

# Focused specificity of intestinal T<sub>H</sub>17 cells towards commensal bacterial antigens

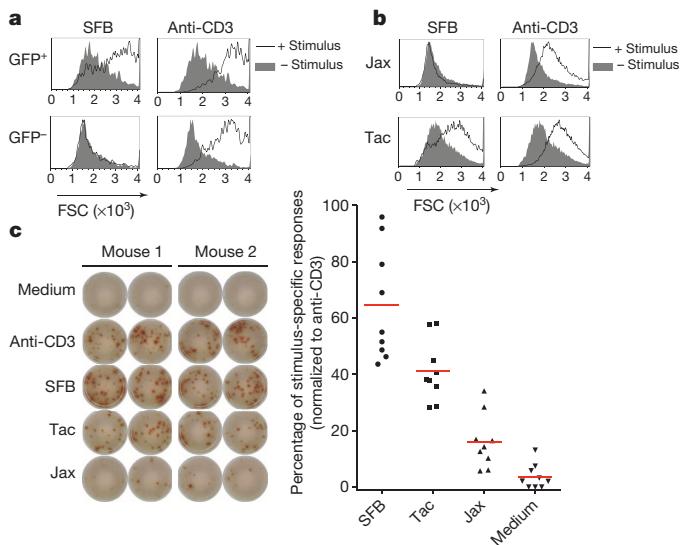
Yi Yang<sup>1</sup>, Miriam B. Torchinsky<sup>1</sup>, Michael Gobert<sup>1</sup>, Huizhong Xiong<sup>1</sup>, Mo Xu<sup>1</sup>, Jonathan L. Linehan<sup>2†</sup>, Francis Alonzo<sup>3</sup>, Charles Ng<sup>1</sup>, Alessandra Chen<sup>1</sup>, Xiyao Lin<sup>1</sup>, Andrew Szesnak<sup>1</sup>, Jia-Jun Liao<sup>1</sup>, Victor J. Torres<sup>3</sup>, Marc K. Jenkins<sup>2</sup>, Juan J. Lafaille<sup>1</sup> & Dan R. Littman<sup>1,4</sup>

**T-helper-17 (T<sub>H</sub>17) cells have critical roles in mucosal defence and in autoimmune disease pathogenesis<sup>1–3</sup>. They are most abundant in the small intestine lamina propria, where their presence requires colonization of mice with microbiota<sup>4–7</sup>. Segmented filamentous bacteria (SFB) are sufficient to induce T<sub>H</sub>17 cells and to promote T<sub>H</sub>17-dependent autoimmune disease in animal models<sup>8–14</sup>. However, the specificity of T<sub>H</sub>17 cells, the mechanism of their induction by distinct bacteria, and the means by which they foster tissue-specific inflammation remain unknown.** Here we show that the T-cell antigen receptor (TCR) repertoire of intestinal T<sub>H</sub>17 cells in SFB-colonized mice has minimal overlap with that of other intestinal CD4<sup>+</sup> T cells and that most T<sub>H</sub>17 cells, but not other T cells, recognize antigens encoded by SFB. T cells with antigen receptors specific for SFB-encoded peptides differentiated into ROR $\gamma$ t-expressing T<sub>H</sub>17 cells, even if SFB-colonized mice also harboured a strong T<sub>H</sub>1 cell inducer, *Listeria monocytogenes*, in their intestine. The match of T-cell effector function with antigen specificity is thus determined by the type of bacteria that produce the antigen. These findings have significant implications for understanding how commensal microbiota contribute to organ-specific autoimmunity and for developing novel mucosal vaccines.

How SFB induces T<sub>H</sub>17 cells and how these cells contribute to self-reactive pathological responses remain key unanswered questions. A recent study, using mice with monoclonal TCRs, suggested that induction of T<sub>H</sub>17 cells by SFB or other microbiota is independent of cognate antigen recognition<sup>15</sup>. To further evaluate mucosal effector T-cell induction in a physiological setting, we undertook an examination of the repertoire and specificity of naturally arising T<sub>H</sub>17 cells. To facilitate analysing live T<sub>H</sub>17 cells, we used *Il23r*<sup>GFP</sup> reporter mice<sup>16</sup>, as among CD4<sup>+</sup> T cells, only this subset expresses IL-23R. We first asked if small intestine lamina propria (SILP) T<sub>H</sub>17 cells are in general responsive to gut luminal commensal antigens. GFP<sup>+</sup> (T<sub>H</sub>17) and GFP<sup>-</sup> (non-T<sub>H</sub>17) CD4<sup>+</sup> T cells, purified from *Il23r*<sup>GFP/+</sup> C57BL/6 (B6) mice that had been colonized with SFB, were incubated with splenic antigen-presenting cells (APCs) and autoclaved small intestinal luminal content of mice from the Jackson laboratory (Jackson) and Taconic Farms (Taconic). We used the measure of forward scatter (FSC) as a surrogate readout for T-cell activation. Intriguingly, only T<sub>H</sub>17 cells mounted a detectable response to Taconic antigens (Extended Data Fig. 1a). SFB is one of the bacteria unique to Taconic flora<sup>8</sup>. Thus we repeated the assay with faecal material from SFB-monoassociated mice (SFB-mono antigens) and detected a robust response only among GFP<sup>+</sup> cells (Fig. 1a). These cells did not respond to major histocompatibility complex class II (MHCII)-deficient APCs loaded with SFB-mono antigens, indicating that the activation was dependent on antigen presentation (Extended Data Fig. 1b). SFB-mono antigens selectively stimulated total CD4<sup>+</sup> T cells from B6 Taconic mice, but not those from B6 Jackson mice, consistent with

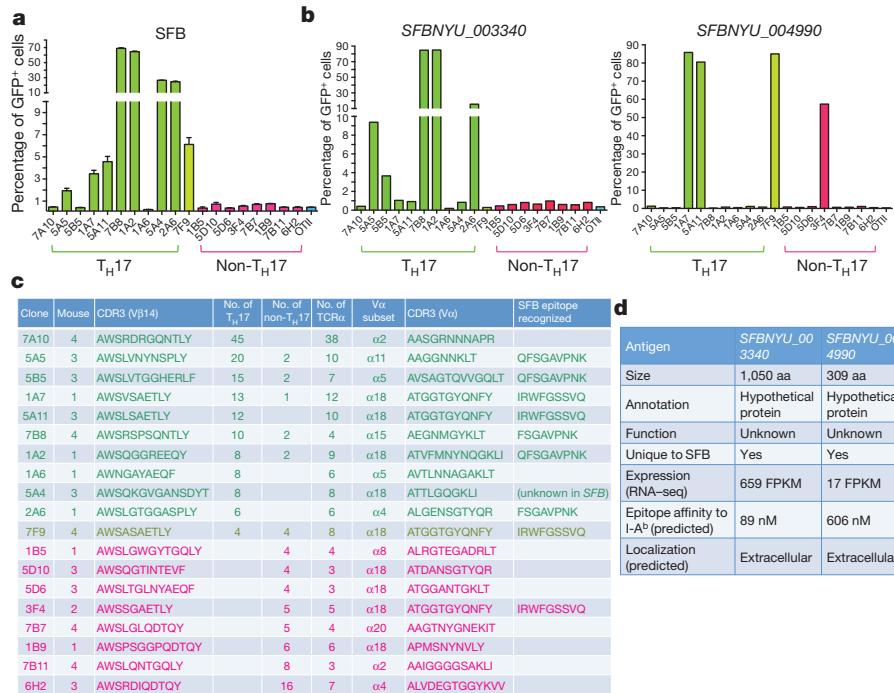
*in vivo* priming of SFB-specific T<sub>H</sub>17 cells (Fig. 1b), and any bystander effect in this assay was negligible (Extended Data Fig. 1c). Next, we used an IL-17A ELISPOT assay to quantify the percentage of T<sub>H</sub>17 cells from SFB-colonized mice responding to commensal antigens. GFP<sup>+</sup> cells had a relatively weak response towards Jackson antigens, but had a robust response towards Taconic antigens. Significantly, SFB mono-associated mouse faecal antigens stimulated over 60% of the T<sub>H</sub>17 cells (Fig. 1c). In contrast, there was no response of T<sub>H</sub>17 cells to faecal material from germ-free mice (data not shown). Thus, the majority of T<sub>H</sub>17 cells in the SILP of SFB-colonized mice react with SFB-derived antigens, whereas a small proportion respond to non-SFB antigen, indicating that most T<sub>H</sub>17 cells are specific for bacteria in the intestinal lumen.

We wished to compare the TCR repertoire of T<sub>H</sub>17 cells and those of non-T<sub>H</sub>17 cells. Using antibodies against a panel of TCR V $\beta$ s, we observed a higher proportion of V $\beta$ 14<sup>+</sup> T cells in T<sub>H</sub>17 cells than in non-T<sub>H</sub>17 cells from the SILP (Extended Data Fig. 2a, b). This bias was



**Figure 1 | Intestinal T<sub>H</sub>17 cells are specific for SFB- and other microbiota-derived antigens.** **a**, Selective activation of intestinal GFP<sup>+</sup> CD4<sup>+</sup> T cells from *Il23r*<sup>GFP/+</sup> mice by faecal extract from SFB-monoassociated mice. Forward scatter (FSC) was evaluated after 2 days. **b**, Activation of SILP CD4<sup>+</sup> T cells from B6 Taconic mice and B6 Jackson mice with faecal extract from SFB-monoassociated mice. Jax, Jackson mice; Tac, Taconic mice. **c**, IL-17A ELISPOT assay of intestinal GFP<sup>+</sup> CD4<sup>+</sup> T cells from SFB-colonized *Il23r*<sup>GFP/+</sup> mice treated with indicated stimuli. Left, representative ELISPOT images. Right, compilation of results from multiple animals. Each symbol represents cells from a separate animal.

<sup>1</sup>The Kimmel Center for Biology and Medicine of the Skirball Institute, New York University School of Medicine, New York, New York 10016, USA. <sup>2</sup>Department of Microbiology, Center for Immunology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA. <sup>3</sup>Department of Microbiology, New York University School of Medicine, New York, New York 10016, USA. <sup>4</sup>Howard Hughes Medical Institute, New York University School of Medicine, New York, New York 10016, USA. <sup>†</sup>Present address: Mucosal Immunology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0485, USA.



**Figure 2 | Most T<sub>H</sub>17 TCR hybridomas recognize SFB unique proteins.**

**a**, Responses of the TCR hybridomas, prepared from T<sub>H</sub>17 and non-T<sub>H</sub>17 intestinal CD4<sup>+</sup> T cells, to faecal material from SFB-monoassociated mice. **b**, Responses of the TCR hybridomas to *Escherichia coli* clones expressing full-length SFBNYU\_003340 and SFBNYU\_004990. Note that a non-T<sub>H</sub>17 TCR

hybridoma also responded to the clone expressing SFBNYU\_004990. **c**, Summary of the nineteen dominant clonotypic TCR clones. Ten T<sub>H</sub>17-biased clones are highlighted in green, and eight non-T<sub>H</sub>17-biased clones are highlighted in red. **d**, Features of the two antigenic proteins of SFB. FPKM, fragments per kilobase of transcript per million mapped reads.

recapitulated when the CD4<sup>+</sup> T cells were stained with antibodies specific for ROR $\gamma$ t and IL-17A, two other characteristic markers of T<sub>H</sub>17 cells (Extended Data Fig. 2c). However, intracellular staining for IFN $\gamma$  and FOXP3 indicated no V $\beta$ 14<sup>+</sup> cell bias among T<sub>H</sub>1 and T regulatory cells (Extended Data Fig. 2c). To determine if the V $\beta$ 14 enrichment of T<sub>H</sub>17 cells is influenced by microbiota, we compared SFB-free B6 Jackson mice with SFB-colonized B6 Taconic mice. The Jackson mice had few ROR $\gamma$ t<sup>+</sup> T<sub>H</sub>17 cells, and there was no enrichment of V $\beta$ 14<sup>+</sup> cells among them. In contrast, Jackson mice cohoused with Taconic mice had increased numbers of lamina propria T<sub>H</sub>17 cells, which were enriched for V $\beta$ 14<sup>+</sup> TCRs (Extended Data Fig. 2d), indicating that the T<sub>H</sub>17 repertoire is shaped by specific microbiota.

We chose to focus on V $\beta$ 14<sup>+</sup> cells to further elucidate the gut CD4<sup>+</sup> T-cell repertoire. First, we used pyrosequencing to examine the repertoire of V $\beta$ 14<sup>+</sup> SILP T<sub>H</sub>17 and non-T<sub>H</sub>17 cells from SFB-colonized mice. The complementarity-determining region 3 (CDR3) of V $\beta$ 14 was characterized for each cell population from eight *Il23r*<sup>GFP/+</sup> mice. Each sample contained a minimum of several hundred unique CDR3 sequences (Extended Data Fig. 3a). Interestingly, the ten most frequently used unique CDR3 sequences accounted for 60% of the T<sub>H</sub>17 and only 40% of the non-T<sub>H</sub>17 repertoire (Extended Data Fig. 3b). Furthermore, the dominant CDR3 sequences in individual mice exhibited a clear bias towards either T<sub>H</sub>17 or non-T<sub>H</sub>17 cells (Supplementary Table 1). Many of these CDR3 sequences were shared between mice and were enriched either in T<sub>H</sub>17 or in non-T<sub>H</sub>17 cells in individual mice (Extended Data Fig. 3c).

The finding that intestinal T<sub>H</sub>17 cells have a distinct repertoire prompted us to further determine their antigen specificity. Thus, we sorted single T cells from four mice and sequenced their V $\beta$ 14 and paired V $\alpha$  chains (Extended Data Fig. 4a). Notably, each mouse carried some V $\beta$ 14 sequences that were present in several sorted cells, and these sequences strongly biased towards T<sub>H</sub>17 or non-T<sub>H</sub>17 cells (Extended Data Fig. 4b), corroborating our findings from the high-throughput sequencing

analysis. To define the antigen specificity of the TCRs from intestinal T<sub>H</sub>17 and non-T<sub>H</sub>17 cell clones, we expressed a cohort of nineteen predominant clonotypic TCRs (ten T<sub>H</sub>17 clones, eight non-T<sub>H</sub>17 clones, and one neutral clone) in a NFAT-GFP<sup>+</sup> hybridoma that can report on TCR signalling<sup>17</sup>. Upon co-culture of the hybridomas with splenic APCs and heat-inactivated mouse intestinal luminal content, several T<sub>H</sub>17-TCR hybridomas, but not non-T<sub>H</sub>17-TCR hybridomas, responded to Taconic antigens, and not to Jackson antigens (Extended Data Fig. 4c). Furthermore, when SFB-mono antigens were used, we detected responses from 7/10 T<sub>H</sub>17-TCR and the neutral TCR hybridoma, but none of the non-T<sub>H</sub>17 cell hybridomas (Fig. 2a). These responses were abrogated if the APCs were from MHCII-deficient mice (Extended Data Fig. 4d).

We next sought to identify epitopes recognized by T<sub>H</sub>17 cell TCRs using a whole-genome shotgun cloning and expression screen, an unbiased approach previously used to identify T-cell antigens from other bacteria<sup>18</sup> (Extended Data Fig. 5a). One bacterial clone, designated 3F12-E8, stimulated 7B8 and four other T<sub>H</sub>17-TCR hybridomas (Extended Data Fig. 5b–d). Based on the recent annotation of the SFB genome<sup>19,20</sup>, we assigned the 672-base pair 3F12-E8 insert to an SFB gene (SFBNYU\_003340<sup>19</sup>). We confirmed the specificity by cloning the full-length gene and demonstrating that its product stimulated the aforementioned five TCRs, but not any other TCRs (Fig. 2b, left). We further mapped a minimal epitope that stimulated all five TCRs and a shorter 8-amino-acid epitope that stimulated only the 7B8 and 2A6 hybridomas (Extended Data Fig. 5e).

Another expression screen was performed using the 1A7 hybridoma, which along with three other TCRs formed a distinct cluster with an identical V $\alpha$  and highly similar V $\beta$ 14 CDR3 sequences (Extended Data Fig. 6a). A stimulatory clone, designated 2D10-A10 (Extended Data Fig. 6b, c), contained the amino-terminal sequence of another SFB gene (SFBNYU\_004990<sup>19</sup>). We mapped the epitope for the 1A7 hybridoma to 9 amino acids (Extended Data Fig. 6d). Both the full-length gene product and a 9-amino-acid peptide stimulated all four TCRs, indicating that

these TCRs indeed recognize the same epitope (Fig. 2b, right). However, the single TCR derived from non-T<sub>H</sub>17 cells (3F4) displayed a much weaker dose-response to peptide antigen than the other TCRs (Extended Data Fig. 6e).

Thus, eight out of eleven V $\beta$ 14 $^+$  T<sub>H</sub>17-TCR hybridomas recognized two distinct antigens encoded by SFB (Fig. 2c). Both proteins are unique to SFB, expressed at a medium to high level, and predicted to be secreted or at the cell surface (Fig. 2d). Importantly, primary V $\beta$ 14 $^+$  T<sub>H</sub>17 cells responded to the two immunodominant SFB epitopes (Extended Data Fig. 7a). Although V $\beta$ 14 $^+$  cells consistently responded slightly better, V $\beta$ 14 $^-$  T<sub>H</sub>17 cells were also stimulated by SFB (Extended Data Fig. 7b), indicating that these cells respond to other SFB epitopes. An *in silico* search was conducted for potential epitopes within the SFB proteome (Extended Data Fig. 7c, d), which yielded several more stimulatory peptides (Extended Data Fig. 7e). Among these, peptide N5, also derived from *SFBNYU\_003340*, was a strong stimulator of intestinal T<sub>H</sub>17 cells, activating both V $\beta$ 14 $^+$  cells and V $\beta$ 14 $^-$  cells (Extended Data Fig. 7f). Thus, in the small intestine, SFB is the dominant antigen source for poly-clonal T<sub>H</sub>17 cells, but for few, if any, non-T<sub>H</sub>17 cells.

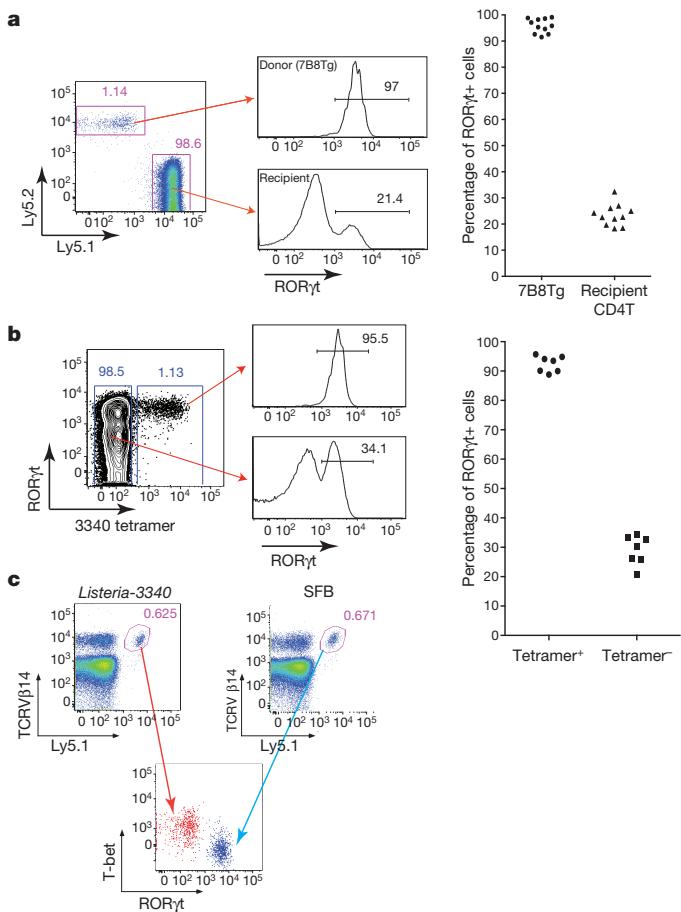
We then asked what happens to T cells expressing SFB-specific TCRs. We generated 7B8 TCR transgenic mice (7B8Tg)<sup>21</sup>, and transferred naive T cells from these mice into isotype-marked congenic B6 mice<sup>22</sup>. After one week, we readily detected donor-derived T cells in the SILP of mice that had been exposed to SFB, whereas they were completely absent in SFB-deficient recipients (Extended Data Fig. 8a). Remarkably, almost all donor-derived cells became positive for ROR $\gamma$ t (Fig. 3a). Similar results were obtained upon transfer of T cells from two other TCR transgenic strains (1A2Tg and 5A11Tg) into SFB-colonized recipient mice (Extended Data Fig. 8b). The donor-derived T cells lacked expression of the transcription factors associated with alternative CD4 T-cell programs (for example, FOXP3, GATA3, and T-bet) (Extended Data Fig. 8c).

To visualize endogenous SFB-antigen-specific T cells, we produced MHCII-tetramers containing peptide A6 from *SFBNYU\_003340* (3340-A6 tetramer)<sup>23</sup>. The I-A $b$ /3340-A6 tetramer specifically stained GFP $^+$  SILP CD4 $^+$  T cells from SFB-colonized *Il23r<sup>GFP/+</sup>* mice (Extended Data Fig. 8d). Furthermore, a sizable population of I-A $b$ /3340-A6 tetramer-positive cells was present in B6 Taconic, but not in B6 Jackson mice (Extended Data Fig. 8e), and these cells were uniformly ROR $\gamma$ t-positive, indicating that they were SFB-elicited T<sub>H</sub>17 cells (Fig. 3b).

We next aimed to determine whether polarization of the antigen-specific T<sub>H</sub>17 cells in response to SFB colonization is dictated by the nature of the antigenic protein or properties of the microbe. *Listeria monocytogenes*, an enteric pathogenic bacterium that also colonizes the small intestine, typically elicits a T<sub>H</sub>1 response<sup>24</sup>. Mice were orally infected with *L. monocytogenes* expressing *SFBNYU-003340* (*Listeria-3340*) (Extended Data Fig. 8f) or SFB before intravenous transfer of 7B8Tg T cells. 7B8Tg T cells accumulated in the SILP of both sets of mice, but, importantly, they expressed T-bet rather than ROR $\gamma$ t when the hosts were colonized with *Listeria-3340* (Fig. 3c).

To further investigate a relationship between the fate of SILP T helper cells and the bacterial origins of antigens, we transferred 7B8Tg T cells into mice that were colonized with both SFB and *Listeria* and simultaneously tracked CD4 $^+$  T-cell responses specific for both bacteria in the SILP using the Ly5.1 $^+$  congenic marker for 7B8Tg cells and listeriolysin O (LLO)-tetramers that stain endogenous *Listeria*-specific T cells derived from the host (Extended Data Fig. 9a). In the presence of both T<sub>H</sub>17- and T<sub>H</sub>1-inducing bacteria, 7B8Tg T cells expressed ROR $\gamma$ t, but not T-bet, whereas LLO-tetramer $^+$  cells expressed T-bet, but not ROR $\gamma$ t (Fig. 4a and Extended Data Fig. 9b, c). This result is in contrast to the T<sub>H</sub>1 polarization of TCR transgenic T cells specific for the commensal CBir1 flagellin antigen observed upon infection with the protozoan parasite *Toxoplasma gondii*<sup>25</sup>, a T<sub>H</sub>1-inducing intestinal pathogen. This suggests that, unlike CBir1-encoding *Clostridia*, SFB has the ability to direct a dominant signal specialized for induction of T<sub>H</sub>17 cells.

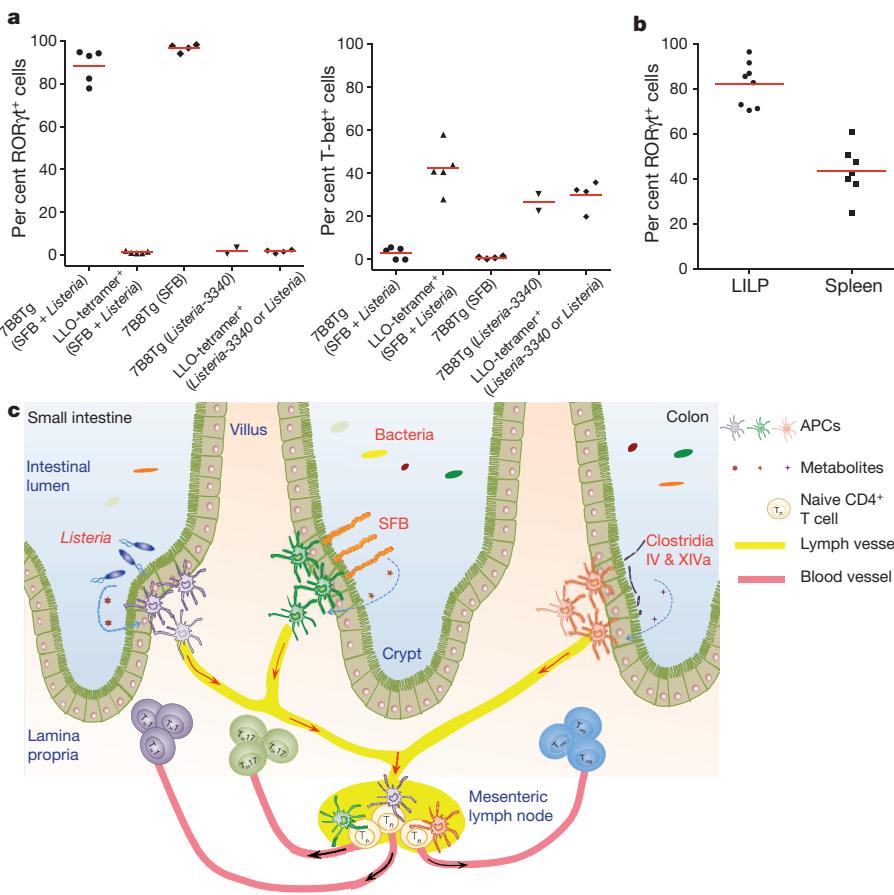
SFB colonization of the small intestine is potentially beneficial, attenuating pathogenic bacteria-induced colitis<sup>8</sup>, but it can also trigger or



**Figure 3 | SFB-specific T cells become T<sub>H</sub>17 cells in the SILP.** **a**, 7B8Tg cells (Ly5.2) were transferred into SFB-colonized mice (Ly5.1), and SILP T cells were analysed after 8–15 days. Left, representative FACS plots. Right, analysis of multiple animals (one symbol per animal). **b**, I-A $b$ /3340-A6 tetramer stain of SILP T cells from SFB-colonized B6 mice. Left, representative FACS plots. Right, analysis of multiple animals (one symbol per animal). **c**, 7B8Tg cells (Ly5.1) were transferred into Ly5.2 congenic hosts orally colonized with *Listeria-3340* or SFB. Seven days after transfer, donor-derived cells in the SILP were analysed. Analyses of cells from different mice are shown in the composite FACS plot. The results are representative of three experiments.

exacerbate systemic autoimmune disease<sup>10,11</sup>, raising the question as to whether SFB-specific T<sub>H</sub>17 cells can circulate beyond the small intestine. We examined the colons and spleens of SFB-positive recipients of 7B8Tg naive T cells, and found these cells in both organs. Importantly, more than 80% of these SFB-specific T cells in colon and 40% in spleen expressed ROR $\gamma$ t (Fig. 4b). Consistent with this result, staining of endogenous T cells from Taconic mice revealed 3340-A6 tetramer-positive cells in the large intestine and most of these cells expressed ROR $\gamma$ t (Extended Data Fig. 10a, b).

Our results therefore indicate that intestinal antigen-specific CD4 $^+$  T cells differentiate to become either T<sub>H</sub>1 or T<sub>H</sub>17 cells, depending on which luminal bacterium delivers the antigen. We propose a deterministic model for T helper cell differentiation whereby the bacterial context of cognate antigen delivery dictates the fate of the antigen-specific T cells (Fig. 4c). Our work opens the way towards elucidating the mechanisms of T<sub>H</sub>17 cell induction by microbiota and of how gut-induced T<sub>H</sub>17 cells can contribute to distal organ-specific autoimmune disease. In addition, it serves as a guide for future studies of human commensal-specific pro-inflammatory T cells that are thought to contribute to autoimmune diseases such as rheumatoid arthritis<sup>26</sup>. Finally, the demonstration of controlled polarized T-cell responses towards commensal bacteria offers the potential for novel approaches towards mucosal vaccination.



**Figure 4 | TCR specificity for distinct luminal bacteria underlies divergent T helper cell differentiation in the SILP.** **a**,  $T_{H17}$  ( $ROR\gamma t$ ) versus  $T_{H1}$  (T-bet) differentiation of SFB- (7B8Tg) and *Listeria* (LLO-tetramer)-specific CD4<sup>+</sup> T cells in mice colonized with either or both bacteria. Each symbol represents cells from one animal.

**b**, Proportions of donor-derived 7B8Tg T cells that express  $ROR\gamma t$  in the colon and spleen of SFB-colonized mice. LILP, large intestine lamina propria. **c**, Model for intestinal niches that promote diverse microbiota-dependent CD4<sup>+</sup> effector T-cell programs. Microbial signals may induce polarizing cytokines or preformed niche-specific antigen-presenting cells may interact with different T-cell-inducing bacteria.

## METHODS SUMMARY

**Mice.** All mice were housed in the animal facility of The Skirball Institute of Biomolecular Medicine at the New York University School of Medicine. Experimental protocols were approved by the Institutional Animal Care and Use Committee. C57BL/6 mice were purchased from Taconic Farm (B6 Taconic) or the Jackson Laboratory (B6 Jackson). *Il23r<sup>GFP</sup>* mice<sup>16</sup>, a gift from M. Oukka, were maintained by breeding with B6 Taconic mice. 7B8Tg, 1A2Tg and 5A11Tg SFB-specific TCR transgenic (Tg) mice were generated as previously described<sup>21</sup> and kept with SFB-minus flora. For adoptive transfer, naive Tg T cells (CD62L<sup>hi</sup> CD44<sup>lo</sup> V $\beta$ 14<sup>+</sup> CD4<sup>+</sup> CD3<sup>+</sup>) were sorted from the spleen and were injected intravenously into congenic recipient mice<sup>22</sup>.

**Generation of TCR hybridomas.** Retroviruses carrying an expression cassette encoding TCR $\alpha$ , TCR $\beta$ , and CD4 were used to infect the NFAT-GFP 58 $\alpha$ - $\beta$ -hybridoma cell line<sup>17</sup>.

**Construction and screen of whole-genome shotgun library of SFB.** The shotgun library was prepared with a procedure modified from a previous study<sup>18</sup>. The library is estimated to contain  $10^4$  clones. The expression of exogenous proteins was induced by isopropylthiogalactoside for 4 h. For antigen screening, pools of heat-killed bacteria ( $\sim 30$  clones per pool) were added to a co-culture of APCs and hybridomas.

**MHCII tetramer production and staining.** MHCII/3340-A6 tetramer was produced as previously described<sup>23</sup>. SILP T cells were incubated at room temperature for 60 min with fluorochrome-labelled tetramer (10 nM) before staining with relevant antibodies at 4 °C.

**Heterologous expression of SFBNYU\_003340 in *Listeria monocytogenes*.** The entire coding region of SFBNYU\_003340, including its predicted signal sequence, was sub-cloned into the *Listeria* expression vector pIMK2<sup>27</sup>. The resultant plasmid was transformed into electrocompetent *Listeria monocytogenes* strain 10403S-*inLA*<sup>m</sup> and plated on selective medium containing kanamycin (50 µg ml<sup>-1</sup>)<sup>28</sup>.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 December 2013; accepted 25 March 2014.

Published online 13 April 2014.

- Bettelli, E., Korn, T., Oukka, M. & Kuchroo, V. K. Induction and effector functions of  $T_{H17}$  cells. *Nature* **453**, 1051–1057 (2008).

- McGeachy, M. J. & Cua, D. J. The link between IL-23 and  $T_{H17}$  cell-mediated immune pathologies. *Semin. Immunol.* **19**, 372–376 (2007).
- Littman, D. R. & Rudensky, A. Y.  $T_{H17}$  and regulatory T cells in mediating and restraining inflammation. *Cell* **140**, 845–858 (2010).
- Ivanov, I. I. et al. The orphan nuclear receptor  $ROR\gamma t$  directs the differentiation program of proinflammatory IL-17<sup>+</sup> T helper cells. *Cell* **126**, 1121–1133 (2006).
- Ivanov, I. I. et al. Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4**, 337–349 (2008).
- Atarashi, K. et al. ATP drives lamina propria  $T_{H17}$  cell differentiation. *Nature* **455**, 808–812 (2008).
- Atarashi, K. et al.  $T_{reg}$  induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232–236 (2013).
- Ivanov, I. I. et al. Induction of intestinal  $T_{H17}$  cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
- Gaboriau-Routhiau, V. et al. The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677–689 (2009).
- Wu, H. J. et al. Gut-residing segmented filamentous bacteria drive autoimmune arthritis via  $T_{H17}$  cells. *Immunity* **32**, 815–827 (2010).
- Lee, Y. K., Menezes, J. S., Umesaki, Y. & Mazmanian, S. K. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl. Acad. Sci. USA* **108** (Suppl 1), 4615–4622 (2011).
- Ivanov, I. I. & Honda, K. Intestinal commensal microbes as immune modulators. *Cell Host Microbe* **12**, 496–508 (2012).
- Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).
- Schnupf, P., Gaboriau-Routhiau, V. & Cerf-Bensussan, N. Host interactions with Segmented Filamentous Bacteria: An unusual trade-off that drives the postnatal maturation of the gut immune system. *Semin. Immunol.* **25**, 342–351 (2013).
- Lochner, M. et al. Restricted microbiota and absence of cognate TCR antigen leads to an unbalanced generation of  $T_{H17}$  cells. *J. Immunol.* **186**, 1531–1537 (2011).
- Awasthi, A. et al. Cutting edge: IL-23 receptor GFP reporter mice reveal distinct populations of IL-17-producing cells. *J. Immunol.* **182**, 5904–5908 (2009).
- Ise, W. et al. CTLA-4 suppresses the pathogenicity of self antigen-specific T cells by cell-intrinsic and cell-extrinsic mechanisms. *Nature Immunol.* **11**, 129–135 (2010).

18. Sanderson, S., Campbell, D. J. & Shastri, N. Identification of a CD4<sup>+</sup> T cell-stimulating antigen of pathogenic bacteria by expression cloning. *J. Exp. Med.* **182**, 1751–1757 (1995).
19. Sczesnak, A. et al. The genome of Th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe* **10**, 260–272 (2011).
20. Prakash, T. et al. Complete genome sequences of rat and mouse segmented filamentous bacteria, a potent inducer of Th17 cell differentiation. *Cell Host Microbe* **10**, 273–284 (2011).
21. Kouskoff, V., Signorelli, K., Benoist, C. & Mathis, D. Cassette vectors directing expression of T cell receptor genes in transgenic mice. *J. Immunol. Methods* **180**, 273–280 (1995).
22. Kearney, E. R., Pape, K. A., Loh, D. Y. & Jenkins, M. K. Visualization of peptide-specific T cell immunity and peripheral tolerance induction *in vivo*. *Immunity* **1**, 327–339 (1994).
23. Moon, J. J. et al. Naive CD4<sup>+</sup> T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).
24. Hsieh, C. S. et al. Development of TH1 CD4<sup>+</sup> T cells through IL-12 produced by Listeria-induced macrophages. *Science* **260**, 547–549 (1993).
25. Hand, T. W. et al. Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. *Science* **337**, 1553–1556 (2012).
26. Scher, J. U. et al. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2**, e01202 (2013).
27. Monk, I. R., Gahan, C. G. & Hill, C. Tools for functional postgenomic analysis of *Listeria monocytogenes*. *Appl. Environ. Microbiol.* **74**, 3921–3934 (2008).
28. Xayarath, B., Marquis, H., Port, G. C. & Freitag, N. E. *Listeria monocytogenes* CtaP is a multifunctional cysteine transport-associated protein required for bacterial pathogenesis. *Mol. Microbiol.* **74**, 956–973 (2009).

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

**Acknowledgements** We thank S. Yong Kim for generating TCR transgenic mice, A. Viale for 454 pyrosequencing, R. Myers for RNA-seq, N. Freitag for providing the *Listeria* strain and expression vector, Y. Umesaki for SFB samples, and K. Murphy for providing the 58 $\alpha$  5 $\beta$  hybridoma line. The Immune Monitoring Core New York University (NYU) is supported in part by grant UL1 TR00038 from the National Center for Advancing Translational Sciences and grant 5P30CA016087-32 from the National Cancer Institute; the NYU Histology Core is supported in part by grant 5P30CA016087-32 from the National Cancer Institute. M.K.J. was supported by grant R01 AI039614 from the NIH. Y.Y. was supported by the Arthritis National Research Foundation. M.X. is supported by the Irvington Institute fellowship program of the Cancer Research Institute. D.R.L. is a Howard Hughes Medical Institute Investigator.

**Author Contributions** Y.Y. and D.R.L. designed the experiments and wrote the manuscript with input from the co-authors. Y.Y., M.B.T., M.X., C.N., A.C., X.L. and J.-J.L. performed most analyses. M.B.T. constructed TCR hybridomas. M.X. developed SFB-specific antibodies. M.G., H.X. and J.J.L. did TCR pyrosequencing analysis. J.L.L. and M.K.J. developed tetramers. F.A. and V.J.T. generated transgenic *Listeria*. A.S. performed RNA-seq analysis of SFB.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R.L. ([dan.littman@med.nyu.edu](mailto:dan.littman@med.nyu.edu)).

## METHODS

**Mice.** C57BL/6 mice were purchased from Taconic Farm (B6 Taconic) or Jackson Laboratory (B6 Jackson). *Il23r*<sup>GFP</sup> mice<sup>16</sup> were provided by M. Oukka and maintained by breeding with B6 Taconic mice. Ly5.1 mice (*B6.SJL-Ptprca Pepcb/BoyJ*) and MHCII-deficient mice (*B6.129S2-H2<sup>d<sup>l</sup>ab1-Ea</sup>/J*) were from Jackson Laboratory.

**Antibodies and flow cytometry.** The following antibodies were from eBiosciences, BD Pharmingen or BioLegend: V $\beta$ 2 (B20.6), V $\beta$ 3 (KJ25), V $\beta$ 4 (KT4), V $\beta$ 5 (MR9-4), V $\beta$ 6 (RR4-7), V $\beta$ 7 (TR310), V $\beta$ 8 (F23.1), V $\beta$ 8.1/8.2 (MR5-2), V $\beta$ 8.3 (8C1), V $\beta$ 10 (B21.5), V $\beta$ 11 (CTVB11), V $\beta$ 12 (MR11-1), V $\beta$ 14 (14-2), CD3 (145-2C11), CD4 (RM4-5), CD25 (PC61), Ly5.1 (A20), Ly5.2 (104), MHCII (M5/114), ROR $\gamma$ T (AFKJS-9 or B2D), FOXP3 (FJK-16s), T-bet (eBio4B10), GATA3 (TWAJ), IL-17A (eBio17B7) and IFN- $\gamma$  (XM61.2). Flow cytometric analysis was performed on an LSR II (BD Biosciences) or an Aria II (BD Biosciences) and analysed using FlowJo software (Tree Star). DAPI (Sigma) was used to exclude dead cells.

**T-cell preparation and staining.** Small intestine lamina propria were minced and then incubated for 30 min at 37 °C with collagenase D (1 mg ml<sup>-1</sup>; Roche), dispase (0.05 U ml<sup>-1</sup>; Worthington) and DNase I (100 µg ml<sup>-1</sup>; Sigma). Lymphocytes were collected at the interface of a 40%/80% Percoll gradient (GE Healthcare). Cells were stained for surface markers, followed by fixation and permeabilization (eBioscience).

**Calculating enrichment scores.** An enrichment score for a given V $\beta$  in IL-23R (GFP)<sup>+</sup> cells is defined as the equation of (per cent of V $\beta$ <sup>+</sup> cells in the GFP-positive fraction)/(per cent of V $\beta$ <sup>+</sup> cells in the GFP-negative fraction). For example, for Extended Fig. 2b, V $\beta$ 14 enrichment was calculated as (7.45/(7.45 + 26.2))/(4.48/(4.48 + 61.8)) or 3.3. A score > 1 means a positive enrichment and a score ≈ 1 means no enrichment.

**High-throughput TCR sequencing.** The SILP cells from *Il23r*<sup>GFP/+</sup> mice were stained for surface markers and V $\beta$ 14<sup>+</sup> CD4<sup>+</sup> T cells were sorted on the Aria II. For each sample, we collected about 2 × 10<sup>4</sup> cells (2.17 ± 0.43 × 10<sup>4</sup> cells for GFP<sup>+</sup> T<sub>H</sub>17 cells and 2.38 ± 0.54 × 10<sup>4</sup> cells for GFP<sup>-</sup> non-T<sub>H</sub>17 cells). Cells were lysed in TRIzol reagent (Invitrogen) and RNA was extracted following the manufacturer's instruction. RNA precipitation was aided with GlycoBlue (Invitrogen). Complementary DNAs were prepared with a reverse transcription kit (USB). V $\beta$ 14 PCRs were performed using barcoded oligonucleotides. PCR products from 16 samples were quantified on NanoDrop. Equal amounts of barcoded PCR product were mixed and sequenced using a 454 GS Junior system (Roche). The raw sequencing data was first aligned using the high-throughput analysis tool provided by IMGT<sup>29</sup>. We obtained 6,647 ± 954 reads for T<sub>H</sub>17 cells and 5,573 ± 889 reads for non-T<sub>H</sub>17 cells. CDR3 usage was further computed with Perl-based scripts developed in-house. The T<sub>H</sub>17 samples had 340–772 unique V $\beta$ 14 CDR3 sequences and the non-T<sub>H</sub>17 samples had 849–2,148 unique V $\beta$ 14 CDR3 sequences.

**Single-cell TCR sequencing.** The SILP cells from *Il23r*<sup>GFP/+</sup> mice were stained for surface markers. GFP<sup>+</sup> and GFP<sup>-</sup> V $\beta$ 14<sup>+</sup> CD4<sup>+</sup> T cells were sorted on the BD Aria II and deposited at one cell per well into 96-well PCR plates preloaded with 5 µl reverse transcription mix (USB). Immediately after sorting, whole plates were incubated at 50 °C for 60 min for cDNA synthesis. Half of the cDNA was used for V $\beta$ 14 PCR using forward primer 5'-ACGACCAATTCTCATCTAACGAC-3' and reverse primer 5'-AAGCACACGAGGGTAGCCT-3'. To retrieve V $\alpha$  sequences, the other half of cDNA was pre-amplified for 16 cycles using a mix of twenty-one forward primers<sup>30</sup> (each modified by adding a 5' extended anchor sequence: TAATAGCACTACTATAGGG) and a reverse primer 5'-CATGTCCAGCACA GTTTTGTCAGT-3'. The primary V $\alpha$  PCR products were diluted and subjected to a second round PCR using forward primer 5'-TAATAGCACTACTATAG GG-3' and reverse primer 5'-GTCAAAGTCGGTGAAACAGGC-3'. PCRs were performed in a LightCycler 480 (Roche). PCR products were cleaned up with ExoSap-IT reagent (USB) and Sanger sequencing was performed by Macrogen. In nearly all cases, for cells with the same V $\beta$ 14 sequence, we retrieved a single unique V $\alpha$  sequence, indicating that these cells were clonotypically identical.

**Generation of TCR hybridomas.** The NFAT-GFP 58 $\alpha$ - $\beta$ -hybridoma cell line<sup>17</sup> was provided by K. Murphy. To reconstitute TCRs, we used a self-cleavage sequence of 2A to link cDNAs of TCR $\alpha$  and TCR $\beta$  generated from annealing of overlapping oligonucleotides (TCR $\alpha$ -p2A-TCR $\beta$ ) and shuttled the cassette into a modified MigR1 retrovector in which IRES-GFP was replaced with IRES-mCD4. Then retroviral vectors were transfected into Phoenix E packaging cells using Lipofectamine 2000 (Invitrogen). Hybridoma cells were transduced with viral supernatants in the presence of polybrene (8 µg ml<sup>-1</sup>) by spin infection for 90 min at 32 °C. Transduction efficiencies were monitored by checking mCD3 surface expression on day 2. We generated nineteen hybridomas for predominant clonotypic TCRs whose V $\alpha$  and V $\beta$  sequences were retrieved from single-cell TCR sequencing (no. of V $\beta$  ≥ 6 for T<sub>H</sub>17 biased clones, and no. of V $\beta$  ≥ 4 for non-T<sub>H</sub>17 biased clones, no. of V $\alpha$  ≥ 3. Note that a single unique V $\alpha$  was identified for every V $\beta$ ). We also generated the OTII hybridoma using TCR sequences provided by F. Carbone (chicken ovalbumin antigen-specific, I-A $b$ -restricted).

**Assay for hybridoma activation.** To prepare antigen-presenting cells, splenocytes from B6 mice that were injected intraperitoneally with 8 × 10<sup>6</sup> FLT3-B16 melanoma cells 10 days before were positively enriched for CD11c<sup>+</sup> cells using MACS LS columns (Miltenyi). 10<sup>4</sup> hybridoma cells were incubated with 2 × 10<sup>5</sup> APCs and autoclaved antigens from intestinal luminal contents or faecal material for two days. GFP induction in the hybridomas (CD3<sup>+</sup> fraction) was analysed by flow cytometry.

**Construction and screen of whole-genome shotgun library of SFB.** The shotgun library was prepared with a procedure modified from a previous study<sup>18</sup>. In brief, genomic DNA was purified from the feces of SFB-monoassociated mice by phenol:chloroform extraction. DNA was subjected to whole-genome amplification with the REPLI-g kit (Qiagen) following the manufacturer's instructions. Amplified materials were partially digested with Sau3A (NEB), then ligated with the BamHI-linearized pGEX-4T1 expression vector (GE Healthcare). Ligation products were introduced into competent Stbl3 cells (Invitrogen). To ensure the quality of the library, we sequenced the inserts of randomly picked colonies. All the sequences were mapped to the SFB genome. The library is estimated to contain 10<sup>4</sup> clones. We grew bacteria in 96-well deepwell plates (VWR) with AirPort micro-porous cover (Qiagen). The expression of exogenous proteins was induced by isopropylthiogalactoside for 4 h. Then bacteria were heat-killed by incubating at 70 °C for 1 h, and stored at -20 °C until use. For antigen screens, pools of bacterial clones (~30 clones per pool) were added to a co-culture of APCs and hybridomas. Clones within the positive pools were screened individually against the hybridoma bait. Finally, the inserts of positive clones were subjected to Sanger sequencing. The sequences were blasted against the SFB genome and aligned to annotated open reading frames.

**Epitope mapping.** We expressed overlapping fragments spanning the active ORF using the pGEX-4T1 bacterial expression system, and colonies were used to stimulate the relevant hybridoma. This process was repeated until we identified minimal fragments conferring antigenicity. The mapping was further verified by stimulating hybridomas with synthetic peptides (Genescrypt).

**RNA-seq analysis of the SFB transcriptome.** Wild-type B6 mice from Jackson Laboratory or Taconic Farm, confirmed for the presence or absence of SFB by quantitative PCR<sup>19</sup>, were used for microbiome transcriptome analysis. Within five minutes after sacrifice, the terminal ileum of each mouse was resected and luminal contents were squeezed with sterile forceps into a mortar cooled with liquid nitrogen. 1 ml nuclease-free TE was washed through the ileum into the mortar. The total luminal contents and washing were then ground to a fine powder with a pestle cooled with liquid nitrogen and kept on dry ice. The powder was then transferred to 15 ml TRIzol (Life Technologies) in a 50 ml falcon tube and vortexed. The manufacturer's protocol was then used to obtain RNA. The resulting RNA was extracted twice with acid phenol-chloroform, precipitated, treated with Ambion Turbo DNA-free, and cleaned-up with an RNeasy column to yield RNA with an undetectable concentration of DNA by Qubit. A portion of this RNA was treated once with Epicentre's Ribo-Zero rRNA removal kit, using equal volumes of specific oligonucleotides from the Meta-bacteria and Human/mouse/rat kits. An Illumina RNA-seq library was prepared from these samples using a previously-described strand-specific Nextera protocol. The resulting reads were aligned to the SFBNYU genome with Bowtie<sup>31</sup> and transcript abundance was estimated using Cufflinks<sup>32</sup> with default parameters.

**Production of anti-SFB antibody and immunostaining.** The cDNA fragments corresponding to amino acids 43–359 (3340N) and 734–1060 (3340C) of SFBNYU\_003340 were cloned into the pGEX6p1 expression vector. Recombinant proteins fused to N-terminal glutathione-S-transferase (GST) were expressed in *E. coli* BL21, purified with glutathione Sepharose 4B (GE), and were released with PreScission protease (GE). The flow-through fractions containing polypeptides without the GST tag were collected as immunogen. Rabbit polyclonal antibodies against both polypeptides were raised by Covance. For immunostaining, bacteria were fixed with 2% paraformaldehyde, followed by washing with 0.5% TritonX-100. Bacteria were incubated sequentially with primary antibody (1:1 mix of the two rabbit-anti-3340 antibodies) and phycoerythrin-conjugated goat anti-rabbit antibody.

**Activation of polyclonal SILP T<sub>H</sub>17 cells.** GFP<sup>+</sup> and GFP<sup>-</sup> SILP CD4<sup>+</sup> T cells sorted from *Il23r*<sup>GFP/+</sup> mice were incubated with 2 × 10<sup>5</sup> APCs (CD11c<sup>+</sup> cells purified from the spleen) and indicated stimuli in complete RPMI medium supplemented with IL-2 (10 U ml<sup>-1</sup>) and IL-7 (5 ng ml<sup>-1</sup>) for 2–3 days. Cells were collected and stained with V $\beta$ -specific antibodies. Forward scatter increment, as readout for cell activation, was analysed by FACS.

**IL-17A ELISPOT assay.** IL-17A ELISPOT was performed with a mouse/rat IL-17A ELISPOT Ready-SET-Go! kit (eBioscience). Dots were automatically enumerated with ImmunoSpot software (Version 5.0).

**MHCII tetramer production and staining.** I-A $b$ /3340-A6 tetramer was produced as previously described<sup>23</sup>. Briefly, QFSGAVPNKTD, an immunodominant epitope from SFBNYU\_0033400, covalently linked to I-A $b$  via a flexible linker, was produced in *Drosophila* S2 cells. Soluble pMHCII monomers were purified, biotinylated,

and tetramerized with phycoerythrin- or allophycocyanin-labelled streptavidin. To stain endogenous cells, SILP cells were first resuspended in FACS buffer with FcR block, 2% mouse serum and 2% rat serum. Then tetramer was added (10 nM) and incubated at room temperature for 60 min. Cells were washed and followed by regular staining at 4 °C. I-A<sup>b</sup>/2W and I-A<sup>b</sup>/LLO tetramers were previously described<sup>23,33</sup>.

**Generation of T<sub>H</sub>17-TCRTg mice.** TCR sequences of 7B8, 1A2 and 5A11 were cloned into the pT $\alpha$  and pT $\beta$  vectors kindly provided by D. Mathis<sup>21</sup>. TCR transgenic animals were generated by the Rodent Genetic Engineering Core at the New York University School of Medicine. Positive pups were genotyped by PCR and kept on SFB-minus flora.

**Adoptive transfer.** Spleens from 7B8Tg mice were collected and disassociated. Red blood cells were lysed using ACK lysis buffer (Lonza). Naive Tg T cells (CD62L<sup>hi</sup> CD44<sup>lo</sup> V $\beta$ 14<sup>+</sup> CD4<sup>+</sup> CD3<sup>+</sup>) were sorted on a BD Aria II. Cells were transferred into congenic Ly5.1 recipient mice by retro-orbital injection. In some experiments, we used Ly5.1/Ly5.2 TCRTg mice as donor and transferred naive Tg T cells to congenic Ly5.2 recipient mice.

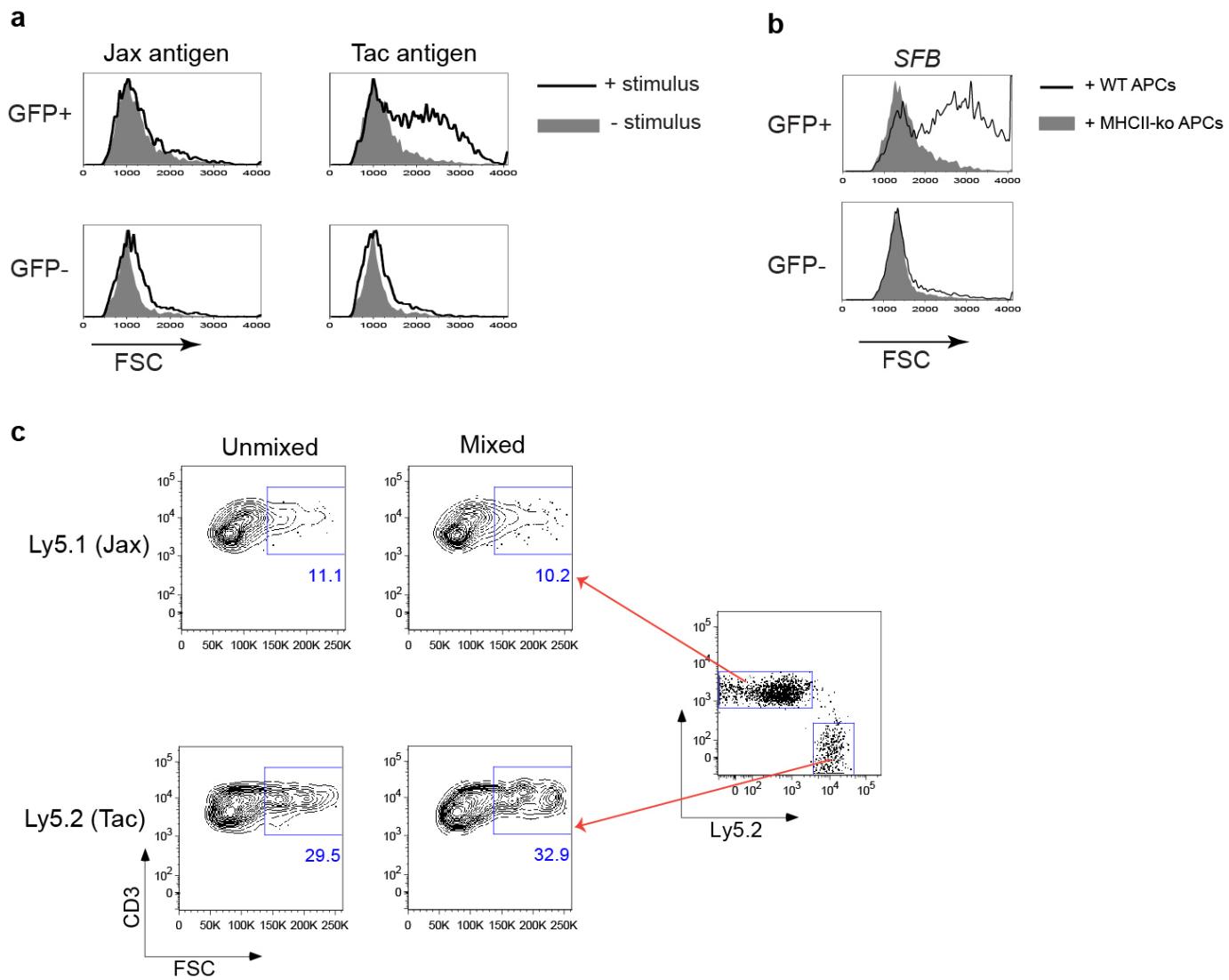
**Heterologous expression of SFBNYU\_003340 in Listeria monocytogenes.** To generate strains of *L. monocytogenes* that express the SFBNYU\_003340 antigen, the entire coding region including its predicted signal sequence was PCR-amplified from a plasmid containing the SFBNYU\_003340 gene. The resultant PCR product was digested and sub-cloned into the *Listeria* expression vector pIMK2 (provided by C. Hill), allowing the gene to be expressed under the synthetic promoter P<sub>help</sub> (High expression promoter in *L. monocytogenes*)<sup>27</sup>. The resultant plasmid designated pIMK2-3340 was transformed into electrocompetent *Listeria monocytogenes* strain 10403S-*inlA*<sup>m</sup> (provided by N. E. Freitag) and plated on selective medium containing kanamycin (50 µg ml<sup>-1</sup>)<sup>28</sup>. pIMK2 is a derivative of the plasmid pPL2 and stably integrates in single copy within the tRNA<sup>Arg</sup> gene following electroporation<sup>34</sup>. The integrity of the SFBNYU\_003340 gene was validated by PCR and expression confirmed by Coomassie staining of *L. monocytogenes* exoproteins.

**Oral infection with SFB and *L. monocytogenes*.** For SFB colonization, we dissolved in sterile PBS fresh faecal pellets collected from *Il23r*<sup>GFP/GFP</sup> *Rag2*<sup>-/-</sup> mice that have highly elevated levels of SFB, and infected mice by oral gavage. For *L. monocytogenes* colonization, we grew *Listeria*-3340 and *Listeria*-empty in brain heart infusion medium and infected mice orally with 10<sup>9</sup> colony forming units.

**Bioinformatic analysis.** Protein predictions were made by bioinformatic tools, including Psort (Version 3.0)<sup>35</sup> and Cello (Version 2.5)<sup>36</sup> for localization prediction, and IEDB (Immune Epitope Database) for MHCII binding affinity prediction.

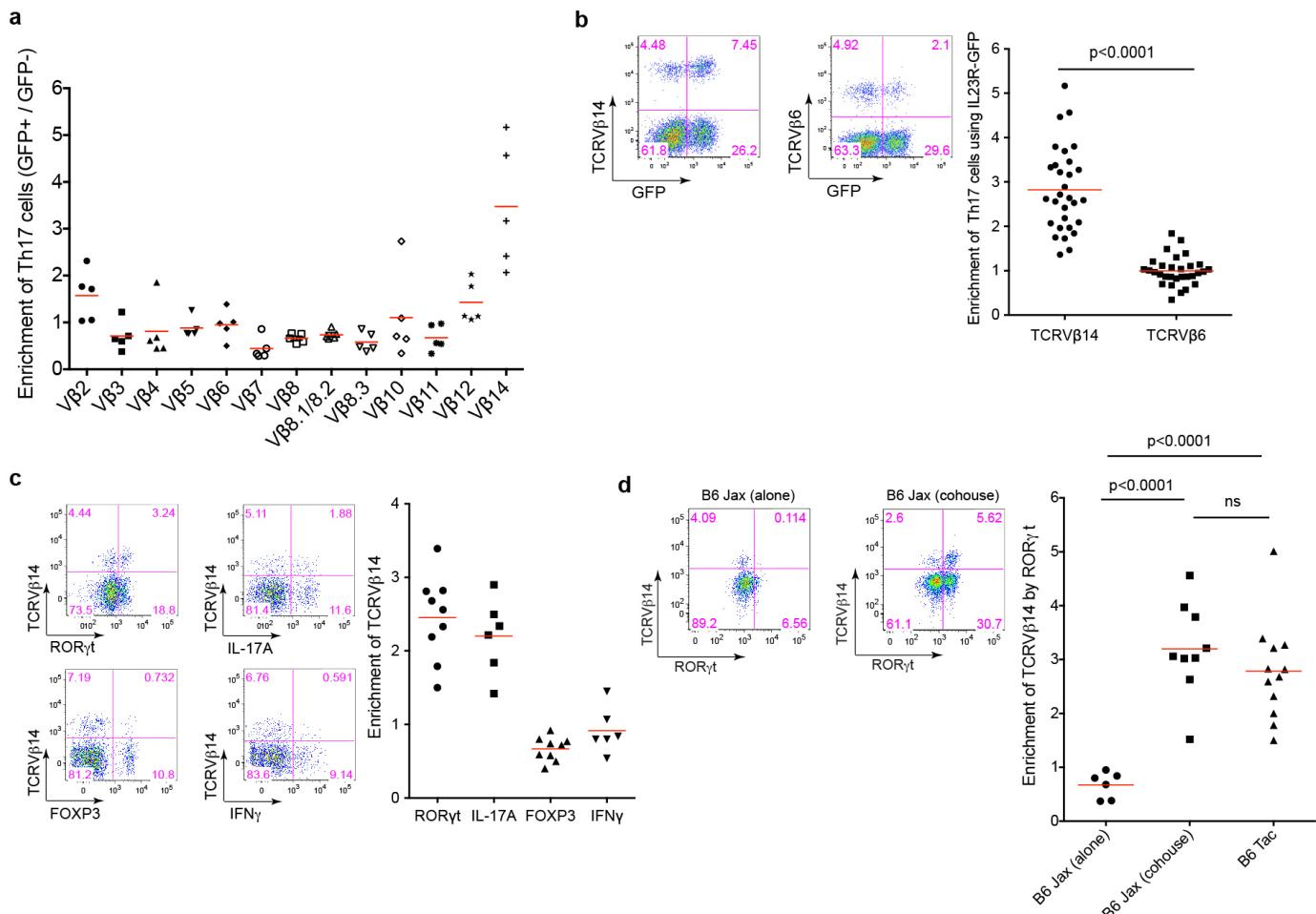
**Statistical analysis.** All analyses were performed using GraphPad Prism (Version 6.0). Differences were considered to be significant at *P* values <0.05.

29. Alamyar, E., Giudicelli, V., Li, S., Duroux, P. & Lefranc, M. P. IMGT/HighV-QUEST: the IMGT(R) web portal for immunoglobulin (Ig) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* **8**, 26 (2012).
30. Currier, J. R. & Robinson, M. A. Spectratype/immunoscope analysis of the expressed TCR repertoire. *Curr. Protocols Immunol.* **Chapter 10**, Unit 10.28 (2001).
31. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
32. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325–2329 (2011).
33. Tubo, N. J. et al. Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. *Cell* **153**, 785–796 (2013).
34. Lauer, P., Chow, M. Y., Loessner, M. J., Porthoy, D. A. & Calendar, R. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J. Bacteriol.* **184**, 4177–4186 (2002).
35. Yu, N. Y. et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615 (2010).
36. Yu, C. S., Chen, Y. C., Lu, C. H. & Hwang, J. K. Prediction of protein subcellular localization. *Proteins* **64**, 643–651 (2006).



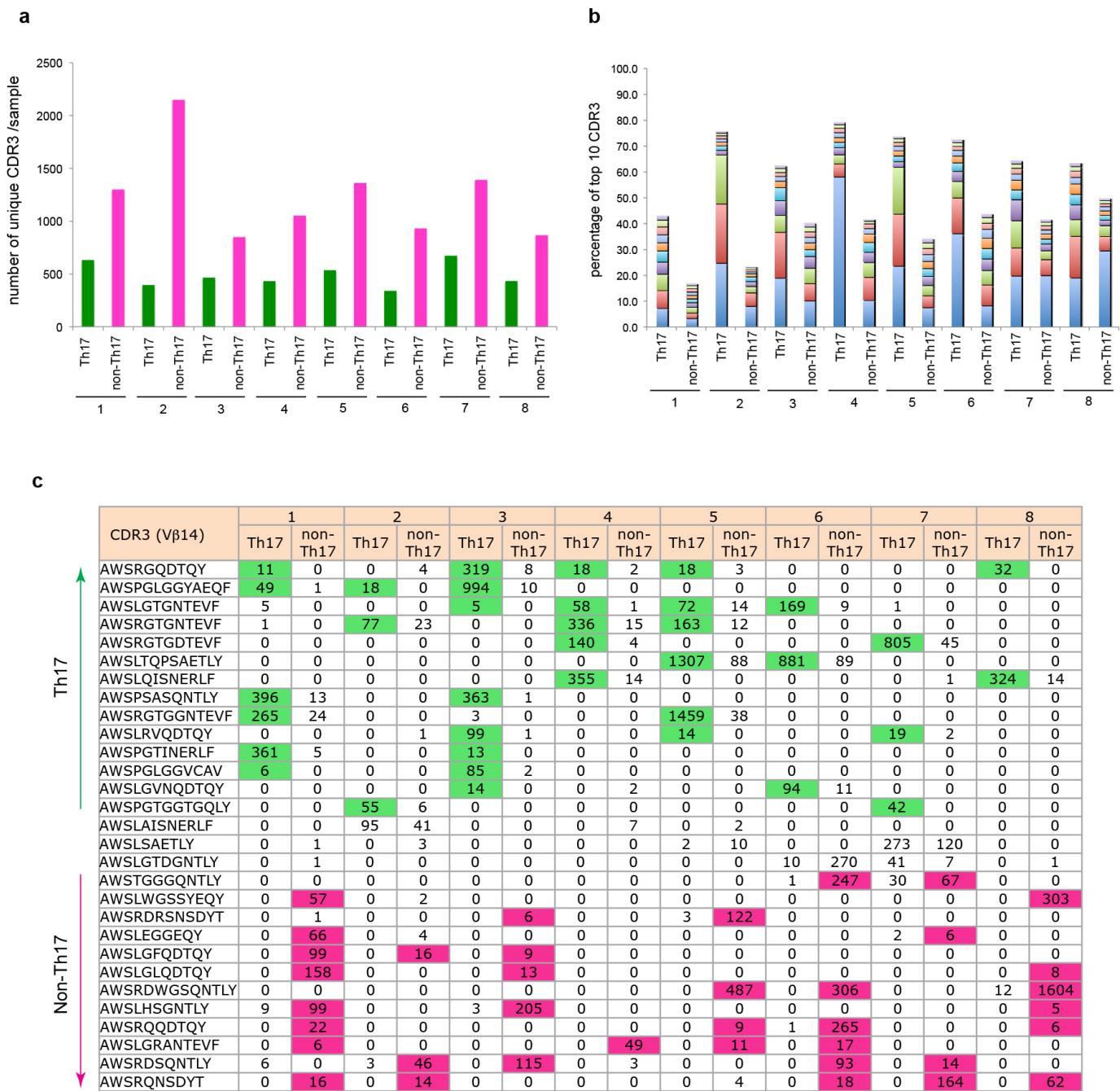
**Extended Data Figure 1 | Stimulation of SILP T<sub>H</sub>17 cells requires intestinal microbiota antigen presentation.** **a**, Intestinal GFP<sup>+</sup> CD4<sup>+</sup> T cells from *I123r*<sup>GFP/+</sup> mice stimulated with faecal material from Jackson and Taconic mice in the presence of syngeneic splenic APCs. Forward scatter was evaluated after 2 days. **b**, T<sub>H</sub>17 cell activation by faecal material from SFB-monoassociated

mice in the presence of APCs sufficient (WT) or deficient (KO) for MHC class II. **c**, Evaluation of potential activation of bystander CD4<sup>+</sup> T cells upon stimulation with SFB antigen. SILP CD4<sup>+</sup> T cells from mice with Jackson flora (Ly5.1) and Taconic flora (Ly5.2) were co-cultured or stimulated separately with APCs and SFB-monoassociated faecal material, and FSC was evaluated.



**Extended Data Figure 2 | Microbiota-dependent TCR usage bias among SILP T<sub>H</sub>17 cells.** **a**, SILP CD4<sup>+</sup> T cells from *Il23r*<sup>GFP/+</sup> mice were analysed for utilization of Vβs in T<sub>H</sub>17 cells versus non-T<sub>H</sub>17 cells. Ratios of the percentage of each TCR Vβ in GFP<sup>+</sup> vs GFP<sup>-</sup> cells are shown. Each symbol represents one mouse. **b**, Relative expression of Vβ14 and Vβ6 TCRs by SILP T<sub>H</sub>17 versus non-T<sub>H</sub>17 CD4<sup>+</sup> T cells from *Il23r*<sup>GFP/+</sup> mice. Left, representative FACS plots. Right, analysis of multiple animals. Each symbol represents one mouse. **c**, Specific enrichment of Vβ14 TCRs in

CD4<sup>+</sup> T cells expressing RORγt and IL-17A, but not FOXP3 or IFNγ. Left, representative FACS plots. Right, analysis of multiple animals. Each symbol represents one mouse. **d**, Correlation of Vβ14 enrichment in T<sub>H</sub>17 cells with the presence of specific commensal microbiota. B6 Jackson mice were housed alone or cohoused with B6 Taconic mice for two weeks. Left, representative FACS analyses. Right, analysis of multiple animals.

**Extended Data Figure 3 | T<sub>H</sub>17 TCR repertoire analysis by pyrosequencing.**

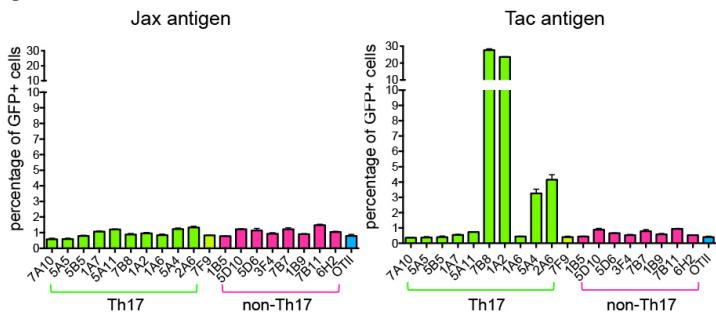
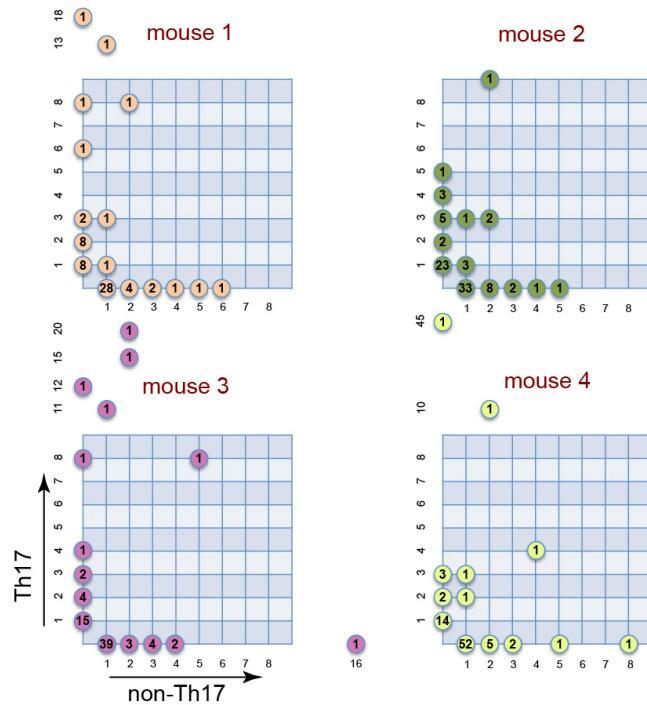
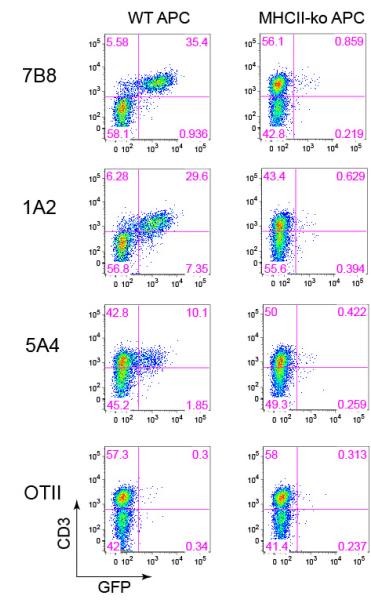
**a**, Numbers of unique V $\beta$ 14 CDR3 sequences of individual SILP T $H$ 17 and non-Th17 samples. The sequences were normalized for numbers of cells and total reads. **b**, Preferential expansion of V $\beta$ 14<sup>+</sup> clones in the T $H$ 17

compartment in the SILP. The proportions of the 10 most abundant V $\beta$ 14 CDR3 sequences from T $H$ 17 and non-Th17 cells from 8 mice are shown.

**c**, T $H$ 17-non Th17 bias of unique V $\beta$ 14 CDR3 sequences in the SILP of multiple mice.

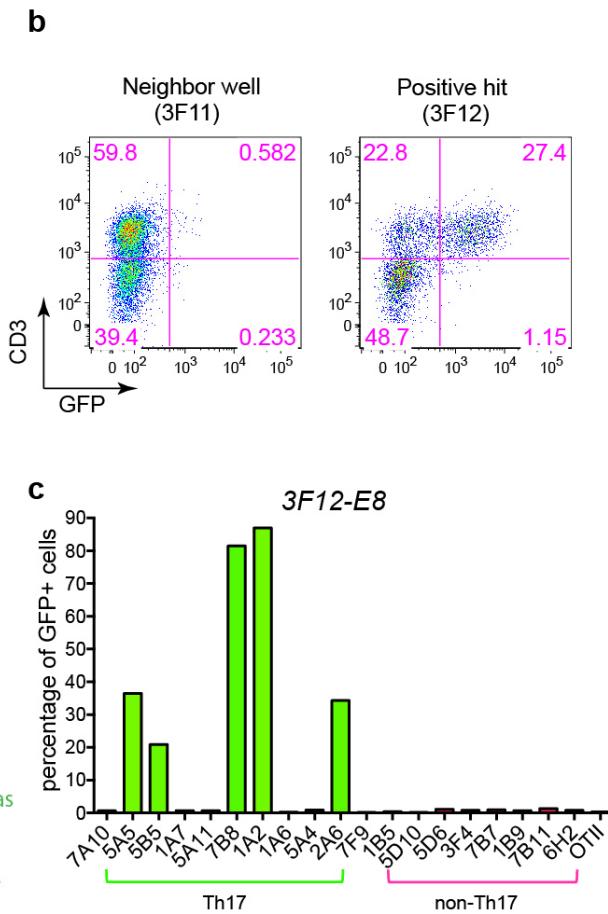
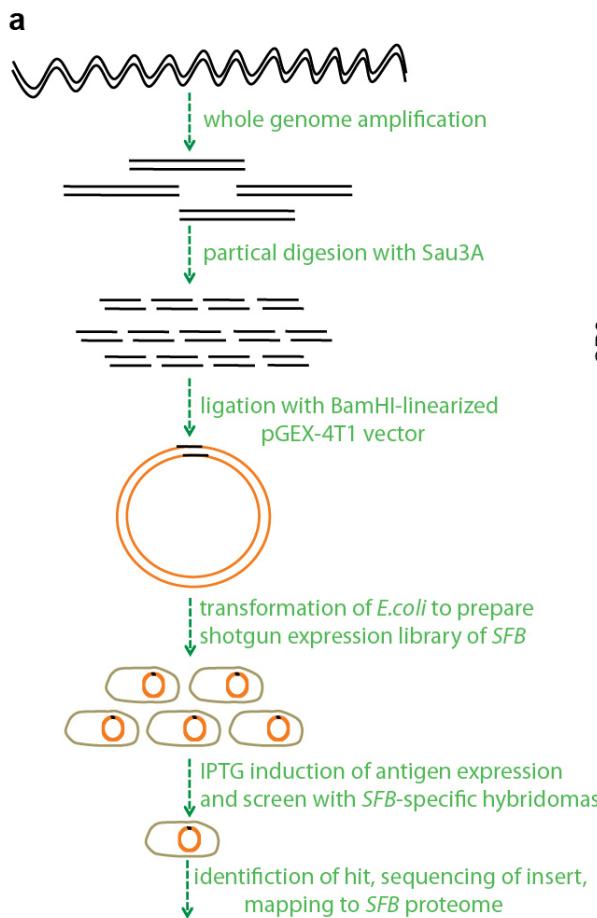
**a**

Mouse	V $\beta$ 14+ Th17 cells			V $\beta$ 14+ non-Th17 cells			# of cloned CDR3	# of unique CDR3
	# of wells	# of cloned CDR3	Cloning efficiency (%)	# of wells	# of cloned CDR3	Cloning efficiency (%)		
1	84	78	92.8	84	61	72.6	139	57
2	88	80	90.9	88	73	82.9	153	85
3	112	107	95.5	112	91	81.2	198	77
4	112	91	81.2	112	89	79.5	180	85

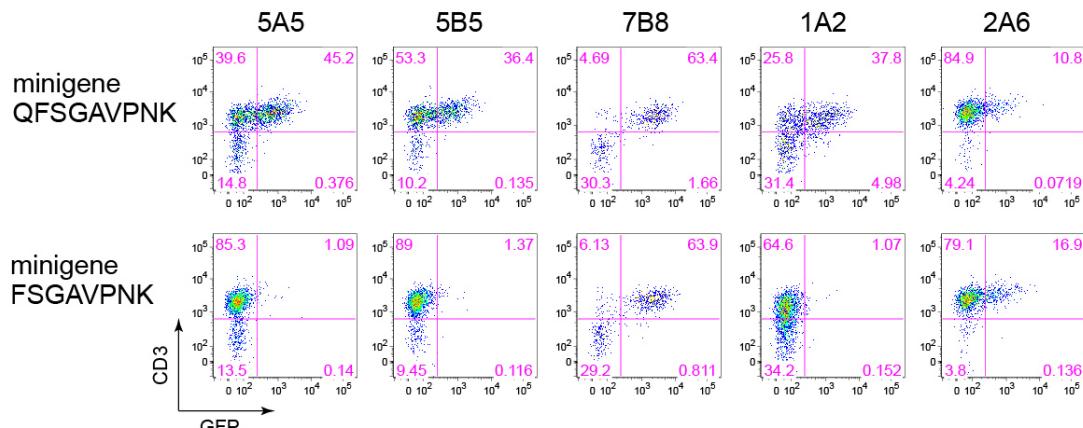
**c****b****d**

**Extended Data Figure 4 | Single-cell TCR cloning and TCR hybridoma screen.** **a**, Efficiency of single-cell V $\beta$ 14 cloning from SILP T<sub>H</sub>17 and non-T<sub>H</sub>17 cells of multiple mice. **b**, Distributions of unique V $\beta$ 14 sequences in T<sub>H</sub>17 and non-T<sub>H</sub>17 cells within the SILP. Each plot represents one mouse shown in **a**, and x axes represent numbers of T<sub>H</sub>17 cells and non-T<sub>H</sub>17 cells for

each unique V $\beta$ 14 sequence. Numbers of unique sequences are shown in coloured circles. **c**, Responses of T<sub>H</sub>17 and non-T<sub>H</sub>17 TCR hybridomas to small intestinal luminal contents from B6 Taconic and B6 Jackson mice. **d**, Stimulation of T<sub>H</sub>17 TCR hybridomas by SFB-monoassociated antigens in the presence of APCs sufficient (WT) or deficient (KO) for MHC class II.

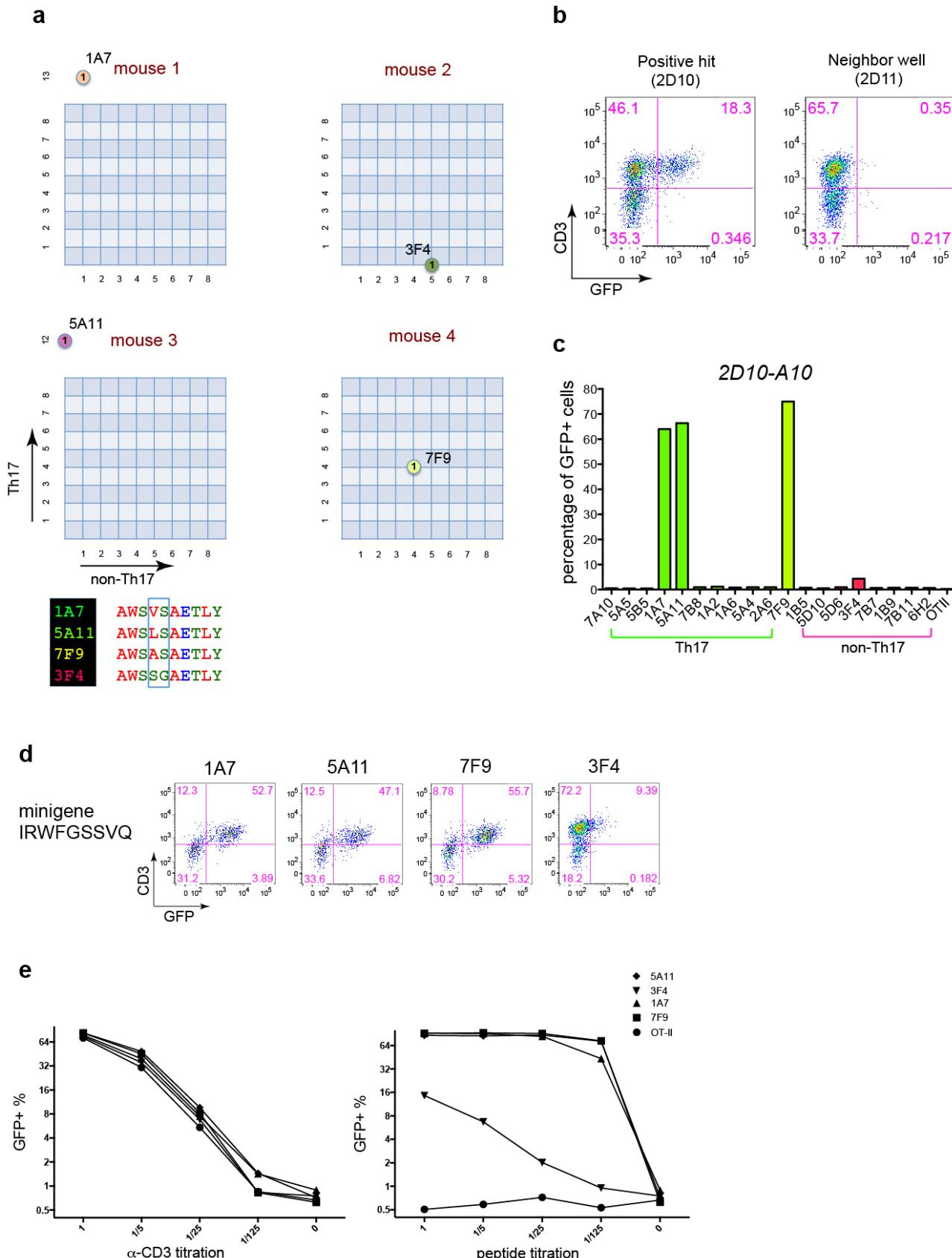
**d**

CLONE	CDR3 (V $\beta$ 14)	V $\alpha$ subset	CDR3 (V $\alpha$ )
5A5	AWSLVNYNSPLY	$\alpha$ 11	AAGGNNNKLT
5B5	AWSLVTGGHERLF	$\alpha$ 5	AVSAGTQVVGQLT
7B8	AWSRSRSPSQNTLY	$\alpha$ 15	AEGNMGYKLT
1A2	AWSQGGREEQY	$\alpha$ 18	ATVFMNYYNQGKLI
2A6	AWSLGTGGASPLY	$\alpha$ 4	ALGENSGTYQR

**e**

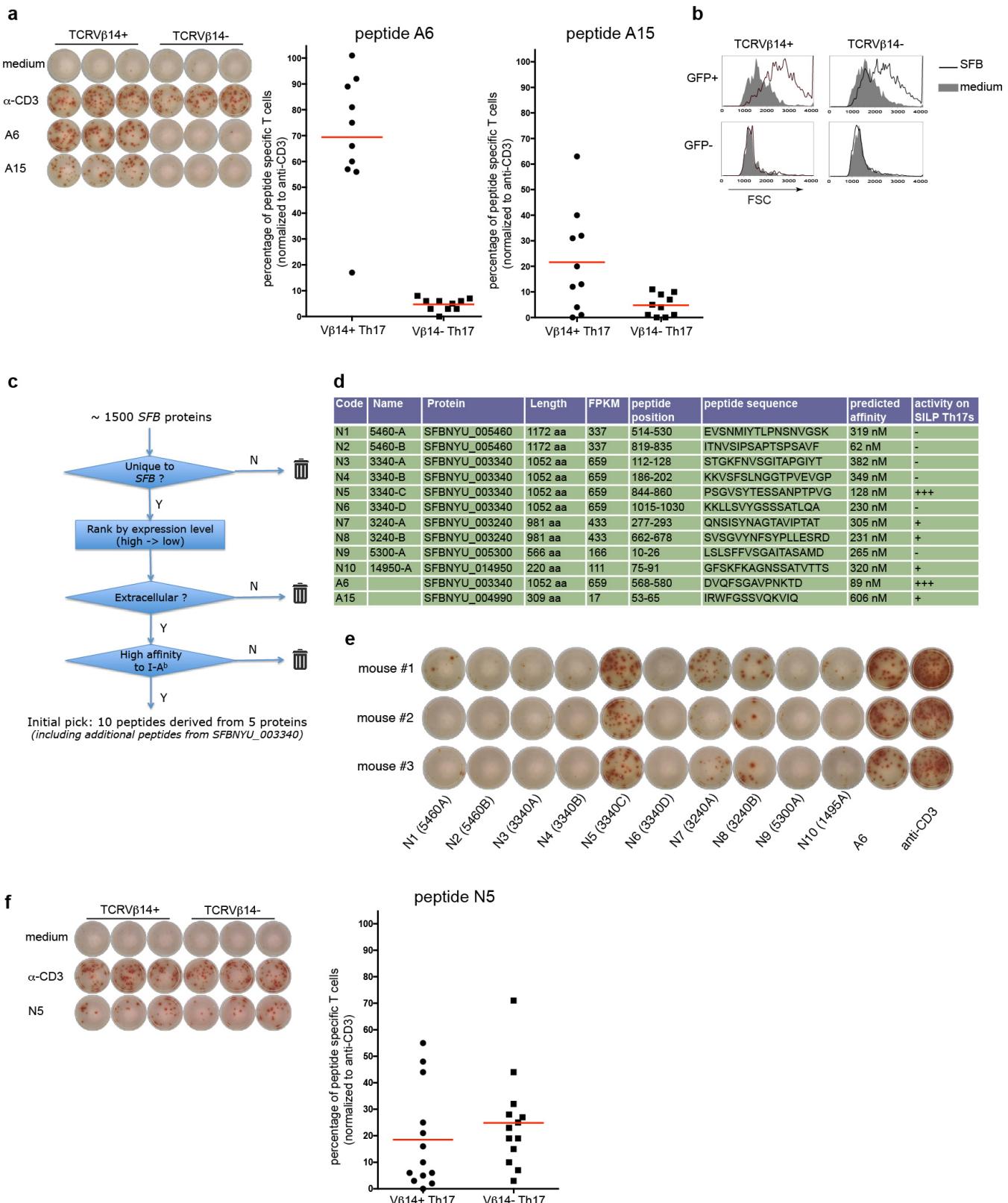
**Extended Data Figure 5 | Identification of *SFBNYU\_003340* epitopes recognized by a subset of the  $T_{H17}$  TCR hybridomas.** **a**, Schematic representation of the antigen screen using a whole-genome shotgun SFB library. **b**, Stimulation of the 7B8 hybridoma by bacterial pool 3F12. **c**, Reactivity of 7B8 and four other TCR hybridomas with bacterial clone

3F12-E8. **d**, Diversity of the CDR3 sequences of TCRs specific for 3F12-E8. Note that they belong to different V $\alpha$  subsets and have distinct V $\beta$ 14 CDR3 sequences. **e**, Responses of the 3F12-E8-specific TCR hybridomas to core epitopes encoded by minigenes expressed in *E. coli*.



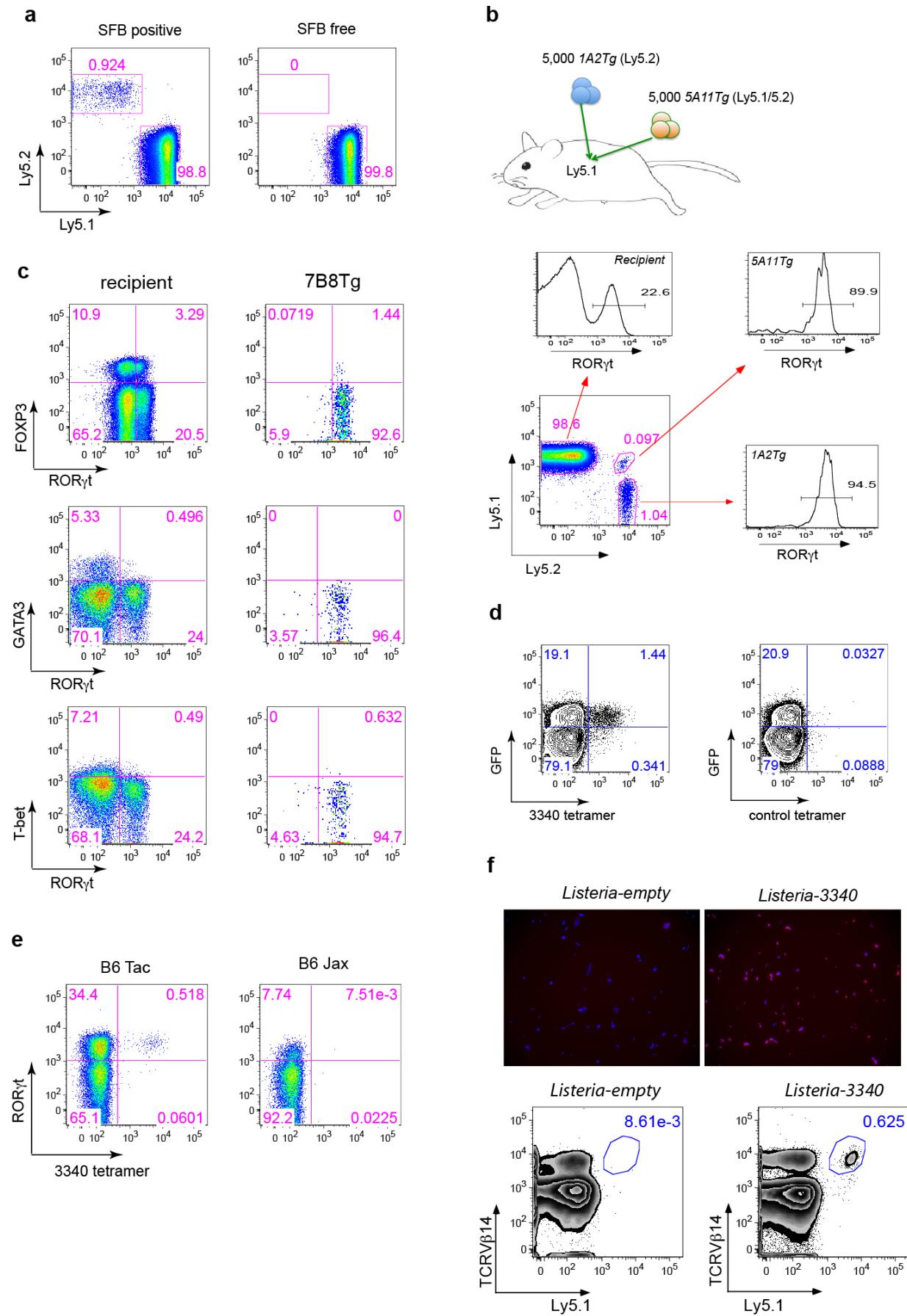
**Extended Data Figure 6 | Identification of SFBNYU\_004990 epitopes recognized by related TCRs.** **a**, Top, the distribution in  $T_{H17}$  and non- $T_{H17}$  cells of four TCRs that share an identical TCR $\alpha$  chain. Bottom, amino acid alignment of the V $\beta$ 14 CDR3 sequences. The green box highlights the sequence differences. **b**, Stimulation of the 5A11 hybridoma by bacterial pool 2D10 in

the SFB antigen screen. **c**, Responses of 4 TCR hybridomas, including a non- $T_{H17}$  hybridoma, to bacterial clone 2D10-A10. **d**, Responses of the 2D10-A10-specific TCR hybridomas to core epitopes encoded by minigenes expressed in *E. coli*. **e**, TCR hybridoma responses to titrated synthetic peptide (IRWFGSSVQKV) in the presence of APCs.



**Extended Data Figure 7 | SFB epitopes recognized by diverse T<sub>H</sub>17 cell TCRs.** **a**, The epitopes recognized by the V $\beta$ 14 $^{+}$  TCR hybridomas stimulate only V $\beta$ 14 $^{+}$  T $\beta$ 17 cells from the SILP. T $\beta$ 17 cells sorted from *Il23r*<sup>GFP/+</sup> mice were stimulated with indicated peptides (listed in **d**) in the presence of APCs. Left, representative IL-17A ELISPOT assay with triplicates. Right, normalized peptide-specific T $\beta$ 17 responses. Each dot represents one mouse. **b**, Polyclonal responses of V $\beta$ 14 $^{+}$  and V $\beta$ 14 $^{-}$  SILP T $\beta$ 17 cells to SFB antigens. Representative FACS plots from five experiments are shown. **c**, Bioinformatics

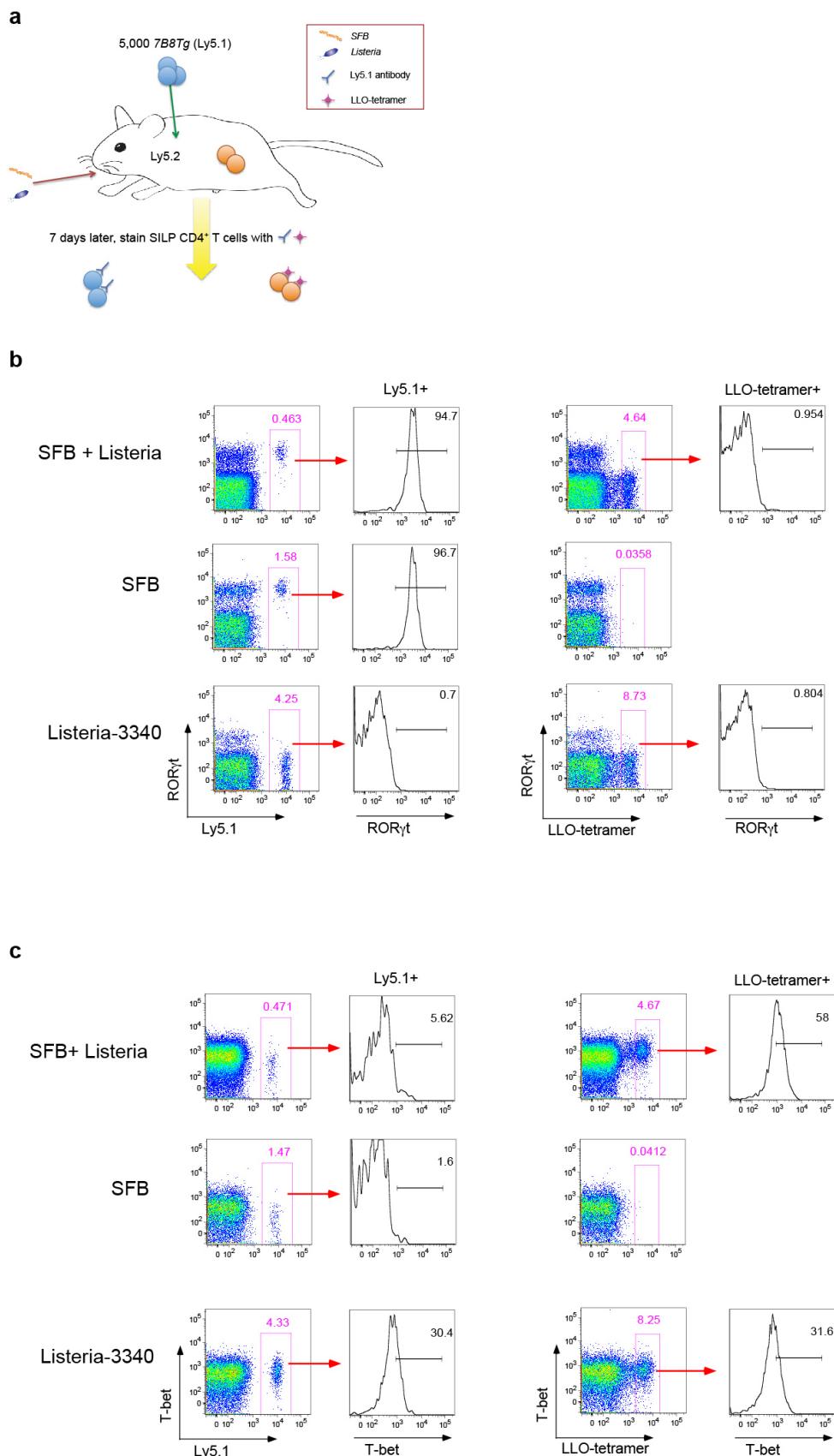
filtering approach to select candidate SFB epitopes. **d**, Summary of newly selected and the known A6 and A15 SFB peptides. **e**, IL-17A ELISPOT screen for indicated peptides using SILP T $\beta$ 17 cells sorted from SFB-colonized *Il23r*<sup>GFP/+</sup> mice. The A6 peptide from SFBNYU\_003340 and anti-CD3 served as positive controls. **f**, V $\beta$ 14 usage in T $\beta$ 17 cells specific for peptide N5. Left, representative IL-17A ELISPOT assay with triplicates for peptide N5, using V $\beta$ 14 $^{+}$  and V $\beta$ 14 $^{-}$  SILP T $\beta$ 17 cells sorted from *Il23r*<sup>GFP/+</sup> mice. Right, normalized N5-specific T $\beta$ 17 responses. Each dot represents one mouse.



**Extended Data Figure 8 | SFB-specific T cells become T<sub>H</sub>17 cells in SFB-colonized mice.** **a**, SFB-dependent 7B8Tg T cell accumulation in the SILP.  $2 \times 10^4$  naive 7B8Tg T cells were transferred into congenic Ly5.1 recipient mice that were SFB-colonized or SFB-free. CD4<sup>+</sup> T cells in the SILP were examined for donor and recipient isotype markers after 13 days. **b**, Top, strategy for co-transfer of congenic 1A2Tg and 5A11Tg T cells into SFB-colonized recipient mice. Bottom, FACS analysis of ROR $\gamma$ t expression in host- and donor-derived CD4<sup>+</sup> T cells in the SILP at 7 days after transfer. **c**, FACS analysis of transcription factors in host- and donor-derived SILP CD4<sup>+</sup> T cells after transfer of naive 7B8Tg T cells as in **a**. **d**, FACS analysis of SILP T cells from *Ii23r*<sup>GFP/+</sup> mice, stained with I-A<sup>b</sup>/3340-A6 tetramer and

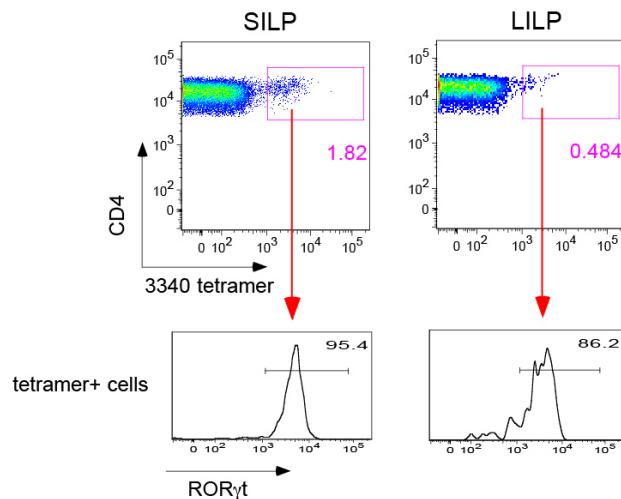
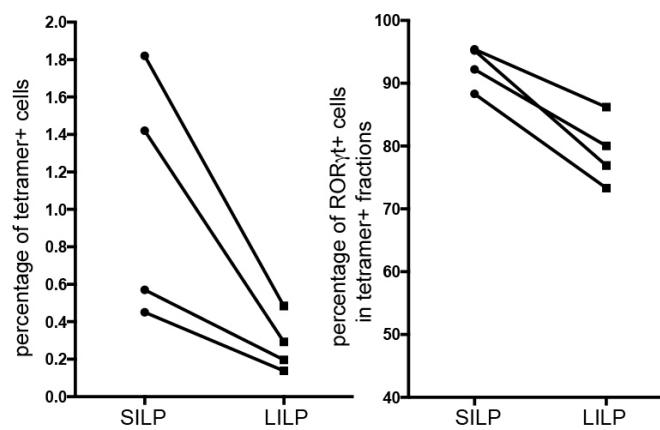
control tetramer (2W). **e**, FACS analysis of SILP T cells of B6 mice from colonies with different microbiota, stained with I-A<sup>b</sup>/3340-A6 tetramer and intracellular ROR $\gamma$ t antibody. **f**, Expansion of 7B8Tg T cells in mice colonized with *Listeria monocytogenes* expressing *SFBNYU\_003340*.

Top, immunofluorescence microscopic visualization of the expression of SFB protein by *L. monocytogenes*. *Listeria-3340* and *Listeria-empty* were stained with anti-3340 rabbit polyclonal antibody. Red, anti-3340 antibody staining. Blue, DAPI staining. Bottom, naive Ly5.1<sup>+</sup> 7B8Tg cells were transferred into congenic mice infected with *Listeria-3340* or *Listeria-empty*. Seven days after transfer, donor-derived CD4<sup>+</sup> T cells in the SILP were analysed by FACS.



**Extended Data Figure 9 | Transcription factor expression in SFB-specific and Listeria-specific T cells in co-infected mice.** Representative of data plotted in Fig. 4b. **a**, Experimental design for tracking both SFB- and Listeria- specific CD4<sup>+</sup> T cells following intestinal colonization with both bacteria. Ly5.2 B6 mice were colonized with *Listeria monocytogenes*, SFB, or

both bacteria, and 7B8Tg T cells from Ly5.1 mice were injected intravenously. Expression of T<sub>H</sub>1 and T<sub>H</sub>17 transcription factors in the SFB-specific 7B8Tg cells and LLO tetramer-specific recipient T cells was evaluated. **b**, Intracellular stain for ROR $\gamma$ t. **c**, Intracellular stain for T-bet.

**a****b**

**Extended Data Figure 10 | SFB-specific T<sub>H</sub>17 cells are present in both SILP and large intestine lamina propria (LILP) of SFB-colonized mice.** T cells were stained with I-A<sup>b</sup>/3340-A6 tetramer and antibody to intracellular ROR $\gamma$ t. **a**, Representative FACS plots (gated on CD4<sup>+</sup> T cells). **b**, Analysis of multiple

animals. Left, per cent of tetramer-positive cells among total CD4<sup>+</sup> T cells in each region of the intestine. Right, per cent of ROR $\gamma$ t<sup>+</sup> cells among the tetramer-positive cells. Each symbol represents cells from a separate animal.