was allowed for index sequence identification.

Raw RNA-seq data of another SBP-encoding strain, *Hungatella hathewayi* (*Clostridium hathewayi*) DSM 13479, in the mid-exponential growth phase were downloaded from the National Center for Biotechnology Information (PRJNA531520)[58]. The RNA-seq data of the two strains were processed to remove ribosomal RNA, low quality bases and adapter sequences using KneadData v.0.10.0 (--bypass-trf). The processed data were then quantified against the protein coding sequences of each strain using Salmon-quant (--validateMappings)[59]. The resulting transcriptional levels of each gene were expressed in transcripts per million, which was log-transformed using the formula $\log_{10}(\mathrm{TPM} + 1)$. RNA-seq raw data of *E. bolteae* have been deposited in the Sequence Read Archive under BioProject accession number PRJNA904383.

In vivo transcriptional analysis of SBP and TPRL homologues

Raw metagenomics and metatranscriptomics data from a human study were downloaded from the National Center for Biotechnology Information (PRJNA354235)[38]. This study sampled data from a cohort of 307 healthy men in the Health Professional Follow-Up Study. Each dataset was processed to remove human-associated reads, low quality bases and adapter sequences using KneadData v.0.10.0 (default setting). The processed metagenomics and metatranscriptomics reads were mapped onto the genomes of select SBP and TPRL-encoding strains (SBP, *T. nexilis* DSM 1787 and *C. bolteae* DSM 15670; TPRL, *Bacteroides stercoris* ATCC 43183, *Bacteroides eggerthii* DSM 20697 and *Bacteroides cellulosilyticus* DSM 14838) using Bowtie 2 (ref. 60) (default setting). The mapped reads were then used to calculate the DNA and RNA level of each gene normalized to reads per kilobase million using htseq-count[61] (default setting) and a custom script. To account for gene copy number variation in vivo, RNA abundance was divided by DNA abundance derived from the matching metagenomic sample. Bioinformatics analyses were done on the Harvard Faculty of Arts & Sciences Research Computing Cluster.

Statistics

All statistical analyses were performed with Graphpad Prism9 v.9.5.1 (GraphPad Software) and Microsoft Excel v.16.73. $P < 0.05$ was considered statistically significant. The actual P values are listed in Supplementary Table 4.

Inclusion and ethics

All the authors of this study have met the authorship criteria specified by Nature Portfolio journals and have been included in recognition of their indispensable contributions to the study's design and implementation. Collaborators confirmed their respective roles and responsibilities before undertaking the research. This research was conducted without significant limitations or prohibitions in the researchers' setting and does not pose any risks of stigmatization, incrimination, discrimination or personal harm to participants.

Reporting summary

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

Data availability

The data have been deposited as follows: Cell Ranger out files, https://doi.org/10.5281/zenodo.8008419; raw sequence data: scTCR-seq of T cells from the small intestine of germ-free mice, https://doi.org/10.5281/zenodo.8007653; scTCR-seq of T cells from the small intestine of

hCom1 colonized mice, https://doi.org/10.5281/zenodo.8007692; scTCR-seq of T cells from the small intestine of hCom2-colonized mice, https://doi.org/10.5281/zenodo.8007696; scTCR-seq of T cells from the large intestine of germ-free mice, https://doi.org/10.5281/zenodo.8007700; scTCR-seq of T cells from the large intestine of hCom1 colonized mice, https://doi.org/10.5281/zenodo.8007704; scTCR-seq of T cells from the large intestine of hCom2-colonized mice, https://doi.org/10.5281/zenodo.8007709; scRNA-seq of T cells from the small intestine of germ-free mice, https://doi.org/10.5281/zenodo.8007733; scRNA-seq of T cells from the small intestine of hCom1 colonized mice, https://doi.org/10.5281/zenodo.8007735; scRNa-seq of T cells from the small intestine of hCom2-colonized mice, https://doi.org/10.5281/zenodo.8007737; scRNA-seq of T cells from the large intestine of germ-free mice, https://doi.org/10.5281/zenodo.8007740; scRNA-seq of T cells from the large intestine of hCom1 colonized mice, https://doi.org/10.5281/zenodo.8007742; scRNA-seq of T cells from the large intestine of hCom2-colonized mice, https://doi.org/10.5281/zenodo.8007744. Source data are provided with this paper.

Code availability

The custom code for scRNA-seq and scTCR-seq analyses is deposited at Zenodo (https://doi.org/10.5281/zenodo.8011206). The following software packages were used: R v.3.6.2 (2019-12-12), platform x86_64-apple-darwin15.6.0 (64-bit); macOS v.10.16; the R packages Seurat v.3.1.5 and scRepertoire v.1.0.0; Microsoft Excel v.16.73; FlowJo (TreeStar) v.10.8.1; Prism 9 for macOS (graphpad) v.9.5.1; PyMOL (Schrödinger, Inc.) v.2.5.

51. Curran, M. A. & Allison, J. P. Tumor vaccines expressing flt3 ligand synergize with ctla-4 blockade to reject preimplanted tumors. *Cancer Res.* **69**, 7747–7755 (2009).
52. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
53. Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: an R-based toolkit for single-cell immune receptor analysis. *F1000Res.* **9**, 47 (2020).
54. Sanderson, S., Campbell, D. J. & Shastri, N. Identification of a $CD4^+$ T cell-stimulating antigen of pathogenic bacteria by expression cloning. *J. Exp. Med.* **182**, 1751–1757 (1995).
55. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
56. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.04.463034 (2021).
57. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
58. Centanni, M., Sims, I. M., Bell, T. J., Biswas, A. & Tannock, G. W. Sharing a β-glucan meal: transcriptomic eavesdropping on a *Bacteroides ovatus-Subdoligranulum variabile-Hungatella hathewayi* consortium. *Appl. Environ. Microbiol.* **86**, e01651-20 (2020).
59. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
60. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
61. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

Acknowledgements We are deeply indebted to D. Bousbaine and other members of the Fischbach Group for helpful suggestions and comments on the manuscript. We thank H. Takayanagi, S. Sawa, R. Muro, N. A. Bracey, M. M. Davis and members of their laboratories for useful discussions. We are grateful to the Stanford Shared FACS Facility for access to flow cytometry and the Stanford Genomics Facility for constructing and sequencing libraries for scRNA-seq and scTCR-seq experiments, and the Harvard Faculty of Arts & Sciences Research Computing Cluster for computational resources. This work was supported by the Stanford Microbiome Therapies Initiative, the Human Frontier Science Program grant no. LT000493/2018-L (K.N.), a Fellowship from the Astellas Foundation for Research on Metabolic Disorders (K.N.), a research grant from Kanae Fundation for the Promotion of Medical Science (K.N.), NIH grant no. K99AI173524 (K.N.), an HHMI-LSRF Award (J.E.B.), a fellowship from the Kwanjeong Educational Foundation (M.B.), the Howard Hughes Medical Institute (E.P.B.), the Alan T. Waterman Award from the National Science Foundation (E.P.B., grant no. CHE-20380529), an HHMI-Simons Faculty Scholar Award (M.A.F.), the Leona M. and Harry B. Helmsley Charitable Trust (M.A.F.), NIH grant no. DK110174 (M.A.F.), the Chan Zuckerberg Biohub (M.A.F.), Stand Up to Cancer (M.A.F.) and the MAC3 Impact Philanthropies (M.A.F.).

Author contributions K.N., M.B., E.P.B. and M.A.F. conceived and designed the experiments. K.N., A.Z., K.A., M.B., J.E.B., A.W., S.J., X.M., A.G.C., M.W., S.H., A.D., J.J.M. and P.M. performed the experiments. K.N., M.B. and M.A.F. analysed data and wrote the manuscript. M.B., E.S.S. and E.P.B. edited the manuscript. All authors discussed the results and commented on the manuscript.

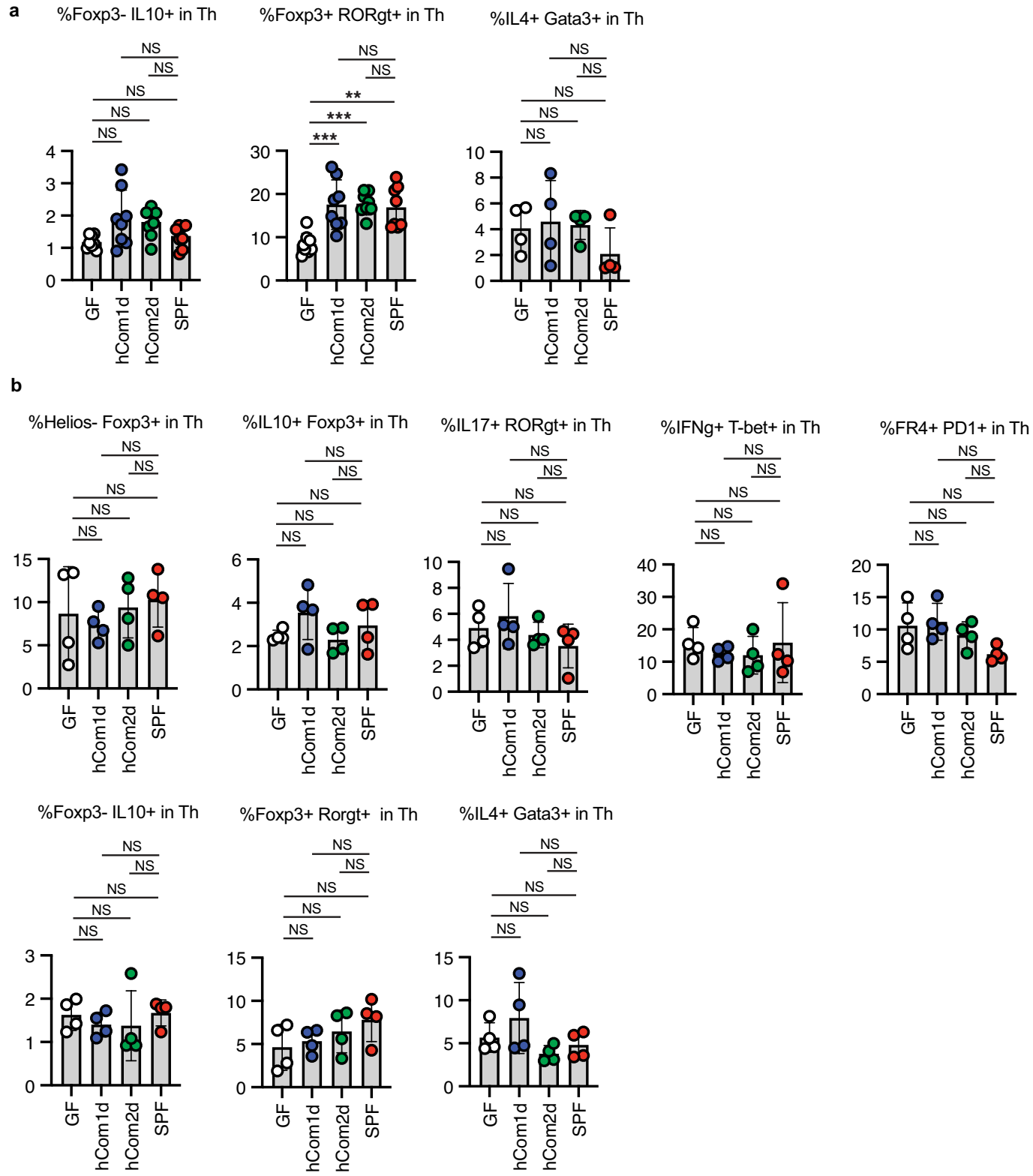
Competing interests Stanford University and the Chan Zuckerberg Biohub have patents pending for microbiome technologies on which the authors are co-inventors. M.A.F. is a cofounder and director of Federation Bio and Kelonia, a cofounder of Revolution Medicines and a member of the scientific advisory boards of the Chan Zuckerberg Initiative and NGM Bio. The remaining authors declare no competing interests.

Additional information

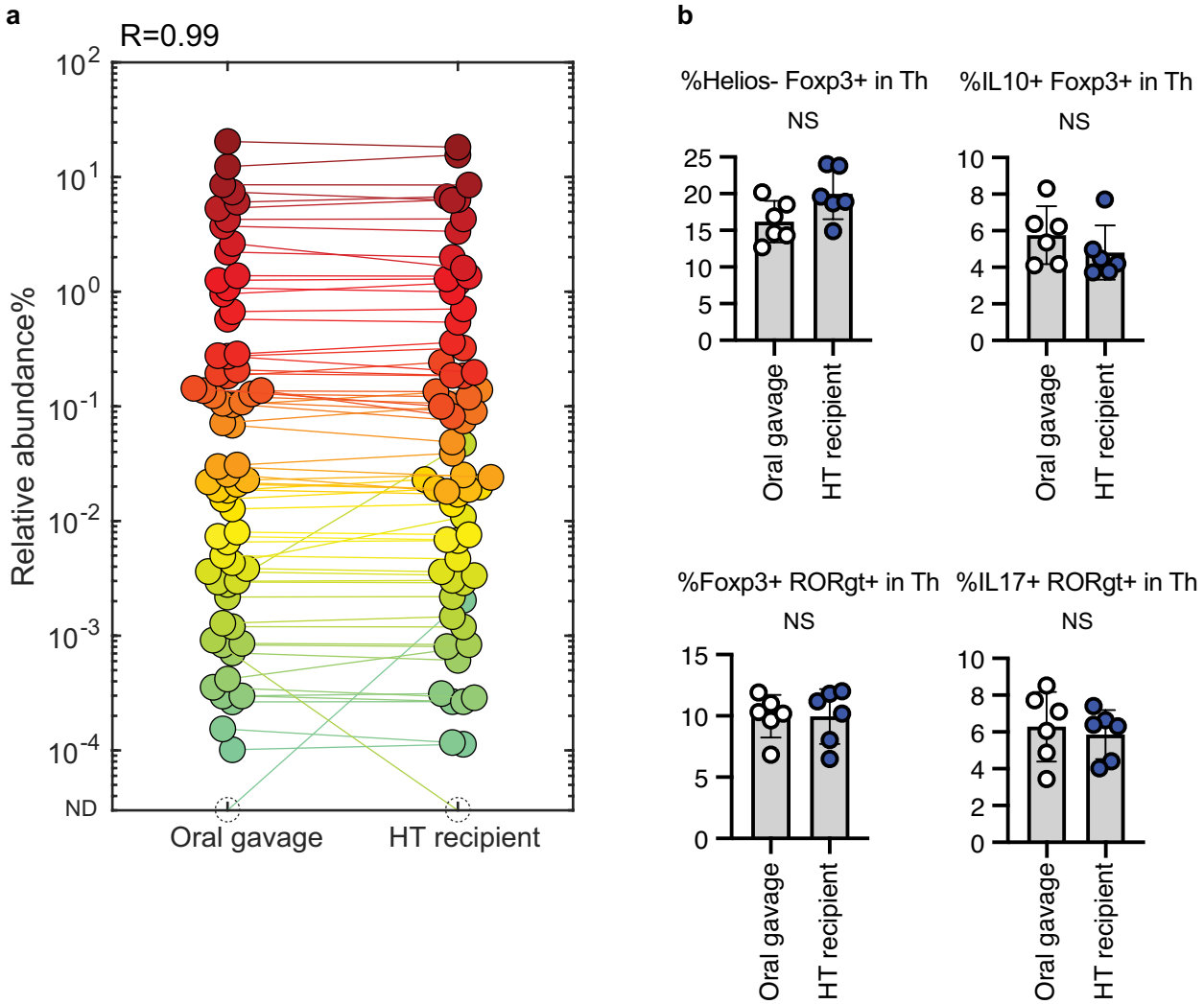
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-06431-8.

Correspondence and requests for materials should be addressed to Michael A. Fischbach.

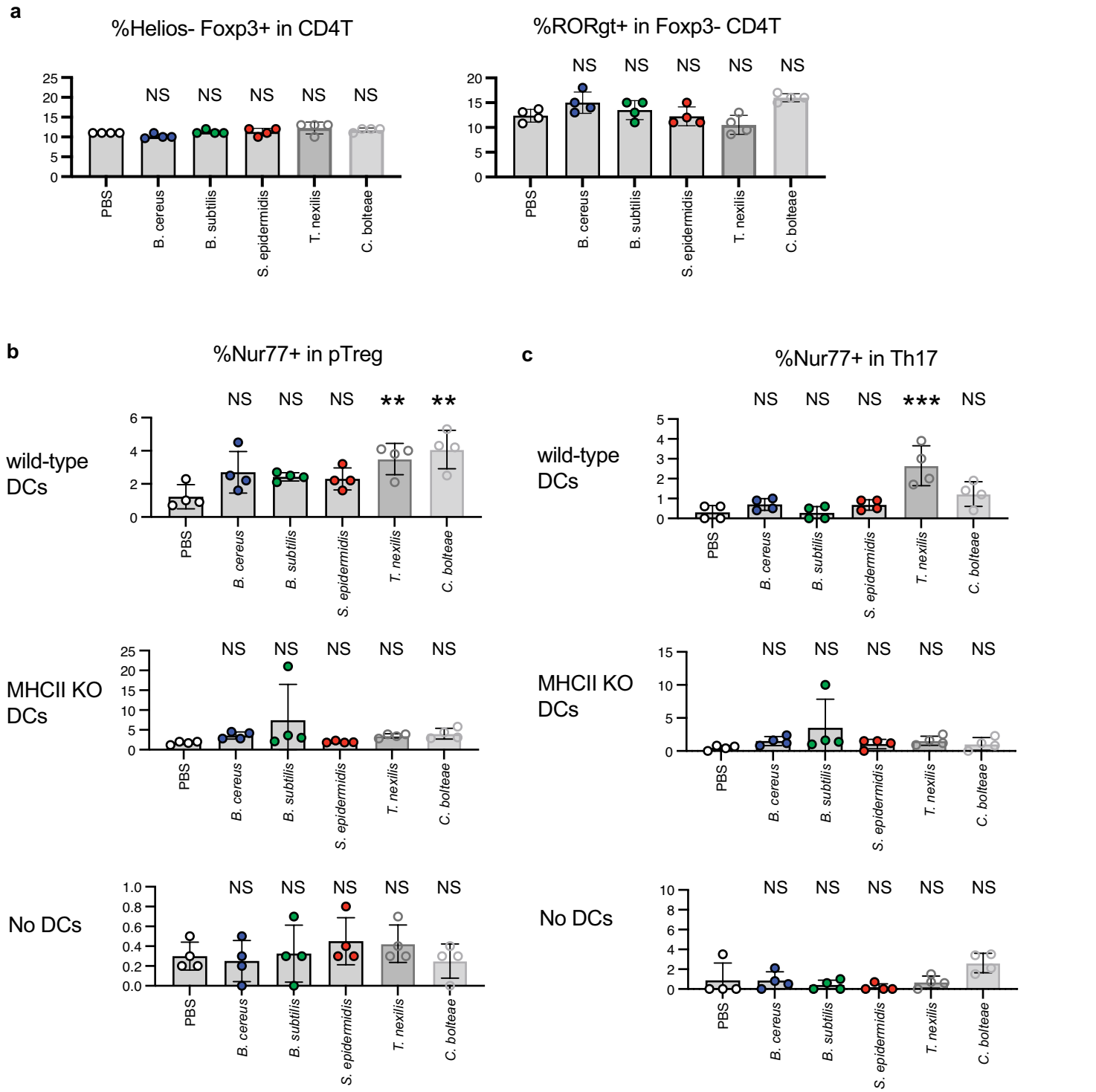
Peer review information *Nature* thanks Jeremiah Faith, Kylie James and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



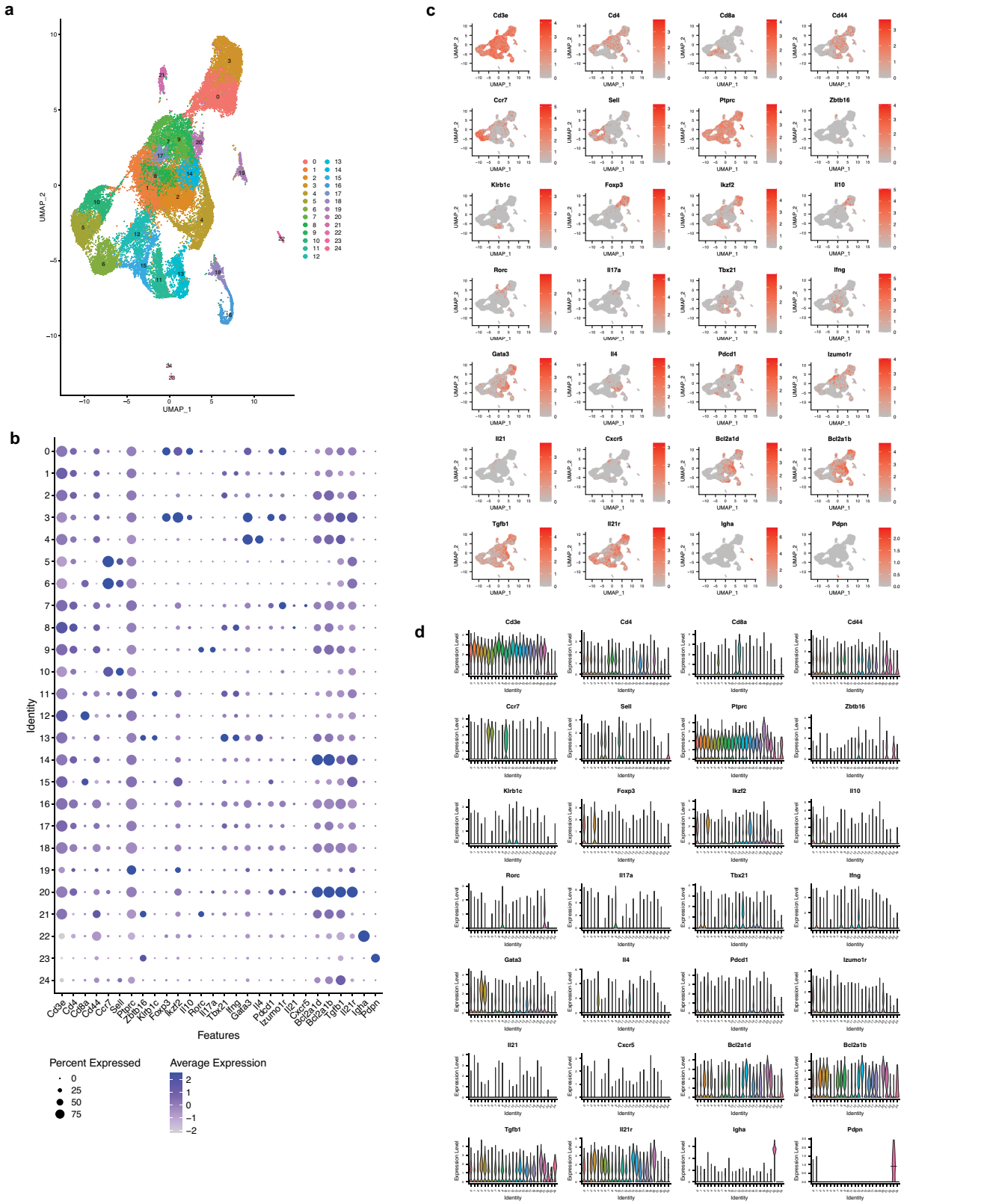
Extended Data Fig. 1 | T cell profiling for hCom1d- and hCom2d-colonized mice. a,b, hCom1d or hCom2d were used to colonize germ-free C57BL/6 mice by oral gavage. Mice were housed for two weeks before sacrifice. Intestinal immune cells were extracted, stimulated by PMA/ionomycin and analyzed by flow cytometry. Th cell subtypes, as a percentage of the total Th cell pool, were analyzed in the large intestine (**a**) or in the small intestine (**b**). See Supplementary Fig. 1b for the gating strategy. Statistical significance was assessed using a one-way ANOVA (NS > 0.05; $^{**}p < 0.01$; $^{***}p < 0.001$). Data shown are mean ± standard deviations. $n = 8, 8, 4$ mice per group from 2 independent experiments (**a**). $n = 4$ mice per group from one experiment (**b**).



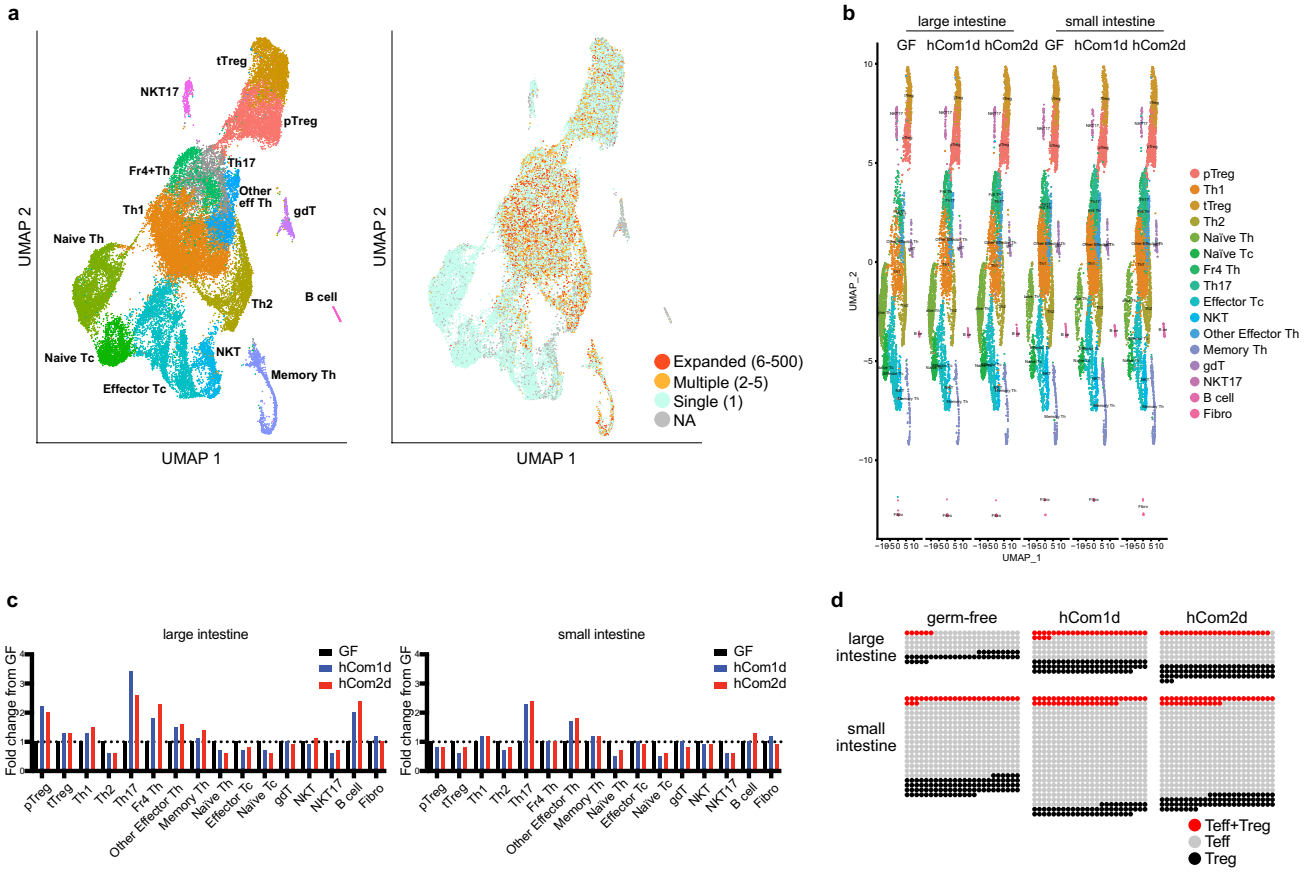
Extended Data Fig. 2 | Introduction of a gut bacterial community to germ-free mice by horizontal transfer. a, Metagenomic analysis of germ-free mice colonized with hCom2d by oral gavage or cohousing (horizontal transfer, HT). Faecal samples were collected 2 weeks after colonization and subjected to metagenomic sequencing; the resulting data were analyzed by NinjaMap to measure the composition of each community. The average of 4 mice per group is displayed above. Each dot is an individual strain; the collection of dots in a column represents the community at a single time point. Strains are coloured according to their rank-order abundance in the oral gavage sample. Relative abundances in the oral gavage and HT samples are highly correlated, indicating that coprophagy (horizontal transfer) results in a similar community architecture to that of oral gavage. **b**, Intestinal T cells from mice colonized with hCom2d by oral gavage and HT show similar phenotypes. Mice were sacrificed after two weeks of colonization. Immune cells were isolated from the large intestine, stimulated by PMA/ionomycin and analyzed by flow cytometry. See Supplementary Fig. 1b for the gating strategy. Statistical significance was assessed using a two-sided t-test (NS > 0.05). Data shown are mean ± standard deviations. n = 6 mice per group from one experiment.



Extended Data Fig. 3 | The mixed lymphocyte assay detects TCR activation while preserving T cell phenotype. a-c, We colonized germ-free mice with hCom1d, waited two weeks, and then sacrificed the mice. Immune cells from the colon were co-cultured for 4 h with three non-community bacterial strains (*Bacillus cereus, Bacillus subtilis*, or *Staphylococcus epidermidis*) or two strains from hCom1d (*Tyzzerella nexilis* and *Clostridium bolteae*). For antigen presentation, we used either MHCII+ wild-type DCs, MHCII-deficient DCs or a no DC control. pTreg (Helios- Foxp3+) cells and Th17 (RORgt+ Foxp3-) cells were analyzed by flow cytometry. **a**, Co-culture with strains and wild-type DCs doesn't affect the number of pTreg cells and Th17 cells. **b,c**, Nur77 expression is upregulated when T cells were cocultured with strains in hCom1d and wild-type DCs in pTreg cells (**b**) and in Th17 cells (**c**). p-values were calculated by comparison to PBS treatment as a negative control using a one-way ANOVA (NS > 0.05; $^{**}p < 0.01$; $^{***}p < 0.001$). Data shown are mean ± standard deviations. $n = 4$ mice per group from one experiment.



Extended Data Fig. 4 | Analysis of scRNA-seq data by unbiased clustering. **a**, Uniform manifold approximation and projection (UMAP) plot of all the cells. The data is generated by merging three groups: hCom1d-colonized, hCom2d-colonized, and germ-free mice. Cells were clustered into 25 groups using the FindClusters function of the Seurat R package. **b**–**d**, Gene expression profiling of cell clusters. To investigate the identity of each cell cluster, expression levels of cell subset markers were visualized by dot plots (**b**), feature plots (**c**) and violin plots (**d**) using Seurat.



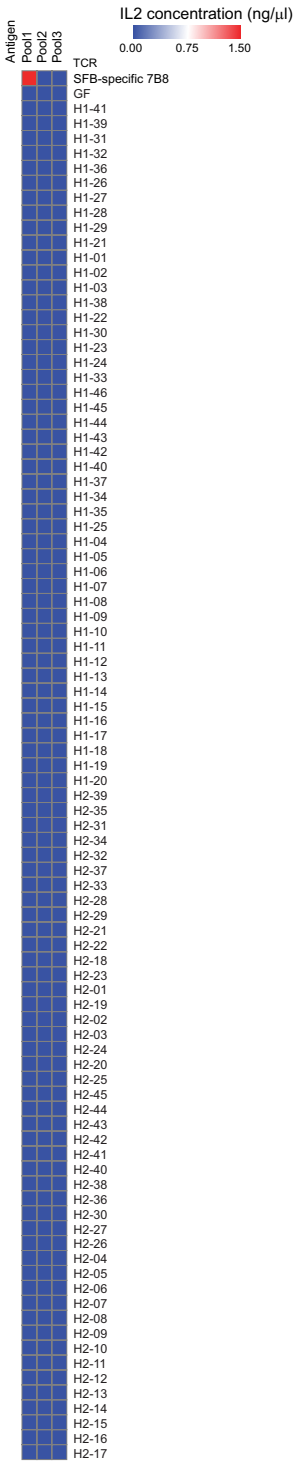
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | scRNA-seq analysis of immune modulation by synthetic community colonization. a, Left panel: Uniform manifold approximation and projection (UMAP) plot. These data represent three merged samples: hCom1d-colonized, hCom2d-colonized and germ-free mice. Right panel: Frequency of TCR clonotypes on the UMAP plot. Expanded TCRs (red) represent clonotypes observed in more than five cells, multiple (orange) are clonotypes found in 2–5 cells, and single (light blue) were seen in only one cell. Most of the expanded TCR clonotypes have an expression profile consistent with effector T cells, whereas naïve T cells are rich in unique (that is, non- expanded) TCR clonotypes. **b**, UMAP plot of intestinal immune cell clusters in each colonization condition. Immune cells were isolated from the large and small intestine from three groups of mice: hCom1d-colonized, hCom2d-colonized, and germ-free. **c**, Analysis of the frequency of T cell subsets in each group. The percentage of each T cell subset on the UMAP plot was calculated by the Seurat R package; fold changes compared to GF mice are shown. Colonization of germ-free mice with hCom1d and hCom2d increased pTreg, Th17, Fr4 Th and other effector T cells in the large intestine, and Th17 and other effector T cells in the small intestine. **d**, Analysis of expanded TCR clonotypes in each sample. Each dot represents one TCR clonotype found in multiple T cells (red, shared between effector T cells and pTreg cells; grey, effector T cell; black, pTreg). **e**, Differentially expressed genes in T cell subsets upon colonization with hCom1d and hCom2d. The FindMarkers function of Seurat was used to find differentially expressed genes. The two-sided non-parametric Wilcoxon rank sum test was used to calculate the adjusted p-value. White bars show mean values. Each dot represents one cell. ***$p < 0.001$. **f**, Criteria used to select TCR clonotypes for making hybridoma cells. Red genes: Upregulated by hCom1d and hCom2d colonization. Black genes: T cell subset markers.

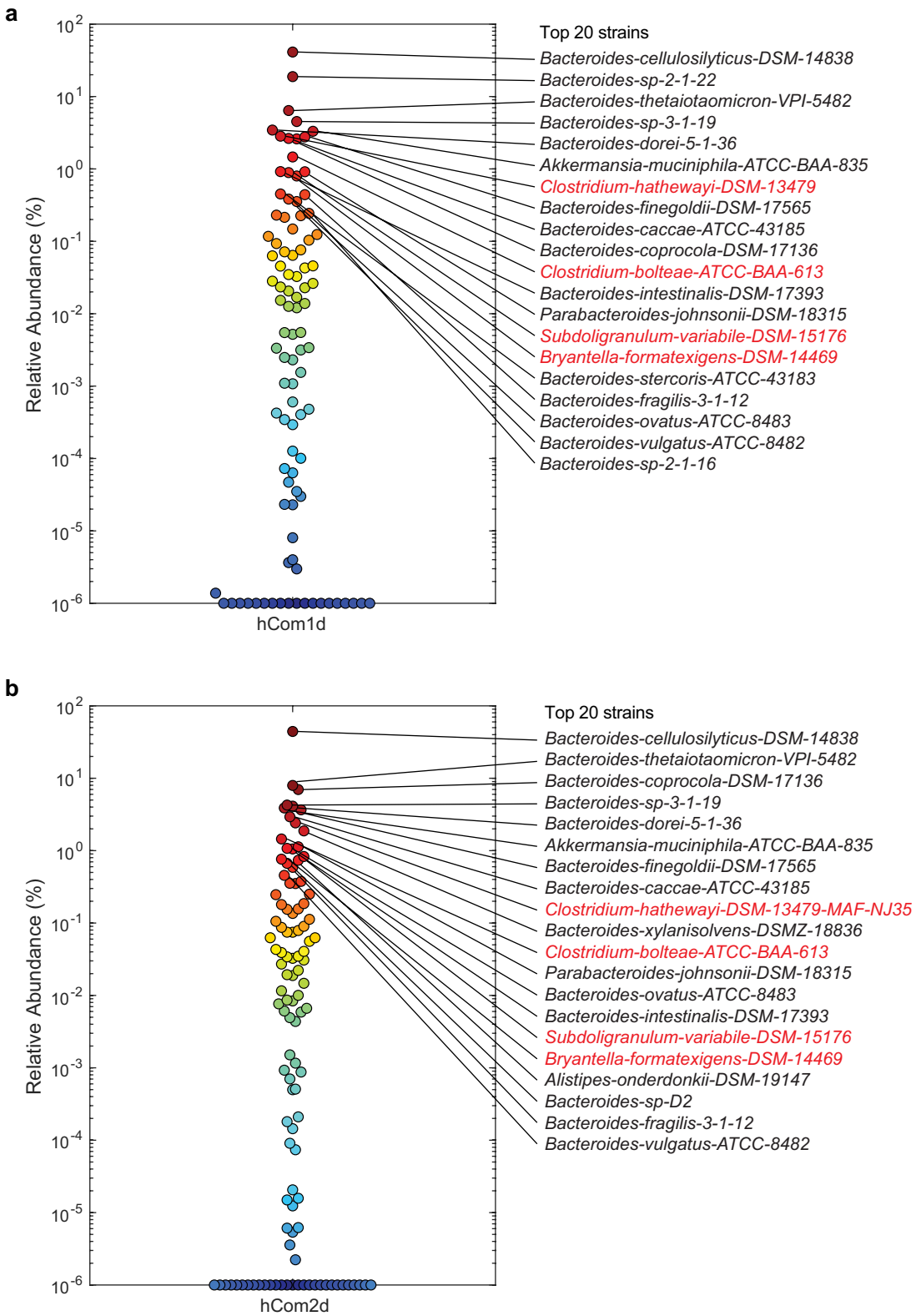
a

#	Peptides	Pool	Bacteria
1	QFSGAVPNKTD	Pool1	SFB
2	GNAYISVLAHYGKNG	Pool1	Helicobacter hepaticus
3	QESPRIAAAYTIKGA	Pool1	Helicobacter hepaticus
4	ESMGNEEFNLPTTNGGHATA	Pool1	Bacteroides thetaiotaomicron
5	KWDYKGSRVWLNDQ	Pool1	Bacteroidetes spp.
6	TLYIGSGAILSGN	Pool1	Akkermansia muciniphila
7	AENLSVTLPAASHAIPHLT	Pool1	E. coli
8	DMATEMVKYSNANILSQAGQ	Pool2	Clostridia
9	VQNRFNSAITNLGNT	Pool2	Salmonella enterica
10	YYETADAFKV	Pool2	Salmonella enterica
11	ISAASYVVPPLPNKR	Pool2	Akkermansia muciniphila
12	NKRSEANPANPPSFA	Pool2	Akkermansia muciniphila
13	HPSRLHAPAITQNAL	Pool2	Akkermansia muciniphila
14	LPGYGLPLCAGPGAS	Pool2	Akkermansia muciniphila
15	SLGAYLSQVALQAAL	Pool3	Lactobacillus murinus
16	TAYLYGIADPRGTRS	Pool3	Lactobacillus murinus
17	SSYRYSSSPLLGIAL	Pool3	Clostridium propionicum
18	SRCAFTASASFN	Pool3	Clostridium propionicum
19	AGLIPIASPIAVAVP	Pool3	Eubacteriurn plexicaudatum
20	LMMFYKSSMSLTAWI	Pool3	Eubacteriurn plexicaudatum

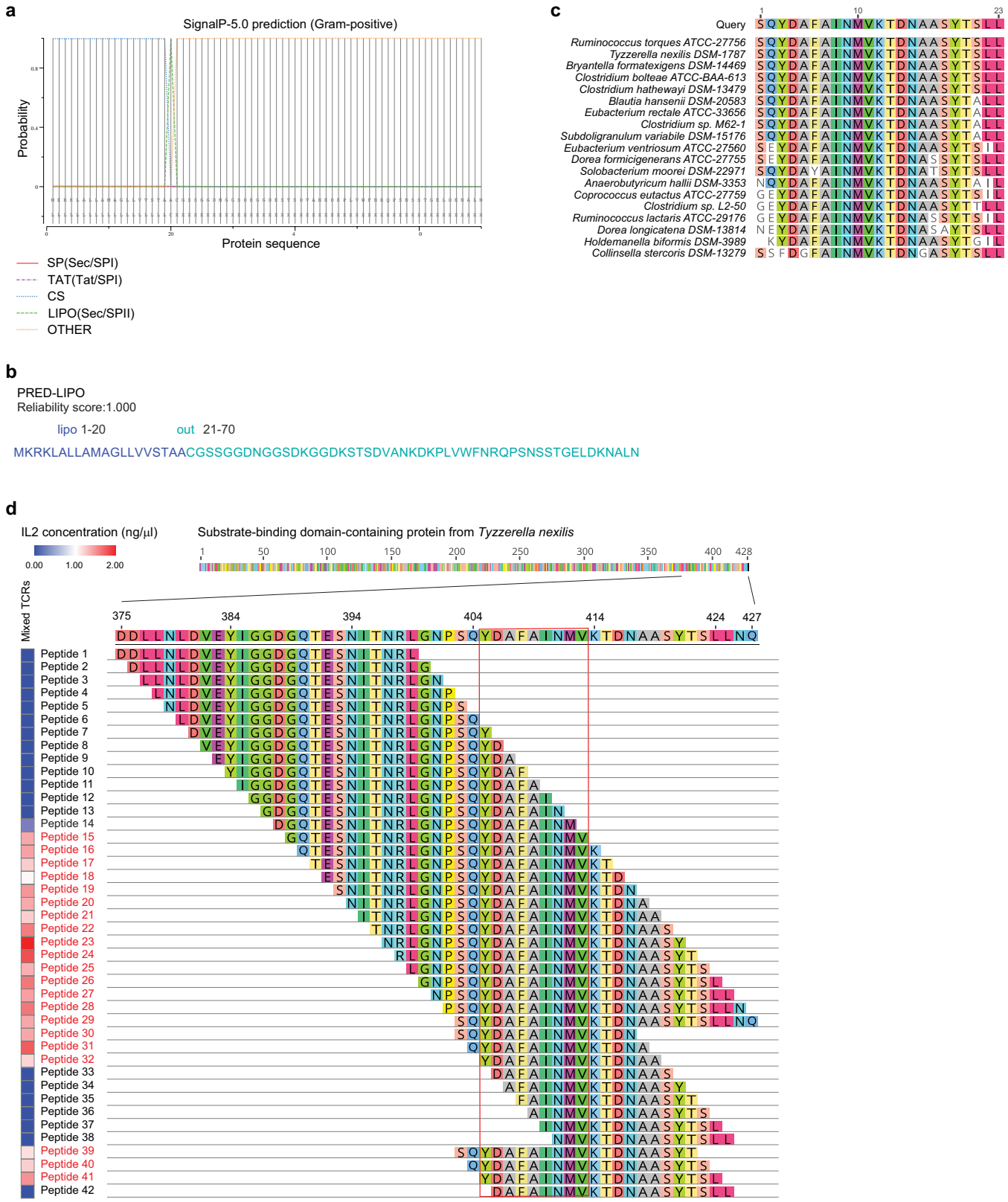
b



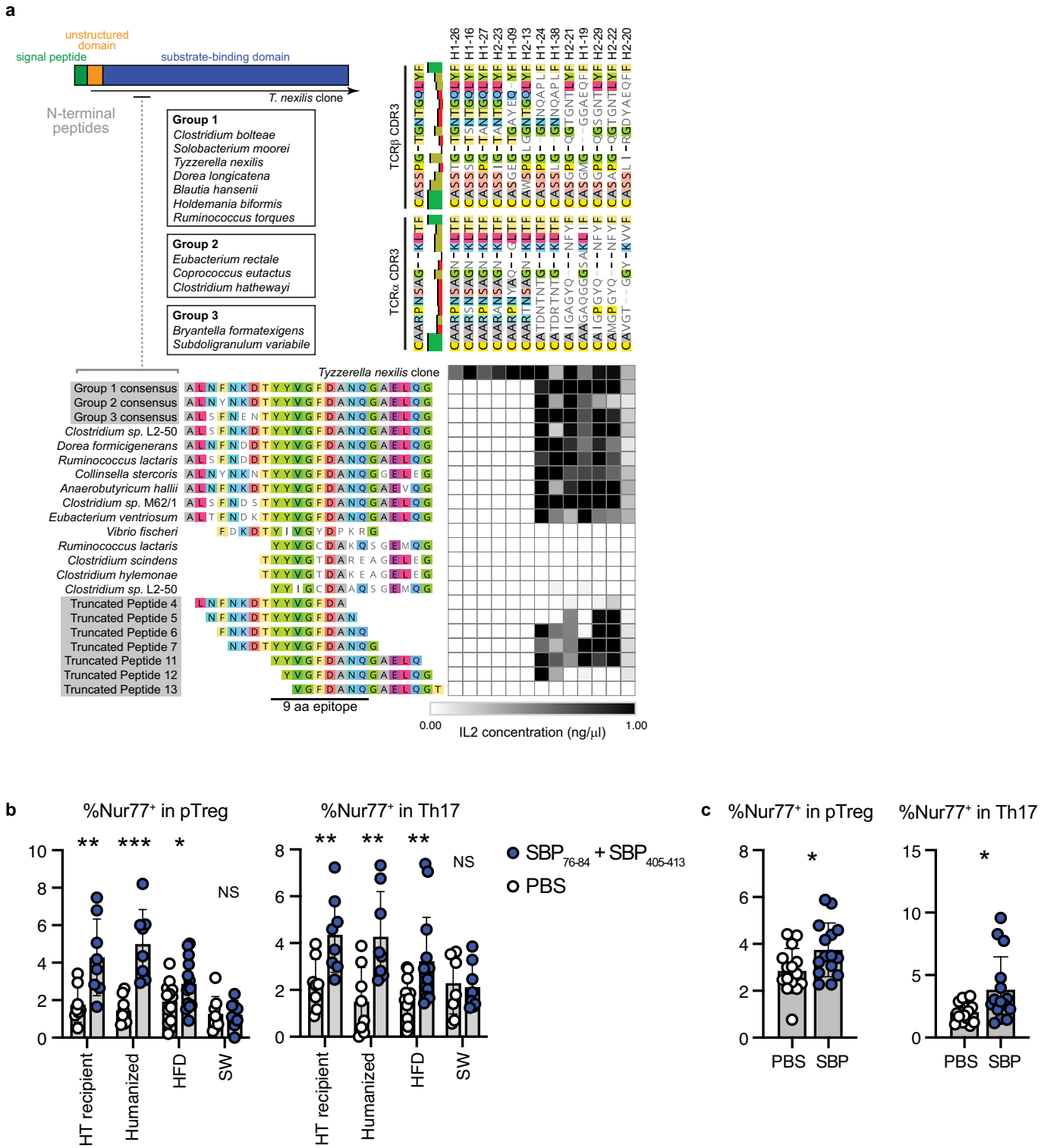
Extended Data Fig. 6 | Reactivity of TCR hybridomas against reported commensal epitopes. a, The list of previously reported commensal epitopes tested in this experiment. Epitopes were pooled into sets of 6-7 for testing the ability to stimulate TCR hybridomas. **b**, Previously reported commensal epitopes were co-cultured with TCR hybridomas and dendritic cells. The SFB-specific 7B8 TCR transgenic T cells, a positive control, showed a response to Pool1, which includes the SFB antigen SFB3340. The 92 TCR hybridomas generated in this work were not responsive to any of the previously reported commensal epitopes.



Extended Data Fig. 7 | Metagenomic analysis for hCom1d and hCom2d strains in the colonization of the mouse intestine. a,b, Each dot is an individual strain; the collection of dots in a column represents the community at a single time point in mice colonized by hCom1d (**a**) or hCom2d (**b**). Strains are colored according to their rank-order relative abundance. Strain names colored red harbor the conserved substrate-binding protein (SBP). Germ-free mice were colonized with synthetic communities, and fecal pellets were collected two weeks after community colonization. The results are an average of 5 mice.

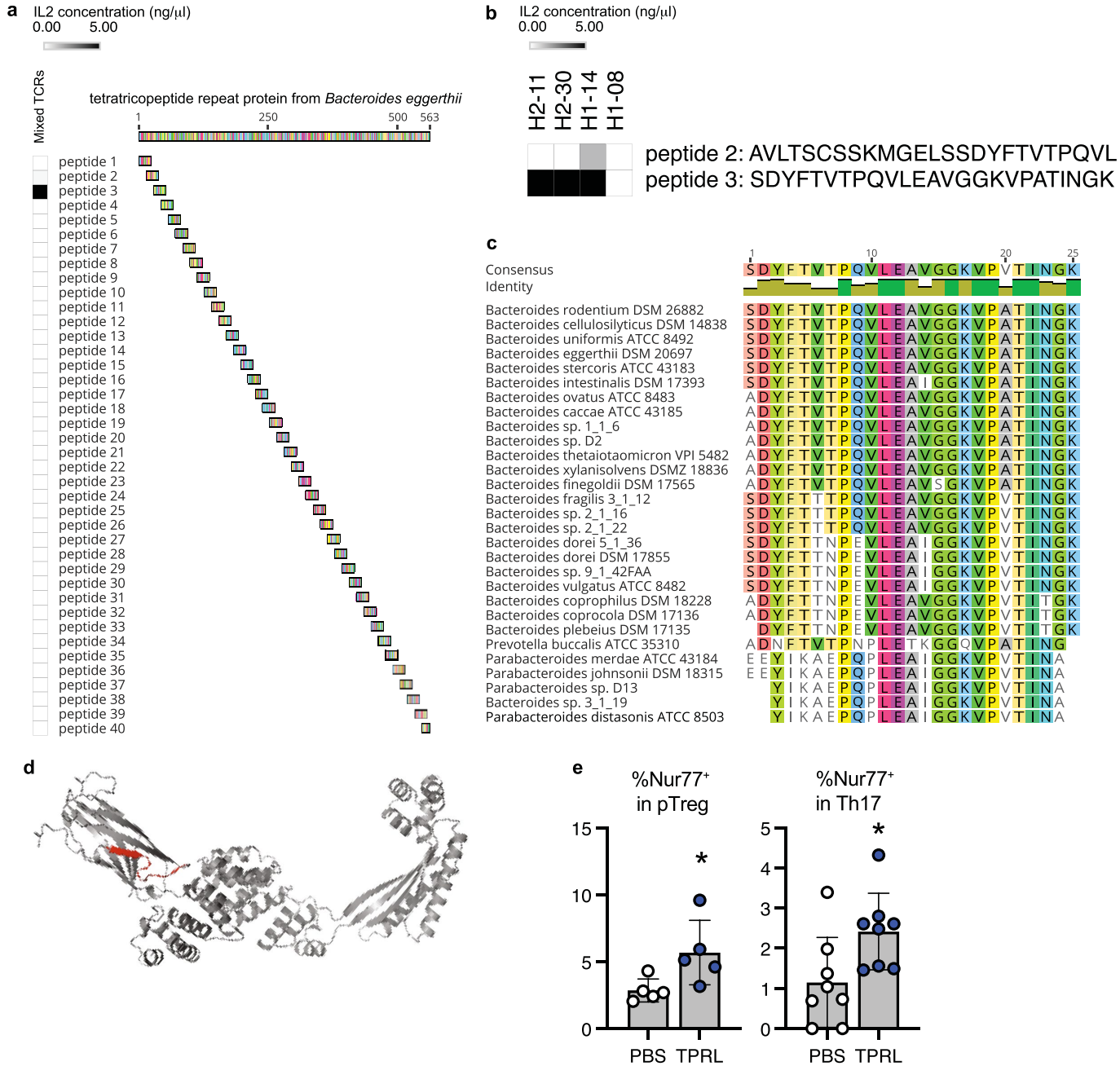


Extended Data Fig. 8 | Identification of the minimal antigen epitopes of SBP using a truncated peptide library. a,b, Prediction of signal sequences and subcellular localization of SBP. SignalP-5.0 (**a**) and PRED-LIPO (**b**) predict that the SBP has a lipoprotein signal peptide. **c**, BLAST search for the antigenic epitopes of SBP in strains from hCom1d and hCom2d. 19 strains were found to harbor homologs of the SBP. **d**, To search for the minimal antigenic epitope in SBP, a library of truncated peptides was synthesized. 13 Firmicutes-reactive TCR hybridomas were mixed and co-cultured with truncated peptides and dendritic cells. The degree of TCR stimulation was estimated by assaying the concentration of IL-2 in the culture supernatant by ELISA. Truncated peptides containing the 9-mer YDAFAINMV stimulated the mixed TCR hybridomas.

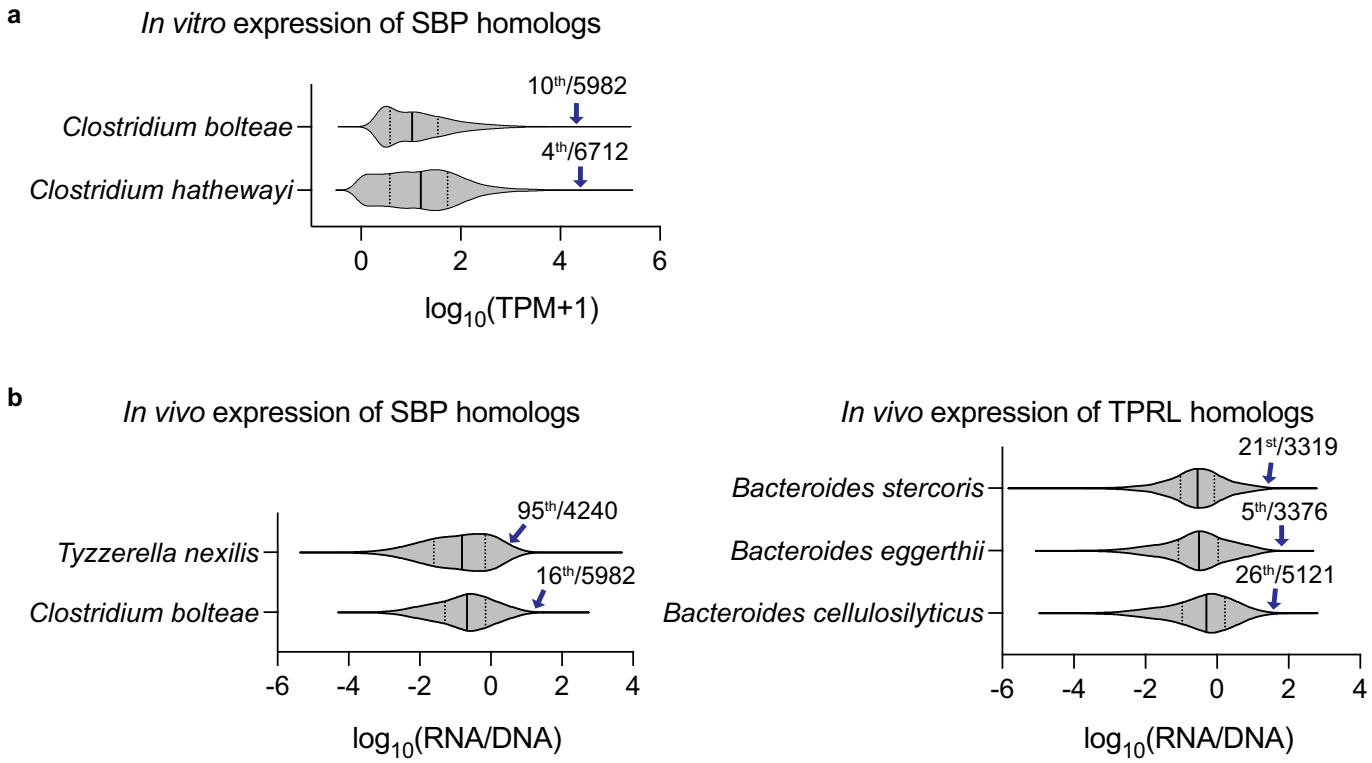


Extended Data Fig. 9 | See next page for caption.

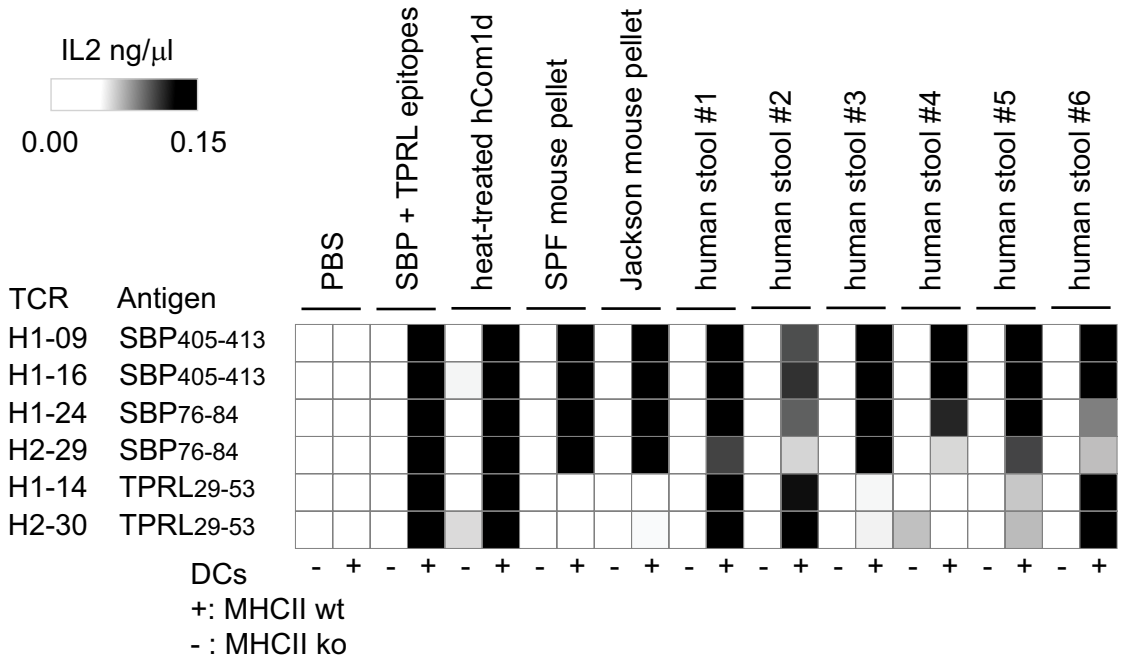
Extended Data Fig. 9 | Discovery of an N-terminal epitope from SBP. **a**, Identification of the N-terminal epitope $SBP_{76\text{-}84}$. N-terminal SBP peptides from strains from hCom1d and hCom2d were synthesized and co-cultured with the T cell hybridomas. Truncated peptides were also tested to identify the minimal epitope. 7 of the 13 TCRs were responsive to the synthetic peptides. The remaining 6 were stimulated by C-terminal peptide $SBP_{405\text{-}413}$ as shown in Fig. 4f. There is a strong correlation between the reactivity of TCRs and the sequences of TCR CD3 regions. **b**, The T cell response to SBP is conserved in different murine settings. We colonized: (1) germ-free C57BL/6 mice with hCom2d by co-housing (horizontal transfer recipient), (2) germ-free C57BL/6 mice with a human fecal community (humanized), (3) germ-free C57BL/6 mice with oral gavage of hCom2d under a condition of a high-fat diet (HFD), and (4) germ-free Swiss Webster mice (SW) with oral gavage of hCom2d. After two weeks, intestinal immune cells were isolated and cocultured with a mix of $SBP_{76\text{-}84}$ and $SBP_{405\text{-}413}$, or PBS as a negative control, using dendritic cells for antigen presentation. Nur77 expression in pTreg or Th17 cells was analyzed by FACS to monitor TCR stimulation (see Supplementary Fig. 1a, c for the gating strategy). SBP-specific T cells were detected in pTreg and Th17 from HT recipient, humanized, and HFD groups, while T cells from SW mice failed to respond to $SBP_{76\text{-}84}$ and $SBP_{405\text{-}413}$. **c**, We hypothesized that T cells from SW mice recognize a distinct epitope in SBP because of the difference in MHC haplotype between C57BL/6 and SW mice. To test this hypothesis, we co-cultured immune cells from hCom2d-colonized SW mice with a mixture of peptides that tile the whole SBP. We detected SBP-specific pTreg and Th17 cells in hCom2-colonized SW mice. This suggests that SBP-specific T cell induction is preserved across different genetic backgrounds of mice, but—as expected—the epitope recognized is distinct because of the difference in MHC haplotype. p-values were calculated using a two-sided t-test by comparison to PBS treatment as a negative control. *p = 0.05. **p = 0.01. ***p = 0.005. NS > 0.05. Data shown are mean ± standard deviations. n = 8, 8, 14, 8 mice per group from 2 independent experiments (**b**). n = 15 mice per group from one experiment (**c**).



Extended Data Fig. 10 | Identification of TPRL as an antigen from *Bacteroides* species. a, To search for the antigenic epitope in TPRL from *Bacteroides eggerthii*, a library of truncated peptides was synthesized. 4 *Bacteroides*-reactive TCR hybridomas were mixed and co-cultured with truncated peptides and dendritic cells. The degree of TCR stimulation was estimated by assaying the concentration of IL-2 in the culture supernatant by ELISA. $TPRL_{29\text{-}53}$ (Peptide 3, SDYFTVTPQVLEAVGGKVPATINGK) stimulated the mixed TCR hybridomas. **b**, We found that TCR H2-11, H2-30, and H1-14 are reactive to Peptide 3 from TPRL by coculturing each *Bacteroides*-reactive TCR hybridoma with Peptides 2 and 3. **c**, Results of a BLAST search using the antigenic epitope $TPRL_{29\text{-}53}$ as a query; the results shown are from strains in hCom1d and hCom2d. **d**, Predicted crystal structure of the TPRL from AlphaFold2. The $TPRL_{29\text{-}53}$ epitope, shown in red, lies within a beta-sheet in the N-terminal domain. **e**, Induction of TPRL-specific T cells in vivo. Germ-free C57BL/6 mice were colonized with hCom2d. After two weeks, intestinal T cells were isolated and cocultured with $TPRL_{29\text{-}53}$ and dendritic cells. Nur77 expression in T cell subsets was analyzed by FACS to monitor TCR stimulation (see Extended Data Fig. 1a, c for the gating strategy). In hCom1d and hCom2d-colonized mice, Th17 and pTreg cells showed an antigen-specific response to SBP. p-values were calculated using a two-sided t-test by comparison to PBS treatment as a negative control. *$p < 0.05$. Data shown are mean ± standard deviations. $n = 5, 8$ mice per group from one experiment.



Extended Data Fig. 11 | T cell targets are highly expressed *in vitro* and *in vivo*. **a**, The SBP is highly expressed in *in vitro* transcriptomic data from two species in our community: *Clostridium bolteae* and *Clostridium hathewayi*. In *C. bolteae*, it is the 10th most highly expressed gene out of 5,982, and in *C. hathewayi*, it is the 4th most highly expressed gene out of 6,712. The expression level was normalized to transcripts per million (TPM). **b**, The SBP and TPRL are highly expressed *in vivo*. Reads were recruited from 378 metatranscriptomic samples[38] and mapped to the whole genomes of select SBP and TPRL-encoding strains in our community (*Tyzzerella nexilis* DSM 1787 and *Clostridium bolteae* DSM 15670 (SBP); *Bacteroides stercoris* ATCC 43183, *Bacteroides eggerthii* DSM 20697, and *Bacteroides cellulosilyticus* DSM 14838 (TPRL)). To account for gene copy number variation, we measured expression as the ratio between RNA and DNA level from paired metagenomic and metatranscriptomic samples. In all cases, we found that the protein is very highly expressed *in vivo*.



Extended Data Fig. 12 | SBP and TPRL are present in human stool samples. We tested whether SBP and TPRL exist at a detectable level in human stool samples. We cocultured SBP- or TPRL-reactive TCR hybridomas with: (1) PBS as a negative control, (2) a mix of SBP and TPRL epitopes as a positive control, (3) heat-treated hCom1d, (4) heat-treated SPF mouse fecal pellets, 5) heat-treated mouse fecal pellets from Jackson, (5) 6 human fecal communities. For antigen presentation, we used MHCII+ wild-type DCs, or MHCII-deficient DCs as a negative control. IL-2 concentration was measured as a readout for TCR stimulation. We found that 6/6 human fecal communities restimulate the $SBP_{405-413}$ and SBP_{76-84}-specific hybridomas and 5/6 restimulate the $TPRL_{29-53}$-specific hybridomas. These data suggest that the SBP and TPRL are expressed by human gut isolates under native conditions.

nature portfolio

Corresponding author(s): Michael A. Fischbach

Last updated by author(s): Jun 7, 2023

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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.

n/a	Confirmed
☐	☒ 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	Chromium controller (10x Genomics) was used for scRNA/TCRseq. The libraries were sequenced on an Illumina HiSeq 4000.
Data analysis	Custom codes for scRNAseq and scTCRseq analysis were deposited (DOI: 10.5281/zenodo.8011206). R version 3.6.2 (2019-12-12) Platform: x86_64-apple-darwin15.6.0 (64-bit). Running under: macOS 10.16. The R packages Seurat version 3.1.5 and scRepertoire version 1.0.0. Microsoft Excel: Version 16.73. FlowJo (TreeStar): Version 10.8.1. Prism 9 for macOS (graphpad): Version 9.5.1. PyMOL (Schrödinger, Inc): Version 2.5.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw sequence data:
scTCRseq of T cells from the small intestine of germ-free mice, DOI: 10.5281/zenodo.8007653
scTCRseq of T cells from the small intestine of hCom1 colonized mice, DOI: 10.5281/zenodo.8007692
scTCRseq of T cells from the small intestine of hCom2 colonized mice, DOI: 10.5281/zenodo.8007696

scTCRseq of T cells from the large intestine of germ-free mice, DOI: 10.5281/zenodo.8007700
scTCRseq of T cells from the large intestine of hCom1 colonized mice, DOI: 10.5281/zenodo.8007704
scTCRseq of T cells from the large intestine of hCom2 colonized mice, DOI: 10.5281/zenodo.8007709
scRNAseq of T cells from the small intestine of germ-free mice, DOI: 10.5281/zenodo.8007733
scRNAseq of T cells from the small intestine of hCom1 colonized mice, DOI: 10.5281/zenodo.8007735
scRNaseq of T cells from the small intestine of hCom2 colonized mice, DOI: 10.5281/zenodo.8007737
scRNAseq of T cells from the large intestine of germ-free mice, DOI: 10.5281/zenodo.8007740
scRNAseq of T cells from the large intestine of hCom1 colonized mice, DOI: 10.5281/zenodo.8007742
scRNAseq of T cells from the large intestine of hCom2 colonized mice, DOI: 10.5281/zenodo.8007744

Cell Ranger out files, DOI: 10.5281/zenodo.8008419
Reference mouse genome: mm10

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	Sample sizes were indicated in the figure legends. No statistical test was used to determine sample size. 3-15 mice per group were chosen based on common practice in the field (doi: 10.1126/science.1198469. doi: 10.1016/j.cell.2009.09.033), animal welfare guidelines and availability of animals, while minimizing the use of animals in accordance with animal care guidelines of Stanford University.
Data exclusions	There were no data exclusions.
Replication	Disclosed in the figure legends.
Randomization	In mouse experiments, sexually matched, age matched litter mates were randomly assigned to experimental groups.
Blinding	As with common practices in the filed (doi.org/10.1038/s41586-023-05801-6), the same investigators were involved in planning, processing, and acquiring the samples, and the experiments could not be performed blinded.

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 and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Fixable Viability Dye eFluor 780 (eBioscience, Cat#: 65-0865-18), Helios (22F6, BioLegend, Cat#: 137214), T-bet (4B10, BioLegend, Cat#: 644806), Nur77-PE (12.14, eBioscience, Cat#: 12-5965-82), Foxp3 (FJK-16s, eBioscience, Cat#: 25-5773-82), Nur77-purified (11C1052, LSBio, Cat#: LS-C183978-100), CD45.1 (A20, eBioscience, Cat#: 47-0453-82), CD4 (RM4-5, BioLegend, Cat#: 100559), CD3e (145-2C11 , BioLegend, Cat#: 100351), FR4 (12A5, BD, Cat#: 744122), IL17A (eBio17B7, eBioscience, Cat#: 12-7177-81), RORgt (B2D, eBioscience, Cat#: 17-6981-82 or Q31-378, BD Bioscience, Cat#: 562894), IFNg (XMG1.2, BioLegend, Cat#: 505830), PD-1 (29F.1A12, BioLegend, Cat#: 135208) and IL10 (JES3-9D7, BioLegend, Cat#: 505010). The cells stained with the Nur77-purified antibody (11C1052, LSBio, Cat#: LS-C183978-100) were further stained with a secondary APC Goat anti-Rabbit IgG (Polycolonal, eBioscience, Cat#: A-10931). All antibody staining was performed at the dilution of 1/200 in the presence of purified anti-mouse CD16/32 (clone 93, Cat#: 553141) and 10% fetal bovine serum to block nonspecific binding.

Validation	All the antibodies were commercially available. All the primary antibodies were validated by their manufacturers to be specific to mouse antigens in the flow cytometry experiment. The specificity of secondary APC Goat anti-Rabbit IgG in flow cytometry was validated by Thermo Fisher Scientific.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The Platinum-E (Plat-E) retroviral packaging cell line was purchased from the Cell Biolabs (Cat#: RV-101). The NFAT-GFP 58α −β− hybridoma cell line (doi.org/10.1038/ni.1835) and B16-FLT3L cell lines (PMID: 10866317) were gifts from Daniel Mucida Lab, The Rockefeller University, NY, USA.
Authentication	Cell lines were authenticated by the supplier. No other authentication was performed in our lab.
Mycoplasma contamination	Confirmed mycoplasma free.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Animals and other organisms

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

Laboratory animals	Gnotobiotic mouse experiments were performed on C57BL/6 or Swiss Webster germ-free mice originally obtained from Taconic Biosciences maintained in aseptic isolators. SFB- C57BL/6 mice (Stock#: 000664), CD45.1+ C57BL/6 mice (Stock#: 002014), Nur77-EGFP C57BL/6 mice (Stock#: 016617), MHCII-deficient C57BL/6 mice (Stock#: 003584), and OTII-TCRtg C57BL/6 mice (Stock#: 004194) were purchased from the Jackson Laboratory. All the mice were at 8-15 weeks old on the day of sacrifice. Both male and female mice were used, and sexually matched littermates were selected as a control throughout the study. The animal facility was operated at Stanford University with the following housing conditions: 12:12 Light- Dark cycle, room temperature between 68-79°F, humidity level 40-65%. All mouse experiments were conducted under a protocol approved by the Stanford University Institutional Animal Care and Use Committee.
Wild animals	None
Field-collected samples	None
Ethics oversight	Stanford University APLAC protocol #32872

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	Mice were euthanized by cervical dislocation. The small and large intestine were collected by dissection and the mesenteric lymph nodes, Peyer's patches, and cecum were removed. Intestinal tissue was shaken at 225 rpm for 40 min at 37 -°C in DMEM (Gibco) containing 5 mM EDTA and 1 mM DTT. Tissues were washed with DMEM, manually shaken to detach the epithelial layer and cut into pieces. Tissue fragments were then digested using a mouse Lamina Propria Dissociation Kit (Miltenyi Biotec, #130-097-410) and a gentleMACS dissociator (Miltenyi Biotec). After digestion, debris was removed using a 40% and 80% Percol gradient purification (GE Healthcare). Single cell suspensions were stained with Fixable Viability Dye eFluor 780 (eBioscience, 65-0865-18) before fixation to detect dead cells. Cells were then fixed for 30 min at room temperature using the fixation/permeabilization buffer supplied with the eBioscience Foxp3/Transcription Factor Staining buffer set (ThermoFisher Scientific, 00-5523-00). After fixation, cells were stained with combinations of antibodies in permeabilization buffer for 30 min at room temperature. All staining was performed in the presence of purified anti-mouse CD16/32 (clone 93) and 10% fetal bovine serum to block nonspecific binding.
Instrument	Cells were analyzed using an LSR II flow cytometer (BD Biosciences).
Software	Data were processed using FlowJo (TreeStar, Version 10.8.1).
Cell population abundance	Cell population abundance was monitored with cell count beads.

Gating strategy	All samples were initially gated using forward scatter and side scatter to identify events corresponding to cells, and then using forward scatter height vs. area to enrich for single cells, next alive cells were selected by negativity for viability dye. The same gating strategy was applied across different samples.

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