Thymus–derived regulatory T cells contribute to tolerance to commensal microbiota

Anna Cebula1, Michal Seweryn2,3, Grzegorz A. Rempala2, Simarjot Singh Pabla1, Richard A. McIndoe1, Timothy L. Denning4, Lynn Bry5, Piotr Kraj1, Pawel Kisielow6 & Leszek Ignatowicz1

Peripheral mechanisms preventing autoimmunity and maintaining tolerance to commensal microbiota involve CD4+ Foxp3+ regulatory T (Treg) cells1–3 generated in the thymus or extrathymically by induction of naive CD4+ Foxp3+ T cells. Previous studies suggested that the T-cell receptor repertoires of thymic Treg cells and induced Treg cells are biased towards self and non-self antigens, respectively3–5, but their relative contribution in controlling immunopathology, such as colitis and other untoward inflammatory responses triggered by different types of antigens, remains unresolved. The intestine, and especially the colon, is a particularly suitable organ to study this question, given the variety of self-, microbiota- and food-derived antigens to which Treg cells and other T-cell populations are exposed. Intestinal environments can enhance conversion to a regulatory lineage6–8 and favour tolerogenic presentation of antigens to naive CD4+ T cells9,10, suggesting that intestinal homeostasis depends on microbiota-specific induced Treg cells12–15. Here, to identify the origin and antigen-specificity of intestinal Treg cells, we performed single-cell and high-throughput sequencing of the T-cell receptor repertoires of CD4+ Foxp3+ and CD4+ Foxp3− T cells, and analysed their reactivity against specific commensal species. We show that thymus-derived Treg cells constitute most Treg cells in all lymphoid and intestinal organs, including the colon, where their repertoire is heavily influenced by the composition of the microbiota. Our results suggest that thymic Treg cells, and not induced Treg cells, dominantly mediate tolerance to antigens produced by intestinal commensals.

We used TCR mini mice16, whose limited but diversified repertoire allows for a comprehensive comparison of T-cell receptors (TCRs) in various organs and subpopulations17. In these mice, thymocytes differentiate naturally as CD4+ Foxp3+ and CD4+ Foxp3− T cells, and efficiently repopulate peripheral lymphoid organs18. Furthermore, the TCR-β chain is identical in all TCR mini T cells, enabling detection of TCR diversity through specific analysis of the TCR-α chain19. To identify Foxp3+ cells, we crossed TCR mini mice with Foxp3GFP reporter mice, which express green fluorescent protein (GFP) under the control of Foxp3 regulatory sequences20. The TCR mini Foxp3GFP and B6Foxp3GFP mice had very similar numbers of Foxp3GFP− cells in different intestinal organs (Supplementary Fig. 1), and CD4+ cells in both types of mice expressed comparable levels of αβ and CCR9 molecules that regulate homing to the intestine (Supplementary Fig. 2). We also found that adoptive transfer of naive CD4+ T cells from TCR mini Foxp3GFP mice to lymphopenic, RAG-deficient mice caused inflammation in the colon and wasting disease. The disease could...
be prevented by the co-transfer of TCR\textsuperscript{mini} CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes, without affecting the colonization of the colon by CD4\textsuperscript{+} Foxp3\textsuperscript{+} T cells, indicating that thymus-derived T\textsubscript{reg} cells can control intestinal inflammation at least in these experimental settings (Supplementary Fig. 3).

Thymic T\textsubscript{reg} cells and induced T\textsubscript{reg} cells have similar phenotypes, and overlapping but distinct TCR repertoires relative to their thymic or peripheral origin\textsuperscript{16,18,19}. To compare dominant TCRs on CD4\textsuperscript{+} and overlapping but distinct TCR repertoires relative to their thymic or peripheral origin, we sorted T\textsubscript{reg} cells, including those from the colon, was represented by induced TCR\textsubscript{reg} cells and limited conversion of CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes, suggesting that the intestinal T\textsubscript{reg} cell repertoire includes a considerable proportion of dominant clones of thymic origin (Fig. 1a).

To compare TCR repertoires comprehensively on thymic, peripheral and intestinal CD4\textsuperscript{+} Foxp3\textsuperscript{+} and T\textsubscript{reg} cell clones, we used high-throughput sequencing (Fig. 1b, c), which also minimized the proportion of unique TCRs identified, that is, found only in one organ. In the colon, unique T\textsubscript{reg} cell TCRs comprised just 9% of all TCR sequences retrieved from this organ, with overall 5% of TCRs found on CD4\textsuperscript{+} Foxp3\textsuperscript{+} T cells but not on CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes, suggesting that the colonic population of induced T\textsubscript{reg} cells expressing TCRs specific for the CD4\textsuperscript{+} Foxp3\textsuperscript{+} lineage is limited (data not shown). Accordingly, the remaining 86% of TCRs from colonic T\textsubscript{reg} cells were expressed on CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes, and Fig. 1b shows that a vast majority of dominant TCRs from colonic T\textsubscript{reg} cells (found at least ten times) were shared between both populations. These TCRs accounted for approximately half of all TCRs retrieved from CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes, indicating that these thymocytes are not rare, recirculating mature induced T\textsubscript{reg} cells (Fig. 1c and data not shown). As shown in Fig. 1c, the similarity indices (mutual information index (MII), depicted by the distance between branches of the dendrogram) calculated for the TCR repertoires from various intestinal T\textsubscript{reg} cell and CD4\textsuperscript{+} Foxp3\textsuperscript{+} populations did not reveal higher similarity, which would be expected if a dominant portion of intestinal T\textsubscript{reg} cells, including those from the colon, was represented by induced TCRs from CD4\textsuperscript{+} Foxp3\textsuperscript{+} and T\textsubscript{reg} cells in all analysed organs was asymmetrically skewed, and only a few TCRs were over-represented in both CD4\textsuperscript{+} populations. Dissimilar allocation of abundant TCRs was previously observed in lymphoid organs and was attributed to a separate thymic differentiation pathway for thymic T\textsubscript{reg} cells and limited conversion of CD4\textsuperscript{+} Foxp3\textsuperscript{+} cells\textsuperscript{20}. In addition, approximately half of the dominant TCRs found on intestinal T\textsubscript{reg} cells (including colonic T\textsubscript{reg} cells) were also found on CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes, suggesting that the intestinal T\textsubscript{reg} cell repertoire includes a considerable proportion of dominant clones of thymic origin (Fig. 1a).

Figure 2 | In TCR\textsuperscript{\textgamma} Foxp3\textsuperscript{GFP} transgenic mice most colonic CD4\textsuperscript{+} Foxp3\textsuperscript{+} T cells share TCRs with CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes. a, Dominant TCRs from colonic T\textsubscript{reg} cells and their frequencies on CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes. The hierarchical dendrogram depicts MII indices between TCR repertoires from CD4\textsuperscript{+} Foxp3\textsuperscript{+} and CD4\textsuperscript{+} Foxp3\textsuperscript{+} populations from the indicated organs. For calculation of MII, the data set from CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes was limited, as described in the Methods.

Figure 3 | Antibiotic-induced changes in colonic flora have profound influence on the TCR repertoire of colonic thymic T\textsubscript{reg} cells. a, The effect of antibiotic treatment on the proportion of T\textsubscript{reg} cells in indicated organs. Three mice per group were analysed. b, Fifty dominant TCRs (Supplementary Table 4) of colonic T\textsubscript{reg} cells from untreated (red bars) or antibiotic-treated (black bars) mice and their frequencies in analysed repertoires. Asterisks denote TCRs not found on CD4\textsuperscript{+} Foxp3\textsuperscript{+} thymocytes. c, Diversity index (Rényi entropy function) of T\textsubscript{reg} cells from indicated organs of untreated and antibiotic-treated mice. Frequencies close to ’0’ correlates to diversity of low-abundant TCRs, and values close to ’2’ denote high-abundant TCRs. d, MII indices for TCR repertoires of CD4\textsuperscript{+} Foxp3\textsuperscript{+} and T\textsubscript{reg} cell populations from antibiotic-treated mice.
T_{reg} cells. In fact, not a single repertoire of T_{reg} cells clustered on the same branch of the dendrogram with CD4^+ Foxp3^+ repertoire(s), suggesting that these repertoires remained mostly dissimilar (as also shown for dominant TCRs in Fig. 1a and Supplementary Fig. 4). Limited conversion in the mesenteric lymph nodes or colon of TCR_{min} Foxp3^{GFP} mice was not a result of impaired recruitment of CD4^+ Foxp3^+ cells in this model because the conversion was apparent in the tumour environment, after adoptive transfer of CD4^+ Foxp3^- cells to lymphopenic hosts and in vitro (ref. 21 and data not shown).

In view of reports analysing mice with broader repertoire of TCRs than that of TCR_{min} mice 4, which suggested that induced T_{reg} cells are overwhelmingly abundant in the colon, we also examined whether the extent of TCR diversity could have a role in determining the relative involvement of thymic and induced T_{reg} cells in maintaining tolerance to colonic antigens. To address this question, we analysed TCR-β Foxp3^{GFP} transgenic mice in which the repertoire of TCRs is much larger than that of TCR_{min} mice owing to the natural diversity of the TCR-β chain. The high-throughput sequencing of the TCRV\(\alpha\)2\(\gamma\) chains of thymic, peripheral and colonic CD4^+ Foxp3^+ and CD4^- Foxp3^- TCR \(\alpha\) subpopulations from TCR-β transgenic mice revealed a similar pattern to the respective repertoires collected from TCR_{min} mice. Approximately 75% of all V\(\alpha\)2\(\gamma\) TCRs retrieved from colonic T_{reg} cells from TCR-β Foxp3^{GFP} transgenics were also expressed by CD4^+ Foxp3^+ thymocytes, including many abundant colonic TCRs (found in the colonic T_{reg} cell repertoire more than 20 times; Fig. 2a), which accounted for approximately 20% of all TCRs retrieved from CD4^+ Foxp3^+ thymocytes (data not shown). Furthermore, the MII indices calculated for the TCR repertoires from colonic, mesenteric and thymic CD4^+ Foxp3^+ and CD4^- Foxp3^- cells retained the same hierarchical clustering as originally observed in TCR_{min} mice (Fig. 2b). Thus, we concluded that the extent of TCR diversity does not have important influence on the predominance of thymic T_{reg} cells in the colon.

To investigate whether changes in the composition of colonic microflora influence the repertoire of thymic T_{reg} cells, we treated the TCR_{min} Foxp3^{GFP} mice with a cocktail of antibiotics. This treatment considerably altered the composition of the colonic microbiota (Fig. 3a). As shown in Supplementary Fig. 5, six out of ten dominant commensal species cultured from the caecum of untreated mice, including members of Clostridiales, an abundant anaerobe known to induce colonic induced T_{reg} cells^3, fell to undetectable levels in cultures from treated mice. Of the remaining four species cultured from the caecum, two markedly increased in biomass with antibiotic treatment and two remained unaffected. Antibiotic treatment also broadly altered the frequency of dominant TCRs identified in colonic T_{reg} cells, as some clones became undetectable, whereas others expanded, probably in response to the rebound growth of more antibiotic-resistant species (Fig. 3b). Most TCRs expressed by dominant T_{reg} cell clones that contracted or expanded in antibiotic-treated mice were also found on CD4^+ Foxp3^+ thymocytes, indicating that changes in the intestinal flora influence the repertoire of colonic thymic T_{reg}
cells (Fig 3b). As shown in Fig 3c, the diversity of the TCRs on colonic Treg cells from antibiotic-treated and untreated mice (calculated from high-throughput sequencing) was not significantly affected, despite their strong numerical reduction (Fig 3a). Calculation of the MII index between Treg cells and CD4+ Foxp3+ populations from different organs of antibiotic-treated mice (Fig 3d) showed that the repertoire of colonic Treg cells remained similar (see Fig 1c) to the rest of the CD4+ Foxp3+ repertoires, which argues against the significant recruitment of induced Treg cells in response to the changing composition of bacterial antigens.

To identify Treg cell clones specific to antigens produced by commensal species, we created hybridomas from colonic Treg cells25, and sorted hybridomas that responded to sterile filtrates of caecal contents from untreated TCRmin Foxp3GFP mice (Supplementary Fig. 6). Of these sub-cloned hybridomas, most that responded to caecal filtrates from untreated mice did not respond to filtrates from antibiotic-treated mice, suggesting that most responding hybridomas expressed TCRs specific for microbial antigens present in untreated mice (Fig 4a). We then identified 26 TCRs from hybridomas that responded to caecal filtrates from untreated mice, and examined their expression on CD4+ Foxp3+ thymocytes. Figure 4b shows that more than 90% of sequenced TCRs derived from colonic Treg cells were also expressed by CD4+ Foxp3+ thymocytes. Next, we tested the reactivity of hybridomas, against bacterial sonicates prepared from 13 cultures of individual species identified in the caeca of TCRmin Foxp3GFP mice (Fig 4c). Figure 4d (left column) shows that four hybridomas responded to isolates from Clostridiales (one of these also responded to phylogenetically related Flavonifractor), and another four hybridomas (Fig 4d, right column) responded to sonicates from Bacteroides or Lactobacillus, suggesting that these responses were elicited by unidentified bacterial antigen(s). Overall, these results demonstrate that colonic Treg cells and CD4+ Foxp3+ thymocytes share different TCRs that recognize microbial antigens.

Our study provides evidence for interactions between the host thymic Treg cell population and the complex communities of microbes present in the gut lumen. In-depth analysis of the TCR repertoire of colonic thymic Treg cells demonstrated that it is sufficiently broad to recognize microflora-derived antigens and that conversion of naive CD4+ Foxp3- cells does not appreciably modify its diversity. These conclusions challenge the recent report suggesting that induced Treg cells constitute the vast majority of colonic Treg cells14. In that study, the authors sampled TCR repertoires of colonic and peripheral CD4+ cells from a TCR-β transgenic line, and used retrogenic mice to determine whether TCRs derived from dominant colonic Treg cell clones support thymic selection of thymus-derived Treg cells. None of the colonic TCRs examined supported thymic Treg cell development, but constitutive expression of TCR in retrogenic mice can compromise thymic Treg cell selection and skew thymocyte commitment to the Foxp3+ lineage, irrespective of the TCR origin15. The number of TCRs examined in that study was about one order of magnitude smaller than that sequenced here, which would preclude detection of colonic TCRs on low-abundant clones in other organs. Nevertheless, half of the most abundant colonic TCRs were found on Treg cells in lymphoid organs14,15, in which 93% of these clones were estimated to represent thymic Treg cells26. Therefore, in both TCR-β transgenic lines14,15, a large proportion of colonic Treg cells can be of thymic origin.

The results of our study are consistent with the findings that thymic Treg cells recognize non-self antigens22,24, become activated after colonization of germ-free mice with standardised microbial flora (Schaedler flora)22, and prevent colitis in CNS1- deficient mice exclusively lacking induced Treg cells24 or in lymphopenic mice that received wild-type, naive CD4+ cells27. We conclude that induced Treg cells can participate in maintaining tolerance to intestinal antigens, but that thymic Treg cells play the dominant part in this process.

METHODS SUMMARY

Mice. The TCRmin, TCRmin Foxp3GFP and TCR-β Foxp3GFP mice were described previously14,17. C57BL/6 (B6), TCR-α-deficient, and RAG-deficient mice were purchased from Jackson Laboratories. All animals were 8–12 weeks old.

Flow cytometry, single-cell sorting and single-cell RT–PCR. Cell-surface staining with monoclonal antibodies was done by standard procedures. Complementary DNAs from single CD4+ Foxp3GFP+ and CD4+ Foxp3GFP+ T cells were synthesized followed by two rounds of PCR, and CD3 Vα chains were sequenced as previously described26. The relative frequency with which a given TCR was found in a particular organ was calculated by dividing its number by the sum of all sequences from that organ.

CDR3 high-throughput sequencing. Analysis of the TCRmin Foxp3GFP V(2)22/V(2)26/2 or V(2)21/26 CDR3 regions (from TCR-β Foxp3GFP transgenic mice) was performed from flow-cytometry-purified CD4+ Foxp3GFP- and Foxp3GFP+ T cells (purity >99%). RNA was isolated (RNeasy Mini Kit, Qiagen) and converted to cDNA (SuperScript III, Invitrogen) with a C2-specific primer. TCR-α CDR3 regions were amplified using primers with incorporated barcodes, and the PCR product was sequenced by EdgeBio/BioServ. The results obtained by this procedure may not precisely reflect the original frequency of different templates present in a given preparation because preferential amplification of some templates could occur. This could affect the accuracy of the frequency estimates of some individual sequences but not our conclusion emerging from high-throughput sequencing, corroborated by the results of single-cell sequencing.

Statistics. MII similarity index measures pairwise similarities between populations by considering the overlap and relative abundances of TCRs24. For calculation of MII, the data set from CD4+ Foxp3+ thymocytes was limited to 2.4 x 106 randomly selected TCRs to match approximately the number of TCRs retrieved from CD4+ Foxp3+ thymocytes.

Antibiotic treatment. TCRmin Foxp3GFP mice were given metronidazole (2.5 mg ml^-1), vancomycin (0.5 mg ml^-1), and ciprofloxacin (0.66 mg ml^-1) (all from Sigma) in drinking water for 6 weeks.

Hybridoma assays. Polyclonal hybridomas responding to caecal lysates were sorted based on expression of a nuclear factor of activated T cells (NFAT)-GFP reporter. Response of cloned hybridomas was measured using the HT-2 assay25.

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

Received 6 March 2012; accepted 15 March 2013.

Published online 28 April 2013.

1. Belkaid, Y. & Rouse, B. T. Natural regulatory T cells in infectious disease. Nature Immunol. 6, 353–360 (2005).
2. Sakaguchi, S., Powrie, F. & Ransohoff, R. M. Re-establishing immunological self-tolerance in autoimmune disease. Nature Med. 18, 54–58 (2012).
3. Jordan, M. S. et al. Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. Nat. Immunol. 2, 301–306 (2001).
4. Ribot, J., Romagnoli, P. & van Meerwijk, J. P. Agonist ligands expressed by thymic epithelium enhance positive selection of regulatory T lymphocytes from precursors with a normally diverse TCR repertoire. J. Immunol. 177, 1101–1107 (2006).
5. Coutinho, A. et al. Thymic commitment of regulatory T cells is a pathway of TCR-dependent selection that isolates repertoires undergoing positive or negative selection. Curr. Top. Microbiol. Immunol. 293, 43–71 (2005).
6. Hsieh, C. S. et al. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. Nature Immunol. 7, 401–410 (2006).
7. Curotto de Lafaille, M. A. & Lafaille, J. J. Natural and adaptive Foxp3+ regulatory T cells: more of the same or a division of labor? Immunity 30, 626–635 (2009).
8. Baines, M. J. & Powrie, F. Regulatory T cells reinforce intestinal homeostasis. Immunity 31, 401–411 (2009).
9. Mucida, D. et al. Regulatory T cell function: the role of retinoic acid. Science 317, 256–260 (2007).
10. Coombes, J. L. et al. Functionally specialized population of mucosal CD103+ DCs induces Foxp3+ regulatory T cells via a TGF-β and retinoic acid-dependent mechanism. J. Exp. Med. 204, 1757–1764 (2007).
11. Denning, T. L. et al. Lamina propria macrophages and dendritic cells differentially induce regulatory and interleukin-17-producing T cell responses. Nature Immunol. 8, 1096–1099 (2007).
12. Round, J. L. & Mazmanian, S. K. Inducible Foxp3+ regulatory T cell development by a commensal bacterium of the intestinal microbiota. Proc. Natl Acad. Sci. USA 107, 12204–12209 (2010).
13. Atarashi, K. et al. Induction of regulatory T cells by indigenous Clostridium species. Science 331, 337–341 (2011).
14. Lathrop, S. K. et al. Peripheral education of the immune system by colonic commensal microbiota. Nature 478, 250–254 (2011).
15. Haribhai, D. et al. A requisite role for induced regulatory T cells in tolerance based on expanding antigen receptor diversity. Immunity 35, 109–122 (2011).
16. Pacholczyk, R. et al. Origin and T cell receptor diversity of Foxp3 \(^+\) CD4 \(^+\) CD25 \(^+\) T cells. *Immunity* **25**, 249–259 (2006).
17. Kuczma, M. et al. Foxp3-deficient regulatory T cells do not revert into conventional effector CD4 \(^+\) T cells but constitute a unique cell subset. *J. Immunol.* **183**, 3731–3741 (2009).
18. Hsieh, C. S. et al. Recognition of the peripheral self by naturally arising CD25 \(^+\) CD4 \(^+\) T cell receptors. *Immunity* **21**, 267–277 (2004).
19. Wong, J., Mathis, D. & Benoist, C. TCR-based lineage tracing: no evidence for conversion of conventional into regulatory T cells in response to a natural self-antigen in pancreatic islets. *J. Exp. Med.* **204**, 2039–2045 (2007).
20. Lathrop, S. K. et al. Antigen-specific peripheral shaping of the natural regulatory T cell population. *J. Exp. Med.* **205**, 3105–3117 (2008).
21. Kuczma, M. et al. Intratumoral convergence of the TCR repertoires of effector and Foxp3 \(^+\) CD4 \(^+\) T cells. *PLoS ONE* **5**, e13623 (2010).
22. Pacholczyk, R. et al. Nonself-antigens are the cognate specificities of Foxp3 \(^+\) regulatory T cells. *Immunity* **27**, 493–504 (2007).
23. Bautista, J. L. et al. Intraclonal competition limits the fate determination of regulatory T cells in the thymus. *Nature Immunol.* **10**, 610–617 (2009).
24. Suffia, I. J. et al. Infected site-restricted Foxp3 \(^+\) natural regulatory T cells are specific for microbial antigens. *J. Exp. Med.* **203**, 777–788 (2006).
25. Geuking, M. B. et al. Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunity* **34**, 794–806 (2011).
26. Josefowicz, S. Z. et al. Extrathymically generated regulatory T cells control mucosal T\(_{\text{reg}}\)2 inflammation. *Nature* **482**, 395–399 (2012).
27. Fahle`n, L. et al. T cells that cannot respond to TGF-\(\beta\) escape control by CD4\(^+\) CD25\(^+\) regulatory T cells. *J. Exp. Med.* **201**, 737–746 (2005).
28. Rempala, G. A. & Seweryn, M. Methods for diversity and overlap analysis in T-cell receptor populations. *J. Math. Biol.* http://dx.doi.org/10.1007/s00285-012-0589-7 (25 September 2012).

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

**Acknowledgements** This work was supported by basic research grants from the National Institutes of Health (NIH; AI 5R01AI079277 to L.I., and DMS 1106485 and R01CA152158 to G.A.R.). The Harvard Digestive Disease Center (HDDC) Microbiome Core facility is supported by P30-DK034854 and Brigham and Women’s Hospital in Boston, Massachusetts. Microbiological analyses were performed by M. Delaney, A. Dubois and Q. Liu in the HDDC Microbiome Core, with additional review of findings by A. B. Onderdonk. We thank J. Pihkala and H. Ignatowicz for technical assistance, M. Kuczma, L. Wojciech, E. Szurek, A. Miazek and P. Muranski for discussion and R. Markowicz for editing the manuscript.

**Author Contributions** A.C. performed most experiments and analysed the data; M.S. and G.A.R. performed statistical analyses; S.S.P. and R.A.M. designed the PACE program; T.L.D. provided expertise in colonic T-cell isolation; L.B. performed the microbiological study; P.Kr. established TCRmini and Foxp3GFP mice models; P.Ki. and L.I. designed the study, analysed the data and wrote the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.I. (ignatowicz@gru.edu).
METHODS

Mice. TCR<sup>min</sup> Fopx3<sup>3FGP</sup> and TCR-β Fopx3<sup>3FGP</sup> (Vβ14β2β12β2) mice were obtained by mating B6Fopx3<sup>3FGP</sup> (ref. 17) with TCR<sup>min</sup> (ref. 16) and TCR-β (ref. 16) mice, respectively. The progeny was screened for the co-expression of the Fopx3<sup>3FGP</sup> reporter, plus the TCR<sup>min</sup> Vα2β3β14 β1 dimer or TCR-β Vβ14 chain, respectively. To eliminate expression of endogenous TCR-α chains, all TCR<sup>min</sup> mice were crossed with mice deficient in endogenous TCR-α loci and were heterozygous for TCR-α Vα2β2β6β2 mini-locus to ensure expression of a single TCR-α chain per T cell. All animals were housed in Georgia Regents University animal facility in accordance to the Institutional regulations.

Purification of intestinal lamina propria T cells. Intestinal regions were opened longitudinally and contents were flushed with ice-cold HBSS (Gelgro). Each region was cut into small pieces and washed with HBSS supplemented with 5% FCS (HyClone) and 2 mM EDTA at 37°C. A single-cell suspension was obtained after treatment with collagenase D (1.0 mg ml<sup>-1</sup>) and DNase I (0.1 mg ml<sup>-1</sup>) (both from Roche). A purified and concentrated suspension of lamina propria lymphocytes was obtained after centrifugation on Percoll (GE Healthcare) gradient (45% and 70%). The interface, enriched in leukocytes, was collected and used for experiments.

Isolation of thymocytes and T cells from lymphoid organs. Single-cell suspensions were prepared from the thymus, inguinal and mesenteric lymph nodes by mechanical disruption. Peyers’s patches were excised from the small intestine wall, and lymphocytes were isolated by enzymatic digestion for 20 min, using collagenase D (1.0 mg ml<sup>-1</sup>) and DNase I (0.1 mg ml<sup>-1</sup>) at 37°C.

Flow cytometry, single-cell sorting and single-cell RT–PCR. Thymocytes and T cells were stained with antibodies against CD4, CD8, V<sub>α</sub>2, Vβ, CD99 and zβ; (BD Biosciences or eBioscience), and analysed using a BD FACS Canto (BD Biosciences). Single cells were sorted (MoFlo cell sorter, Beckman Coulter) into 96-well plates from a sorted (purity >99%) population of CD4<sup>+</sup> Fopx3<sup>3FGP</sup> and CD4<sup>+</sup> Fopx3<sup>3FGP</sup> T cells. CDNA was synthesized using MMLV reverse transcriptase (Promega) and random hexamers (IDT), followed by two rounds of PCR via Perfect Tag Polymerase (5 PRIME). Products of the CDRA V<sub>α</sub>2 chain obtained in the second PCR reaction were sequenced in the Genomic Core Facility at the University of Illinois. All necessary precautions were taken to avoid PCR contamination, as previously described<sup>16</sup>

High-throughput CDR3 sequencing. The following C<sub>α</sub>-specific primer was used for cDNA synthesis: 5′-TGCGGACAATTATTGGTGAGTC-3′. The following primers with incorporated tags for Ion Torrent high-throughput sequencer were used: Vα2IT, 5′-CCATCTACTCCTGGGCTGGCCAGCGTTCCAGCTACCTCCGCCGTGCCAGAGTCGATTTAGGCAGCA CAGGCTTCTTCCGGTCTG-3′. The accuracy of the high-throughput CDR3 sequencing was ensured by the use of high-fidelity DNA polymerase with a low intrinsic error rate (AccuPrime Taq DNA Polymerase High Fidelity; Invitrogen). In addition, Ion Torrent suite software filters were used during data processing to exclude low-quality reads and erroneous sequences derived from mixed DNA templates. Most common Ion Torrent sequencer errors are base insertions and deletions occurring in homopolymers, which result in frameshifts and stop codons. To identify errors within Vβ2 and Jβ2 (or Jβ26) segments, all sequences were aligned to constant regions. To estimate the amplification-specific error within the CDR3 region, two monoclonal CDR3 regions from TCR<sup>min</sup> mice were amplified (approximately 1×10<sup>7</sup> reads were collected), and the reads that differed from the original template were counted. This approach estimated that less than 5% of CDR3 regions may contain errors (Supplementary Table 6), which is significantly below the threshold adversely affecting the statistical similarity and overlap analysis.<sup>22</sup>

Microbiology. Caeca with content were dissected under sterile conditions, placed in cryovials and immediately snap frozen in liquid nitrogen. Samples were further processed and analysed in the Harvard Digestive Disease Center Microbiome Core facility. Phylogenetic identification and subtyping was based on 16S ribosomal RNA classification. Molecular speciation of bacteria was performed with 16S rRNA gene analyses. Assembled sequences were loaded into the Ribosomal Database Project’s SEQMATCH tool to identify the taxonomic assignment (http://rdp.cme.msu.edu/).

Statistical analysis. The comparison of various TCR repertoires was conducted by means of the assessment of their respective diversities as well as overlap between populations (Figs 1c, 2b, 3c, d and Supplementary Fig. 4a). In the current context, under the term ‘diversity’ we understand both the richness and the abundance patterns of the repertoires.

For the sake of quantifying the TCR diversity for a single repertoire, we have adopted the information-theoretic approach based on the notion of the Renyi entropy function of order ω denoted H<sub>ω</sub> (ref. 30). This function quantifies diversity by means of the formula:

\[ H_\omega = - \frac{1}{1-\omega} \log \left( \sum_i p_i^\omega \right) \]

in which \( p_i \) is the observed frequency of the \( i \)-th species, and the order \( \omega \) is a non-negative exponential weight parameter. The value of \( \omega \) below unity gives more weight to the less abundant (and thus possibly under-sampled) species, whereas the values above unity give more weight to the more abundant species. For \( \omega = 1 \), the above formula is not well-defined but may be obtained by taking the limiting expression as \( \omega \) approaches unity. In this case:

\[ H_1 = - \sum_i p_i \log p_i \]

is the usual Shannon entropy function known in the information theory, which weights equally the contributions of all observed species to the repertoire diversity.

By plotting the values of the Renyi entropy H<sub>ω</sub> against its index, we are able to analyse diversity of TCR populations graphically in terms of their diversity, weighted towards rare (\( \omega < 1 \)) and abundant (\( \omega > 1 \)) species (Fig. 3c). Because the \( p_i \) quantities are the empirical counts of the observed species, for the sake of obtaining bounds on the sampling error in the values H<sub>ω</sub> we apply the computational methods based on
the non-parametric bootstrap as described previously\textsuperscript{28}. To improve the robustness of \( H_a \) against unseen species (that is, the possible under-sampling of the species richness) we have also considered a version of the analysis in which the 'Chao–Shen' correction was applied to compute the Shannon entropy, as described\textsuperscript{28}. In our particular case, it turned out that the results of this alternative analysis differed only marginally from the original ones.

The pairwise overlap analysis as described in Figs 1c, 2b, 3d and Supplementary Fig. 4a was conducted on the basis of the hierarchical clustering of the repertoires with an appropriately chosen dissimilarity function. Similarly, as for the analysis of the single repertoire diversity, we have adopted an information theoretical approach to pairwise compare TCR repertoires. Because the pairs of observed frequencies of different TCR species may be arranged in a two-way contingency table, the entropy-based index, known as the mutual information index (MII), can be applied to measure the association between observed TCR frequencies and the corresponding class (repertoire) labels. If \( MI \) is the usual mutual information statistic in a two-way contingency table, and \( n_1 \) and \( n_2 \) are the proportions of species observed in different TCR populations (that is, \( n_1 + n_2 = 1 \)), the MII is given by the formula:

\[
MII = \frac{MI}{-n_1 \log n_1 - n_2 \log n_2}
\]

The MII may be shown to take values between 0 and unity, with the 0 value admitted only for repertoires with linearly dependent vectors of TCR frequencies. This property makes MII an appropriate measure of dissimilarity for the current purpose of TCR populations clustering. For the practical purpose of computing the sampling errors in values of MII for all pairs of repertoires, the statistical computational bootstrap methods were applied as explained above. With the MII as the dissimilarity measure we used a canonical clustering procedure based on agglomerative clustering with Ward linkage method. The outcome of the algorithm is presented as a dendrogram or a tree diagram with its leaves representing TCR populations. The leaves are located at the tree distances from each other computed according to their MII values.

29. Singh, N. \textit{et al.} Generation of T cell hybridomas from naturally occurring FoxP3\textsuperscript{+} regulatory T cells. \textit{Methods Mol. Biol.} \textbf{707}, 39–44 (2011).

30. Renyi, A. On measures of information and entropy. \textit{Proc. 4th Berkley Symp. Math. Stat. Prob.} \textbf{1}, 547–561 (1961).