A novel lineage of RORγt$^+$Aire$^+$ antigen presenting cells promotes peripheral generation of intestinal regulatory T cells and tolerance during early life

Authors: Blossom Akagbosu[1†], Zakieh Tayyebi[2,3†], Gayathri Shibu[1,4], Yoselin A. Paucar Iza[1,4], Deeksha Deep[4,5,6], Yollanda Franco Parisotto[1], H. Amalia Pasolli[7], Rasa Elmentaite[8], Maximilian Knott[9], Saskia Hemmers[4,5,10], Lorenz Jahn[11], Christin Friedrich[12], Jacob Verter[5], Marcel van den Brink[4,11,13], Georg Gasteiger[12], Thomas G. P. Grünewald[9,14,15], Christina Leslie[2], Alexander Y. Rudensky[4,5]*, Chrysothemis C. Brown[1,4,5,16,17,18]*

Affiliations:

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, USA

[2]Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[3]Tri-Institutional Program in Computational Biology and Medicine, Weill Cornell Graduate School, New York, NY, USA

[4]Immunology and Microbial Pathogenesis Program, Weill Cornell Medicine Graduate School of Medical Sciences, New York, NY, USA

[5]Howard Hughes Medical Institute and Immunology Program, Sloan Kettering Institute, and Ludwig Center at Memorial Sloan Kettering Cancer Center, New York, NY, USA

[6]Tri-Institutional MD-PhD Program, Weill Cornell/Rockefeller/Sloan Kettering, New York, NY USA

[7]Electron Microscopy Resource Center, The Rockefeller University, New York, NY, USA

[8]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK.

[9]Institute of Pathology, Faculty of Medicine, LMU Munich, Munich, Germany.

[10]Department of Immunology, Duke University, Durham, North Carolina, USA

[11]Department of Immunology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[12]Würzburg Institute of Systems Immunology, Max Planck Research Group, Julius-Maximilians-Universität Würzburg, Würzburg, Germany

[13]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[14]Hopp-Children's Cancer Center Heidelberg (KiTZ), Heidelberg, Germany

[15]Division of Translational Pediatric Sarcoma Research, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany

[16]Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[17]Infection, Inflammation and Rheumatology Section, UCL Great Ormond Street Institute of Child Health, London, UK

[18]Parker Institute for Cancer Immunotherapy, Memorial Sloan Kettering Cancer Center, New York, NY, USA

* Correspondence to: rudenska@mskcc.org (A.Y.R) or brownc10@mskcc.org (C.C.B).

[†] These authors contributed equally

Abstract

Tolerance to self- or innocuous foreign antigens is vital for preservation of organismal health. Within the thymus, medullary thymic epithelial cells (mTECs) expressing *A*uto*I*mmune *R*egulator, Aire, play a critical role in self-tolerance through deletion of autoreactive T cells and promotion of thymic regulatory T (Treg) cell development. A second wave of Treg cell differentiation occurs in the periphery, upon exposure to dietary and commensal microbiota derived antigens within the first few weeks of life, yet the cell types responsible for the generation of peripheral Treg (pTreg) cells are not known. Here we identified a new lineage of tolerogenic RORγt$^+$ Aire$^+$ antigen-presenting cells (APC) with a hybrid dendritic cell (DC)-mTEC phenotype, dubbed Thetis cells (TCs), comprising 4 major sub-groups (TC I-IV). We uncovered a developmental wave of TCs within intestinal lymph nodes during a critical early life window, coincident with the wave of pTreg cell differentiation. While TC I bore remarkable homology with Aire$^+$ mTECs, TC IV were specifically enriched for critical molecules required for pTreg generation, including the TGF-β activating integrin αvβ8. Loss of either MHCII or Itgb8 expression by TCs led to a profound impairment in intestinal pTreg differentiation, with onset of intestinal inflammation. In contrast, MHCII expression by RORγt$^+$ group 3 innate lymphoid cells (ILC3) and classical DCs was dispensable for pTreg generation, further implicating TCs as the critical tolerogenic RORγt$^+$ APC. Our studies reveal parallel pathways for establishment of tolerance within the thymus and periphery, marked by involvement of shared cellular and transcriptional programs.

Main

Within the thymus, a distinct lineage of epithelial cells establishes tolerance to self-antigens through deletion of autoreactive T cells and promotion of thymic regulatory T (Treg) cell differentiation. These functions of medullary thymic epithelial cells (mTECs) are mediated in part through the expression of Aire, which regulates the ectopic expression of tissue-restricted antigens (TRAs)[1]. Another major site of tolerance induction resides within intestinal lymphoid tissue where infants are exposed to a slew of new dietary components and colonizing microbes upon weaning. Establishing a harmonious host-microbiota relationship in this early life developmental window is critical to prevent later onset of immune-mediated disorders[2,3]. Central to the establishment of tolerance to intestinal microbes is the differentiation of naive T cells into peripherally generated Treg (pTreg) cells upon encounter with commensal-derived antigens[4–7]. Yet the antigen presenting cell (APC) that promotes pTreg cell differentiation is not known. The narrow time window for establishing intestinal immune homeostasis suggests the presence of a developmentally restricted tolerogenic APC within the neonatal intestinal niche.

Antigen presentation by RORγt$^+$ APCs is critical for pTreg induction in early life

Extra-thymic pTreg cells, distinguished from their thymic counterparts by expression of the orphan nuclear receptor RORγt, arise in the mesenteric lymph nodes (mLN) in response to commensal bacterial antigens and play a critical role in suppressing inflammatory immune responses against gut microbes. In an intriguing symmetry, mice deficient in MHC class II-restricted antigen presentation by RORγt$^+$ cells (*MHCII*$^{\Delta RORγt}$), develop severe intestinal inflammation due to a failure to establish tolerance to commensal bacteria, suggesting a potential connection between RORγt$^+$ APCs and RORγt$^+$ pTreg cell generation. To address this possibility, we analyzed *MHCII*$^{\Delta RORγt}$ mice at 3 weeks of age, when pTreg cells accumulate within the intestine. We observed a dramatic reduction in RORγt$^+$ pTreg cells within the mLN and colonic lamina propria (**Fig. 1a,b**), along with expansion of effector CD44^{hi} T cells (**Extended data Fig. 1a**). At 9 weeks of age these mice exhibited a sustained, severe reduction in RORγt$^+$ pTreg cells along with expansion of colonic Th17 cells (**Fig. 1c**, **Extended data 1b**), in line with previous studies demonstrating a prominent role for pathobiont-specific RORγt$^+$ pTreg cells in suppressing inflammatory Th17 cells[5]. Histological analysis demonstrated severe colitis with marked inflammatory cell infiltrate, mucosal ulceration and crypt loss (**Fig. 1d,e**), confirming the critical role of RORγt$^+$ APCs in preventing dysregulated intestinal immune responses[5]. Together these results demonstrate an essential role for an early life RORγt$^+$ APC in pTreg cell generation and raise the question as to the nature of the tolerogenic RORγt$^+$ APC.

Identification of a novel lineage of RORγt$^+$ antigen presenting cells

A number of candidate APC types have been suggested to regulate tolerance to the intestinal microbiota including dendritic cells (DCs) and MHCII$^+$ ILC3 cells, also known as lymphoid tissue inducer (LTi)-like cells. Amongst these, loss of tolerance to commensals in *MHCII$^{\Delta RORγt}$* mice has previously been attributed to ILC3s on the basis that they represent the only RORγt$^+$MHCII$^+$ cell type[8]. Since ILC3 cells do not prime T cells *ex vivo*, it was suggested that ILC3s mediate negative selection of commensal reactive effector T (Teff) cells[8]. However, recent studies have identified RORγt expressing DCs[9,10] as well as 'ILC3-like' cells[11], although the function of these cells is not known. Critically, the spectrum of RORγt$^+$ APCs has not been examined within the mLN at the time when pTreg cells first arise. To fill this gap, we generated a novel *Rorc$^{Venus\text{-}creERT2}$* mouse to aid identification and labeling of RORγt$^+$ cells (**Extended data Fig. 2a,b**). Analysis of *Rorgt^{cre}R26$^{lsl\text{-}tdTomato}$Rorc$^{Venus\text{-}creERT2}$* mice confirmed that expression of Venus protein, translated downstream of exon 11, faithfully reflected RORγt expression within mLN and large intestine (**Extended data Fig. 2c**).

RORγt$^+$ pTreg cells first appeared within the mLN between postnatal days 10 and 14 (P10–14) with rapid accumulation thereafter (**Extended data Fig. 2d**). We therefore performed paired single cell (sc)RNA- and ATAC-sequencing of CD45$^+$Lin$^-$RORγt(Venus)$^+$MHCII$^+$ cells isolated from the mLN at 2 weeks of age (**Fig. 2a**, **Extended data Fig 2e**). After quality filtering, we retained chromatin accessibility and transcriptional profiles on 10,145 cells. Unsupervised clustering of either the RNA or ATAC-seq data revealed two major cell types (**Fig. 2b-d; Extended data Fig. 2f**). The first represented ILC3s spanning their full developmental spectrum including an ILC3 progenitor (ILC3p), proliferating and mature NCR$^+$ ILC3s, and CCR6$^+$ LTi cells (**Fig. 2d, Extended data Fig. 2f,g**). The second cell type did not express canonical ILC genes. These cells were distinguished by an intriguing combination of both epithelial and DC-associated transcription factors and cell surface molecules, and could be further classified into 4-subpopulations and a small proliferating cluster of cells (**Fig. 2b-d**, **Extended data Fig. 2f,g**). Although these cells expressed the DC marker *Zbtb46*, this transcript was also unexpectedly highly expressed by MHCII$^+$ ILC3s, a finding confirmed by analysis of *Zbtb46^{GFP}Rorgt^{cre}R26$^{lsl\text{-}tdTomato}$* mice (**Extended data Fig. 2j**). To elucidate the identity of non-ILC3 RORγt$^+$ APCs, we compared similarity of pseudo-bulk transcriptomes across a comprehensive database of immune and stromal cells (ImmGen). As expected, ILC3 scRNA clusters aligned with ILC3s, whereas the remaining clusters exhibited surprisingly high correlation with both mTECs and DCs (**Fig 2e**), including specific expression of Aire, the signature gene for a subset of mTECs (mTEC II), within two of the clusters (I and III; **Fig 2f)**. In light of the 'shape-shifting' hybrid phenotype of this lineage of RORγt$^+$ APCs, we refer to these cells as Thetis cells (TCs).

Comparison of TC cluster identity defined by chromatin accessibility or gene expression revealed near perfect congruence (**Extended data Fig. 2j**), confirming that TCs comprise four distinct cell types (TC I-IV). Analysis of pseudo-bulk transcriptomes for TC sub-groups alongside published single-cell thymic epithelial transcriptomes[12], further corroborated the transcriptional overlap between TC and mTEC subsets (**Fig. 2g**). Overall, these data demonstrated the existence of a novel lineage of RORγt$^+$Aire$^+$ cells, distinct from ILC3s, present within intestinal lymph nodes during early life.

Characterization of the TC phenotypic landscape

Extra-thymic Aire expression has previously been reported in CCR7$^+$ DCs[13]. Of note, the gene expression signature that distinguishes CCR7$^+$ DCs from their CCR7$^-$ counterparts is not DC lineage defining, but rather represents a particular transcriptional program that can be acquired by both classical DC subsets (cDC1 and cDC2) as well as other APC types[14,15], reflecting enhanced cell migration, T cell priming capacity and expression of immune regulatory molecules[15]. The shared expression of Aire in TCs and CCR7$^+$ DCs prompted us to examine the relationship between these two cell types. Analysis of *Rorc*$^{Venus\text{-}creERT2}$*Aire*^{GFP} mice confirmed wide-spread Aire(GFP) expression by Lin$^-$CD11c$^+$MHCII$^+$CCR7$^+$ cells encompassing both DC1 and DC2 (**Extended Data Fig. 3a,b**); however, < 4% of CCR7$^+$ cells expressed RORγt. Moreover, CXCR6$^-$RORγt$^+$MHCII$^+$ TCs were also found amongst CCR7$^-$ and CD11c$^-$ MHCII$^+$ cell populations (**Extended Data Fig. 3a,b**), suggesting that TCs were distinct from CCR7$^+$ DCs. To gain further insight into the distinguishing features of Aire expressing TCs, DCs, and mTECs we performed orthogonal SMART-seq2 (SS2) scRNA-seq analysis of Lin$^-$RORγt$^+$MHCII$^+$ cells isolated from the mLN of 3-week-old *Rorc*$^{Venus\text{-}creERT2}$ mice in parallel with mLN Aire$^+$ DCs and Aire$^+$ mTECs from age-matched Aire^{GFP} mice (**Fig. 3a; Extended data Fig. 3c**). Clustering analysis combined with mapping of SS2 transcriptomes to the droplet 10X dataset and visualization of the integrated datasets, confirmed the presence of LTi-like ILC3 and TC I-IV (**Fig. 3b**, **Extended data Fig. 3d,e**). Within the SS2 dataset, Aire$^+$ mTECs clustered with TC I (**Extended data Fig. 3d,e**). Nevertheless, a direct comparison revealed unique expression of epithelial genes (*Foxn1, Krt17, Krt8*), the thymic marker gene *Tbata*, and *Fezf2* by mTECs, whereas TC I expressed genes associated with DCs (*Ccr7*, *Cd83, Dpp4*) (**Fig. 3c, Supplementary Table 2)**. In addition, TCs exclusively expressed *Ptprc* (CD45) and RORγt (**Extended data Fig. 3f**), a finding confirmed by analysis of mTECs from *Rorc*$^{Venus\text{-}creERT2}$ mice (**Extended data Fig. 3g**). Despite overlapping markers, TCs clustered separately from Aire$^+$ DCs (**Extended data Fig. 3d,e**), distinguished by a number of immune regulatory genes (**Fig. 3d, Supplementary Table 3**), underscoring the distinct identity of these two cell types. Index sorting analysis of cell surface markers revealed that TCs spanned a spectrum from CD11c$^{-/lo}$ (TC I) to CD11c^{hi} (TC II-IV) cells (**Fig. 3e,f)**. In addition, TC IV was distinguished by high levels of CD11b expression. These findings suggest that

RORγt$^+$CD11c$^+$CD11b$^+$ cells, previously identified amongst Tbet$^-$ cDC2B[10], represent TC IV. Of note, ILC3s, traditionally identified as CD90(Thy1)$^+$ cells, encompassed both CD90$^-$ and CD90$^+$ cells and did not express CD11c (**Fig. 3e, Extended data Fig. 3h**). To assess whether the partially overlapping transcriptional features of TCs with DC and mTECs are coupled to similar or distinct morphological attributes, we analyzed TC I and TC IV cells by electron microscopy. DCs, Aire$^+$ mTECs, and MHCII$^+$ ILC3 cells served as morphological reference populations. Both TC I and TC IV cells were distinct from classical CCR7$^+$ and CCR7$^-$ DCs, as well as mTECs, and featured distinctive mitochondria with rounded, condensed cristae **(Extended data Fig. 3i**). In contrast to DCs and TCs, MHCII$^+$ ILC3s had a typical lymphoid appearance. To further probe TC localization and morphology, we examined their spatial distribution in the mLN of *Rorgt^{cre}R26$^{lsl\text{-}tdtomato}$*Aire^{GFP} mice at P18. Both TC I and TC IV cells were preferentially located within the deep T cell zone, co-localizing with Foxp3$^+$ Treg cells which were similar concentrated within the deep T zone (**Extended data Fig. 3j,k**).

To address the developmental relationship between TCs and DCs, we analyzed DC fate-mapping *Clec9a$^{cre/wt}$R26$^{lsl\text{-}tdTomato}$Rorc$^{Venus\text{-}creERT2}$*Aire^{GFP} reporter mice in which both cDC1 and cDC2 are labelled by virtue of Clec9a expression in DC progenitors[16]. Surprisingly, TCs were not Clec9a fate-mapped, in contrast to RORγt$^-$Aire(GFP)$^+$CCR7$^+$ DC2s (**Fig. 3g**). Given reports of a lymphoid pathway for cDC differentiation[17], we tested whether TCs represented DCs derived from this alternative pathway by using a novel *Rag1$^{RFP\text{-}creERT2}$R26$^{lsl\text{-}YFP}$* mouse model to fate-map progeny of lymphoid progenitor cells following neonatal 4-hydroxytamoxifen (4-OHT) administration. In contrast to efficient labeling of lymphoid subsets, YFP$^+$ cells were absent amongst Lin$^-$MHCII^{hi}CXCR6$^-$ cells encompassing TCs (**Extended data Fig. 3l**). These findings suggested that TCs are ontogenically and transcriptionally distinct from classical DCs. Thus, the overlapping phenotype of TCs and DCs, in particular CCR7$^+$ DCs, likely reflects shared functions related to cell migration and antigen presentation rather than shared ontogeny (**Fig. 3h**).

To determine the transcription factors (TF) that regulate TC differentiation and heterogeneity, we turned to our scATAC-seq data, integrating differential TF motif activity with gene expression, in order to pinpoint candidate transcriptional regulators. This analysis identified canonical ILC3 TFs in ILC3s including RORα, GATA3 and TCF1, as well T-bet in NCR$^+$ ILC3s (**Fig. 3i**), validating our approach. All TC subsets were distinguished by a unique group of TFs including Spi-B, a critical regulator of mTEC differentiation, as well as core TFs governing myeloid cell differentiation (PU.1, BCL11A, IRF8 and IRF4) (**Fig. 3i**), in agreement with their transcriptional overlap with both mTECs and DCs. Notably, several of the signature TC TFs have been shown to regulate Aire expression in mTECs[18], suggesting a conserved transcriptional network shared by these two cell types. Together these findings establish the unique identity of TCs,

delineating their shared and distinct features with both mTECs and DCs, and reveal phenotypic heterogeneity within the TC lineage.

TC subsets are enriched for distinct tolerance-related gene expression programs

To gain further insight into the function of TC subsets, we examined their distinguishing transcriptional features (**Figure 4a**). TC I expressed canonical Aire$^+$ mTEC II markers, including *Aire*, *CD80*, *CD86*, *Tnfrsf11b* (OPG), and genes associated with neuronal adhesion, signaling and growth (*Nrxn1*, *Nrn1, Ncam1*). Of note, a recent study identified a population of 'mTEC-like' Aire$^+$RORγt$^+$ cells within lymph nodes, similarly distinguished by neuronal genes[19], likely representing TC I. TC II was distinguished by exclusive expression of several distinctive genes including *pIgR*, *Cldn7* and *Cfp*. Interestingly, pIgR and claudin-7 are signature molecules for a group of mTECs with a history of Aire expression (mTEC III), highlighting further parallels between mTEC and TC subsets. TC III expressed high levels of Aire as well as *Nlrc5*, *H2-M2,* and *Apobec1*, suggesting unique roles in self-antigen presentation. TC IV expressed immune-regulatory genes (*Cd274*) as well as genes associated with cell migration (*Marcks*, *Cxcl16).* Given the critical role of Aire in promiscuous expression and presentation of tissue-restricted antigens (TRAs)[20], we analyzed TRA expression levels amongst TCs. This revealed significant enrichment of TRA genes in TCs, relative to MHCII$^+$ ILC3s, most pronounced amongst TC I-III (**Figure 4b**). Analogous to TRA expression in mTECs[12], TRA expression in TCs was stochastic and spanned antigens unique to organs that are specifically affected by autoimmune inflammation in Aire deficiency, including pancreas, testis, adrenal and liver (**Extended data Fig. 4a,b, Supplementary Table 3**), suggesting potential roles for Aire$^+$ TCs in peripheral tolerance to self-antigens. Analysis of differential chromatin accessibility and motif enrichment across TC subsets identified several subset-specific TFs, further underpinning the observed TC heterogeneity (**Fig. 4c**).

To determine if TC subsets had similar or distinct functional properties, we used Latent Dirichlet Allocation (LDA), a probabilistic topic model, to capture shared and unique gene expression programs. This revealed distinct functional modules, enriched within each TC subset, as well as one shared module between TC III and IV (topic 2; **Extended Data Fig. 4c**,**d)**. Notably, TC IV was enriched for a 'Treg' module encompassing critical molecules including IL-2, the TGF-β activating integrin Itgβ8, CD40 and Ccl22 (**Fig. 4d**), suggesting that TC IV promotes pTreg generation. To validate the observed TC phenotypes, we devised a panel of markers for flow cytometry (**Fig. 4e**). MHCII^{hi}RORγt$^+$ ILC3s were distinguished from TCs by expression of CXCR6 (**Fig. 4e,f**). Amongst Lin$^-$CXCR6$^-$RORγt$^+$MHCII^{hi} cells, we confirmed the presence of TC subsets as well as expression of key immune-regulatory molecules (**Fig. 4f**). Together, these analyses suggest that TC subsets are molecularly and functionally distinct, and

highlight potential roles for TCs in peripheral immune tolerance, in particular a role for TC IV in pTreg differentiation.

Antigen presentation by ILC3s or DCs is not required for intestinal tolerance

Given the overlapping phenotype of TCs with mTECs and DCs, two professional APCs with known roles in thymic T cell tolerance, we hypothesized that TCs were the relevant RORγt$^+$MHCII$^+$ cell type for instructing pTreg cell differentiation. A direct comparison of TC and ILC3 transcriptomes, as well as cell surface protein expression, confirmed that TCs were enriched for molecules associated with antigen presentation, T cell activation and cell migration, in contrast to MHCII$^+$ ILC3 cells (**Extended data Fig. 5a,b**). Furthermore, in contrast to TCs, we did not observe CCR7 expression on MHCII$^+$ ILC3s (**Extended data Fig. 5c**), despite expression of *Ccr7* transcript, making their migration from large intestine to mLN unlikely. The unique expression of RORα by MHCII$^+$ ILC3s (**Extended data Fig. 5d**), suggested the utility of *Rora^{cre}* mice as a means to selectively ablate MHCII on ILC3s. Since RORα is expressed by ILC progenitors (ILCp), we first analyzed RORα fate-mapped cells using *Rora^{cre}R26$^{lsl\text{-}tdomato}$* mice, to determine the lineage relationship between ILCps and TCs. Within the mLN, all MHCII-expressing ILCs represent RORγt$^+$ ILC3s (**Extended data Fig. 5e**) and close to 90% of these cells were tagged with tdTomato vs <0.5% Lin$^-$CXCR6$^-$MHCII$^+$ cells encompassing DCs and TCs (**Extended data Fig. 5f**), confirming that TCs are not developmentally related to ILCs. Importantly, ILC3s were the only MHCII$^+$ RORα fate-mapped cell type within the mLN (**Extended Data Fig. 5f**), establishing *MHCII$^{\Delta ROR\alpha}$* mice as a selective model for studying the functional role of ILC3 antigen presentation. Indeed, analysis of 3-week-old *MHCII$^{\Delta ROR\alpha}$* mice confirmed a complete loss of MHCII expression on ILC3s, with no changes in TCs (**Extended data Fig. 5g,h**). Surprisingly, the intestinal T cell composition was not perturbed in *MHCII$^{\Delta ROR\alpha}$* mice, with equivalent proportions and numbers of CD4$^+$ Teff and Treg cells, including RORγt$^+$ pTregs cells residing within the mLN and large intestine (**Extended data Fig. 5i,j**). Besides ILC3s, antigen presentation by sub-immunogenic DCs is thought to favor T cell tolerance. Although TCs may have been inadvertently targeted by studies utilizing “DC-specific” Cre drivers due to expression of both CD11c and Zbtb46, the absence of Clec9a fate-mapped TCs, allowed us to revisit a role for classical DC in pTreg differentiation through analysis of *Clec9a$^{cre/cre}$H2-Ab1$^{fl/fl}$* (*MHCII$^{\Delta DC}$*) mice in which DCs but not TCs were rendered MHCII-deficient (**Extended Data Fig. 5k**). Surprisingly, we did not observe changes in RORγt$^+$Foxp3$^+$ cells in these mice (**Extended Data Fig. 5l**). Overall, these findings demonstrate that MHCII antigen presentation by ILC3s or cDCs is dispensable for pTreg cell differentiation, leaving TCs as the likely pTreg-inducing RORγt$^+$ APC.

A developmental wave of TCs during early life induces pTreg cells in an Itgb8 dependent manner

Given the narrow temporal window of opportunity for establishing intestinal immune tolerance, we hypothesized that the presence of TCs might determine this developmental window. Our analysis of TC abundance in mice ranging from 7 days to 6 weeks of age revealed their striking enrichment between 1 and 3 weeks of age, with rapid decline thereafter (**Fig. 5a**). Notably, TCs, in particular TC IV cells, were specifically enriched within the mLN compared to skin-draining peripheral LN (pLN) (**Fig. 5b**). To determine the dynamics of neonatal TC differentiation, we used *Rorc*$^{Venus\text{-}creERT2}$*R26*$^{lsl\text{-}tdtomato}$ mice to label RORγt-expressing cells and their progeny. Following treatment of mice with 4-OHT at P1, over 60% of TCs remained tdTomato^{+} at P7. This proportion fell to 15% by P14, although total numbers of both tdtTomato^{-} and tdTomato^{+} TCs increased between P7–14 reflecting *de novo* TC differentiation or proliferation during this critical developmental window. Both the proportion of tdTomato^{+} TCs and total cell numbers declined from P14 reflecting waning differentiation beyond this age (**Extended data Fig. 6a,b**). In contrast, the proportion of fate-mapped MHCII^{+} ILC3s declined between P7 and 14 but remained stable thereafter (**Extended data Fig. 6a**), consistent with the notion that ILC3s are maintained by self-renewal. Thus, TCs or putative RORγt^{+} TC progenitors, are present at birth, and are prominently enriched within the mLN at the time of intestinal tolerance induction during early life. The developmental window for TCs, their enrichment within mLN, and their role in pTreg induction, suggested a role for the microbiota in instructing TC differentiation. However, the proportion and number of TCs was similar in germ-free (GF) mice compared to age-matched progeny that had been conventionalized with a normal microbiota (**Extended data Fig. 6c**), indicating that the differentiation of TCs is independent of microbial cues. Aire is highly conserved between mouse and human with similar roles in immune tolerance. To determine if TC counterparts exist in humans, we analyzed a recent single-cell atlas encompassing second trimester to adult intestine and mLN[22]. Within a group of cells annotated as 'lymphoid' DCs we identified a cluster of cells distinct but closely related to CCR7^{+} DCs, expressing signature TC genes (*TNFRSF11B, SPIB*) including high levels of *AIRE* (**Extended data Fig. 6d-g**). Analysis of orthologous signature TC subset genes confirmed enrichment within the putative human TC cluster, most prominently for TC III and TC IV defining genes (**Extended data Fig. 6h**). In contrast to CCR7^{+} DCs, human TCs were almost exclusively present within the mLN (**Extended data Fig. 6i**) and were highly enriched within fetal samples (32% vs 3.8%; **Extended data Fig. 6j**), further implying a conserved role for TCs in early life intestinal tolerance.

The close temporal and spatial relationship between TC IV and pTreg cells supported a key role for this TC subset in pTreg differentiation. Indeed, our earlier droplet RNA-seq analysis had highlighted expression of Itgb8, a TGF- β activating integrin, by TCs. TGF-β signaling is known to be critical for pTreg cell differentiation. Activation of latent extracellular TGF-β requires physical interaction with either integrin

αvβ6 or αvβ8 and loss of either TGF-β signaling on T cells, or Itgβ8 expression by hematopoietic cells, leads to impaired pTreg differentiation and development of autoimmunity and colitis[23,24]. Analysis of TGF-β signaling pathway genes in TC transcriptomes profiled by SS2, confirmed high expression of both *Itgav* and *Itgb8* by TC IV along with unique expression of *Tgfb2* (**Fig. 5c**). To address the role of TCs in TGF-β mediated pTreg differentiation, we generated mice with conditional loss of the *Itgb8* gene in RORγt$^+$ cells (*Itgb8*$^{\Delta ROR\gamma t}$). Importantly, although ILC3s expressed Itgav, they did not express Itgb8 (**Fig. 5d, Extended data Fig. 7a**). Furthermore, ATAC-seq analysis showed that *Itgb8* locus was inaccessible in ILC3s (**Extended data Fig. 7a**), confirming that *Itgb8*$^{\Delta ROR\gamma t}$ mice have a specific deficiency of Itgb8 in TCs and T cells. Analysis of mLN and large intestine from 3-week-old mice, revealed a dramatic reduction in pTreg cell frequency and numbers (**Fig. 5d-f**), mirroring the loss of pTreg cells observed in *MHCII*$^{\Delta ROR\gamma t}$ mice. Differentiation of pTreg cells was normal in *Cd4*^{cre}*Itgb8*$^{fl/fl}$ mice (**Extended Data Fig. 7b**), indicating that TGF-β activation by TCs, not T cells, is a critical mechanism for intestinal pTreg cell differentiation. Overall, our findings suggest that TCs, prominently present in the mLN during a critical early life window, are a highly specialized lineage of tolerogenic APCs that play an essential role in establishing intestinal tolerance through the generation of pTreg cells.

Discussion

Our study identified a novel lineage of RORγt$^+$ cells, enriched in the intestine during a critical early life period when host-microbiota symbiosis is first established. Our finding that TCs instruct extra-thymic Treg cell generation provides a cellular basis for the reported time-window for establishment of intestinal immune tolerance. Contrary to the view that 'neonatal immune privilege', first demonstrated in the 1950s, results from the presence of immunosuppressive or 'immature' DCs with an inferior stimulatory capacity, our results suggest the existence of a dedicated lineage of tolerogenic neonatal APCs with a hybrid mTEC-DC phenotype. Intriguingly, pTreg cell abundance in adulthood is determined by cues sensed within the first week of life[25], coincident with the observed wave of TC differentiation, suggesting that modulation of early life TC development may have lasting effects on intestinal immune tolerance.

A defining feature of TCs was their expression of RORγt – a transcription factor shared by ILC3s. A number of critical roles have been attributed to MHCII$^+$ ILC3s, including regulation of cancer immunity, autoimmunity and tolerance to microbiota[8,26–28]. These conclusions are based on experimental MHC class II gene ablation in RORγt-expressing cells as a lineage-specific approach for studying ILC3 function. The finding that MHCII$^+$ ILC3s are dispensable for pTreg generation and mucosal tolerance, raises the question as to whether other ILC3 ascribed functions are mediated by TCs. Although we cannot

formally exclude the possibility that TCs and ILC3s work in concert to establish intestinal tolerance, our data strongly supports a role for TCs as the primary tolerogenic cell type in early life, equipped with critical molecules required for antigen presentation, T cell priming and pTreg differentiation. Future studies using models of ILC3-restricted MHCII deficiency will help to discern the functional significance of MHCII expression by ILC3s. as well as the contribution of both ILC3s and TCs to tolerance in later life.

While the precise ontogeny of TCs remains to be established, a remarkable finding was their convergent differentiation with mTECs, challenging our current view of boundaries between hematopoietic and non-hematopoietic cells. The transcriptional overlap with Aire^{+} mTECs suggests conserved functions and highlights a previously unappreciated symmetry between thymic and peripheral tolerance pathways. Thus, we propose that TCs represent a broadly tolerogenic lineage with roles in tolerance to both self and foreign antigens. Given the prominent role of TCs in pTreg cell generation and intestinal tolerance, future exploration of their biology may yield key insights into mechanisms of immune tolerance and autoimmune and inflammatory disease pathogenesis.

Acknowledgements

We thank Joris van der Veeken, Clarissa Campbell and all members of the Brown and Rudensky labs for their technical support and helpful discussions. We thank the Single Cell Core Facility at MSK for sample processing and sequencing. We acknowledge the use of the Integrated Genomics Operation Core, funded by the NCI Cancer Center Support Grant (CCSG, P30 CA08748), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology. We thank Jay Gardner for discussions, Mark Anderson for provision of Aire^{GFP} mice, Caetano Reis e Sousa for *Clec9a*^{cre} mice, Dean Sheppard for *Itgb8*$^{fl/fl}$ mice, and Sarah Teichmann for assistance with human single cell datasets.

Funding

This work was supported by the Wellcome Trust, the Ludwig Center at Memorial Sloan Kettering, Parker Institute for Cancer Immunotherapy, NCI Cancer Center Support Grant P30 CA08748, NCI grant U54 CA209975 (A.Y.R. and C.L.), NIAID grant R01AI034206 (A.Y.R.) and the Hilton-Ludwig Cancer Prevention Initiative (Conrad N. Hilton Foundation and Ludwig Cancer Research). D.D. was supported by a Medical Scientist Training Program grant from NIGMS of the National Institutes of Health under award number T32GM007739 to the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program and by a F30 Predoctoral Fellowship from NIAID of the National Institutes of Health under award number F30AI154660-01. The laboratory of T.G.P.G. is supported by the Barbara and Wilfried Mohr foundation. A.Y.R. is an HHMI investigator. C.B. was supported by a Wellcome Trust Fellowship (WT201483/Z/16/Z), a Josie Robertson Investigator Award and a Parker Institute for Cancer Immunotherapy Senior Fellowship.

Author Contributions

C.C.B. and A.Y.R. conceived the study and wrote the manuscript. C.C.B. designed and performed experiments and analyzed data. Z.T. designed and performed computational analyses. B.A., G.S., Y.A.P.I, and Y.F.P. performed experiments and analyzed data; D.D. performed immunofluorescence staining and imaging analyses; S.H. and C.F. performed experiments; R.E. performed initial analysis of human intestinal single-cell data; H.A.P. performed electron microscopy analyses; J.V. processed tissues; M.K. performed immunofluorescence staining and imaging analyses for human tissue; L.J. generated mice under the supervision of M.vdB.; G.G. provided mice, T.G.P.G. supervised human IF experiments and C.L. supervised computational analyses. All authors read and approved the manuscript.

Declarations of interest

M.vdB. has received research support and stock options from Seres Therapeutics and stock options from Notch Therapeutics and Pluto Therapeutics; he has received royalties from Wolters Kluwer; has consulted, received honorarium from or participated in advisory boards for Seres Therapeutics, WindMIL Therapeutics, Rheos Medicines, Merck & Co, Inc., Magenta Therapeutics, Frazier Healthcare Partners, Nektar Therapeutics, Notch Therapeutics, Forty Seven Inc., Priothera, Ceramedix, Lygenesis, Pluto Therapeutics, GlaskoSmithKline, Da Volterra, Vor BioPharma, Novartis (Spouse), Synthekine (Spouse), and Beigene (Spouse); he has IP Licensing with Seres Therapeutics and Juno Therapeutics; and holds a fiduciary role on the Foundation Board of DKMS (a nonprofit organization). A.Y.R. is a member of SAB and has equity in Surface Oncology, RAPT Therapeutics, and holds an IP licensed to Takeda, which is not related to the content of this study.

Materials and Methods

Mice

Rorc$^{Venus\text{-}T2A\text{-}creERT2}$ mice were generated by insertion of a targeting construct into the *Rorc* 3-UTR by homologous recombination in embryonic stem (ES) cells on the C57Bl/6 background. The IRES-Venus-T2A-creER-frt-NeoR-frt cassette targeting construct was created by cloning. Homologous arms were retrieved from BAC clone RP24-209K20. To facilitate ES cell targeting Crispr/cas9 system was used. The gRNA was in vitro transcribed using MEGA shortscript T7 kit (Life Tech Corp AM1354) using recombineering techniques. The targeting vector, cas9 protein (Fisher Scientific A36498 Truecut Cas9 Protein v2) and gRNA were co-electroporated into G1 ES cells derived from an F1 hybrid blastocyst of 129S6 x C57BL/6J. The resulting chimeras were bred with FLPeR mice to excise the NEO cassette. *Rag1*$^{RFP\text{-}creERT2}$ (C57BL/6-Tg(Rag1-RFP,-cre/ERT2)33Narl) mice, obtained from the Rodent Model Resource Center (RMRC), were generated by insertion of a BAC transgene comprising the Rag1 promoter and RFP-IRES-creERT2 into ES cells from C57Bl/6 mice.

Adig(Aire^{GFP}), *Clec9a*^{cre}, *Rora*^{cre}, *Itgb8*$^{fl/fl}$ and *Cd4*^{cre} mice have been previously described[16,29–32]. *Rorgt*^{cre}, *H2-Ab1*$^{fl/fl}$, *R26*$^{lsl\text{-}tdTomato}$, *R26*$^{lsl\text{-}YFP}$, *Zbtb46*^{GFP}, C57Bl/6 (CD45.2) mice were purchased from Jackson Laboratories. Generation and treatments of mice were performed under protocol 21-05-007 and 08-10-023, approved by the Sloan Kettering Institute (SKI) Institutional Animal Care and Use Committee. All mouse strains were maintained in the SKI animal facility in specific pathogen free (SPF) conditions in accordance with institutional guidelines and ethical regulations. Both male and female mice were included in the study and we did not observe sex-dependent effects. All mice analyzed were age matched unless otherwise specified. All animals used in this study had no previous history of experimentation and were

naïve at the time of analysis.

Germ Free mice

Germ-free (GF) *Rorc*$^{Venus\text{-}creERT2}$Aire^{GFP} mice were delivered by sterile cesarean section and fostered by germ-free dams. GF animals were bred in flexible isolators (Class Biologically Clean; USA) at Weill Cornell Medicine and fed with autoclaved 5KA1 chow. GF status was checked by aerobic and anaerobic cultures of fecal samples for bacteria and fungi and by PCR of fecal DNA samples for bacterial 16S and fungal/yeast 18S genes.

SPF Conventionalization

Freshly collected feces from C57Bl/6J mice were used to generate experimental specific pathogen free (SPF) mice for experiments comparing immune cell populations in litters from GF or SPF mice. Feces were manually homogenized in 1mL of sterile PBS. Solid material from the fecal homogenate was removed by brief centrifugation at low speed. 100uL of a 2-fold dilution of this fecal slurry was administered to *Rorc*$^{Venus\text{-}creERT2}$Aire^{GFP} mice by gavage. SPF animals were maintained on the same diet as GF mice and housed in the same manner.

Tissue processing

Mice were euthanized by CO_2 inhalation. Organs were harvested and processed as follows. Lymphoid organs were digested in collagenase in RPMI1640 supplemented with 5% fetal calf serum, 1% L-glutamine, 1% penicillin–streptomycin, 10 mM HEPES, 1 mg/ml collagenase A (Sigma, 11088793001) and 1U/mL DNase I (Sigma, 10104159001) for 45 min at 37°C, 250 rpm. Large intestine was removed, flushed with PBS and incubated in PBS supplemented with 5% fetal calf serum, 1% L-glutamine, 1% penicillin–streptomycin, 10 mM HEPES, 1 mM dithiothreitol, and 1 mM EDTA for 15 min to remove the epithelial layer. Samples were washed and incubated in digest solution for 30 min. 1/4 inch ceramic beads (MP Biomedicals, 116540034) were added to large intestine samples (3 per sample) to aid in tissue dissociation. Digested samples were filtered through 100-µm strainers and centrifuged to remove collagenase solution. Thymus samples were minced with scissors followed by enzymatic digestion in RPMI1640 supplemented with 10% fetal calf serum, 1% L-glutamine, 10mM HEPES, 62.5ug/ml Liberase™ and DNase I 0.4mg/ml. Density-gradient centrifugation using a three-layer Percoll gradient with specific gravities of 1.115, 1.065 and 1.0 was used to enrich for stromal cells for flow cytometric analysis. For sorting of mTECs, single cell suspension of digested thymocytes were depleted of CD45^{+} cells using CD45 microbeads (Miltenyi Biotec).

Flow cytometry

For flow cytometric analysis, dead cells were excluded either by staining with LIVE/DEAD Fixable Violet or Zombie NIR in PBS for 10 minutes at 4°C, prior to cell-surface staining. Cells were then incubated with anti-CD16/32 in staining buffer (2% FBS, 0.1% Na azide, in PBS) for 10 minutes at 4°C to block binding to Fc receptors. Extracellular antigens were stained for 20-30 minutes at 4°C or RT (CCR7 staining) in staining buffer. For transcription factor analysis, cells were fixed and permeabilized with Ebioscience Foxp3 kit per manufacturer instructions. Intracellular antigens were stained for 30min at 4°C in 1x Perm/Wash buffer. Live cells were treated with DNase (0.08 U/ml) for 10min at RT and washed with staining buffer prior to acquisition on a BD LSR or Cytek Aurora™. 123count eBeads™ were added to quantify absolute cell numbers. The antibodies used for flow cytometry and FACS are listed in Supplementary Table 4.

Histological analysis of intestinal inflammation

Mice were euthanized by CO_2 inhalation and large intestines were harvested and immediately placed into 10% formalin. Histopathological assessment for inflammation scoring in the intestine was performed on H&E stained sections based on established scoring systems for intestinal inflammation in mouse models[33]. Assessment includes severity and extent of inflammatory cell infiltrates, epithelial changes and mucosal architecture changes. Briefly, severity and extent of inflammatory cell infiltrate in the mucosa and if extending to submucosa and muscularis were evaluated histologically. Other evaluations include proliferation of epithelial cells lining the mucosa villous atrophy, crypts, loss of goblet cells, crypt abscesses, erosions and ulceration.

Multiome scRNA and scATAC-sequencing

For single cell RNA and ATAC seq of RORγt$^+$MHCII$^+$ cells, mLN from 2 w old (P14–17) *Rorc* $^{Venus\text{-}creERT2}$ mice were pooled from 16 biological replicates and processed as described earlier. Cells were depleted of Lineage (TCRb, TCR$\gamma\delta$, CD19, B220, NK1.1)$^+$ cells via staining with biotinylated antibodies followed by magnetic bead negative selection. Cells were incubated with anti-CD16/32 in sorting buffer (2% FBS in PBS) for 10 minutes at 4°C to block binding to Fc receptors. Extracellular antigens were stained for 30 minutes at 4°C in sorting buffer (2% FBS, 2mM EDTA, in PBS). Cells were washed and resuspended in sorting buffer with SYTOX blue (Invitrogen) for exclusion of dead cells. Live, CD45$^+$Lin(TCRb, TCR$\gamma\delta$ CD19)$^-$RORγt(Venus)$^+$ MHCII$^+$ cells were then sort purified. Cells were sorted into cRPMI, before being pelleted and resuspended in RPMI-2% FBS. Single Cell Multiome ATAC + Gene Expression was performed with 10X genomics system using Chromium Next GEM Single Cell Multiome Reagent Kit A

(catalog no. 1000282) and ATAC Kit A (catalog no. 1000280) following Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Kits User Guide and demonstrated protocol - Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing. Briefly, >50,000 cells (viability 95%) were lysed for 4min and resuspended in Diluted Nuclei Buffer (10x Genomics, PN- 2000207). Lysis efficiency and nuclei concentration was evaluated on Countess II automatic cell counter by trypan blue staining. 9,660 nuclei were loaded per transposition reaction, targeting recovery of 6,000 nuclei after encapsulation. After the transposition reaction, nuclei were encapsulated and barcoded. Next-generation sequencing libraries were constructed following User Guide, which were sequenced on an Illumina NovaSeq 6000 system.

Plate-based Smart-seq2 sequencing

RORγt$^+$MHCII$^+$ cells were enriched from a pool of MLN from 3-week-old (P21) *Rorc*$^{Venus\text{-}creERT2}$ mice. Cells were depleted of Lineage (TCRb, TCR$\gamma\delta$, CD19, B220, NK1.1)$^+$ cells via staining with biotinylated antibodies followed by magnetic bead negative selection. Live, Lin(CD3, TCRβ, TCR$\gamma\delta$ CD19, B220, NK1.1)$^-$CD64$^-$Ly6C$^-$MHCII$^+$RORγt(Venus)$^+$ cells were then sorted into single wells. Cells were also stained for CD90, CD11c and CD11b for acquiring index sorting information on cell surface expression. Aire$^+$ mTECs were enriched from a pool of thymi from 3-week old mice via staining with biotinylated antibodies against CD45 followed by magnetic bead negative selection. CD45$^-$Epcam$^+$MHCII$^+$Aire(GFP)$^+$ cells were sorted into single wells. Aire$^+$ DCs were enriched from a pool of MLN from the same 3-week old mice. Cells were depleted of Lin$^+$ cells as described above and live, Lin(CD3, TCRβ, TCR$\gamma\delta$ CD19, B220, NK1.1)$^-$CD90$^-$CD64$^-$Ly6C$^-$CD11c$^+$MHCII$^+$Aire(GFP)$^+$ cells were then sorted into single wells. Retrospective index sorting analysis confirmed that Aire(GFP)$^+$ cells were CD11c^{lo}MHCII^{hi}, representing CCR7$^+$ DCs.

Single cells were sorted into Buffer RLT (Qiagen). Cell lysates were immediately sealed and spun down before transferring to dry ice and storing at -80 ºC. RNA was purified using the Agencourt RNAClean XP beads (Beckman Coulter) at a 2.2X ratio. First-strand cDNA synthesis was achieved using Maxima H Minus Reverse Transcriptase (ThermoFisher) according to the manufacturer's protocol using oligo dT primers, with the addition of a custom template-switch oligo in a 1mM final concentration. cDNA was amplified for 24 cycles using KAPA HiFi HotStart ReadyMix (Kapa Biosystems KK2601). After PicoGreen quantification, 0.1-0.2ng of cDNA was used to prepare libraries with the Nextera XT DNA Library Preparation Kit (Illumina) in a total volume of 6.25µL with 12 cycles of PCR. Indexed libraries were pooled by volume and cleaned by aMPure XP beads (Beckman Coulter) at a 1X ratio. Pools were sequenced

on a HiSeq 4000 in a PE50 or PE100 run using the HiSeq 3000/4000 SBS Kit (Illumina). An average of 1.8 million paired reads were generated per sample and the percent of mRNA bases per sample averaged 63%.

Mouse single-cell RNA-seq and single-cell ATAC-seq computational analysis

Pre-processing of the 10X multiome scRNA-seq and scATAC-seq for RORγt$^+$MHCII$^+$ cells

Single-cell RNA-seq and ATAC-seq FASTQ files were aligned to mm10 (Cell Ranger mouse reference genome mm10-2020-A-2.0.0) and counted by Cell Ranger ARC v2.0.0 with default parameters. The barcodes were filtered based on the number of RNA-seq transcripts (>1k and <50k), the number of detected genes (>500 and <6k), and the fraction of mitochondrial transcripts (<15%). Barcodes were further filtered based on the number of ATAC-seq fragments (3.5 < log10{nFrags} < 4.5) and TSS enrichment score (>4). Arrow files were created from the scATAC-seq fragments using ArchR v1.0.1[34], and doublets were identified and removed with default parameters. Finally, any genes detected in <2 cells in the scRNA-seq data were discarded, leaving 20779 genes. After clustering the scRNA-seq data (described in 'Dimensionality reduction, cell clustering, and visualization'), and based on the expression of marker genes, we identified 5 minor *Rorc*$^-$ or *Ptprc*$^-$ contaminant clusters (glial cells; cluster 17, pDCs; cluster 18, Rorc$^-$DCs; cluster 19, and mono/macrophages; clusters 20-21) which were excluded from downstream analyses. In total, 10145 cells remained, with a median scRNA-seq library-size of 3150 and a median of 13885 in the number of ATAC-seq fragments.

Pre-processing of the Smart-seq2 scRNA-seq dataset

Smart-seq2 sequencing data from demultiplexed samples was aligned to the mouse reference genome using STAR v2.7.7a[35] with '--twopassMode Basic --outFilterMultimapNmax 1 --quantMode TranscriptomeSAM'. Sequence reads were aligned and annotated using a STAR index created from GENCODE GRCm38 (mm10) release M25 primary assembly genome and gene annotations[36]. Alignment files were individually name-sorted using Samtools v1.11[37], and then used to create a cell-by-gene count matrix using featureCounts[38] (subread v2.0.1). The count matrix was filtered based on the number of transcripts (>50K), number of detected genes (>1300), and the fraction of mitochondrial transcripts (<8%). Finally, genes detected in <2 cells were discarded. A total of 481 cells remained, with a median library size of 924319 from 27195 genes.

Dimensionality reduction, cell clustering, and visualization

For each scRNA-seq dataset, the filtered count matrix was library-size normalized, log-transformed ('log-normalized' expression values) and then centered and scaled ('scaled' expression values) using Seurat v4.0.4. Principal component analysis (PCA) was performed on the scaled data (npcs=50). PhenoGraph clustering[39] was performed using the first *N* principal components (PCs) with *k* nearest neighbors (*N*=30 and *k*=30 for the multiome scRNA-seq data; *N*=20 and *k*=30 for the Smart-seq2 dataset; *N*=30 and *k*=20 for the human gut DCs). Cell clustering was visualized using UMAP[40], computed from the nearest neighbor graph built by PhenoGraph.

The multiome scATAC-seq data analysis was restricted to the cells in clusters 1-16 of the scRNA-seq results, as previously described for pre-processing. Latent Semantic Indexing (LSI) was performed on 100000 top variable tiles (500 bp genomic bins) identified after 10 iterations of 'IterativeLSI' by ArchR. Tiles from non-standard chromosomes, chrM, and chrY were not included in this analysis. Cells were clustered (method=Seurat, k.param=30, resolution=1.2) and visualized with UMAP (nNeighbors=30) using 30 LSI components. In both the scRNA-seq and scATAC-seq data, we identified several clusters of LTi cells (scRNA clusters 9-16 and scATAC clusters 7-13). These clusters showed weak pairwise matchings between scRNA and scATAC; therefore, they were combined as one group of LTi cells for downstream analyses, where stated.

Differential gene expression tests

Differentially expressed genes (DEGs) between groups of cells were identified with MAST[41], performed using Seurat functions. MAST was run on the log-normalized expression values. In all tests, genes were only considered if they were detected in at least 10% of the cells in at least one of the two groups compared (min.pct=0.1, logfc.threshold=0). In one-vs-rest DE tests comparing multiple groups, each group was compared to all the cells from other groups. Specific DE comparisons are described in the results. DEGs were reported according to their fold change (>1.5) and adjusted *p*-value (<0.01). Ribosomal and mitochondrial genes were removed from the final list of genes reported/visualized. Where stated, the top DEG markers were subsequently selected for each group, based on fold change.

Data imputation for scRNA-seq data

MAGIC imputation[42] was applied to the log-normalized expression values for the multiome scRNA-seq dataset to further de-noise and recover missing values. Imputed gene expression values were only used for data visualization on UMAP overlays and heatmaps, where stated.

Cell-cycle scores

Using standard Seurat functions, we computed cell cycle scores for known S-phase and G2/M-phase marker genes[43] to identify proliferating cells.

Topic modeling for scRNA-seq data

'Topics' were identified by fitting a Latent Dirichlet Allocation (LDA) model, also known as a Grade of Membership (GoM) model[44], to the raw gene expression count matrix for TCs (clusters 1-5 of the multiome scRNA-seq data) using CountClust v1.18.0[45]. Genes that were detected in fewer than 10 TCs were not included. The optimal number of topics (K=8) was selected among values ranging from 3 to 15 with the maximum Bayes Factor (BF). The role of a topic in each cell is measured by the degree to which it represents that topic, and the topic weights sum up to 1 in each cell. The importance of a gene for each topic is measured by how distinctively differentially expressed it is in that topic, by measuring the KL-divergence of its relative gene expression to other topics, assuming a Poisson distribution. One topic, defined by ribosomal and mitochondrial genes and shared across all clusters, was removed from the topic model visualizations.

Integrating the Smart-seq2 dataset with the multiome dataset

RORγt$^+$MHCII$^+$ transcriptomes (based on cell-type as sorted) from the SS2 dataset were integrated with transcriptomes from the 10X multiome scRNA-seq data, using Seurat[46]. Based on the variability of genes in both datasets, 5000 top scoring genes were selected by Seurat functions to identify 'integration anchors' with Canonical Correlation Analysis (CCA). Expression values for these genes were integrated, scaled, and used for PCA. A UMAP embedding was computed from the first N=30 PCs (k=30). Additionally, using Seurat functions, the RORγt$^+$MHCII$^+$ cells from the SS2 dataset (query) were mapped to multiome scRNA-seq clusters (reference) by projecting the PCA from the reference onto the query to identify 'transfer anchors', and then assigning a prediction score for each reference cluster to query cells. The cluster identity with the highest score is chosen as the predicted label for each cell.

Single-cell enrichment scores for gene sets

Given a set of genes, we standardized the log-normalized expression values of each gene across cells and then averaged these values for all genes in the set, assigning an enrichment score to each cell. Where stated, these scores were standardized across cells and reported as z-scores.

Creating pseudo-bulk samples from scRNA-seq data

Pseudo-bulk samples were created by averaging the unimputed log-normalized gene expression values for each cluster. In cases where scaled values were used for downstream analyses, these average expression values were standardized across the pseudo-bulk samples.

Tissue-restricted antigen (TRA) enrichment in multiome scRNA-seq data

An established list of 6611 TRA genes was taken from previously published data[20] and filtered to 4587 TRA genes detected in the multiome scRNA-seq dataset. We computed an enrichment score of this gene set for each cell, excluding proliferating cells (cluster 6 and 7) and ILC3p cells (cluster 1).

Identifying TC-enriched TRAs by gene set enrichment analysis (GSEA)

To avoid potential noise from single-cell data in GSEA, we created pseudo-bulk samples for clusters in the multiome scRNA-seq data. GSEA v4.1.0[47,48] was performed to compare the enrichment of TRAs in non-proliferating TC samples (clusters 2-5) versus LTi samples (clusters 9-16), with 'log2_Ratio_of_Classes' as the gene ranking metric and otherwise default settings. 1554 TRAs with 'core enrichment' were selected as top TRAs for TCs.

Similarity of multiome scRNA-seq clusters to bulk microarray ImmGen samples

The RMA-normalized and log2-transformed gene expression data of 224 bulk microarray samples from a publicly available ImmGen dataset was downloaded from https://www.haemosphere.org[49]. For each gene, the probeset with the highest mean expression was retained. We included all cell types isolated from naïve, untreated mice. Pseudo-bulk samples were generated from the multiome scRNA-seq data for each TC subset, and non-proliferating MHCII$^+$ ILC3s (NCR$^+$ ILC3 and LTi cells). The gene expression vectors were scaled across bulk/pseudo-bulk samples within each dataset, and their pairwise cosine similarities were used to compare the samples. These similarity scores were computed from the expression of 2399 DEGs (FC >1.3, adjusted P <0.01) comparing the scRNA-seq clusters in a one-vs-rest test, that were also expressed in the microarray data. The proliferating and progenitor clusters were excluded from the DE test, and the LTis were grouped together.

Similarity of TCs to thymic epithelial cells

scRNA-seq profiles of CD45$^-$ thymic epithelial cells were downloaded from a publicly available dataset (GSE103967)[12]. The raw counts were library-size normalized, log-transformed, and used to create pseudo-bulk samples for each thymic epithelial cluster. Pseudo-bulk samples were also generated to represent the multiome scRNA-seq TC clusters (2-5). These pseudo-bulk gene expression vectors were

scaled across samples within each dataset, and their pairwise cosine similarities were used to compare clusters from the two datasets. These similarity scores were computed from the expression of 1740 DEGs (FC >1.3, adjusted *P* <0.01) identified in a one-vs-rest DE test for non-proliferating TC clusters (2-5), that were also expressed in the thymic epithelial cells.

scRNA-seq dataset of human gut DCs

Dendritic cells (annotated as cDC1, cDC2, or Lymphoid DC) within the myeloid dataset from the human gut atlas[22] were re-clustered. From the gene markers for each TC subset (one-vs-rest DE test for non proliferating TC scRNA-seq clusters, FC >1.5, adjusted *P* <0.01, we identified orthologous human genes that were uniquely mapped by gprofiler2 and computed enrichment scores for TC sub-set gene signatures for each human cell.

Peak calling for the multiome scATAC-seq data

For peak-calling of the scATAC-seq data, clusters for similar cell types were grouped: C1 (TC IV), C2-4 (TC I,II,III), C5-6 (NCR+ ILC3), and C7-13 (LTi). Filtered ATAC-seq fragments for each group were extracted from ArchR arrow files. We performed MACS2 v2.2.7.1 on fragments of each group with '--gsize mm --qval 0.01 --nomodel --ext 200 --shift -100 --call-summits'. The peak summits were extended by 100 bp in each direction. Regions extending outside of mm10 chromosomes, arising from chrY or chrM, overlapping with blacklist regions precompiled by ArchR (merged from the ENCODE mm10 v2 blacklist regions from https://github.com/Boyle-Lab/Blacklist/blob/master/lists/mm10-blacklist.v2.bed.gz and mitochondrial regions that are highly mappable to the mm10 nuclear genome from https://github.com/caleblareau/mitoblacklist/blob/master/peaks/mm10_peaks.narrowPeak), or containing 'N' nucleotides (>0.001 of the sequence) were filtered. Regions from all groups were compiled and overlapping regions were merged to their union, resulting in a non-overlapping set of 176942 peaks. A peak-by-cell count matrix was created by ArchR with a 'ceiling' value of 4 for the counts to avoid strong biases.

Transcription factor (TF) motif enrichment with chromVAR

The peaks that were accessible in <10 cells were filtered from the peak insertion counts, created as described in the previous section, and the resulting 176898 x 10145 peak-by-cell count matrix was used for motif enrichment with chromVAR v1.14.0[50]. Mouse motif PWMs were downloaded from the CIS-BP database[51] ('Mus_musculus_2022_01_14_6-40_pm'), and the missing PWMs were extracted from 'mouse_pwms_v1' in chromVARmotifs v0.2.0. The GC content of the peaks was computed with chromVAR, and motifs were matched to them by motifmatchr v1.14.0. Then, chromVAR 'deviations' of

the motifs were computed for the peak-by-cell count matrix. The 'top motif' for each TF was selected by correlating its log-normalized gene expression values (from multiome scRNA-seq) with the deviation *z*-scores of its motifs, in the same cells, and picking the motif with the highest Pearson correlation coefficient. Finally, TF-motif pairs with a correlation higher than 0.1 were selected. This resulted in 56 top TFs, out of 739 CIS-BP TFs that were expressed (i.e. had any transcripts detected) in the multiome scRNA-seq profiles. The same process was repeated for the 139528 x 1552 peak-by-cell count matrix of TCs (multiome scATAC-seq clusters 1-4) and the peaks accessible in at least 10 TCs. Out of 652 CIS-BP TFs that were expressed in TCs, 68 had a TF-motif correlation higher than 0.1 and were selected as top TFs for TCs.

Neonatal 4-OH Tamoxifen administration

For labeling of RORγt+ cells, *Rorc*$^{Venus\text{-}creERT2}$Aire^{GFP} mice were injected intra-peritoneally (i.p.) on P1 with 25μg 4-OH-tamoxifen (4-OHT) and analyzed on P8, 15 and 21. For RAG1 fate-mapping, *Rag1*$^{RFP\text{-}creERT2}$*R26*$^{lsl\text{-}YFP}$ mice were injected with 25μg 4-OHT i.p. on P3, 5 and 7 and analyzed on P15.

Electron microscopy

RORγt^{+}MHCII^{+} cells were enriched from a pool of mLN from P18 *Rorc*$^{Venus\text{-}creERT2}$ mice (for TC IV or reference CCR7^{-} and CCR7^{+} DCs) and P18 Aire^{GFP} mice (for TC I). Cells were depleted of Lineage (TCRb, TCRgd, CD19, B220, NK1.1)$^{+}$ cells via staining with biotinylated antibodies followed by magnetic bead negative selection. Live, Lin(TCRb, TCRgd, CD19, B220, NK1.1)$^{-}$CD64^{-}Ly6C^{-} MHCII^{+}RORγt(Venus)$^{+}$CD11c^{+}CD11b^{+}cells (TC IV), Lin^{-}RORγt(Venus)$^{-}$CD11c^{lo}MHCII^{hi} (CCR7^{+} DC), Lin^{-}RORγt(Venus)$^{-}$CD11c^{hi}MHCII^{-} (CCR7^{-} DC), or Lin^{-}CXCR6^{-}CD11c$^{-/lo}$Aire(GFP)^{hi} (TC I) were then sorted directly into 2% glutaraldehyde, 4% PFA, and 2 mM $CaCl_2$ in 0.1 M sodium cacodylate buffer (pH 7.2), fixed for >1 h at room temperature, postfixed in 1% osmium tetroxide, dehydrated in acetone, and processed for Epon embedding. Ultrathin sections (60–65 nm) were counterstained with uranyl acetate and lead citrate. Images were taken with a transmission electron microscope (Tecnai G2-12; FEI, Hillsboro, OR) equipped with a digital camera (AMT BioSprint29).

Immunofluorescence: tissue preparation, microscopy, and image analysis

Mesenteric lymph nodes (mLNs) were dissected from 2–3 week-old *Rorgt*^{cre}*R26*$^{lsl\text{-}tdt}$AIRE^{GFP} mice and trimmed of fat using a dissection scope and forceps. mLNs were fixed in 2% paraformaldehyde for 4 hours in 4°C, washed 3x in PBS, and dehydrated in 30% sucrose in 0.1 M phosphate buffer overnight (16-20 hours). mLNs were embedded in optimal cutting temperature (OCT) compound, frozen on dry ice,

and stored at -80°C. 15-20μ sagittal sections were placed on Superfrost Plus microscopy slides and stored at -20C until staining. mLN sections were permeabilized using 0.2% Triton X-100 for 15 minutes at room temperature, washed 3x with PBS, blocked in 5% rabbit and donkey serum for 1 hour at room temperature, and washed 3x with PBS. Next, the sections were incubated with the following primary antibodies in PBS overnight at 4°C: CD11c BV421 (Biolegend 117329 Clone N418, 1:50), CD4 BV480 (BD Biosciences 565634, Clone RM4-5, 1:50), Foxp3 FITC (ThermoFisher 11-5773-82, Clone FJK-16s, 1:50), goat anti-dsRed (MyBioSource #MBS448092, polyclonal, 1:300), GFP AF647 (ThermoFisher A31852, polyclonal, 1:100), and MHCII AF700 (Biolegend 107621, Clone M5/114.15.2, 1:200) antibodies. The samples were washed 3x in PBS the next day and incubated with the following secondary antibody at room temperature for 2 hours: donkey anti-goat IgG (H+L), Cross-Adsorbed, AF555 (ThermoFisher A-21432, 1:1000). After washing 3x with PBS, samples were mounted in SlowFade Diamond antifade reagent (ThermoFisher S36967). No. 1.5 coverglass was used to seal the slide and all subsequent imaging was done on Leica SP8 microscope. Analysis was performed by histo-cytometry methods using Imaris software. Image segmentation was performed in Imaris using the “Surface Object Creation” module, which employs a seeded region growing, k-means, and watershed algorithm to define individual cells. Once cells coordinates were identified, all analysis was done in R using the spatstat package and ggplot2 for visualization of spatial density patterns[52]. AIRE(GFP) cells were identified based on staining of Anti-GFP AF647 on reporter mice and negative control no reporter mice.

Statistical analysis

Analysis of all data was done with unpaired two-tailed t test, one or two-way ANOVA with a 95% confidence interval, or Mann-Whitney U test, as specified in the text or legends. $P < 0.05$ was considered significant: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. Details as to number of replicates, sample size, significance tests, and value and meaning of *n* for each experiment are included in the Methods or Figure legends. Statistical tests were performed with Prism (GraphPad Software). scATAC and RNA-sequencing experiments were carried out once. Unless otherwise stated, all other experiments were carried out independently at least twice. Mice were non-randomly allocated to experimental groups to ensure equal distribution of genotypes between treatments. Researchers were not blinded as to genotype or treatment during the experiments. No measures were taken to estimate sample size of to determine whether the data met the assumptions of the statistical approaches used. Significance (α) was defined as < 0.05 throughout, after correcting for multiple comparisons.

Data and material availability

The mouse sequencing data are available through the Gene Expression Omnibus under accession GSExxxx.

References

1. Anderson, M. S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science (80-.).* **298**, 1395–1401 (2002).
2. Olszak, T. *et al.* Microbial exposure during early life has persistent effects on natural killer T cell function. *Science (80-.).* **336**, 489–493 (2012).
3. Al Nabhani, Z. *et al.* A Weaning Reaction to Microbiota Is Required for Resistance to Immunopathologies in the Adult. *Immunity* (2019) doi:10.1016/j.immuni.2019.02.014.
4. Sefik, E. *et al.* Individual intestinal symbionts induce a distinct population of RORγ+ regulatory T cells. *Science (80-.).* **349**, 993–997 (2015).
5. Xu, M. *et al.* C-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature* (2018) doi:10.1038/nature25500.
6. Kim, K. S. *et al.* Dietary antigens limit mucosal immunity by inducing regulatory T cells in the small intestine. *Science (80-.).* **351**, 858–863 (2016).
7. Ohnmacht, C. *et al.* The microbiota regulates type 2 immunity through RORγt+ T cells. *Science (80-.).* **349**, 989–993 (2015).
8. Hepworth, M. R. *et al.* Innate lymphoid cells regulate CD4+ T-cell responses to intestinal commensal bacteria. *Nature* (2013) doi:10.1038/nature12240.
9. Papaioannou, N. E. *et al.* Environmental signals rather than layered ontogeny imprint the function of type 2 conventional dendritic cells in young and adult mice. *Nat. Commun. 2021 121* **12**, 1–20 (2021).
10. Brown, C. C. *et al.* Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell* **179**, 846-863.e24 (2019).
11. Yamano, T. *et al.* Aire-expressing ILC3-like cells in the lymph node display potent APC features. *J. Exp. Med.* (2019) doi:10.1084/jem.20181430.
12. Bornstein, C. *et al.* Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. *Nature* (2018) doi:10.1038/s41586-018-0346-1.
13. Park, J. E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science (80-.).* (2020) doi:10.1126/science.aay3224.
14. Nirschl, C. *et al.* Core skin DC signatures control immune tolerance to skin cancer and limit anti-tumor immunity. *J. Immunother. Cancer* (2015) doi:10.1186/2051-1426-3-s2-p205.
15. Maier, B. *et al.* A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature* (2020) doi:10.1038/s41586-020-2134-y.
16. Schraml, B. U. *et al.* XGenetic tracing via DNGR-1 expression history defines dendritic cells as a hematopoietic lineage. *Cell* **154**, (2013).

17. Salvermoser, J. *et al.* Clec9a-mediated ablation of conventional dendritic cells suggests a lymphoid path to generating dendritic cells in vivo. *Front. Immunol.* **9**, 699 (2018).
18. Herzig, Y. *et al.* Transcriptional programs that control expression of the autoimmune regulator gene Aire. *Nat. Immunol. 2016 182* **18**, 161–172 (2016).
19. Wang, J. *et al.* Single-cell multiomics defines tolerogenic extrathymic Aire-expressing populations with unique homology to thymic epithelium. *Sci. Immunol.* **6**, 5053 (2021).
20. Sansom, S. N. *et al.* Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res.* (2014) doi:10.1101/gr.171645.113.
21. Mackley, E. C. *et al.* CCR7-dependent trafficking of RORγ+ ILCs creates a unique microenvironment within mucosal draining lymph nodes. *Nat. Commun.* **6**, 5862 (2015).
22. Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space and time. *Krishnaa T. Mahbubani* **597**, 16 (2021).
23. Travis, M. A. *et al.* Loss of integrin αvβ8 on dendritic cells causes autoimmunity and colitis in mice. *Nat. 2007 4497160* **449**, 361–365 (2007).
24. Lacy-Hulbert, A. *et al.* Ulcerative colitis and autoimmunity induced by loss of myeloid αv integrins. *Proc. Natl. Acad. Sci.* **104**, 15823–15828 (2007).
25. Ramanan, D. *et al.* An Immunologic Mode of Multigenerational Transmission Governs a Gut Treg Setpoint. *Cell* (2020) doi:10.1016/j.cell.2020.04.030.
26. Hepworth, M. R. *et al.* Group 3 innate lymphoid cells mediate intestinal selection of commensal bacteria-specific CD4+ T cells. *Science (80-.).* **348**, 1031–1035 (2015).
27. Goc, J. *et al.* Dysregulation of ILC3s unleashes progression and immunotherapy resistance in colon cancer. *Cell* **184**, 5015-5030.e16 (2021).
28. Grigg, J. B. *et al.* Antigen-presenting innate lymphoid cells orchestrate neuroinflammation. *Nat. 2021 6007890* **600**, 707–712 (2021).
29. Gardner, J. M. *et al.* Deletional tolerance mediated by extrathymic aire-expressing cells. *Science (80-.).* (2008) doi:10.1126/science.1159407.
30. Zembrzycki, A., Chou, S. J., Ashery-Padan, R., Stoykova, A. & O'Leary, D. D. M. Sensory cortex limits cortical maps and drives top-down plasticity in thalamocortical circuits. *Nat. Neurosci. 2013 168* **16**, 1060–1067 (2013).
31. Lee, P. P. *et al.* A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15**, 763–774 (2001).
32. Proctor, J. M., Zang, K., Wang, D., Wang, K. & Reichardt, L. F. Vascular Development of the Brain Requires β8 Integrin Expression in the Neuroepithelium. *J. Neurosci.* **25**, 9940 (2005).

33. Erben, U. *et al.* A guide to histomorphological evaluation of intestinal inflammation in mouse models. *Int. J. Clin. Exp. Pathol.* **7**, 4557–4576 (2014).
34. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet. 2021 533* **53**, 403–411 (2021).
35. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
36. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
37. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
38. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
39. Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184–197 (2015).
40. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).
41. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
42. van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729.e27 (2018).
43. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–96 (2016).
44. Taddy, M. A. On Estimation and Selection for Topic Models. (2012).
45. Dey, K. K., Hsiao, C. J. & Stephens, M. Visualizing the structure of RNA-seq expression data using grade of membership models. *PLOS Genet.* **13**, e1006599 (2017).
46. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
47. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
48. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–50 (2005).
49. Choi, J. *et al.* Haemopedia RNA-seq: A database of gene expression during haematopoiesis in mice and humans. *Nucleic Acids Res.* **47**, D780–D785 (2019).
50. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* (2017) doi:10.1038/nmeth.4401.

51. Weirauch, M. T. *et al.* Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158**, 1431–1443 (2014).
52. Baddeley A, Rubak E, T. R. *Spatial Point Patterns: Methodology and Applications with R*. (Chapman and Hall/CRC, 2015).

Figure 1. RORγt⁺ APCs promote pTreg differentiation and intestinal tolerance during early life

a, Representative flow cytometry of RORγt and Foxp3 expressing T cell subsets and **b**, summary graphs for frequencies of total T_{reg} (Foxp3⁺) and frequency of RORγt⁺ cells among CD4⁺Foxp3⁺ Treg cells in mesenteric lymph nodes (mLN) and large intestine lamina propria (LI) of 3 w old *MHCII*$^{\Delta ROR\gamma t}$ (n=4) and control (*H2-ab1*$^{fl/fl}$) (n=4) mice**. c**, Adult (9 week) *MHCII*$^{\Delta ROR\gamma t}$ (n=5) and control (n=3) mice were analyzed for frequencies of total T_{reg} (Foxp3⁺), pTreg (RORγt⁺Foxp3⁺), and T_H17 (Foxp3⁻RORγt⁺) among T cells in indicated tissues. **d,** Histological analysis of H&E stained sections of the colon. Scale bars represent 200 μm. **e,** Histological colitis score in 12-week-old *MHCII*$^{\Delta ROR\gamma t}$ (n=5) and control (n=3) mice. Error bars: means ± s.e.m. Each symbol represents an individual mouse. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; unpaired two-sided *t*-test.

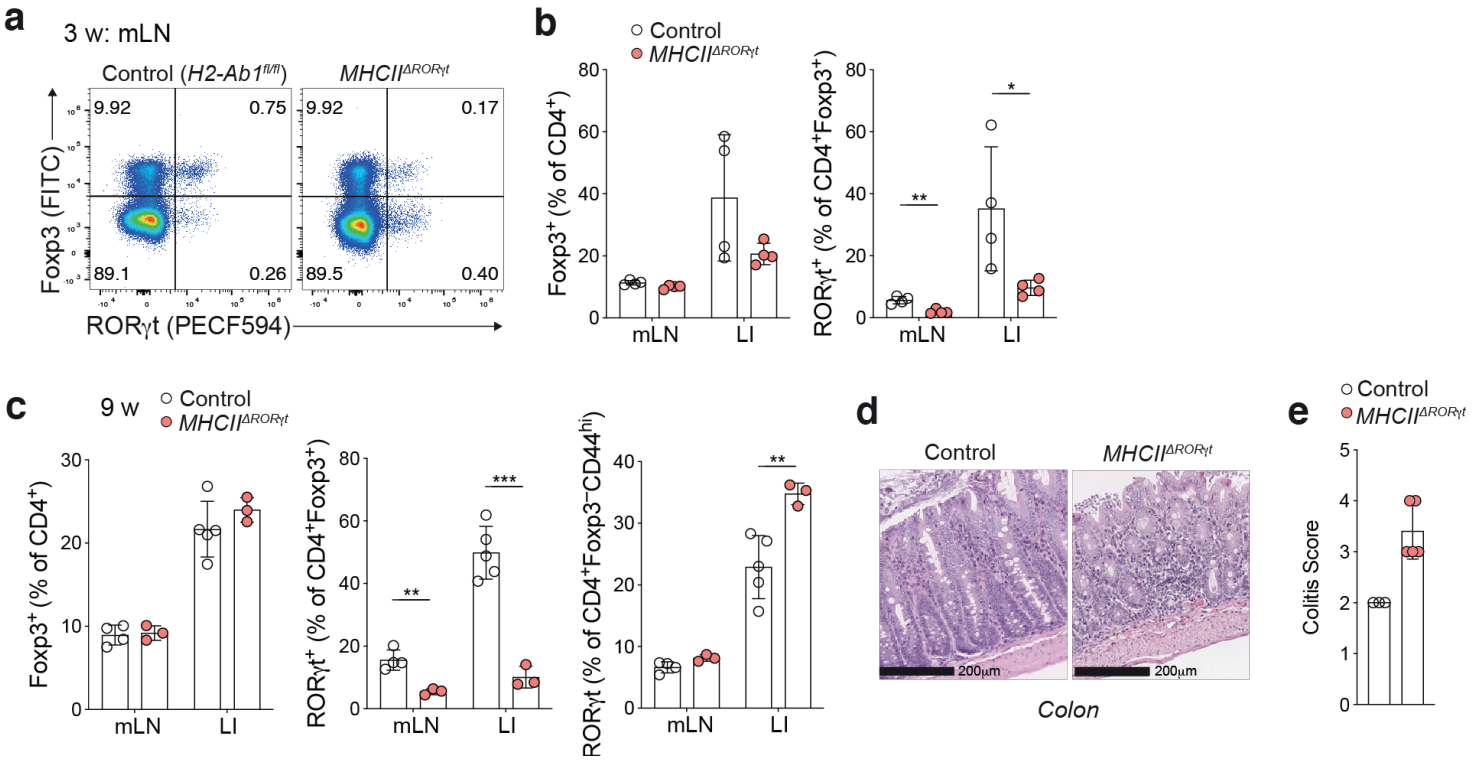


Figure 2. Identification of a novel RORγt⁺ Aire expressing antigen presenting cell lineage

a, Schematic of paired single cell transcriptome and epigenome profiling of Lin⁻RORγt⁺MHCII⁺ cells from mLN of 2-week-old *Rorc*$^{Venus\text{-}creERT2}$ mice (pooled from 16 biological replicates). UMAP visualization of 10,145 cells profiled by scRNA-seq (**b**) or scATAC-seq (**c**), colored by cluster annotation. **d**, Dot plot showing the expression of canonical ILC3 or cluster I-IV marker genes **e**, Similarity between cell types identified in (b) and ImmGen bulk microarray profiles for immune and stromal cells **f**, scRNA-seq UMAP colored by imnputed expression of *Aire*. **g**, Similarity between TC subsets in (b) and mTEC subsets.

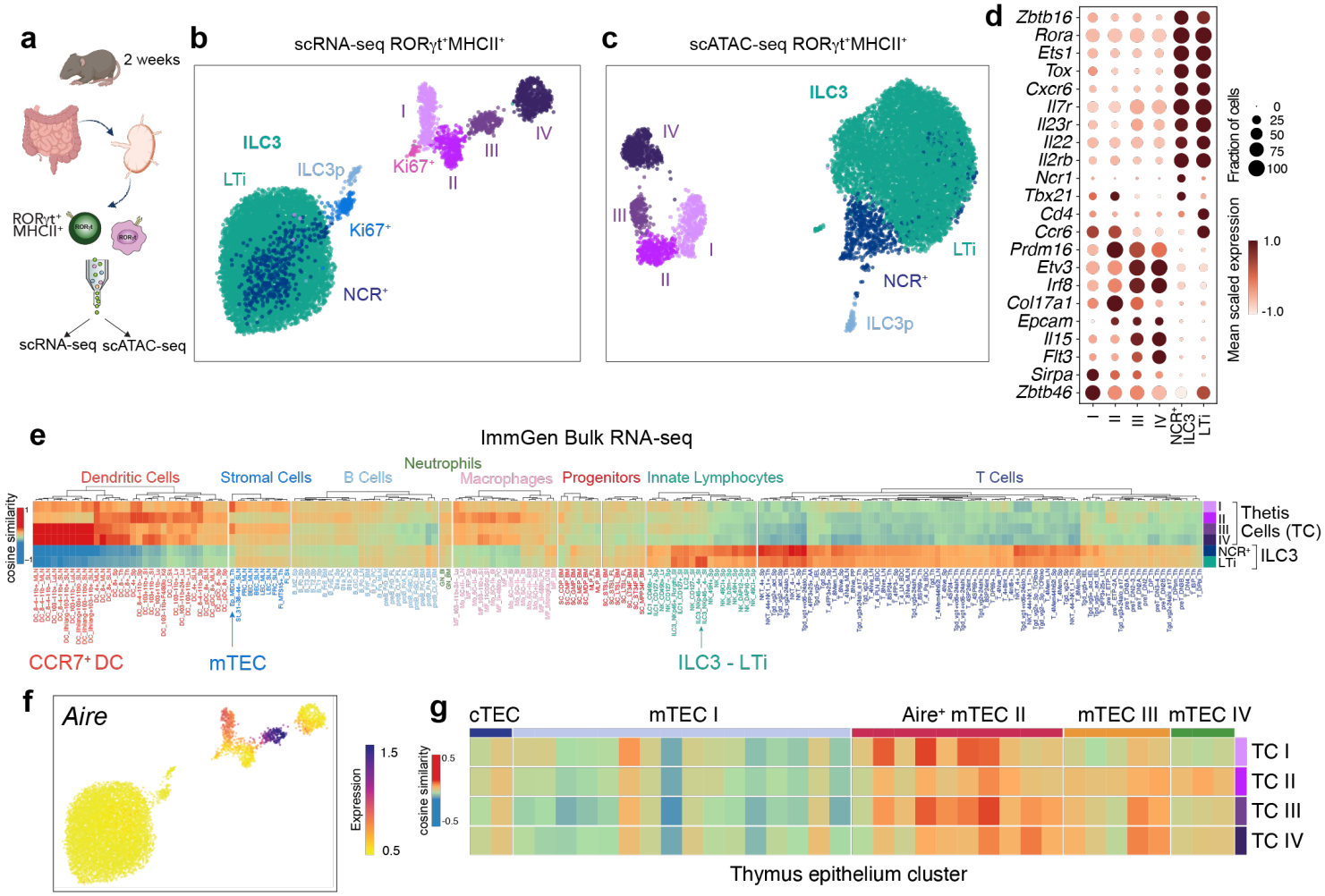


Figure 3. **TCs are ontogenically, transcriptionally and phenotypically distinct from cDCs.**

a, Schematic of single cell transcriptome profiling of RORγt^{+}MHCII^{+} cells from mLN of P21 *Rorc*$^{Venus\text{-}creERT2}$ mice encompassing TCs and MHCII^{+} ILC3s, alongside reference mLN Aire^{+}MHCII^{hi} (CCR7^{+}) DCs and thymic Aire^{+} mTECs from P21 Aire^{GFP} mice. **b**, UMAP visualization of integrated 10X Genomics and Smart-Seq2 (SS2) scRNA-seq analysis for RORγt^{+}MHCII^{+} TCs, colored by SS2 TC transcriptomes or 10X cluster annotation. **c**, Heatmap reporting scaled expression values for top differentially expressed genes (FC>1.5, adj. P<0.01) between Aire^{+} mTEC and TC I. **d**, Heatmap reporting scaled expression values for top differentially expressed genes (FC>1.5, adj. P<0.01) between indicated SS2 clusters (Extended data Fig. 3d). **e**, Flow cytometry of index sorted mLN RORγt^{+}MHCII^{+} cells and **f**, summary graphs for CD11c and CD11b cell surface protein median fluorescence intensity. **g**, tdTomato labeling in cDC and TC from mLN of DC fate-mapping RORγt and Aire double reporter (*Clec9a*$^{Cre/+}$*R26*$^{lsl\text{-}tdtomato}$*Rorc*$^{Venus\text{-}creERT2}$Aire^{GFP}) mice at P18. **h**, Schematic of DC and TC ontogeny demonstrating distinct and overlapping transcriptional regulators and cell surface markers. **i**, Heatmap reporting TF motif activity scores for top TF-motif and gene expression pairs in scATAC/RNA-seq data (Fig. 1b). Color indicates chromVAR motif deviation score. Error bars (**g**): means ± s.e.m. Box plots (**f**) indicate the median (center lines) and interquartile range (hinges), and whiskers represent min and max, dots represent outliers. *P < 0.05; **P < 0.01; ****P < 0.0001; two way ANOVA.

Figure 3

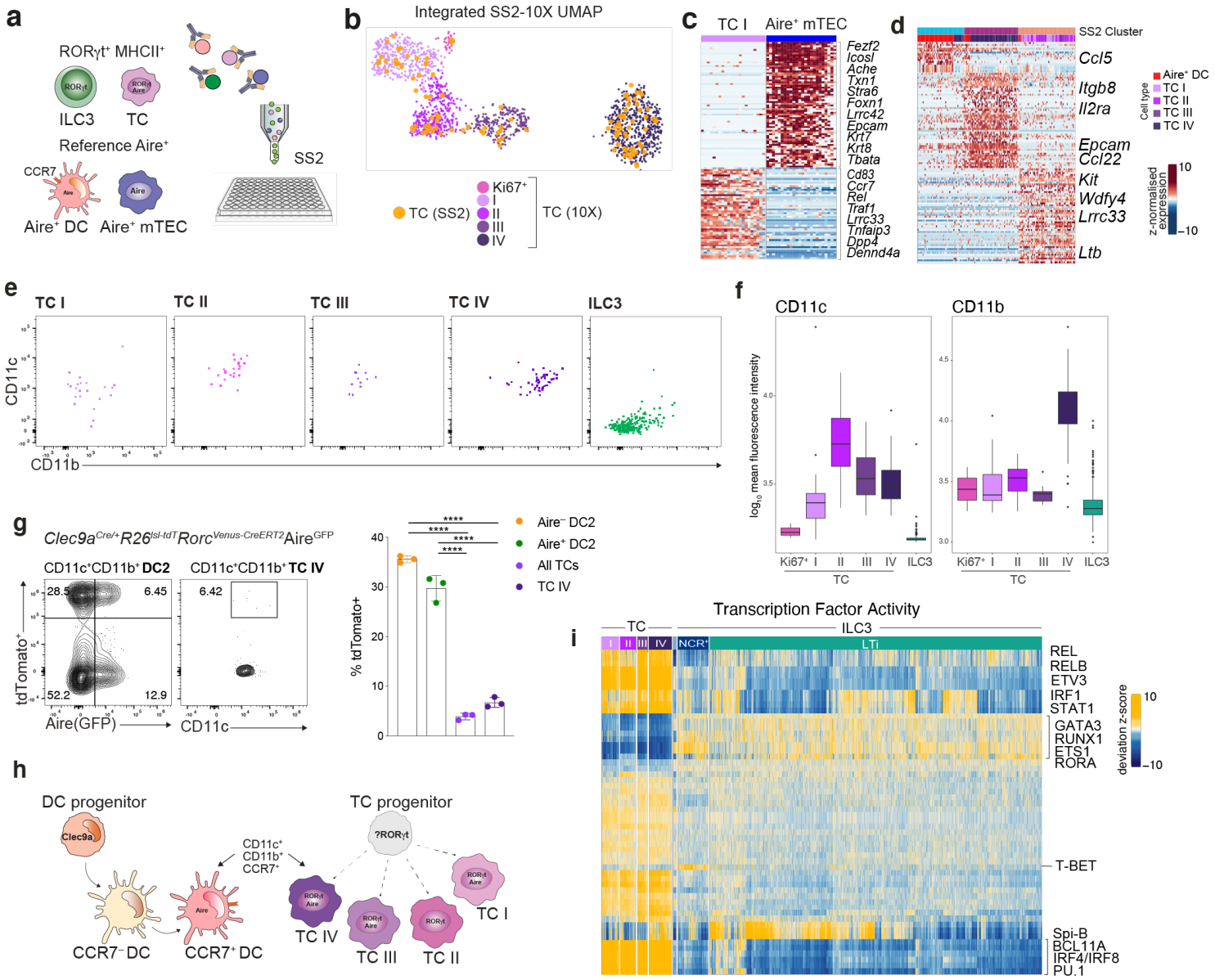

a
RORγt+ MHCII+
ILC3
TC
Reference Aire+
CCR7
Aire+ DC
Aire+ mTEC
SS2
b
Integrated SS2-10X UMAP
TC (SS2)
Ki67+
I
II
III
IV
TC (10X)
c
TC I
Aire+ mTEC
Fezf2
Icosl
Ache
Txn1
Stra6
Foxn1
Lrrc42
Epcam
Krt7
Krt8
Tbata
Cd83
Ccr7
Rel
Traf1
Lrrc33
Tnfaip3
Dpp4
Dennd4a
d
SS2 Cluster
Ccl5
Itgb8
Il2ra
Epcam
Ccl22
Kit
Wdfy4
Lrrc33
Ltb
Cell type
Aire+ DC
TC I
TC II
TC III
TC IV
z-normalised expression
10
−10
e
TC I
TC II
TC III
TC IV
ILC3
CD11c
CD11b
f
CD11c
CD11b
log10 mean fluorescence intensity
Ki67+
I
II
III
IV
ILC3
TC
g
Clec9aCre/+R26lsl-tdTRorcVenus-CreERT2AireGFP
CD11c+CD11b+ DC2
CD11c+CD11b+ TC IV
28.5
6.45
52.2
12.9
6.42
tdTomato+
Aire(GFP)
CD11c
% tdTomato+
Aire− DC2
Aire+ DC2
All TCs
TC IV
****
h
DC progenitor
Clec9a
CCR7− DC
CCR7+ DC
CD11c+
CD11b+
CCR7+
TC progenitor
?RORγt
TC IV
TC III
TC II
TC I
i
Transcription Factor Activity
TC
ILC3
I
II
III
IV
NCR+
LTi
REL
RELB
ETV3
IRF1
STAT1
GATA3
RUNX1
ETS1
RORA
T-BET
Spi-B
BCL11A
IRF4/IRF8
PU.1
deviation z-score
10
−10

Figure 4. Transcriptional and epigenetic features of TC subsets

a, Heatmap showing scaled, imputed expression of top 125 differentially expressed genes (one vs the rest, FC>1.5, adj. *P*<0.01) for each TC cluster. **b**, Enrichment of tissue-restricted antigen (TRA) genes across TC and ILC3 subsets identified in Fig. 1a. **c**, Heatmap reporting scaled chromVAR deviation TF motif scores (left) and corresponding TF gene expression values (right) for top TF gene-motif pairs in TCs in scATAC-seq data. **d**, Topic modeling of scRNA-seq TC transcriptomes infers functional programs. Dot plot showing expression of top 20 genes for topic 7 across TC clusters and UMAP overlay of topic 7 weight demonstrating enrichment in TC IV. **e**, Dot plot showing expression of select cell surface markers, differentially expressed between TC subsets. **f**, Gating strategy for identification of TC subsets and expression of signature molecules. Plots are representative of 6 mice from two independent experiments. Box plots (**b**) indicate the median (center lines) and interquartile range (hinges), and whiskers represent min and max, dots represent outliers. *P*<0.0001 ****; Mann Whitney U test.

Figure 4

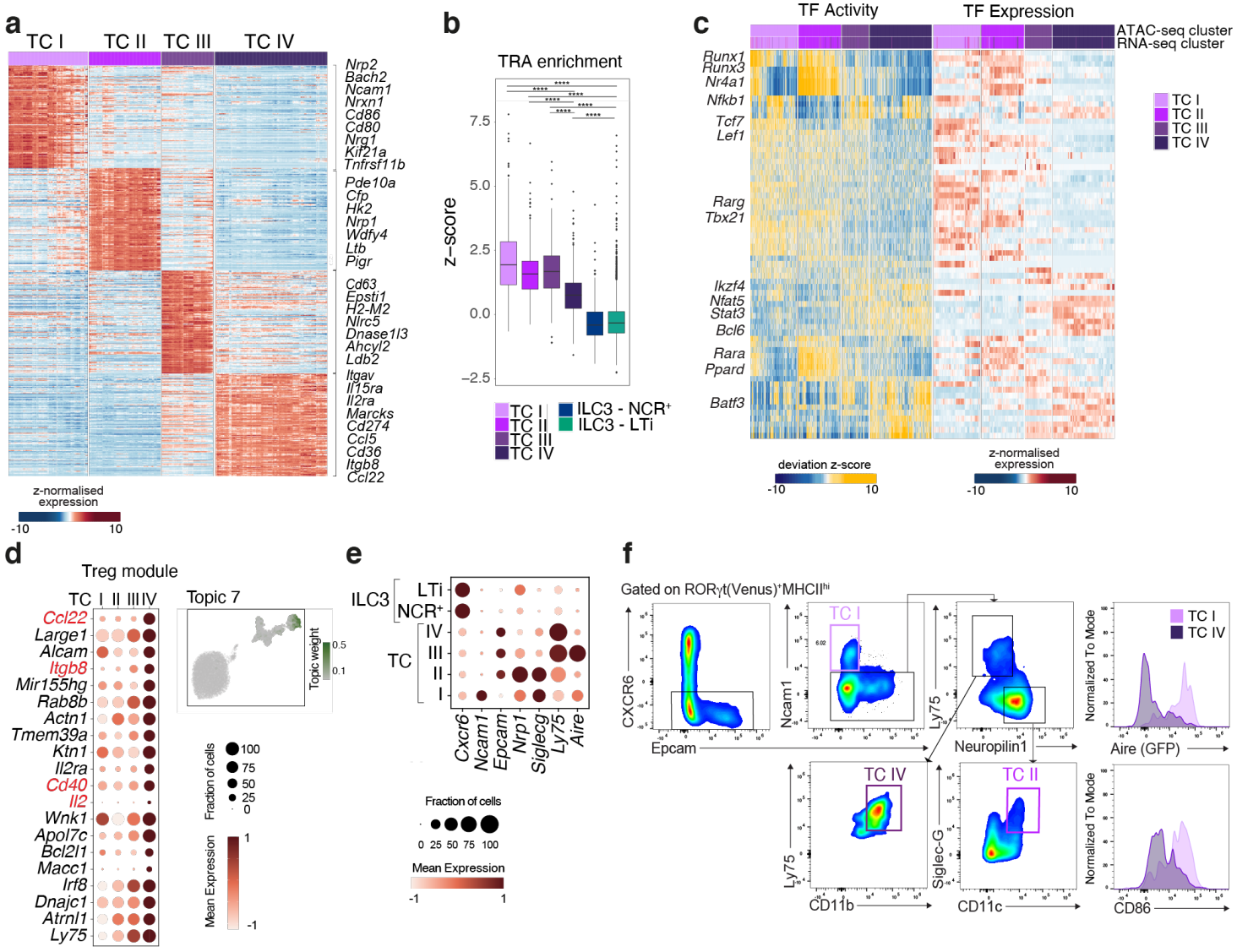
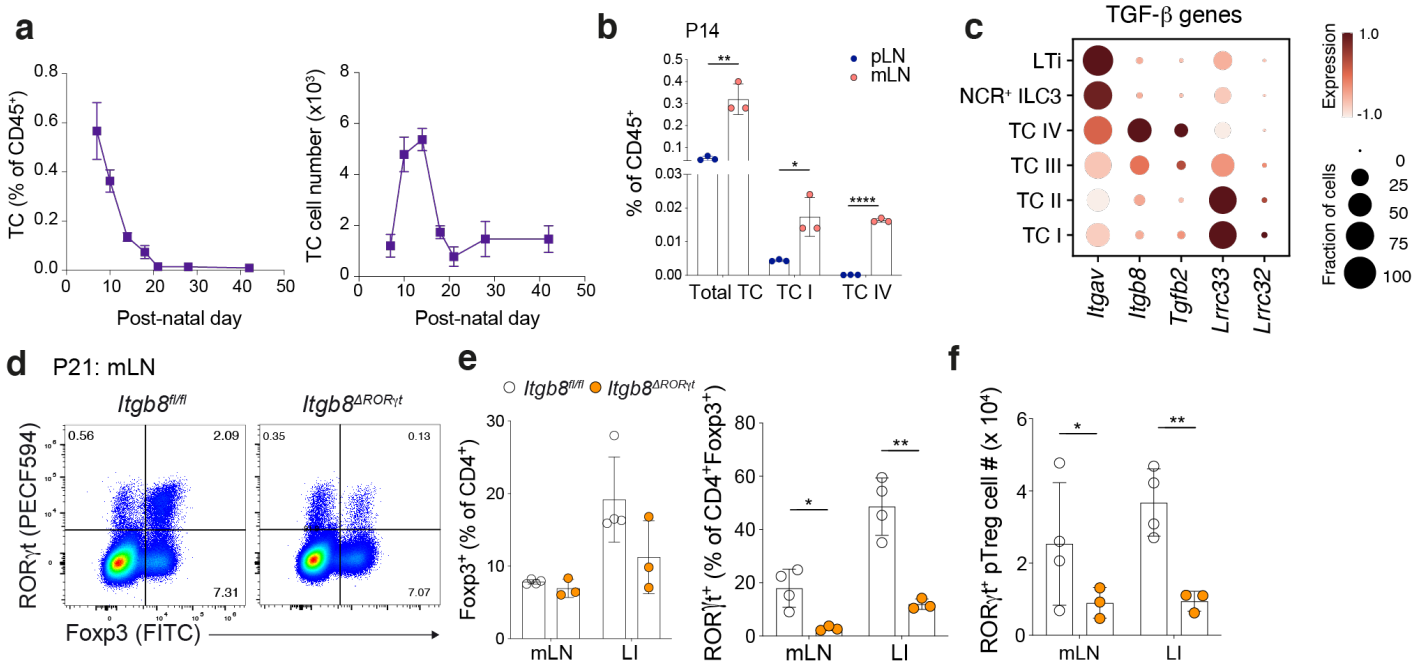

a
TC I
TC II
TC III
TC IV
Nrp2
Bach2
Ncam1
Nrxn1
Cd86
Cd80
Nrg1
Kif21a
Tnfrsf11b
Pde10a
Cfp
Hk2
Nrp1
Wdfy4
Ltb
Pigr
Cd63
Epsti1
H2-M2
Nlrc5
Dnase1l3
Ahcyl2
Ldb2
Itgav
Il15ra
Il2ra
Marcks
Cd274
Ccl5
Cd36
Itgb8
Ccl22
z-normalised expression
-10
10
b
TRA enrichment
z−score
7.5
5.0
2.5
0.0
−2.5
TC I
TC II
TC III
TC IV
ILC3 - NCR+
ILC3 - LTi
c
TF Activity
TF Expression
ATAC-seq cluster
RNA-seq cluster
Runx1
Runx3
Nr4a1
Nfkb1
Tcf7
Lef1
Rarg
Tbx21
Ikzf4
Nfat5
Stat3
Bcl6
Rara
Ppard
Batf3
TC I
TC II
TC III
TC IV
deviation z-score
-10
10
z-normalised expression
-10
10
d
Treg module
TC I II III IV
Ccl22
Large1
Alcam
Itgb8
Mir155hg
Rab8b
Actn1
Tmem39a
Ktn1
Il2ra
Cd40
Il2
Wnk1
Apol7c
Bcl2l1
Macc1
Irf8
Dnajc1
Atrnl1
Ly75
Topic 7
Topic weight
0.5
0.1
Fraction of cells
100
75
50
25
0
Mean Expression
1
-1
e
ILC3
LTi
NCR+
TC
IV
III
II
I
Cxcr6
Ncam1
Epcam
Nrp1
Siglecg
Ly75
Aire
Fraction of cells
0 25 50 75 100
Mean Expression
-1
1
f
Gated on RORγt(Venus)+MHCIIhi
CXCR6
Epcam
TC I
Ncam1
Ly75
Neuropilin1
Normalized To Mode
TC I
TC IV
Aire (GFP)
TC IV
Ly75
CD11b
TC II
Siglec-G
CD11c
Normalized To Mode
CD86

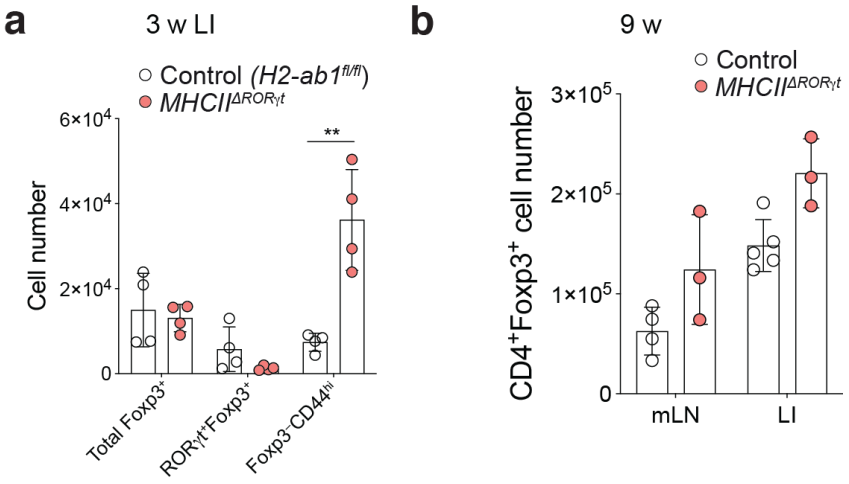
Figure 5. A developmental wave of TCs promotes early life pTreg differentiation in an Itgb8 dependent manner

a Percentages and numbers of TCs within mLN from postnatal day 7 to 6-weeks-of age**.** *n* = 3–8 mice per data point, one to two experiments per timepoint **b**, Frequency of TCs in skin draining lymph nodes (pLN) and mLN of *Rorc*$^{Venus\text{-}creERT2}$Aire^{GFP} mice at P14; *n* = 3 mice per group. **c**, Dot plot showing expression of TGF-β pathway genes in TCs and ILC3s. **d**, Representative flow cytometry of RORγt and Foxp3 expressing T cell subsets and summary graphs for frequencies (**e**) and numbers (**f**) of total T_{reg} (Foxp3$^+$) and pTreg (RORγt$^+$Foxp3$^+$) cells in mLN and large intestine lamina propria of 3 w old *Itgb8*$^{\Delta RORγt}$ (*n* = 3) and *Itgb8*$^{fl/fl}$ (*n* = 4) mice. Error bars: means ± s.e.m. **P* < 0.05; ***P* < 0.01; *****P* < 0.0001; unpaired two-sided *t*-test.



Extended Data Figure 1. Analysis of pTreg cell generation in mice harboring MHC class II-deficient RORγt⁺ APCs.

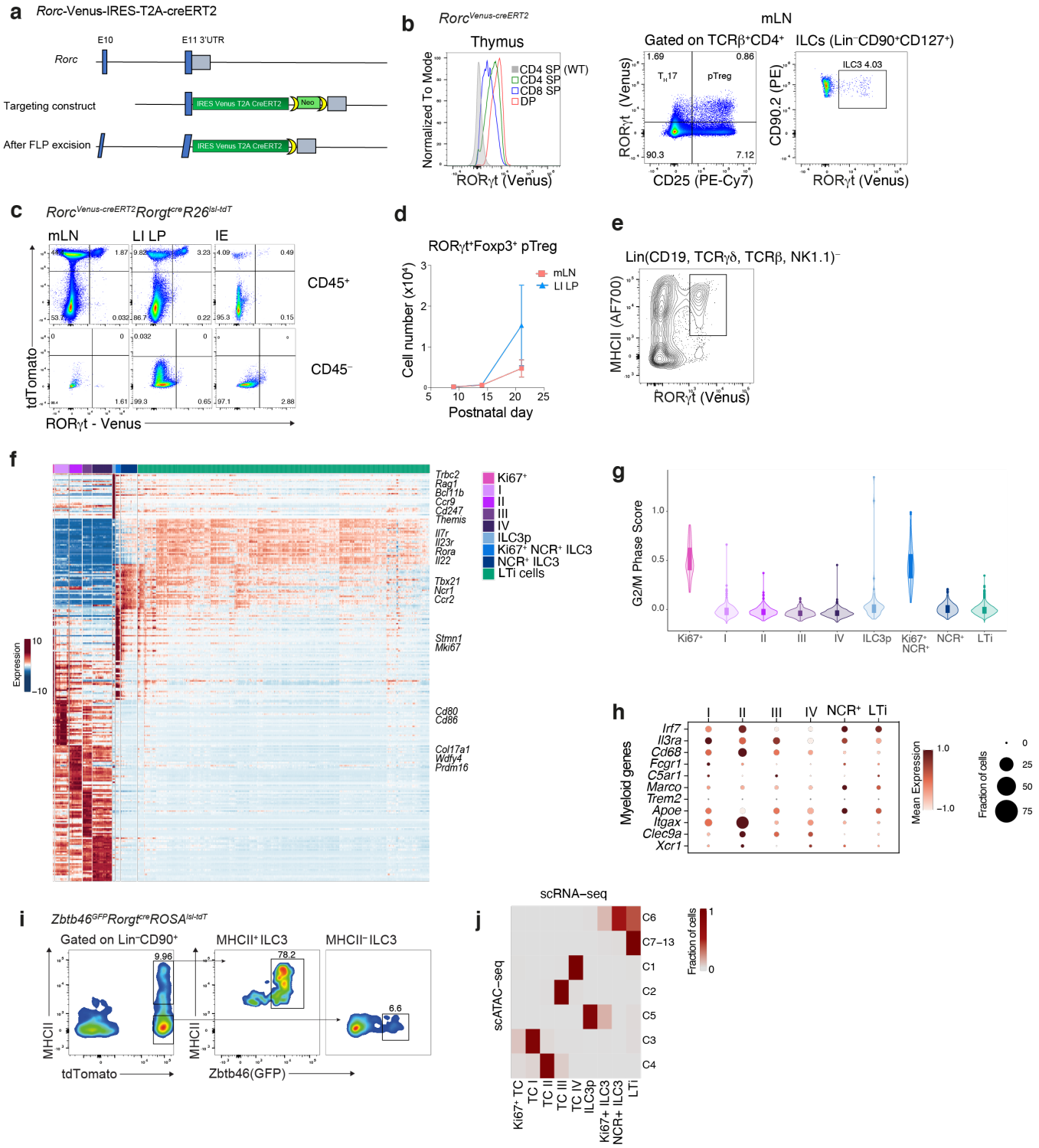
a, Quantification of total T_{reg} (Foxp3$^+$), pTreg (RORγt$^+$Foxp3$^+$), and CD4$^+$ T_{eff} (Foxp3$^-$CD44^{hi}) cells in the large intestine lamina propria (LI) of 3-week-old *MHCII*$^{\Delta ROR\gamma t}$ and control (*H2-Ab1*$^{fl/fl}$) mice (n = 4 mice per group). **b**, Quantification of total T_{reg} (Foxp3$^+$) cells in mLN and LI of 9-week-old *MHCII*$^{\Delta ROR\gamma t}$ (n = 3) and control (n = 5) mice. Each symbol represents an individual mouse. Error bars: means ± s.e.m.. **P < 0.01; unpaired two-sided t-test.



Extended Data Figure 2. Identification of a novel RORγt$^+$Aire$^+$ APC lineage

a, Targeting strategy for the *Rorc* locus. **b,** Flow cytometry of Venus expression in thymocytes (left) or mLN TCRβ$^+$CD4$^+$ T cells (middle) and Lin$^-$CD90$^+$CD127$^+$ innate lymphoid cells (ILC; right) isolated from adult mice. **c,** Flow cytometry of mLN, LI and IEL CD45$^+$ and CD45$^-$ cells in P16 RORγt fate-mapper Rorc reporter (*Rorc*$^{Venus\text{-}creERT2}$*RORgt*^{Cre}*Rosa26*$^{lsl\text{-}tdT}$) mice. Representative of six mice from two independent experiments. **d**, RORγt$^+$Foxp3$^+$ pTreg cell numbers in mLN and LI of *Rorc*$^{Venus\text{-}creERT2}$ mice at indicated postnatal ages, n = 3–4 mice per time-point. **e**, Cell sorting scheme for Lin(TCRβ, TCRγδ, CD19, NK1.1)$^-$ RORγt(Venus)$^+$MHCII$^+$ cells. **f**, Heatmap reporting scaled, imputed expression of top differentially expressed genes for each scRNA-seq cluster (one vs the rest, FC>1.5, P<0.01). **g**, Expression score of cell-cycle genes for each scRNA-seq cluster. **h**, Dot plot showing expression of myeloid genes. **i**, Flow cytometric analysis of Zbtb46 (GFP) expression in ILC3 subsets from mLN of 3-week-old *Zbtb46*^{GFP}*Rorgt*^{cre}*R26*$^{lsl\text{-}tdtomato}$ mice. Representative of four mice from two independent experiments. **j**, Correspondence between cell labels for scATAC-seq and scRNA-seq.

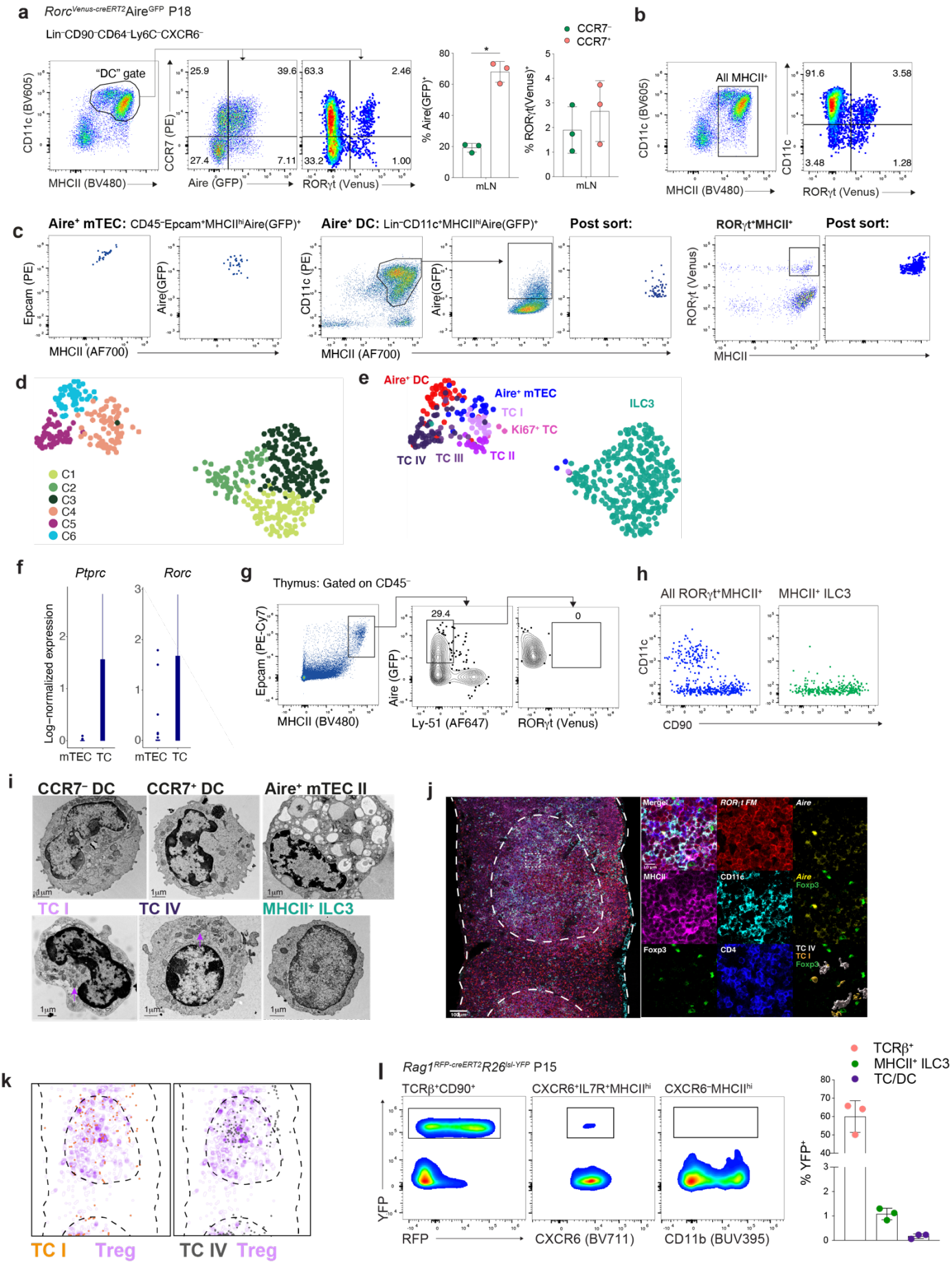
Extended Data Figure 2



Extended Data Figure 3. Distinguishing features of TCs, mTECs and DCs.

a-**b**, Flow cytometric analysis of Lin^{-}CD11c^{+}MHCII^{+} 'DCs' (**a**) and Lin^{-}CXCR6^{-}RORγt(Venus)$^{+}$MHCII^{+} TCs (**b**) in mLN of *Rorc*$^{creERT2\text{-}Venus}$Aire^{GFP} mice at P18. **c**, Flow cytometric analysis of index sorted thymic Aire^{+} mTECs, mLN Aire^{+} DCs isolated from 3-week-old Aire^{GFP} mice and mLN Lin^{-}RORγt^{+}MHCII^{+} cells from 3-week-old *Rorc*$^{Venus\text{-}creERT2}$ mice. **d**-**e**, UMAP visualization of 481 cells (Fig. 3a) colored by (**d**) PhenoGraph cluster or (**e**) reference cell-type or RORγt^{+}MHCII^{+} cell-type as assigned by mapping RORγt^{+}MHCII^{+} SS2 cells to 10X scRNA-seq clusters (Fig. 2b). **f,** Bar graph showing log-normalized expression of *Ptprc* and *Rorc* genes in Aire^{+} mTECs and TCs. **g**, Flow cytometry of RORγt expression in Aire^{+} mTECs isolated from P18 *Rorc*$^{Venus\text{-}creERT2}$Aire^{GFP} mice. Representative plot from one of two independent experiments with n = 3–5 mice. **h**, Index sorting flow cytometric analysis of all RORγt^{+}MHCII^{+} cells (left panel) and cells identified as ILC3 (right panel). **i**, Electron microscopy of CCR7^{-} DCs, CCR7^{+} DCs, Aire^{+} mTECs, TC I, TC IV and MHCII^{+} ILC3 cells. Arrows indicate distinctive mitochondrial cristae in TCs. **j**-**k**, Representative immunofluorescence imaging of TC and Treg markers in mLN sections from 2-week-old *Rorgt*^{Cre}*R26*$^{lsl\text{-}tdtomato}$Aire^{GFP} mice. Images are representative of two independent experiments with similar results. **k**, Immunofluorescence analysis demonstrating cell density heatmaps for Treg cells overlaid with localization of indicated cell type. **l**, Flow cytometry analysis of TCRβ^{+}, MHCII^{+} ILC3, and CXCR6^{-}MHCII^{+} cells encompassing TCs and DCs, from mLN of RAG1 fate-mapped (*Rag1*^{creERT2}*R26*$^{lsl\text{-}YFP}$) mice (n = 3) at P15 following 4-OHT treatment on P3, 5 and 7. Error bars: means ± s.e.m. $^{**}P < 0.01$; unpaired two-sided t-test (**a**).

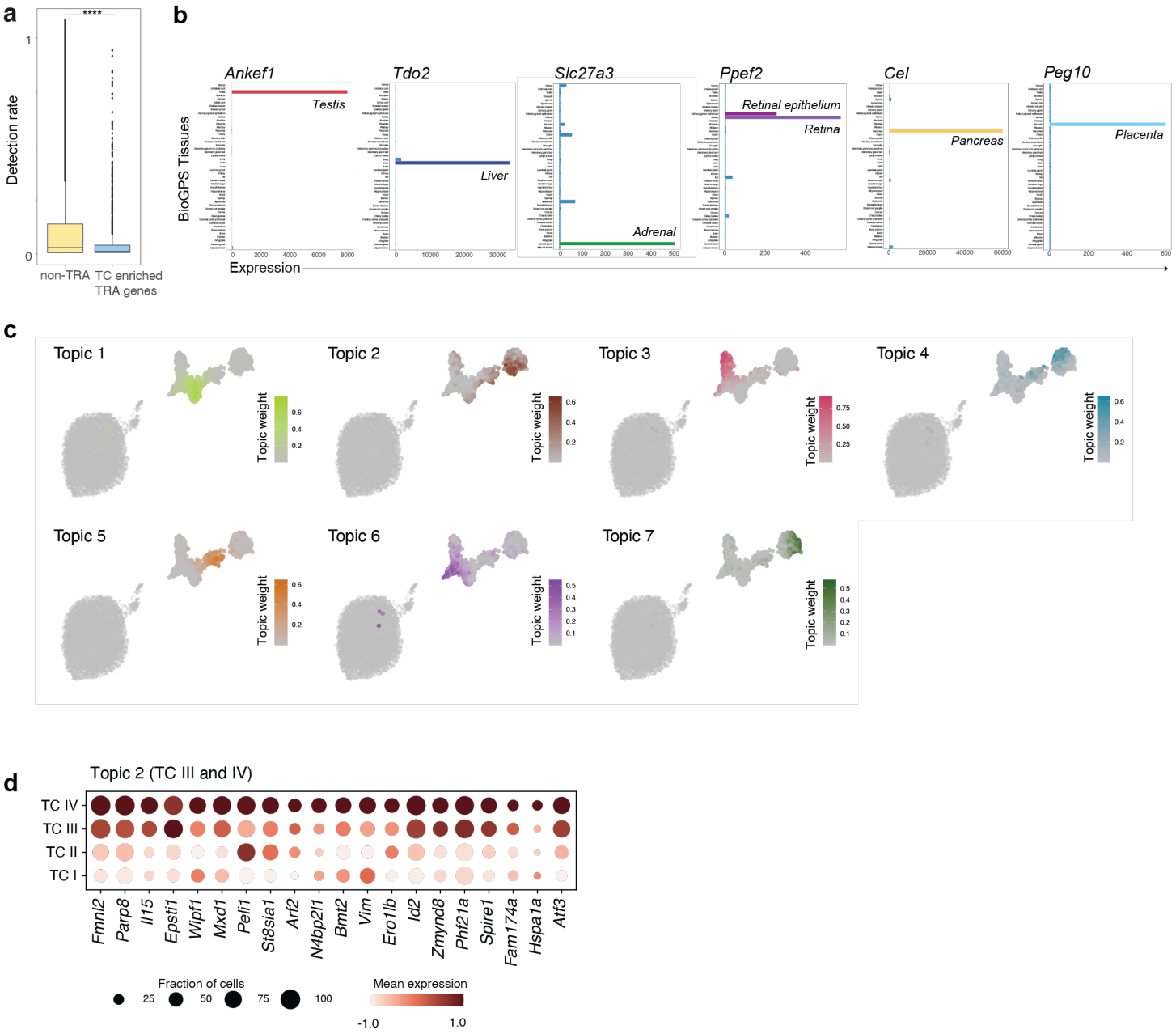
Extended Data Figure 3



Extended Data Figure 4. Characterization of TC subsets

a, Frequency of TRA transcript detection in TCs for TC enriched TRA genes and non-TRA genes. **b**, Expression pattern of example TC-enriched tissue-restricted antigens across tissues (BioGPS). **c**, Topic modeling of scRNA-seq data for TCs identifies distinct and overlapping functional modules between TC subsets. UMAPs colored by the weight of each topic in cells. **d**, Dot plot showing expression of top 20 genes for topic 2 across TC clusters. Box plots (**a**) indicate the median (center lines) and interquartile range (hinges), and whiskers represent min and max, dots represent outliers. $P<0.0001$ ****; Mann Whitney U test.

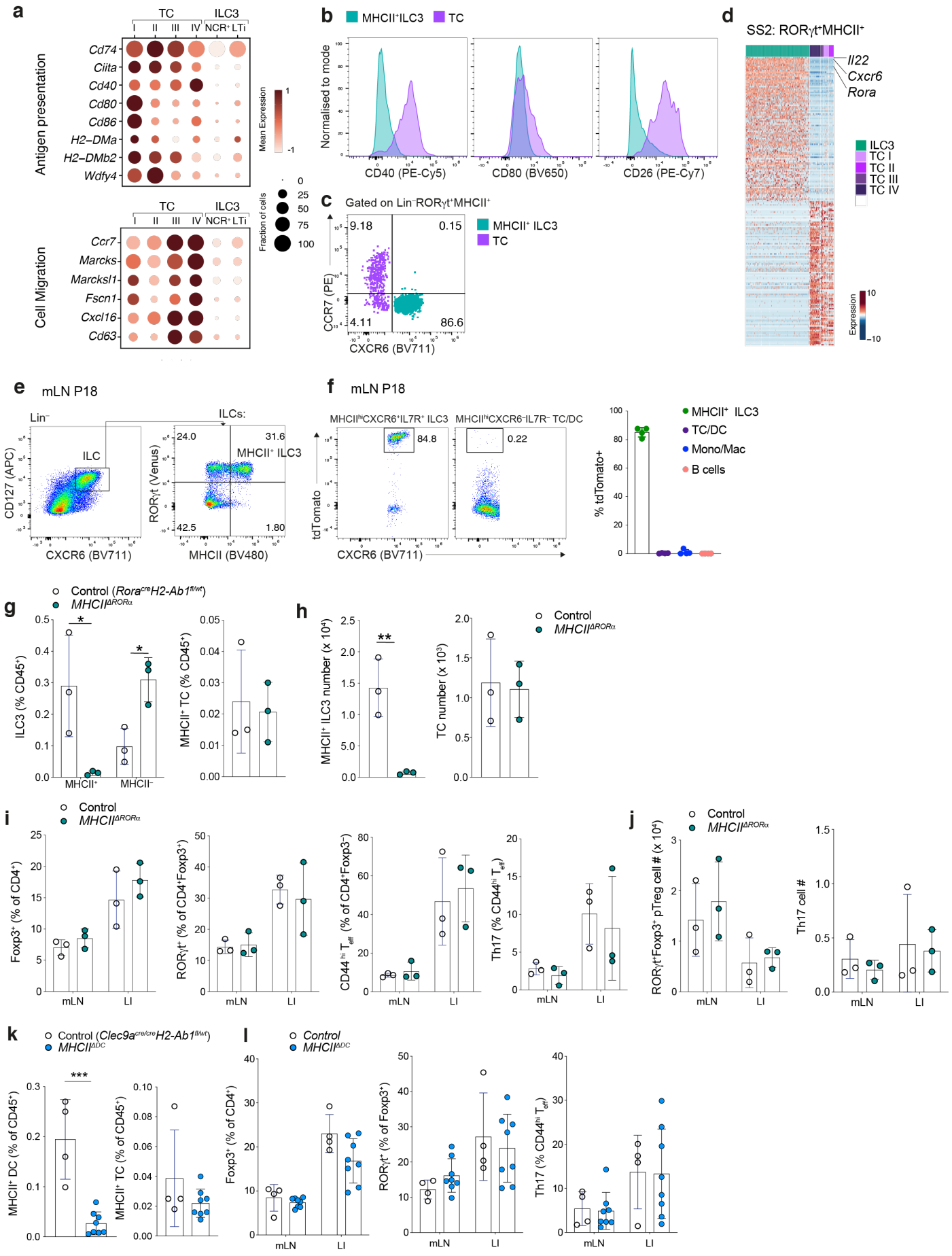
Extended data Figure 4



Extended Data Figure 5. Antigen presentation by ILC3s or DCs is not required for extra-thymic intestinal pTreg differentiation.

a, Dot plot showing expression of genes related to antigen presentation, T cell priming and cell migration across TC and MHCII$^+$ ILC3 clusters (Fig. 2b). **b**, Flow cytometry of mLN ILC3s (CXCR6$^+$RORγt$^+$MHCII$^+$) and TCs (CXCR6$^-$RORγt$^+$MHCII$^+$) showing expression of indicated chemokine receptors, co-stimulatory and immune-regulatory molecules by TC and ILC3 in mLN of P18 *Rorc*$^{Venus\text{-}creERT2}$Aire^{GFP} mice. **c**, Heatmap showing expression of top differentially expressed genes between TCs and MHCII$^+$ ILC3s profiled by SS2 (Fig. 3a), identifying *Rora* as an ILC3/TC distinguishing gene. **d**, Flow cytometric analysis demonstrating that all MHCII$^+$ ILCs (Lin$^-$CXCR6$^+$IL7R$^+$) are RORγt$^+$ ILC3s. **e**, Flow cytometric analysis of RORγt(Venus) and MHCII expression amongst Lin$^-$CXCR6$^+$IL7R$^+$ ILCs within mLN of P18 mice. Representative of >20 mice. **f**, Representative flow cytometric analysis of tdTomato expression in MHCII$^+$ILC3 and TCs/DCs isolated from mLN of *Rora*^{cre}*R26*$^{lsl\text{-}tdTomato}$ fate-mapped mice at P18 and summary bar graphs for all MHCII$^+$ cell types (right panel). n = 4 mice from two independent experiments. **g-j**, Immune cell composition of 3-week-old *MHCII*$^{\Delta ROR\alpha}$ (n = 3) and *Rora*^{cre}*H2-Ab1*$^{fl/wt}$ (n = 3) mice. (**g**) Percentage and (**h**) number of MHCII$^+$ ILC3s and TCs. **i**, Frequency of total T_{reg} (Foxp3$^+$), RORγt$^+$ pTreg cells and Th17 cells in mLN and LI and total pTreg cell number (**j**). **k-l**, Immune cell composition in mLN of P21 *MHCII*$^{\Delta DC}$ (n = 4) and control (*Clec9a*^{cre}*H2-Ab1*$^{fl/wt}$) (n = 8) mice from 2 independent experiments. Frequency of MHCII expressing DCs or TCs within mLN (**k**). Frequency of total Foxp3$^+$ Treg cells, pTreg cells amongst CD4$^+$Foxp3$^+$ cells, and Th17 cells in mLN and LI (**l**). Error bars: means ± s.e.m. *P < 0.05, **P < 0.05, ***P < 0.01, ****P < 0.0001; unpaired two-sided t-test.

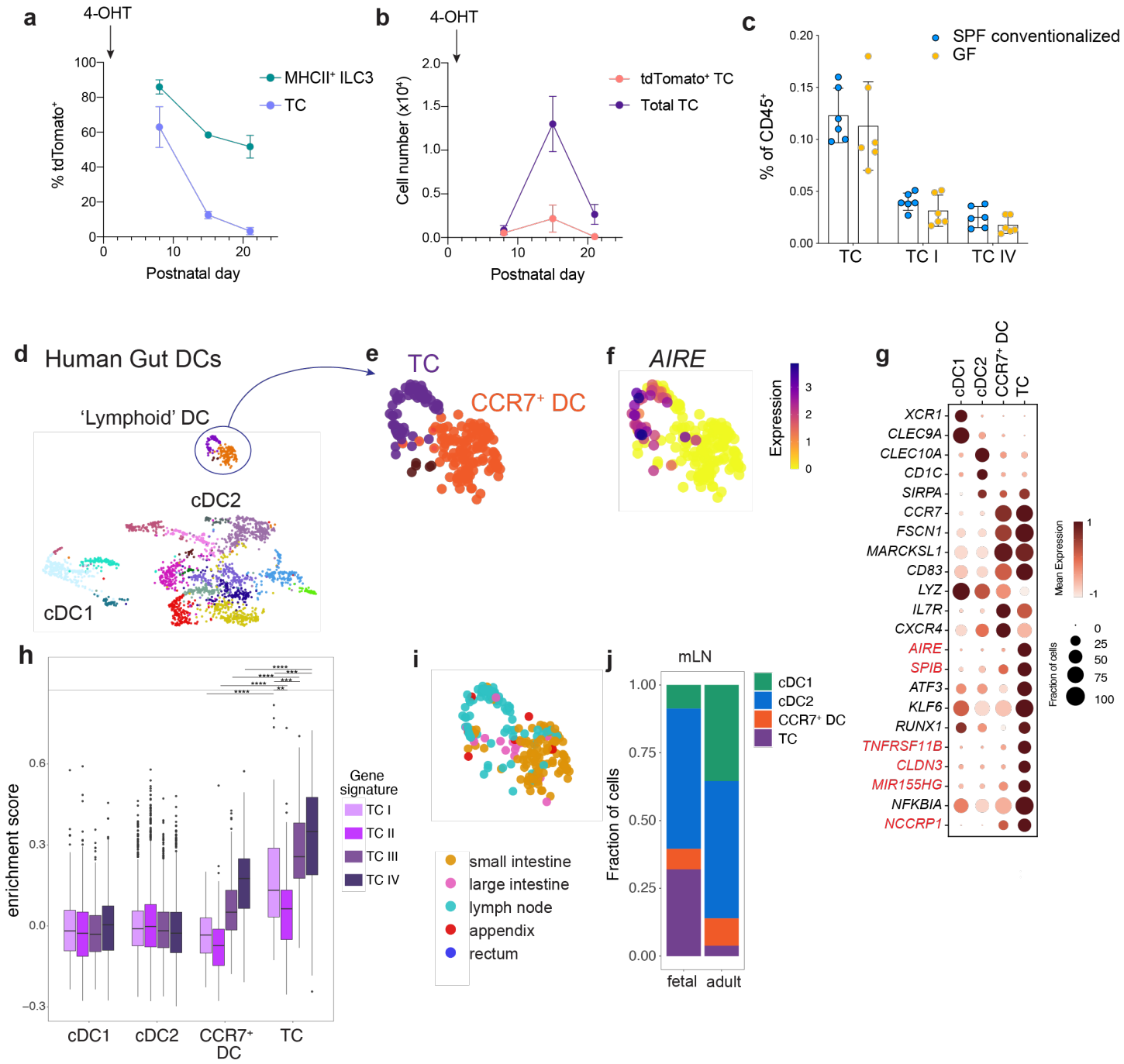
Extended Data Figure 5



Extended Data Figure 6. Identification of a developmental wave of intestinal TCs, conserved across mouse and humans.

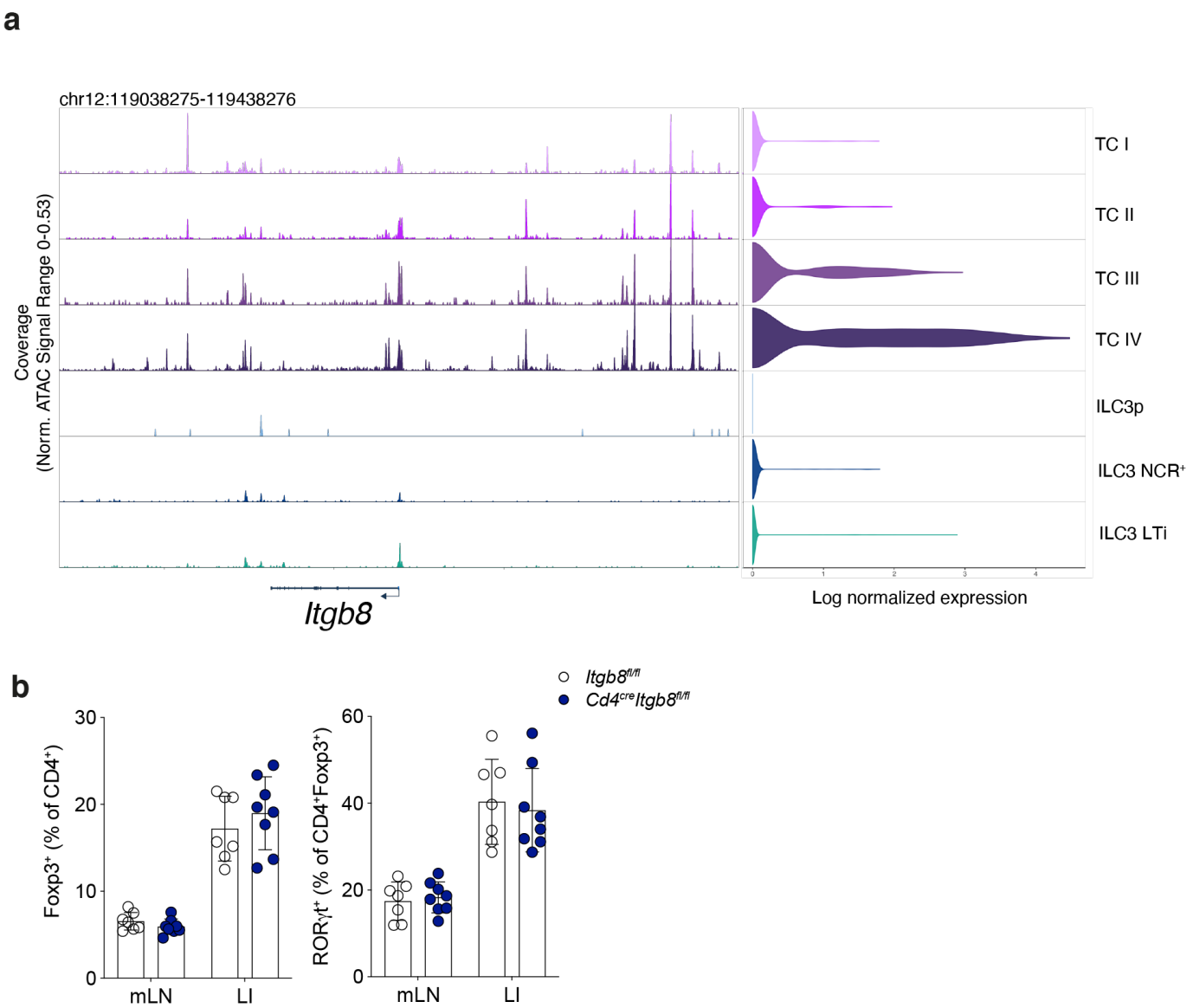
a, Percentage of tdTomato^{+} TCs and MHCII^{+} ILC3s isolated from mLN of *Rorc*$^{Venus\text{-}creERT2}$*R26*$^{lsl\text{-}tdTomato}$AireGFP mice at indicated time intervals following administration of 4-OHT on P1. n = 4–5 mice per timepoint. **b**, Total number of tdTomato^{-} and tdTomato^{+} TCs at indicated ages. **c**, Proportion of TCs in mLN of germ free (GF) and SPF conventionalized *Rorc*$^{Venus\text{-}creERT2}$Aire^{GFP} mice at P18. n = 6 mice per group. **d**-**g**, Human gut atlas single cell transcriptomes. Cells annotated as DCs were reclustered with PhenoGraph and visualized with UMAP (**d**). UMAP of 'lymphoid' DC clusters colored by PhenoGraph cluster (**e**), or unimputed expression of AIRE (**f**). **g**, Dot plot showing select genes differentially expressed between indicated cell subsets. **h**, Enrichment of TC subset signature genes within indicated human APC subsets. **i**, UMAP colored by tissue of origin. **j**, Proportion of indicated DC/TC subsets within mLN samples in fetal *vs* adult samples. Clusters annotated as cDC2 or cDC1 were grouped for analysis. Error bars: means ± s.e.m. (**c**). Box plots (**h**) indicate the median (center lines) and interquartile range (hinges), and whiskers represent min and max, dots represent outliers. ****P < 0.0001, ***P < 0.001, **P < 0.01; Mann Whitney U test (**h**).

Extended Data Figure 6



Extended Data Figure 7. TCs promote intestinal pTreg differentiation in an Itgb8-dependent manner.

a, Chromatin accessibility at the *Itgb8* locus and Itgb8 transcript levels in corresponding TC and ILC3 subsets (cells as in Fig 1c). **b**. Frequency of total Foxp3^{+} Treg cells and percentage of RORγt^{+} pTreg cells in mLN and LI of *Itgb8*$^{\Delta Cd4}$ or Itgb8$^{fl/fl}$ mice. *n* = 7 mice per group pooled from two independent experiments. Error bars: means ± s.e.m. NS, not significant; unpaired two-sided *t*-test.



List of Supplementary Tables

Supplementary Table 1. Table_S1.csv
Excel file containing top 50 marker genes for mTECs vs TCs (SS2).

Supplementary Table 2. Table_S2.csv
Excel file containing top marker genes for SS2 clusters

Supplementary Table 3. Table_S3.csv
Excel file containing list of TC-enriched tissue restricted antigens.

Supplementary Table 4. Table_S4.xlsx
Excel file containing list of antibodies used in this study.