Title: A cell atlas of human thymic development defines T cell repertoire formation

Authors: Jong-Eun Park[1], Rachel A. Botting[2], Cecilia Domínguez Conde[1], Dorin-Mirel Popescu[2], Marieke Lavaert[3,4], Daniel J. Kunz[1,19,20], Issac Goh[2], Emily Stephenson[2], Roberta Ragazzini[9], Elizabeth Tuck[1], Anna Wilbrey-Clark[1], Kenny Roberts[1], Veronika R. Kedlian[1], John R. Ferdinand[5], Xiaoling He[25], Simone Webb[2], Daniel Maunder[2], Niels Vandamme[6,21], Krishnaa Mahbubani[7], Krzysztof Polanski[1], Lira Mamanova[1], Liam Bolt[1], David Crossland[8,24], Fabrizio de Rita[24], Andrew Fuller[2], Andrew Filby[2], Gary Reynolds[2], David Dixon[2], Kourosh Saeb-Parsy[7], Steven Lisgo[8], Deborah Henderson[8], Roser Vento-Tormo[1], Omer A. Bayraktar[1], Roger A. Barker[25], Kerstin B. Meyer[1], Yvan Saeys[6,21], Paola Bonfanti[9,10,11], Sam Behjati[1,22,23], Menna R. Clatworthy[1,5,12], Tom Taghon[3,4,*], Muzlifah Haniffa[1,2,13,*], Sarah A. Teichmann[1,19,*]

Affiliations:

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK
[2]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK
[3]Faculty of Medicine and Health Sciences, Department of Diagnostic Sciences, Ghent University, C. Heymanslaan 10, MRB2, Entrance 38, 9000 Ghent, Belgium
[4]Cancer Research Institute Ghent (CRIG), Ghent University, Ghent, Belgium
[5]Molecular Immunity Unit, Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, United Kingdom, CB2 0QQ
[6]Data Mining and Modeling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium
[7]Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, United Kingdom, CB2 0QQ
[8]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, NE1 3BZ, UK
[9]Epithelial Stem Cell Biology & Regenerative Medicine Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
[10]Great Ormond Street Institute of Child Health, University College London, London, UK
[11]Institute of Immunity and Transplantation, University College London, London, UK
[12]Cambridge University Hospitals NHS Foundation Trust, United Kingdom, CB2 0QQ
[13]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE2 4LP, UK
[19]Theory of Condensed Matter Group, Cavendish Laboratory/Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK
[20]The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, United Kingdom
[21]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium
[22]Department of Haematology and Wellcome and MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, CB2 2XY, UK
[23]Department of Paediatrics, University of Cambridge, Cambridge CB2 0SP, UK
[24]Department of Adult Congenital Heart Disease and Paediatric Cardiology/Cardiothoracic Surgery, Freeman Hospital, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE2 4LP, UK
[25]John van Geest Centre for Brain Repair, WT-MRC Cambridge Stem Cell Institute, University of Cambridge, Forvie Site, Robinson Way, Cambridge CB2 0PY, UK

*Corresponding authors

Abstract: The thymus provides a nurturing environment for the differentiation and selection of T cells, a process orchestrated by their interaction with multiple thymic cell types. We used single-cell RNA-sequencing (scRNA-seq) to create a cell census of the human thymus across the lifespan and to reconstruct T-cell differentiation trajectories and T-cell receptor (TCR) recombination kinetics. Using this approach, we identified and located *in situ* CD8αα$^+$ T-cell populations, thymic fibroblast subtypes and activated dendritic cell (aDC) states. In addition, we reveal a bias in TCR recombination and selection, which is attributed to genomic position and the kinetics of lineage commitment. Taken together, our data provide a comprehensive atlas of the human thymus across the lifespan with new insights into human T-cell development.

One Sentence Summary: We profiled human thymus using single cell RNA-sequencing across development and aging, revealing the diversity and dynamics of human thymic cell types and the kinetics of T-cell receptor recombination.

Main Text:

Introduction

The thymus plays an essential role in the establishment of adaptive immunity and central tolerance as it mediates the maturation and selection of T cells. This organ degenerates early during life and the resulting reduction in T-cell output has been linked to age-related incidence of cancer, infection and autoimmunity (*1*, *2*). T-cell precursors from fetal liver or bone marrow migrate into the thymus, where they differentiate into diverse types of mature T cells (*3*, *4*). The thymic microenvironment cooperatively supports T-cell differentiation (*5*, *6*). While thymic epithelial cells (TECs) provide critical cues to promote T-cell fate (*7*), other cell types are also involved in this process, such as dendritic cells (DC) that undertake antigen presentation, and mesenchymal cells, which support TEC differentiation and maintenance (*8–11*). Seminal experiments in animal models have provided major insights into the function and cellular composition of the thymus (*12*, *13*). More recently, scRNA-seq has revealed new aspects of thymus organogenesis and new types of thymic epithelial cells (TECs) in mouse (*14–16*). However, the human organ matures in a mode and tempo that is unique to our species (*17–19*), calling for a comprehensive genome-wide study for human thymus.

T-cell development involves a parallel process of staged T-cell lymphocyte differentiation accompanied by acquisition of a diverse TCR repertoire for antigen recognition (*20*). This is achieved by the genomic recombination process that selects one V, (D) and J segments from the array of gene segments. Interestingly, this VDJ gene recombination can preferentially include certain gene segments, leading to the skewing of the repertoire (*21–23*). To date, most of our knowledge of VDJ recombination and repertoire biases, has come from animal models and human peripheral blood analysis, with little comprehensive data on the human thymic TCR repertoire (*22*, *24*, *25*).

Here, we applied scRNA-seq to generate a comprehensive transcriptomic profile of the diverse cell populations present in embryonic, fetal, paediatric and adult stages of the human thymus and we combined this with detailed TCR repertoire analysis to reconstruct the T-cell differentiation process.

Cellular composition of the human thymus across life

We performed scRNA-seq on 15 fetal thymi beginning from 7 post-conception weeks (PCW), when the thymic rudiment can be dissected, to 17 PCW, when thymic development is completed (**Fig. 1, A and B**). We also analysed 9 postnatal samples, covering the entire period of active thymic function. Isolated single cells were sorted based on CD45, CD3 or EPCAM expression to sample thymocytes and enrich for non-thymocytes, prior to single-cell transcriptomic analysis coupled with TCRαβ profiling. After quality control including doublet removal, we obtained a total of 138,397 cells from the developing thymus and 117,504 cells from postnatal thymus (**Table S1**). If available, other relevant organs were collected from the same donor. We performed batch correction using the BBKNN algorithm combined with linear regression (**fig. S1)** (*26*)**.**

We have annotated cell clusters into more than 40 different cell types or cell states (**Fig. 1, C and D and Table S2, 3**), which can be clearly identified by the expression of specific marker genes (**fig. S2 and Table S4**). Differentiating T cells are well represented in the dataset, including double negative (DN), double positive (DP), CD4$^+$ single positive (CD4$^+$T), CD8$^+$ single positive (CD8$^+$T), FOXP3$^+$ regulatory (Treg), CD8αα$^+$ and γδ T cells. We also identified other immune cells including B cells, NK cells, innate lymphoid cells (ILCs), macrophages, monocytes and dendritic cells (DCs).

Our dataset also featured diverse non-immune cell types, which constitute the thymic microenvironment. We further classified them into subtypes including thymic epithelial cells (TECs), fibroblasts, vascular smooth muscle cells (VSMCs), endothelial cells and lymphatic endothelial cells (**Fig. 1E**). Thymic fibroblasts were further divided into two subtypes, neither of which has been previously described: Fibroblast type 1 (Fb1) cells (*COLEC11*, *C7, GDF10*) and Fibroblast type 2 (Fb2) cells (*PI16, FN1*, *FBN1)* (**Fig. 1E**). Fb1 cells uniquely express *COLEC11,* which plays an important role in innate immunity (*27*) and *ALDH1A2*, an enzyme responsible for the production of retinoic acid, which regulates epithelial growth (*28*). In contrast, extracellular matrix (ECM) genes and Semaphorins which regulate vascular development (*29*), are specifically detected in Fb2 (**fig. S3A**). To explore the localisation pattern of these fibroblast subtypes, we performed *in situ* smFISH targeting Fb1 and Fb2 markers (*COLEC11* and *FBN1*) together with general fibroblast (*PDGFRA*), endothelial (*CDH5*) and VSMC (*ACTA2*) markers (**Fig. 1F**). The results show that Fb1 cells were peri-lobular, while Fb2 cells were interlobular, often associated with large blood vessels lined with VSMCs, consistent with their transcriptomic profile of genes regulating vascular development. We confirmed the expression of *GDF10* and *ALDH1A2* localised in the peri-lobular area (**Fig. 1F**).

In addition to fibroblasts, we also identified subpopulations of human TECs (**Fig. 1E and fig. S4)**. To maximise the coverage of epithelial cells, we enriched for EPCAM positive cells across several time points (**Fig. 1B**). To annotate human TECs, we compared our human dataset to the published mouse TEC dataset (*15*) (**figs. S5, S6, and S7**). We were able to identify conserved TEC populations across species, including *PSMB11*-positive cTECs, *KRT14*-positive mTEC(I), *AIRE*-expressing mTEC(II), and *KRT1*-expressing mTEC(III) (**Figs. 1E and S4**). Interestingly, cTECs were more abundant during early development (7-8 PCW), and an intermediate population (mcTEC), which are marked by expression of *DLK2,* was evident in late fetal and paediatric human thymi (**fig. S4B**). We identified a very rare population of mTEC(IV) cells in humans, which are similar to tuft-like mTEC(IV) cells described in the mouse thymus. However, *DCLK1* or *POU2F3*, the markers used to define mTEC(IV) cells in the mouse (*15*, *16*), were enriched but not specific to this population in human (**figs. S4B, S5 and S6**). We noted two EPCAM$^+$ cell types which are specific to human: *MYOD1* and *MYOG*-expressing myoid cells (TEC(myo)s), and *NEUROD1, NEUROG1*, *CHGA*-expressing TEC(neuro)s (**Fig. 1E and figs. S6 and S7**). Notably, *CHRNA1*, which has been associated with the autoimmune disease myasthenia gravis (*30*), was specifically expressed by both of these cell types in addition to mTEC(II) cells (**Fig. 1E**), expanding the candidate cell types which may be involved in tolerance induction in myasthenia gravis (*31*, *32*). Supporting this possibility, we detected *MYOD1* and *NEUROG1* expressing cells preferentially located in thymic medulla (**Fig. 1F**).

Lastly, we analysed the expression pattern of genes known to cause congenital T-cell immunodeficiencies to provide insight into when and where these rare disease genes may play a role during thymic development (**fig. S8**).

Coordinated development of thymic stroma and T cells

Next, we investigated the dynamics of the different thymic cell types across development (**Fig. 1G**). In the early fetal samples (7-8 PCW), the lymphoid compartment contained NK cells, γδ T cells and ILC3s, with very few differentiating αβT cells (**Fig. 1G**). Differentiating T cells are mostly found at DN stage in 7 PCW sample, which gradually progress through DP to SP stages thereafter, reaching equilibrium at around 12 PCW (**Fig. 1G**). Conversely, the proportion of innate lymphocytes decreased (**Fig. 1G**).

Of note, the adult sample showed morphological evidence of thymic degeneration (**fig. S9**). Comparison with spleen and lymph nodes taken from the same donor showed the presence of terminally differentiated T cells in the thymus, suggesting re-entry into thymus or contamination with circulating cells (**Fig. 1G and fig. S10**). Notably, cytotoxic CD4$^+$T lymphocytes (CD4$^+$CTL) expressing *IL10*, perforin and granzymes were enriched in the degenerated thymus sample (*33*) (**fig. S10C**). The trend of increased memory T cells and B cells are also confirmed in other samples (**Fig. 1G,** p-value: $9.3\text{x}10^{-6}$ for memory T cells and 0.0096 for memory B cells).

The trend in T cell development was mirrored by corresponding changes in thymic stromal cells. We observed temporal changes in TEC populations starting from enriched cTECs towards the balanced representation of cTECs and mTECs (**Fig. 1G,** p-value: 0.0054), aligned with the onset of T-cell maturation. This supports the notion of 'thymic crosstalk' in which epithelial cells and mature T cells interact synergistically to support their mutual differentiation (*34*).

Moreover, fibroblast composition also changed during development. The Fb1 population mentioned in the previous section dominated early development, with similar numbers of Fb1 and Fb2 cells observed at later developmental timepoints (p-value: 0.014), and a reduction in the number of cycling cells (**Fig. 1G**). This is also confirmed by thymic fibroblast explant cultures, which showed an increase in Fb2 cell marker PI16 by FACS analysis (**fig. S3, B and C**).

Finally, other immune cells also change dynamically over gestation and in postnatal life. Macrophages were abundant during early gestation, while DCs increased throughout development (**Fig. 1G**). DC1 was dominant after 12 PCW, and pDCs increased in frequency in postnatal life (*p*-values: $2.7\text{x}10^{-8}$ for macrophage, $1.05\text{x}10^{-3}$ for DC1, $4.86\text{x}10^{-5}$ for DC2)

To further investigate the factors mediating the coordinated development of thymic stroma and T cells, we systematically investigated cellular interactions using our public database CellPhoneDB.org (*35*) to predict the ligand-receptor pairs specifically expressed across them (**Table S5**). Among the predicted interactions, we checked the expression pattern of signaling factors known to be involved in thymic development across different cell types and developmental stages (**fig. S11**) (*36–41*). Lymphotoxin signaling (*LTB*:*LTBR*) comes from diverse immune cells and is received by most of the stromal cell states. In contrast, RANKL-RANK (*TNFRSF11*:*TNFRSF11A*) signaling is confined between ILC3 and mTEC(II) cells/lymphatic endothelial cells. FGF signaling (*FGF7*:*FGFR2*) comes from Fibroblasts signaling to TECs, with decreasing expression of *FGFR2* in adult thymus. For Notch signaling, while *NOTCH1* is the the main receptor expressed in ETPs, diverse Notch ligands are expressed by different cell types: cTECs and endothelial cells expressed both *JAG2* and *DLL4*, and other TECs broadly expressed *JAG1* (*42*, *43*).

Conventional T cell differentiation trajectory

As fetal liver is the main haematopoietic organ and source of HSC/MPP when the thymic rudiment develops, we analysed paired thymus and liver samples from the same fetus (*44*), similarly to what has been described for early hematopoietic organs (*45*). We merged the thymus and liver data, and selected clusters including liver HSC/MPP, thymic ETPs and DN thymocytes for data analysis and visualisation (**Fig. 2A, 2B and fig. S12**). This positioned thymic ETPs at the isthmus between fetal liver HSC/MPP and pre/pro B cells. We integrated our liver/thymic hematopoietic progenitor subset with the single-cell transcriptomes of human hematopoietic progenitors sorted from bone marrow using defined markers (*46*) (**Fig. S13**). This analysis positions the ETPs next to the multi-lymphoid progenitor (MLP) from bone marrow and early lymphoid progenitor in fetal liver.

To investigate the downstream T cell differentiation trajectory, we selected the T cell populations and projected them using UMAP and force-directed graph analysis (**Fig. 2C, fig. S14A and Data S1),** which showed a continuous trajectory of differentiating T cells. To confirm the validity of this trajectory, we overlaid hallmark genes of T-cell differentiation: CD4/CD8A/CD8B genes (**Fig. 2D**), cell cycle (*CDK1*) and recombination (*RAG1*) genes (**Fig. 2E**) and fully recombined TCRα/TCRβ (**Fig. 2F**) (*47*). The trajectory started from CD4$^-$CD8$^-$ DN cells, which gradually express CD4 and CD8 to become CD4$^+$CD8$^+$ DP cells, and then transitions through a CCR9^{high} Tαβ(entry) stage to diverge into mature CD4$^+$ or CD8$^+$ SP cells (**Fig. 2D**). We also noted a separate lineage of cells diverging from the DN-DP junction corresponding to γδ T-cell differentiation. Additional T-cell lineages identified in this analysis will be discussed in the following section (**Fig. 2C, grey**). DN and DP cells were separated into two phases by the expression of cell cycle genes (**Fig. 2E**). We designated the early population with strong cell cycle signature as proliferating (P) and the later population quiescent (Q), respectively (**Fig. 2C**). Expression of VDJ recombination genes (*RAG1* and *RAG2*) increased from the late proliferative phase, and peaked at the quiescent phases. This pattern reflects the proliferation of T cells which precedes each round of recombination (*48*, *49*).

Next, we aligned the TCR recombination data to this trajectory (**Fig. 2F**). In the DN stage, recombined TCRβ sequences were detected from the late P phase, which coincides with an increase in recombination signature and the expression of pre-TCR-alpha (*PTCRA*) (**Fig. 2G and fig. S15**). The ratio of non-productive to productive recombination events (non-productivity score) for TCRβ was relatively higher in DN stages, and dropped to a basal level as cells entered DP stages, demonstrating the impact of beta-selection (**Fig. 2H**). Notably, the non-productivity score for TCRβ was highest in the DN(Q) stage, suggesting that cells failing to secure a productive TCRβ recombination for the first allele undergo recombination of the other allele. In the DP stage, recombined TCRα chains were detected from P stage onwards. In contrast to TCRβ, non-productive TCRα chains were not enriched in the DP(Q) cells, but were rather depleted (**Fig. 2H**).

To match the transcriptome-based clustering from this study to a published protein-marker based sorting strategy, we compared our data with repository data from FACS-sorted thymocytes analysed by microarray (*50*) (**fig. S16**). Based on the cell cycle gene signature and marker gene expression, DN(P), DN(Q), DP(P) stages are closely matched to CD34+CD1A+, ISP CD4+, and DP CD3- populations respectively. Both our DP(Q) and Tαβ(entry) stage cell signatures are enriched in the bulk transcriptome data from the DP CD3+ FACS-sorted cells. The enrichment of pre-beta selection cells in DN(Q) cells matches well with the characteristics of ISP CD4+ serving as a checkpoint for beta-selection (**Fig. 2F and fig. S15**).

To model the development of conventional αβT cells in more detail, we performed pseudo-time analysis, which resulted in an ordering of cells highly consistent with known marker genes and transcription factors (**Fig. 2G**). In addition, we identified T-cell developmental markers, including *ST18* for early DN, *AQP3* for DP and *TOX2* for DP to SP transition. To derive further insights into transcription factors that specify T-cell stages and lineages, we created a correlation-based transcription factor network, after imputing gene expression (see Methods), which demonstrated modules of transcription factors specific for lineage commitment (**Fig. 2I**).

Development of Tregs and unconventional T cells

In addition to conventional CD4$^+$ or CD8$^+$ T cells, which comprise the majority of T cells in the developing thymus, our data identified multiple unconventional T cell types, which were grouped by the expression of signature marker genes (**Fig. 3, A, B and Fig. 2I**). Unconventional T cells have been suggested to require agonist selection for development (*3*). In support of this, we observed a lower ratio of non-productive TCR chains for these cells, implying that they reside longer in the thymus compared to conventional T cells (**Fig. 3C**).

Next, we investigated whether development of these unconventional T cells was dependent on the thymus. We reasoned that if a population is thymus-dependent, it would accumulate after thymic maturation (~10 PCW) and be enriched in the thymus compared to other hematopoietic organs. Consistent with this, all unconventional T cells were enriched in the thymus, particularly post-thymic maturation, suggesting that they are thymus-derived (**Fig. 3D**).

Tregs were the most abundant unconventional T cells in the thymus. There was a clear differentiation trajectory connecting αβT cells and Tregs. We defined the connecting population as differentiating Tregs (Treg(diff)) (**Fig. 3A)**. Compared to canonical Tregs, Treg(diff) cells had lower *FOXP3* and *IL2RA* expression, and higher expression of *IKZF4*, *GNG8* and *PTGIR* (**Fig. 3B)**. These genes have been associated with autoimmunity and Treg differentiation (*51*).

We also noted another population which shares expression modules with Treg(diff) cells, but not with terminally differentiated Treg cells. We named this population as T(agonist) defined by the expression of a non-coding RNA, MIR155HG (**Figs. 3, A and B)**. Interestingly, this population expressed IL2RA but has low FOXP3 mRNA. These features are similar to a previously described mouse CD25$^+$FOXP3$^-$ Treg progenitor (*52*) (**fig. S17**). Further analysis showed that the signature of two Treg progenitors (CD25$^+$ and FOXP3^{lo} Treg progenitors) defined in previous studies are expressed at a higher level in T(agonist) and Treg(diff) populations, respectively (**fig. S17B**). The UMAP and force-directed graph showed that both of these populations are linked to mature Tregs (**fig. S17A**), suggesting the possibility of two Treg progenitors in the human thymus.

Other unconventional T cell populations included CD8αα$^+$T cells, NKT-like cells and Th17-like cells (**Fig. 3B**). There were three distinct populations of CD8αα$^+$T cells: *GNG4*$^+$ *CD8aa*$^+$T(I) cells, *ZNF683*$^+$ *CD8aa*$^+$T(II) and a *CD8aa*$^+$ NKT-like population marked by *EOMES* (**Fig. 3E**). *GNG4*$^+$*CD8aa*$^+$T(I) and *ZNF683*$^+$*CD8aa*$^+$T(II) both shared *PDCD1* expression at an early stage, which decreased in their terminally differentiated state (**fig. S14B**). While *GNG4*$^+$ *CD8aa*$^+$T(I) displayed a clear trajectory diverging from late DP stage (αβT SP entry cells), *ZNF683*$^+$*CD8aa*$^+$T(II) cells have a mixed αβ and γδ T cell signatures, and sit next to both *GNG4*$^+$*CD8aa*$^+$T(I) cells and γδ T cells (**Fig. 3A and fig. S14B**).

EOMES$^+$ NKT-like cells have a shared gene expression profile with NK cells (*NKG7*, *IFNG*, *TBX21*) and are enriched in γδ T cells, i.e. their TCRs are γδ rather than αβ (**Fig. 3B and fig. S14B**). Interestingly, previously described gene sets from bulk RNA sequencing of human thymic or cord blood CD8αα$^+$T cells can now be deconvoluted into our three CD8αα$^+$T cell populations using signature genes. These results suggest that our three CD8αα$^+$T cell populations are present in these previously published thymic and cord blood samples at different frequencies, as shown in (**fig. S18**) (*53*).

Finally, we found another fetal specific cell cluster which we named as “Th17-like cells”, based on *CD4*, *CD40LG*, *RORC* and *CCR6* expression (**Fig. 3B**). Th17-like cells and NKT-like cells expressed *KLRB1* and *ZBTB16*, which are hallmarks of innate lymphocytes (*54*, *55*) (**Fig. 3F**).

As described above, many cell clusters contained a mixed signature of αβ and γδ T cells, meaning that a single cluster contained some cells with αβ TCR expression and others with γδ TCR. To classify cells into αβ and γδ T cells, we analysed the TCRα/δ loci, where recombination of TCRα excises TCRδ, making the two mutually exclusive (**Fig. 3G**). This clearly showed that γδ T cells diverging between the DN and DP populations are pure γδ T cells. In contrast, CD8αα$^+$T(II), NKT-like and Th17-like cells include both αβ and γδ T cell populations, suggesting transcriptomic convergence of some αβ and γδ T cells.

Interestingly, *TRDV1* and *TRDV2*, the two most frequently used TCRδ V genes in human, displayed clear usage bias: *TRDV2* was used at an earlier stage (DN), while *TRDV1* was exclusively utilised in later T-cell development (DP(Q) and αβT entry) (**Fig. 3H**). Based on this pattern, we can attribute the stage of origin of γδ T-cell populations, which suggests that CD8αα$^+$T(II) are derived from the late DP stage, while NKT-like/Th17-like cells arise from earlier stages (**Fig. 3H**).

Discovery and characterisation of *GNG4*$^+$ CD8αα T cells in the thymic medulla

Having identified unconventional T cells and their trajectory of origin within thymic T-cell development, we focused on our newly discovered GNG4$^+$CD8αα$^+$T(I) cells, as they have a unique gene expression profile (*GNG4*, *CREB3L3* and *CD72*). This is in contrast to CD8αα$^+$T(II) cells, which express known markers of CD8αα$^+$T cells such as *ZNF683* and *MME (53)*. Moreover, the expression level of *KLF2*, a regulator of thymic emigration, was extremely low in

CD8αα^{+}T(I) cells, suggesting that they may be thymic-resident (**Fig. 3B**). To locate and validate CD8αα^{+}T(I) cells *in situ*, we performed RNA smFISH targeting *GNG4* in fetal thymus tissue sections. The *GNG4* RNA probe identified a distinct group of cells enriched in the thymic medulla, and co-localised with *CD8A* RNA (**Fig. 3I**). *TNFRSF9* (CD137), is a marker shared between CD8αα^{+}T(I) cells and Tregs. When tested *in situ*, GNG4^{+} cells were a subset of TNFRSF9^{+} cells, further confirming the validity of the localisation pattern.

As CD137 is a surface marker of both CD8αα^{+}T(I) cells and Tregs, we enriched these cells using this marker (**fig. S19**). Further refinement using CD3^{+}CD137^{+}CD4^{-} FACS-sorting allowed us to specifically enrich for CD8αα^{+}T(I) cells, and confirm their identity by Smart-seq2 scRNA sequencing, providing additional transcriptomic phenotyping of these cells (**Fig. 3J**).

To compare our findings in human thymus to mouse thymus, we generated a comprehensive mouse thymus single cell atlas of postnatal murine samples (4, 8, 24 weeks old) and combined this data with a published prenatal mouse thymus scRNA-seq dataset (*14*) (**fig. S20**). Integrative analysis of mature T cells from human and mouse shows that cell states are well mixed across species (**fig. S21**). This analysis showed that GNG4+ CD8αα^{+}T(I) cells in humans are most similar to the mouse intraepithelial lymphocytes precursor type A (IELpA) cells (*56*) (**fig. S21**), sharing expression of *HIVEP3*, *NR4A3*, *PDCD1* and *TNFRSF9* (**fig. S22**). However, there were also highly differentially expressed genes between them, including *GNG4* and *XCL1* in human, and *ZEB2* and *CLDN10* in mouse, suggesting a potential difference in function (**fig. S23**). Moreover, human CD8αα^{+}T(I) fully mature into a CD8A$^{[high]}$/CD8B$^{[low]}$ phenotype whereas mouse IELpA cells become triple negative (CD8A$^{[low]}$CD8B$^{[low]}$CD4$^{[low]}$) cells (**fig. S23**). This shows that human and mouse TNFRSF9^{high} agonist selected cells in the thymus take on distinct transcriptional characteristics.

Recruitment and activation of DCs for thymocyte selection

Selection of T cells is coordinated by specialised TECs and DCs. We identified three previously well-characterised thymic DC subtypes: DC1 (XCR1^{+}CLEC9A^{+}), DC2 (SIRPA^{+}CLEC10A^{+}), pDC (IL3RA^{+}CLEC4C^{+}) (*6*, *57*, *58*). We also identified a population that was previously incompletely described, which we term as "activated DCs" (aDCs), characterised by *LAMP3* and *CCR7* expression (**Fig. 4, A and B**) (*59*, *60*). aDCs expressed high level of chemokines and co-stimulatory molecules, together with transcription factors like *AIRE* and *FOXD4*, which we validated *in situ* (**Fig. 4B and fig. S24**), suggesting that they may correspond to the previously described AIRE^{+}CCR7^{+} DCs in human tonsils and thymus (*61*).

Interestingly, our single-cell data revealed three subsets within the aDC group, identified by distinct gene expression profiles: aDC1, aDC2 and aDC3 (**Fig. 4, A and B**). aDC1 and aDC2 subtypes shared several marker genes with DC1 and DC2, respectively. To systematically compare aDC subtypes to canonical DCs, we calculated an identity score for each DC population by summarising marker gene expression. This demonstrated a clear relationship between aDC1-DC1 and aDC2-DC2 pairs, suggesting that each aDC subtype derives from a distinct DC population (**fig S25**). Interestingly, aDC1 and aDC2 displayed distinct patterns of chemokine expression, suggesting functional diversification of these aDCs (**Fig. 4B**). Moreover, aDC3 cells had decreased MHC class II and co-stimulatory molecule expression compared to other aDC subsets, which may reflect a post-activation DC state.

Having identified two canonical TECs and a variety of DC subsets, we used CellPhoneDB analysis to identify specific interactions between these antigen-presenting cells and differentiating T cells (*35*). We focused on interactions mediated by chemokines, which enable cell migration and anatomical co-localisation (**Fig. 4C**). This demonstrated the relay of differentiating T cells from the cortex to the medulla, which is orchestrated by *CCL25*:*CCR9* and *CCL19/21*:*CCR7* interactions between cTEC/mTEC and DP/SP T cells, respectively (*62*). Interestingly, aDC expressed *CCR7*, together with *CCL19*, enabling attraction to and recruitment of T cells into the thymic medulla. Moreover, they strongly expressed the chemokines *CCL17* and *CCL22*, whose receptor *CCR4* was enriched in CD4$^+$ T cells and particularly Tregs. aDCs also potentially recruit other DCs and mature Tregs via *CXCL9/10*:*CXCR3* interactions and are able to provide a strong co-stimulatory signal, which suggests a role in Treg generation. We also noted that GNG4$^+$CD8αα$^+$T(I) T cells expressed XCL1, which may be involved in the recruitment of XCR1-expressing DC1 cells (*63*). Our analysis shows that XCL1 is expressed most highly by CD8αα$^+$T(I) cells and at a lower level by NK cells (**fig. S26**). The location of CD8αα$^+$T(I) in the peri-medullary region suggests a potential relay of signals from CD8αα$^+$T(I) to recruit XCR1$^+$DC1s into the medulla, where these cells are activated and upregulate CCR7. (**Fig. 4D**).

To confirm our *in-silico* predictions, we performed smFISH to identify the anatomical location of CD8αα$^+$T(I) cells (*GNG4*), DC1s (*XCR1*), aDCs (*LAMP3*, *CD80*) and Tregs (*FOXP3*). A generic marker of non-activated DCs (*ITGAX*) and mTECs (*AIRE*) was also used to provide a reference for the organ structure. Imaging of consecutive sections of fetal thymus (15 PCW) revealed the zonation of CD8αα$^+$T(I)/DC1/non-activated DCs located in the peri-medullary region and aDC/Tregs enriched in the center of the medulla (**Fig. 4E-4H**). All localisation patterns are supportive of our *in-silico* model, demonstrating the power of single-cell transcriptomics coupled with CellPhoneDB predictions.

Bias in human TCR repertoire formation and selection

As our data featured detailed T-cell trajectories combined with single-cell resolution TCR sequences, it provided an opportunity to investigate the kinetics of TCR recombination. TCR chains detected from the TCR-enriched 5' sequencing libraries were filtered for full-length recombinants, and were associated with our cell type annotation. This allowed us to analyse patterns in TCR repertoire formation and selection (**Fig. 5, A and B**).

For TCRβ, we observed a strong bias in VDJ gene usage which persisted from the initiation of recombination (DN cells) to the mature T-cell stage (**Fig. 5A**). This bias is not explained by recombination signal sequence (RSS) score (**fig. S27**). The bias does correlate well with genomic position (**fig. S27**), and this is consistent with a looping structure of the locus, which has been observed in the mouse (**Fig. 5C**) (*64*). However, the V gene usage bias that we observe in human is not found in mouse (*25*). We also observed a preferential association of D2 genes with J2 genes, while D1 genes can recombine with J1 and J2 genes with similar frequency (**fig. S28**). There was no clear association between TCRβ V-D or V-J pairs (**fig. S28A**).

While the initial recombination pattern largely shapes the repertoire, selection also contributes to the preference in TCRβ repertoire. We observed that several TRBV genes were depleted or enriched after beta-selection compared (DP cells) to before beta-selection (DN cells). This suggests that there are germline-encoded differences between the different Vβ gene's ability to respond to peptide-MHC (pMHC) stimulation (**fig. S29A**). This result is in line with the molecular finding that Vβ makes the most contacts with pMHC molecule *versus* DJ (and also Vα) (*65*).

For the TCRα locus, we found a clear association between developmental timing and V-J pairing as described before (*66*): Proximal pairs were recombined first, followed by recombination of distal pairs (**Fig. 5B**), which in turn restricts the pairing between V and J genes (**fig. S28B**). This provides direct evidence for progressive recombination of the TCRα locus (**Fig. 5D**). Notably, proximal pairs were relatively depleted in mature T cells compared to DP cells, showing a further bias in the positive selection step (**fig. S28B**).

To investigate whether differential TCR repertoire bias exists between cell types, we compared the TCR repertoire of different cell types by running a principal component analysis (**Fig. 5E**). Notably, we observed a clear separation of $CD8^+$ T cells and other cell types. The trend was consistent in all individual donor samples. Statistical testing of the difference in odds ratios identified several TCR genes responsible for this phenomenon (**fig. S29B**). The observed trend was largely similar to that seen in naive $CD4^+$/$CD8^+$T cells isolated from peripheral blood (*22*, *23*). Notably, the TRAV-TRAJ repertoire of $CD8^+$T cell was biased towards distal V-J pairs compared to other cell types (**Fig. 5F**). Considering that distal repertoires are generated at a later stage of progressive TCRα recombination, this might be due to slower or less efficient commitment towards the $CD8^+$T lineage (**Fig. 5D**). There was also a slight bias towards proximal V-J pairs for $CD8αα^+$T(I) cells that was much more evident in the postnatal thymic sample compared to fetal samples (**fig. S29C**) (*53*).

DISCUSSION

Here we generated a single-cell atlas of the human thymus throughout development *in utero* and in postnatal life alongside complementary *in situ* imaging to provide spatial context for our atlas. We reconstructed the trajectory of human conventional and unconventional T-cell differentiation combined with TCR repertoire information, which revealed a bias in the TCR repertoire of mature conventional T cells. As TCR repertoire bias predisposes our reactivity to diverse pMHC combinations, this may have profound implications for how we respond to antigenic challenges.

Our analysis of the thymic microenvironment revealed the complexity of cell types constituting the thymus, and the breadth of interactions between stromal cells and innate immune cells to coordinate thymic development to support T cell differentiation. The intercellular communication network that we describe between thymocytes and supporting cells can be used to enhance *in vitro* culture systems to generate T cells, as well as future T-cell therapeutic engineering strategies.

References and Notes:

1. S. Palmer, L. Albergante, C. C. Blackburn, T. J. Newman, Thymic involution and rising disease incidence with age. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1883–1888 (2018).
2. H. E. Lynch, G. L. Goldberg, A. Chidgey, M. R. M. Van den Brink, R. Boyd, G. D. Sempowski, Thymic involution and immune reconstitution. *Trends Immunol.* **30**, 366–373 (2009).
3. G. L. Stritesky, S. C. Jameson, K. A. Hogquist, Selection of self-reactive T cells in the thymus. *Annu. Rev. Immunol.* **30**, 95–114 (2012).

4. M. J. Sánchez, J. C. Gutiérrez-Ramos, E. Fernández, E. Leonardo, J. Lozano, C. Martínez, M. L. Toribio, Putative prethymic T cell precursors within the early human embryonic liver: a molecular and functional analysis. *J. Exp. Med.* **177**, 19–33 (1993).

5. L. Sun, C. Sun, Z. Liang, H. Li, L. Chen, H. Luo, H. Zhang, P. Ding, X. Sun, Z. Qin, Y. Zhao, FSP1(+) fibroblast subpopulation is essential for the maintenance and regeneration of medullary thymic epithelial cells. *Sci. Rep.* **5**, 14871 (2015).

6. L. Wu, K. Shortman, Heterogeneity of thymic dendritic cells. *Semin. Immunol.* **17**, 304–312 (2005).

7. G. Anderson, Y. Takahama, Thymic epithelial cells: working class heroes for T cell development and repertoire selection. *Trends Immunol.* **33**, 256–263 (2012).

8. Y.-J. Liu, A unified theory of central tolerance in the thymus. *Trends Immunol.* **27**, 215–221 (2006).

9. J. Gameiro, P. Nagib, L. Verinaud, The thymus microenvironment in regulating thymocyte differentiation. *Cell Adh. Migr.* **4**, 382–390 (2010).

10. W. E. Jenkinson, S. W. Rossi, S. M. Parnell, E. J. Jenkinson, G. Anderson, PDGFRalpha-expressing mesenchyme regulates thymus growth and the availability of intrathymic niches. *Blood*. **109**, 954–960 (2007).

11. S. Inglesfield, E. J. Cosway, W. E. Jenkinson, G. Anderson, Rethinking Thymic Tolerance: Lessons from Mice. *Trends Immunol.* **40**, 279–291 (2019).

12. J. F. A. P. Miller, The golden anniversary of the thymus. *Nat. Rev. Immunol.* **11**, 489–495 (2011).

13. J. Gordon, N. R. Manley, Mechanisms of thymus organogenesis and morphogenesis. *Development*. **138**, 3865–3878 (2011).

14. E. M. Kernfeld, R. M. J. Genga, K. Neherin, M. E. Magaletta, P. Xu, R. Maehr, A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation. *Immunity*. **48**, 1258–1270.e6 (2018).

15. C. Bornstein, S. Nevo, A. Giladi, N. Kadouri, M. Pouzolles, F. Gerbe, E. David, A. Machado, A. Chuprin, B. Tóth, O. Goldberg, S. Itzkovitz, N. Taylor, P. Jay, V. S. Zimmermann, J. Abramson, I. Amit, Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. *Nature*. **559**, 622–626 (2018).

16. C. N. Miller, I. Proekt, J. von Moltke, K. L. Wells, A. R. Rajpurkar, H. Wang, K. Rattay, I. S. Khan, T. C. Metzger, J. L. Pollack, A. C. Fries, W. W. Lwin, E. J. Wigton, A. V. Parent, B. Kyewski, D. J. Erle, K. A. Hogquist, L. M. Steinmetz, R. M. Locksley, M. S. Anderson, Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development. *Nature*. **559**, 627–631 (2018).

17. J. Mestas, C. C. W. Hughes, Of mice and not men: differences between mouse and human immunology. *J. Immunol.* **172**, 2731–2738 (2004).

18. A. M. Farley, L. X. Morris, E. Vroegindeweij, M. L. G. Depreter, H. Vaidya, F. H. Stenhouse, S. R. Tomlinson, R. A. Anderson, T. Cupedo, J. J. Cornelissen, C. C. Blackburn, Dynamics of thymus organogenesis and colonization in early human development. *Development*. **140**, 2015–2026 (2013).

19. B. V. Kumar, T. J. Connors, D. L. Farber, Human T Cell Development, Localization, and Function throughout Life. *Immunity*. **48**, 202–213 (2018).

20. J. Nikolich-Zugich, M. K. Slifka, I. Messaoudi, The many important facets of T-cell repertoire diversity. *Nat. Rev. Immunol.* **4**, 123–132 (2004).

21. R. Jores, T. Meo, Few V gene segments dominate the T cell receptor beta-chain repertoire of the human thymus. *J. Immunol.* **151**, 6110–6122 (1993).

22. J. A. Carter, J. B. Preall, K. Grigaityte, S. J. Goldfless, A. W. Briggs, F. Vigneault, G. S. Atwal, T-cell receptor αβ chain pairing is associated with CD4+ and CD8+ lineage specification. *bioRxiv* (2018), p. 293852.

23. P. L. Klarenbeek, M. E. Doorenspleet, R. E. E. Esveldt, B. D. C. van Schaik, N. Lardy, A. H. C. van Kampen, P. P. Tak, R. M. Plenge, F. Baas, P. I. W. de Bakker, N. de Vries, Somatic Variation of T-Cell Receptor Genes Strongly Associate with HLA Class Restriction. *PLoS One*. **10**, e0140815 (2015).

24. R. L. Warren, J. D. Freeman, T. Zeng, G. Choe, S. Munro, R. Moore, J. R. Webb, R. A. Holt, Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res.* **21**, 790–797 (2011).

25. S. Gopalakrishnan, K. Majumder, A. Predeus, Y. Huang, O. I. Koues, J. Verma-Gaur, S. Loguercio, A. I. Su, A. J. Feeney, M. N. Artyomov, E. M. Oltz, Unifying model for molecular determinants of the preselection Vβ repertoire. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E3206–15 (2013).

26. K. Polański, J.-E. Park, M. D. Young, Z. Miao, K. B. Meyer, S. A. Teichmann, BBKNN: Fast Batch Alignment of Single Cell Transcriptomes. *Bioinformatics* (2019), doi:10.1093/bioinformatics/btz625.

27. S. Hansen, L. Selman, N. Palaniyar, K. Ziegler, J. Brandt, A. Kliem, M. Jonasson, M.-O. Skjoedt, O. Nielsen, K. Hartshorn, T. J. D. Jørgensen, K. Skjødt, U. Holmskov, Collectin 11 (CL-11, CL-K1) is a MASP-1/3-associated plasma collectin with microbial-binding activity. *J. Immunol.* **185**, 6096–6104 (2010).

28. K. M. Sitnik, K. Kotarsky, A. J. White, W. E. Jenkinson, G. Anderson, W. W. Agace, Mesenchymal cells regulate retinoic acid receptor-dependent cortical thymic epithelial cell homeostasis. *J. Immunol.* **188**, 4801–4809 (2012).

29. C. Gu, E. Giraudo, The role of semaphorins and their receptors in vascular development and cancer. *Exp. Cell Res.* **319**, 1306–1316 (2013).

30. H. J. Garchon, F. Djabiri, J. P. Viard, P. Gajdos, J. F. Bach, Involvement of human muscle acetylcholine receptor alpha-subunit gene (CHRNA) in susceptibility to myasthenia gravis. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4668–4672 (1994).

31. L. Mesnard-Rouiller, J. Bismuth, A. Wakkach, S. Poëa-Guyon, S. Berrih-Aknin, Thymic myoid cells express high levels of muscle genes. *J. Neuroimmunol.* **148**, 97–105 (2004).

32. A. Zółtowska, T. Pawełczyk, M. Stopa, J. Skokowski, J. Stepiński, A. Roszkiewicz, W. Nyka, Myoid cells and neuroendocrine markers in myasthenic thymuses. *Arch. Immunol. Ther. Exp.* . **46**, 253–257 (1998).

33. V. S. Patil, A. Madrigal, B. J. Schmiedel, J. Clarke, P. O'Rourke, A. D. de Silva, E. Harris, B. Peters, G. Seumois, D. Weiskopf, A. Sette, P. Vijayanand, Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci Immunol.* **3** (2018).

34. N. Lopes, A. Sergé, P. Ferrier, M. Irla, Thymic Crosstalk Coordinates Medulla Organization and T-Cell Tolerance Induction. *Front. Immunol.* **6**, 365 (2015).

35. R. Vento-Tormo, M. Efremova, R. A. Botting, M. Y. Turco, M. Vento-Tormo, K. B. Meyer, J.-E. Park, E. Stephenson, K. Polański, A. Goncalves, L. Gardner, S. Holmqvist, J. Henriksson, A. Zou, A. M. Sharkey, B. Millar, B. Innes, L. Wood, A. Wilbrey-Clark, R. P. Payne, M. A. Ivarsson, S. Lisgo, A. Filby, D. H. Rowitch, J. N. Bulmer, G. J. Wright, M. J. T.

Stubbington, M. Haniffa, A. Moffett, S. A. Teichmann, Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature*. **563**, 347–353 (2018).

36. D. Elewaut, C. F. Ware, The unconventional role of LT alpha beta in T cell differentiation. *Trends Immunol.* **28**, 169–175 (2007).

37. S. W. Rossi, M.-Y. Kim, A. Leibbrandt, S. M. Parnell, W. E. Jenkinson, S. H. Glanville, F. M. McConnell, H. S. Scott, J. M. Penninger, E. J. Jenkinson, P. J. L. Lane, G. Anderson, RANK signals from CD4(+)3(-) inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. *J. Exp. Med.* **204**, 1267–1272 (2007).

38. J. M. Revest, R. K. Suniara, K. Kerr, J. J. Owen, C. Dickson, Development of the thymus requires signaling through the fibroblast growth factor receptor R2-IIIb. *J. Immunol.* **167**, 1954–1961 (2001).

39. M. J. García-León, P. Fuentes, J. L. de la Pompa, M. L. Toribio, Dynamic regulation of NOTCH1 activation and Notch ligand expression in human thymus development. *Development*. **145** (2018), doi:10.1242/dev.165597.

40. G. E. Desanti, J. E. Cowan, S. Baik, S. M. Parnell, A. J. White, J. M. Penninger, P. J. L. Lane, E. J. Jenkinson, W. E. Jenkinson, G. Anderson, Developmentally regulated availability of RANKL and CD40 ligand reveals distinct mechanisms of fetal and adult cross-talk in the thymus medulla. *J. Immunol.* **189**, 5519–5526 (2012).

41. E. J. Cosway, B. Lucas, K. D. James, S. M. Parnell, M. Carvalho-Gaspar, A. J. White, A. V. Tumanov, W. E. Jenkinson, G. Anderson, Redefining thymus medulla specialization for central tolerance. *J. Exp. Med.* **214**, 3183–3195 (2017).

42. I. Van de Walle, G. De Smet, M. Gärtner, M. De Smedt, E. Waegemans, B. Vandekerckhove, G. Leclercq, J. Plum, J. C. Aster, I. D. Bernstein, C. J. Guidos, B. Kyewski, T. Taghon, Jagged2 acts as a Delta-like Notch ligand during early hematopoietic cell fate decisions. *Blood*. **117**, 4449–4459 (2011).

43. I. Van de Walle, E. Waegemans, J. De Medts, G. De Smet, M. De Smedt, S. Snauwaert, B. Vandekerckhove, T. Kerre, G. Leclercq, J. Plum, T. Gridley, T. Wang, U. Koch, F. Radtke, T. Taghon, Specific Notch receptor-ligand interactions control human TCR-αβ/γδ development by inducing differential Notch signal strength. *J. Exp. Med.* **210**, 683–697 (2013).

44. D.-M. Popescu, R. A. Botting, E. Stephenson, K. Green, S. Webb, L. Jardine, E. F. Calderbank, K. Polanski, I. Goh, M. Efremova, M. Acres, D. Maunder, P. Vegh, Y. Gitton, J.-E. Park, R. Vento-Tormo, Z. Miao, D. Dixon, R. Rowell, D. McDonald, J. Fletcher, E. Poyner, G. Reynolds, M. Mather, C. Moldovan, L. Mamanova, F. Greig, M. D. Young, K. B. Meyer, S. Lisgo, J. Bacardit, A. Fuller, B. Millar, B. Innes, S. Lindsay, M. J. T. Stubbington, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Filby, P. Carey, A.-C. Villani, A. Roy, A. Regev, A. Chédotal, I. Roberts, B. Göttgens, S. Behjati, E. Laurenti, S. A. Teichmann, M. Haniffa, Decoding human fetal liver haematopoiesis. *Nature*. **574**, 365–371 (2019).

45. Y. Zeng, C. Liu, Y. Gong, Z. Bai, S. Hou, J. He, Z. Bian, Z. Li, Y. Ni, J. Yan, T. Huang, H. Shi, C. Ma, X. Chen, J. Wang, L. Bian, Y. Lan, B. Liu, H. Hu, Single-Cell RNA Sequencing Resolves Spatiotemporal Development of Pre-thymic Lymphoid Progenitors and Thymus Organogenesis in Human Embryos. *Immunity*. **51**, 930–948.e6 (2019).

46. D. Pellin, M. Loperfido, C. Baricordi, S. L. Wolock, A. Montepeloso, O. K. Weinberg, A. Biffi, A. M. Klein, L. Biasco, A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat. Commun.* **10**, 2395 (2019).

47. D. K. Shah, J. C. Zúñiga-Pflücker, An overview of the intrathymic intricacies of T cell development. *J. Immunol.* **192**, 4017–4023 (2014).
48. H. T. Petrie, M. Tourigny, D. B. Burtrum, F. Livak, Precursor thymocyte proliferation and differentiation are controlled by signals unrelated to the pre-TCR. *J. Immunol.* **165**, 3094–3098 (2000).
49. M. R. Tourigny, S. Mazel, D. B. Burtrum, H. T. Petrie, T cell receptor (TCR)-beta gene recombination: dissociation from cell cycle regulation and developmental progression during T cell ontogeny. *J. Exp. Med.* **185**, 1549–1556 (1997).
50. W. A. Dik, K. Pike-Overzet, F. Weerkamp, D. de Ridder, E. F. E. de Haas, M. R. M. Baert, P. van der Spek, E. E. L. Koster, M. J. T. Reinders, J. J. M. van Dongen, A. W. Langerak, F. J. T. Staal, New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **201**, 1715–1723 (2005).
51. B. J. Schmiedel, D. Singh, A. Madrigal, A. G. Valdovino-Gonzalez, B. M. White, J. Zapardiel-Gonzalo, B. Ha, G. Altay, J. A. Greenbaum, G. McVicker, G. Seumois, A. Rao, M. Kronenberg, B. Peters, P. Vijayanand, Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*. **175**, 1701–1715.e16 (2018).
52. D. L. Owen, S. A. Mahmud, L. E. Sjaastad, J. B. Williams, J. A. Spanier, D. R. Simeonov, R. Ruscher, W. Huang, I. Proekt, C. N. Miller, C. Hekim, J. C. Jeschke, P. Aggarwal, U. Broeckel, R. S. LaRue, C. M. Henzler, M.-L. Alegre, M. S. Anderson, A. August, A. Marson, Y. Zheng, C. B. Williams, M. A. Farrar, Thymic regulatory T cells arise via two distinct developmental programs. *Nat. Immunol.* **20**, 195–205 (2019).
53. G. Verstichel, D. Vermijlen, L. Martens, G. Goetgeluk, M. Brouwer, N. Thiault, Y. Van Caeneghem, S. De Munter, K. Weening, S. Bonte, G. Leclercq, T. Taghon, T. Kerre, Y. Saeys, J. Van Dorpe, H. Cheroutre, B. Vandekerckhove, The checkpoint for agonist selection precedes conventional selection in human thymus. *Sci Immunol*. **2** (2017), doi:10.1126/sciimmunol.aah4232.
54. J. R. Fergusson, V. M. Fleming, P. Klenerman, CD161-expressing human T cells. *Front. Immunol.* **2**, 36 (2011).
55. E. S. Alonzo, D. B. Sant'Angelo, Development of PLZF-expressing innate T cells. *Curr. Opin. Immunol.* **23**, 220–227 (2011).
56. R. Ruscher, R. L. Kummer, Y. J. Lee, S. C. Jameson, K. A. Hogquist, CD8αα intraepithelial lymphocytes arise from two main thymic precursors. *Nat. Immunol.* **18**, 771–779 (2017).
57. J. Oh, J.-S. Shin, The Role of Dendritic Cells in Central Tolerance. *Immune Netw.* **15**, 111–120 (2015).
58. A.-C. Villani, R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A. Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D. McDonald, A. Filby, W. Li, P. L. De Jager, O. Rozenblatt-Rosen, A. A. Lane, M. Haniffa, A. Regev, N. Hacohen, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. **356** (2017), doi:10.1126/science.aah4573.
59. N. Watanabe, Y.-H. Wang, H. K. Lee, T. Ito, Y.-H. Wang, W. Cao, Y.-J. Liu, Hassall's corpuscles instruct dendritic cells to induce CD4+CD25+ regulatory T cells in human thymus. *Nature*. **436**, 1181–1185 (2005).
60. P. J. Fairchild, J. M. Austyn, Thymic dendritic cells: phenotype and function. *Int. Rev. Immunol.* **6**, 187–196 (1990).

61. J. R. Fergusson, M. D. Morgan, M. Bruchard, L. Huitema, B. A. Heesters, V. van Unen, J. P. van Hamburg, N. N. van der Wel, D. Picavet, F. Koning, S. W. Tas, M. S. Anderson, J. C. Marioni, G. A. Holländer, H. Spits, Maturing Human CD127+ CCR7+ PDL1+ Dendritic Cells Express AIRE in the Absence of Tissue Restricted Antigens. *Front. Immunol.* **9**, 2902 (2018).
62. Z. Hu, J. N. Lancaster, L. I. R. Ehrlich, The Contribution of Chemokines and Migration to the Induction of Central Tolerance in the Thymus. *Front. Immunol.* **6**, 398 (2015).
63. Y. Lei, A. M. Ripen, N. Ishimaru, I. Ohigashi, T. Nagasawa, L. T. Jeker, M. R. Bösl, G. A. Holländer, Y. Hayashi, R. de W. Malefyt, T. Nitta, Y. Takahama, Aire-dependent production of XCL1 mediates medullary accumulation of thymic dendritic cells and contributes to regulatory T cell development. *J. Exp. Med.* **208**, 383–394 (2011).
64. J. A. Skok, R. Gisler, M. Novatchkova, D. Farmer, W. de Laat, M. Busslinger, Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nat. Immunol.* **8**, 378–387 (2007).
65. R. J. Mallis, K. Bai, H. Arthanari, R. E. Hussey, M. Handley, Z. Li, L. Chingozha, J. S. Duke-Cohan, H. Lu, J.-H. Wang, C. Zhu, G. Wagner, E. L. Reinherz, Pre-TCR ligand binding impacts thymocyte development before αβTCR expression. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8373–8378 (2015).
66. Z. M. Carico, K. Roy Choudhury, B. Zhang, Y. Zhuang, M. S. Krangel, Tcrd Rearrangement Redirects a Processive Tcra Recombination Program to Expand the Tcra Repertoire. *Cell Rep.* **19**, 2157–2173 (2017).
67. P. Bullen, D. I. Wilson, The Carnegie staging of human embryos: a practical guide. *Molecular genetics of early human development*, 27–35 (1997).
68. W. M. Hern, Correlation of fetal age and measurements between 10 and 26 weeks of gestation. *Obstet. Gynecol.* **63**, 26–32 (1984).
69. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst*. **8**, 281–291.e9 (2019).
70. L. Haghverdi, M. Büttner, F. A. Wolf, F. Buettner, F. J. Theis, Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods*. **13**, 845–848 (2016).
71. J.-E. Park, K. Polański, K. Meyer, S. A. Teichmann, Fast Batch Alignment of Single Cell Transcriptomes Unifies Multiple Mouse Cell Atlases into an Integrated Landscape. *bioRxiv* (2018), p. 397042.
72. H. Hu, Y.-R. Miao, L.-H. Jia, Q.-Y. Yu, Q. Zhang, A.-Y. Guo, AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33–D38 (2019).

Acknowledgements: We gratefully acknowledge the Sanger Flow Cytometry Facility, Newcastle University Flow Cytometry Core Facility, Sanger Cellular Generation and Phenotyping (CGaP) Core Facility, and Sanger Core Sequencing pipeline for support with sample processing and sequencing library preparation. We thank the MRC/Engineering and Physical Sciences Research Council Newcastle Molecular Pathology Node for support on paraffin embedding fetal tissues. We thank Jana Eliasova for graphical images and Jinwook Choi for helpful discussions, and Sarah Aldridge for editing the manuscript. The human embryonic and fetal material was provided by the Joint MRC / Wellcome (MR/R006237/1) Human Developmental Biology Resource (www.hdbr.org). The material from the deceased organ donor was provided by the Cambridge Biorepository for Translational Medicine. Paediatric thymus material was provided by Ghent University Hospital. We are grateful to the donors and donor families for granting access to the tissue samples. **Funding**: This study was supported by Wellcome Human Cell Atlas Strategic Science Support (WT211276/Z/18/Z); S.A.T. was funded by Wellcome (WT206194), ERC Consolidator (no. 646794) and EU MRG-Grammar awards ; M.H. was funded by Wellcome (WT107931/Z/15/Z), The Lister Institute for Preventive Medicine and NIHR and Newcastle-Biomedical Research Centre; and T.T. was funded by the Research Foundation Flanders (FWO grant G053816N) and Ghent University Special Research Fund (BOF18-GOA-024). J.-E.P. is supported by an EMBO Long-Term Fellowship. D.J.K. was supported by the Wellcome Trust under grants 203828/Z/16/A and 203828/Z/16/Z. R.A/B is an NIHR Senior Investigator and is also supported by NIHR funding of the Cambridge Biomedical Research Centre. Cambridge fetal tissue was collected in part from funding from the WT MRC Cambridge Stem Cell Institute. P.B. received funding from the European Research Council (ERC-Stg 639429), the Rosetrees Trust (M362; M362-F1), the UCL Therapeutic Acceleration Support Fund and was supported by the GOSH BRC. **Author contributions:** J-E.P, S.A.T, M.H, and T.T. designed the experiments. J-E.P; R.A.B; D.M.P.; E.S.; M.L.; C.D.C; J.R.F; N.V.; K.M.; I.G; X.H. performed sampling and library prep with help from S.W.; D.M.; A.F.; A.F.; L.M.; G.R.; D.D.; R.V-T.. J-E.P.; D.J.K; K.P.; M.L; C.D.C; V.R.K. analysed the data. J-E.P.; C.D.C; E.T.; O.B.; K.R.; A.W-C.; R.A.B; R.R. performed validation experiments. J-E.P; C.D.C; S.A.T; M.H.; T.T. wrote the manuscript with contributions from R.A.B; K.B.M; Y.S.; M.R.C.; P.B; S.B.; R.V-T.; D.J.K. All authors read and accepted the manuscript. **Competing interests**: A patent application has been submitted by Wellcome Sanger Institute based on these results. **Data and materials availability**: All sequencing data has been deposited to ArrayExpress (Accession: E-MTAB-8581) and can be accessed online through Interactive web portal (https://developmentcellatlas.ncl.ac.uk). All codes used for data analysis are available from Zenodo repository (DOI: 10.5281/zenodo.3572422)

Supplementary Materials:

Materials and Methods

Supplementary Text

Figures S1 to S29

Tables S1 to S8

Data S1

References 65 - 71

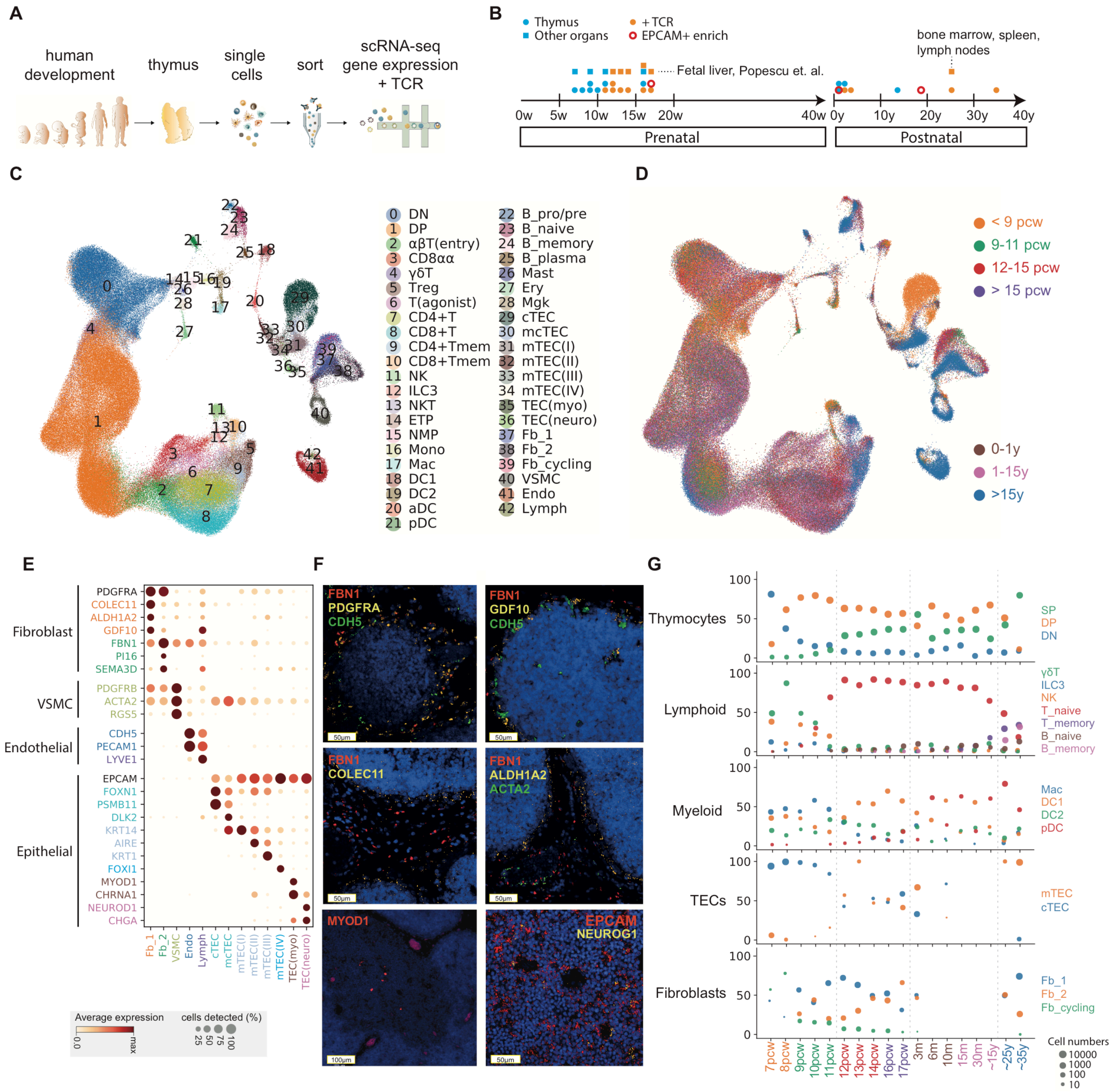

Fig. 1. Cellular composition of the developing human thymus

(A) Schematic of single-cell transcriptome profiling of the developing human thymus.
(B) Summary of gestational stage/age of samples, organs (circle: thymus, rectangle: fetal liver, adult bone marrow, adult spleen and lymph nodes) and 10x Genomics chemistry (colours).
(C) UMAP visualisation of the cellular composition of the human thymus colored by cell type (DN: double-negative T cells, DP: double-positive T cells, ETP: Early thymic progenitor, aDC: activated dendritic cells, pDC: plasmacytoid dendritic cells, Mono: monocyte, Mac: macrophage, Mgk: megakaryocyte, Endo: endothelial cells, VSMC: vesicular smooth muscle cells, Epi: epithelial cells, Fb: fibroblasts, Ery: erythrocytes).
(**D**) Same UMAP plot coloured by age groups, indicated by post-conception weeks (PCW) or postnatal years (y).

(E) Dot plot for marker gene expressions in thymic stromal cell types. Color represents maximum-normalised mean expression of marker genes in each cell group, and size indicates the proportion of cells expressing marker genes. (This scheme is consistently used throughout the manuscript.)
(F) RNA single-molecule FISH in human fetal thymus slides with probes targeting stromal cell populations. Top left: Fb2 population marker *FBN1* (red), general fibroblast markers *PDGFRA* (yellow) and *CDH5* (green). Top right: Fb1 marker *GDF10* (yellow), *FBN1* (red) and *CDH5* (green). Middle left: Fb1 marker *COLEC11* (yellow), *FBN1* (red), Middle right: Fb1 marker *ALDH1A2* (yellow), VSMC marker *ACTA2* (green), *FBN1* (red), Bottom left: TEC(myo) marker *MYOD1* (red). Bottom right: Epithelial cell marker *EPCAM* (red) and TEC(neuro) marker *NEUROG1* (yellow). Data representative of n=2.
(G) Relative proportion of cell types throughout different age groups. Dot size are proportional to absolute cell numbers detected in the dataset. Statistical testing for population dynamics was performed by t-testing using proportions between stage groups. X-axis shows age of samples, which are coloured in the same scheme as Fig. 1D.

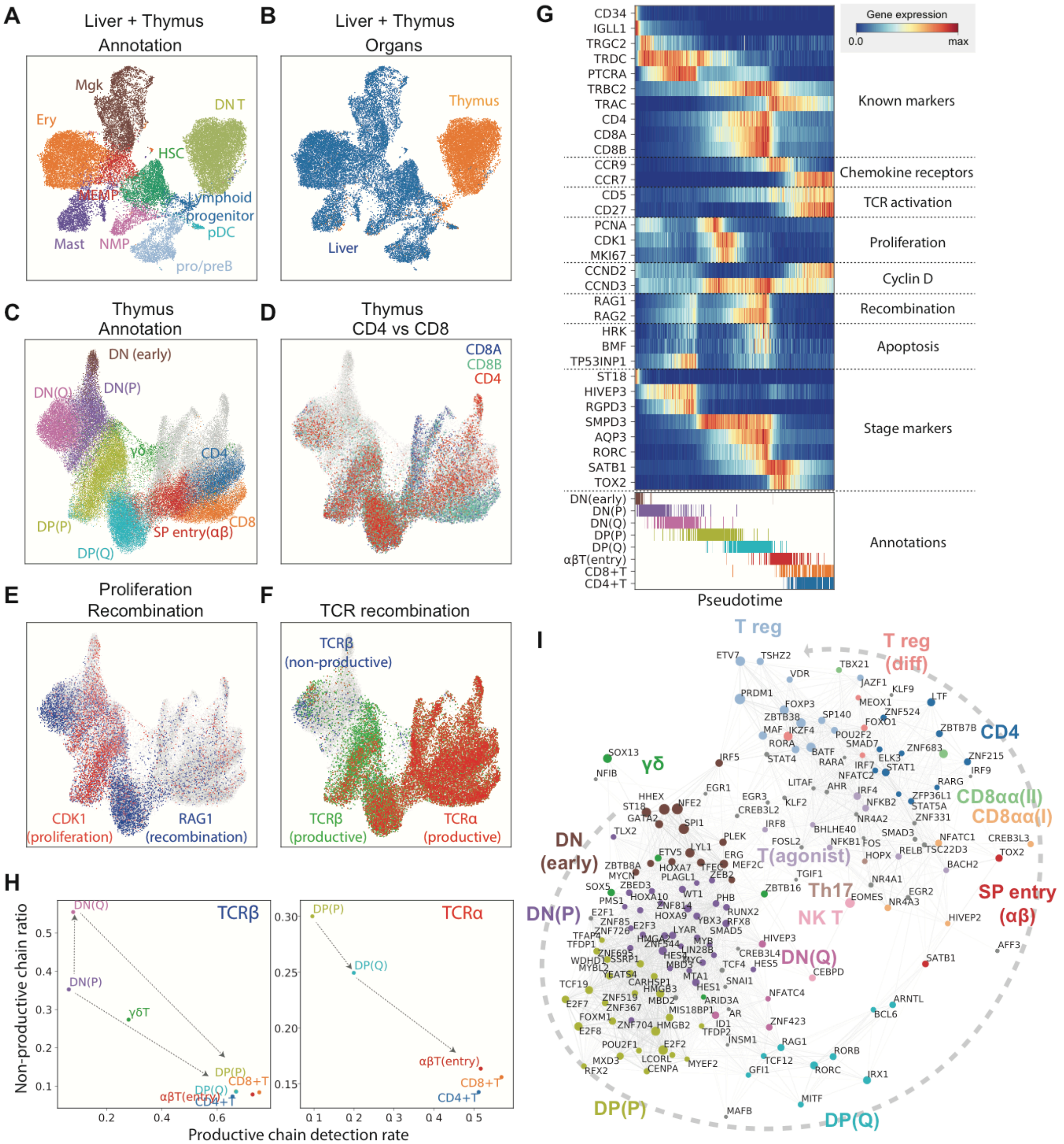

Fig. 2. Thymic seeding of early thymic progenitors (ETPs) and T cell differentiation trajectory

(A) UMAP visualisation of ETP and fetal liver hematopoietic stem cells/early progenitors. (HSC: Hematopoietic stem cells, NMP: Neutrophil-Myeloid progenitors, MEMP: Megakaryocyte-Erythrocyte-Mast cell progenitors). The same UMAP coloured by **(B)** organ (liver in blue and thymus in yellow/red). **(C)** UMAP visualisation of developing thymocytes after batch correction. (DN: double negative T cells, DP: double positive T cells, SP: single positive T cells, P: proliferating, Q: quiescent). The data contains cells from all sampled developmental stages. Cells from abundant clusters are down-sampled for better visualisation. The reproducibility of structure is confirmed across individual sample. Unconventional T cells are marked as grey.

(**D-F**) The same UMAP plot showing *CD4* (red), *CD8A* (blue) and *CD8B* (turquoise) gene expression (D)**,** *CDK1* (red) cell cycle and *RAG1* (blue) recombination gene expression (E), and TCRα (red) and TCRβ (green = productive and blue = non-productive) VDJ genes (F).
(G) Heatmap showing differentially expressed genes across T cell differentiation pseudotime. Upper panel: X-axis represents pseudo-temporal ordering. Gene expression levels across pseudotime axis are maximum-normalised and smoothed. Genes are grouped by their functional categories and expression patterns. Lower panel: Cell type annotation of cells aligned along the pseudotime axis. The same colour schemes are used as (C).
(H) Scatter plot showing the rate of productive chain detection within cells in specific cell types (x-axis) and the ratio between the number of non-productive/productive TCR chains detected in specific cell types (y-axis); TCRβ (left panel) and TCRα (right panel).
(I) Graph showing correlation-based network of transcription factors expressed by thymocytes. Nodes represent transcription factors, and edge widths are proportional to the correlation coefficient between two transcription factors. TFs with significant association to specific cell types depicted in colour. Node size is proportional to the significance of association to specific cell types.

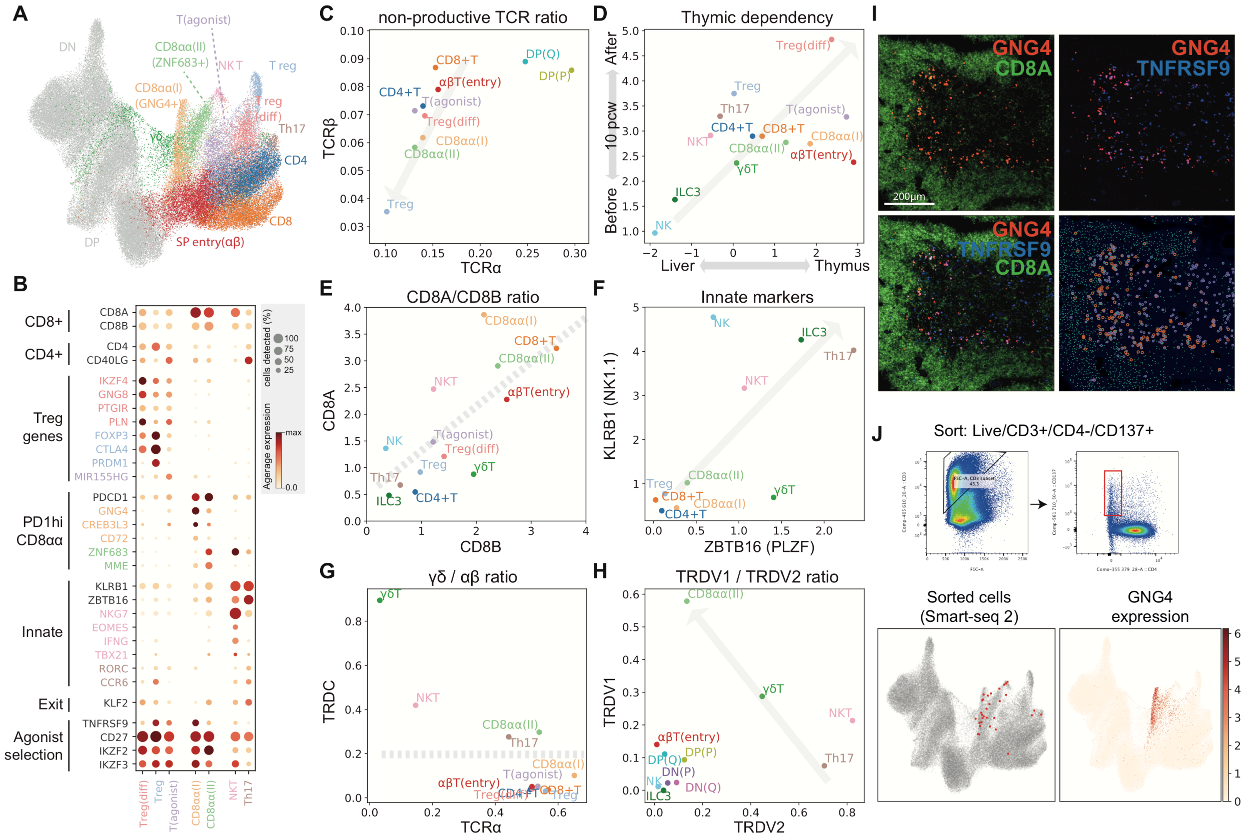

Fig. 3. Identification of GNG4+ CD8aa T cells in the thymic medulla

(A) UMAP visualisation of mature T cell populations in the thymus. Axes and coordinates are as Fig. 2C. (The cell annotation colour scheme used here is maintained throughout this figure.)
(B) Dot plot showing marker gene expression for the mature T cell types. Genes are stratified according to associated cell types or functional relationship.
(C) Scatter plot showing the ratio between the number of non-productive/productive TCR chains detected in specific cell types in TCRα chain (x-axis) and TCRβ chain (y-axis). Same colour schemes apply as in (A). The grey arrow indicates a trendline for decreasing non-productive TCR chain ratio in unconventional *versus* conventional T cells.
(D) Scatter plot showing the relative abundance of each cell type between fetal liver and thymus (x-axis) and before and after thymic maturation (delimited at 10 PCW) (y-axis). Grey arrow indicates trendline for increasing thymic dependency.
(E-H) Scatter plot comparing the characteristics of unconventional T cells based on *CD8A* vs. *CD8B* expression levels (E), *KLRB1* vs *ZBTB16* expression levels (F), TCRα productive chain vs *TRDC* detection ratio (G) and *TRDV1* vs *TRDV2* expression levels (H). Grey arrows or lines are used to set boundaries between groups (E, G, H) or indicate the trend of innate marker gene expression (F).
(I) single-molecule RNA FISH showing *GNG4* (red), *TNFRSF9* (blue) and *CD8A* (green) in a 15 PCW thymus. Right bottom panel shows detected spots from the image on top of the tissue structure based on DAPI signal. Colour scheme for spots are the same as in the image.
(J) FACS gating strategy to isolate CD8aa(I) cells (live/CD3+/CD4-/CD137+) and Smart-seq2 validation of FACS-isolated cells projected to the UMAP presentation of total mature T cells from discovery dataset (bottom left panel). *GNG4* expression pattern is overlaid onto the same UMAP plot (bottom right panel).

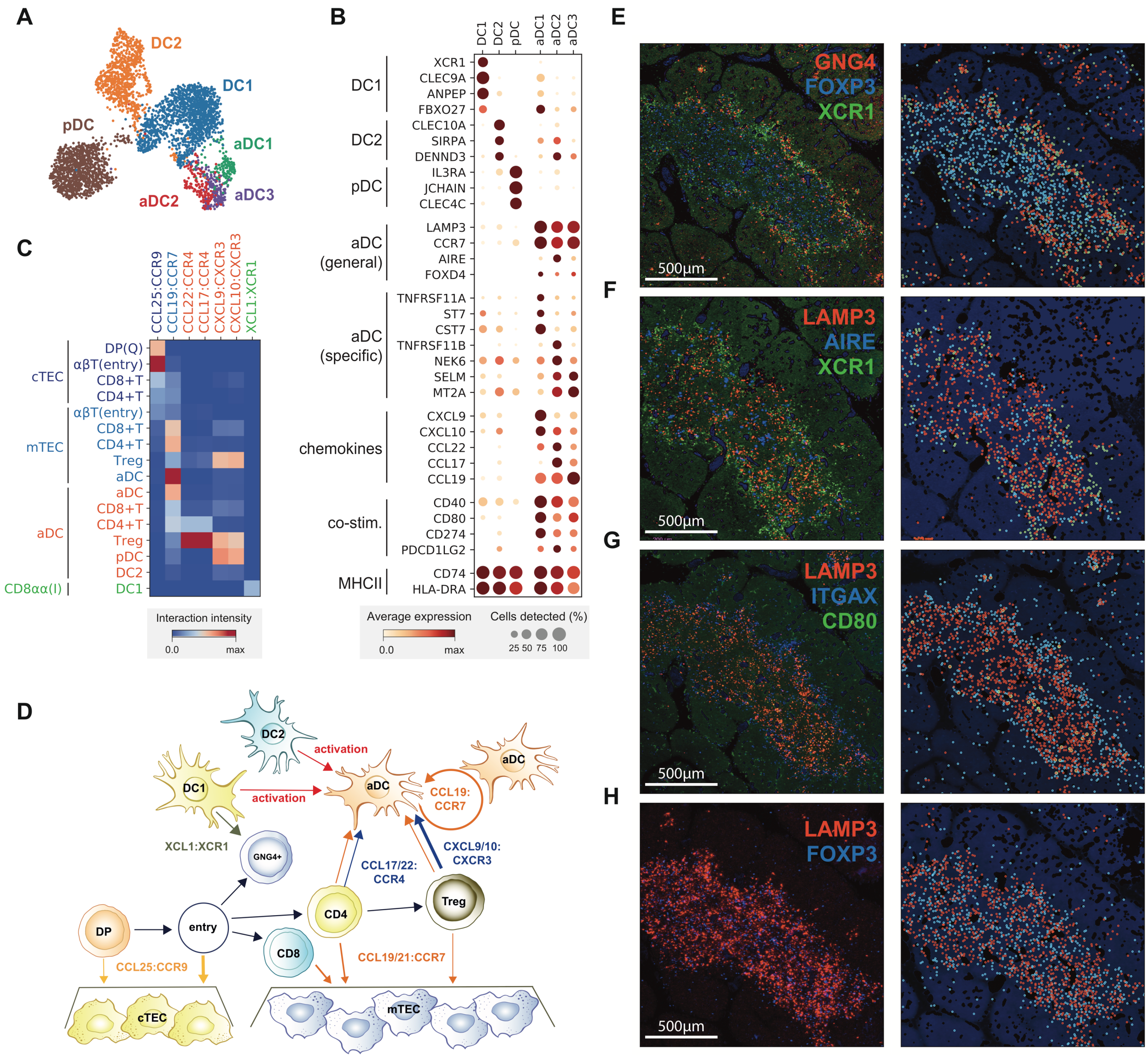

Fig. 4. Recruitment and activation of dendritic cells for thymocyte selection

(A) UMAP visualisation of thymic DC populations and **(B)** dot plot of their marker genes.
(C) Heat map of chemokine interactions between T cells, DCs and TECs, where the chemokine is expressed by the outside cell type and the cognate receptor by the inside cell type.
(**D**) Schematic model summarising the interactions between thymic epithelial cells (TECs), dendritic cells (DCs) and T cells. The ligand is secreted by the cell at the beginning of the arrow, and the receptor is expressed by the cell at the end of the arrow.
(E) Left-hand panels: single molecule RNA FISH detection of *GNG4* (red), *XCR1* (green) and *FOXP3* (blue) in 15 PCW thymus. Right-hand panels: Computationally detected spots are presented as a solid circle over the tissue structure based on DAPI signal. Colour schemes for circles are the same as in the image.
(F-H) Sequential slide sections from the same sample are stained for the detection of *LAMP3* (red), *AIRE* (blue) and *XCR1* (green) (F), *LAMP3* (red), *ITGAX* (blue) and *CD80* (green) (G), *LAMP3* (red), *FOXP3* (blue) (H). Spot detection and representation as in (**E**). Data representative of n = 2.

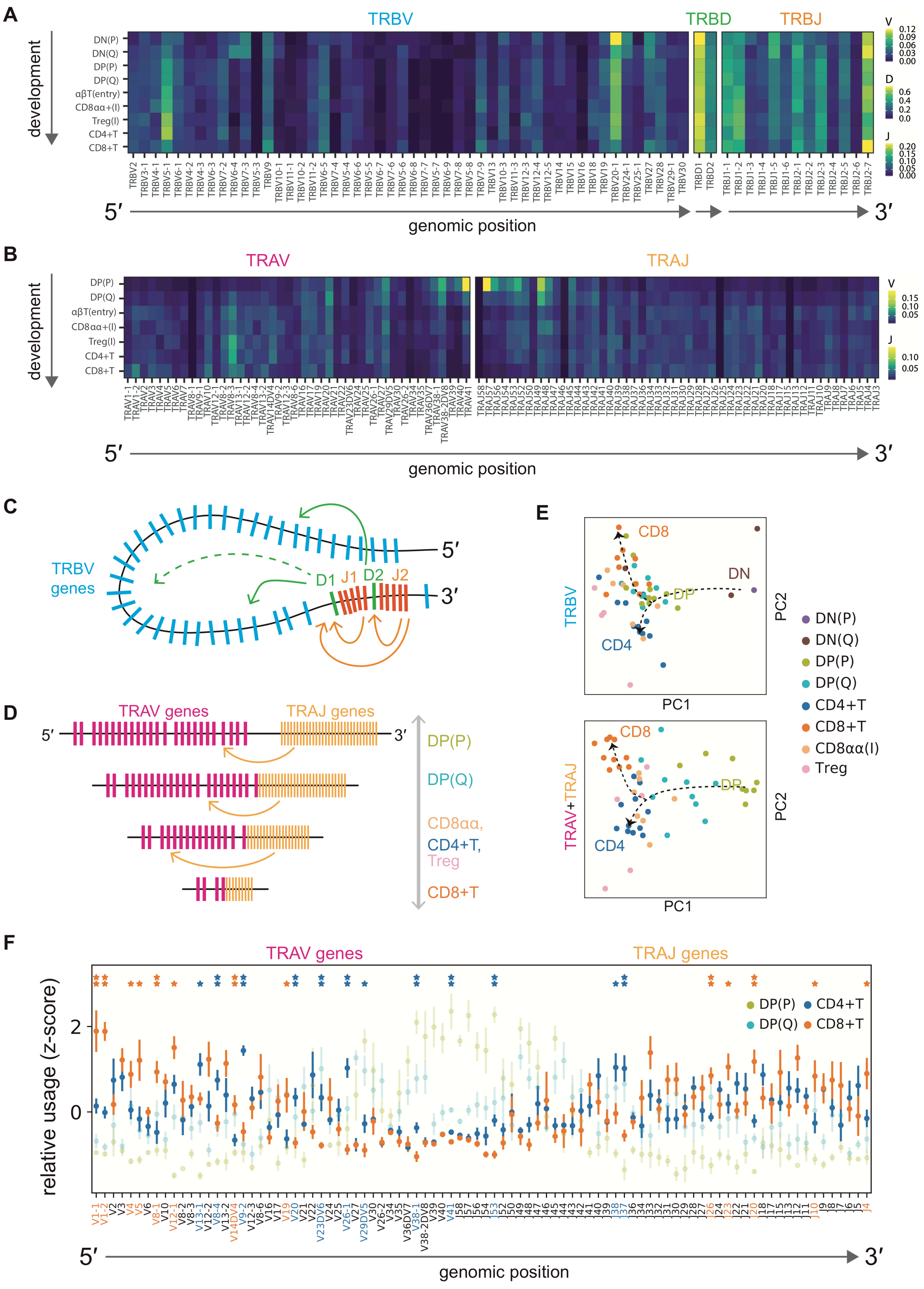
A
TRBV
TRBD
TRBJ
development
DN(P)
DN(Q)
DP(P)
DP(Q)
αβT(entry)
CD8αα+(I)
Treg(I)
CD4+T
CD8+T
TRBV2 TRBV3-1 TRBV4-1 TRBV5-1 TRBV6-1 TRBV4-2 TRBV4-3 TRBV6-3 TRBV7-2 TRBV6-4 TRBV7-3 TRBV5-3 TRBV9 TRBV10-1 TRBV11-1 TRBV10-2 TRBV11-2 TRBV6-5 TRBV7-4 TRBV5-4 TRBV6-6 TRBV5-5 TRBV6-7 TRBV7-6 TRBV5-6 TRBV6-8 TRBV7-7 TRBV5-7 TRBV6-9 TRBV7-8 TRBV5-8 TRBV7-9 TRBV13 TRBV10-3 TRBV11-3 TRBV12-3 TRBV12-4 TRBV12-5 TRBV14 TRBV15 TRBV16 TRBV18 TRBV19 TRBV20-1 TRBV24-1 TRBV25-1 TRBV27 TRBV28 TRBV29-1 TRBV30
TRBD1 TRBD2
TRBJ1-1 TRBJ1-2 TRBJ1-3 TRBJ1-4 TRBJ1-5 TRBJ1-6 TRBJ2-1 TRBJ2-2 TRBJ2-3 TRBJ2-4 TRBJ2-5 TRBJ2-6 TRBJ2-7
V
0.12
0.09
0.06
0.03
0.00
D
0.6
0.4
0.2
0.0
J
0.20
0.15
0.10
0.05
0.00
5′
genomic position
3′
B
TRAV
TRAJ
development
DP(P)
DP(Q)
αβT(entry)
CD8αα+(I)
Treg(I)
CD4+T
CD8+T
TRAV1-1 TRAV1-2 TRAV2 TRAV3 TRAV4 TRAV5 TRAV6 TRAV7 TRAV8-1 TRAV9-1 TRAV10 TRAV12-1 TRAV8-2 TRAV8-3 TRAV13-1 TRAV12-2 TRAV8-4 TRAV13-2 TRAV14DV4 TRAV9-2 TRAV12-3 TRAV8-6 TRAV16 TRAV17 TRAV19 TRAV20 TRAV21 TRAV22 TRAV23DV6 TRAV24 TRAV25 TRAV26-1 TRAV27 TRAV29DV5 TRAV30 TRAV26-2 TRAV34 TRAV35 TRAV36DV7 TRAV38-1 TRAV38-2DV8 TRAV39 TRAV40 TRAV41
TRAJ58 TRAJ57 TRAJ56 TRAJ54 TRAJ53 TRAJ52 TRAJ50 TRAJ49 TRAJ48 TRAJ47 TRAJ46 TRAJ45 TRAJ44 TRAJ43 TRAJ42 TRAJ41 TRAJ40 TRAJ39 TRAJ38 TRAJ37 TRAJ36 TRAJ34 TRAJ33 TRAJ32 TRAJ31 TRAJ30 TRAJ29 TRAJ28 TRAJ27 TRAJ26 TRAJ25 TRAJ24 TRAJ23 TRAJ22 TRAJ21 TRAJ20 TRAJ18 TRAJ17 TRAJ15 TRAJ13 TRAJ12 TRAJ11 TRAJ10 TRAJ9 TRAJ8 TRAJ7 TRAJ6 TRAJ5 TRAJ4 TRAJ3
V
0.15
0.10
0.05
J
0.10
0.05
5′
genomic position
3′
C
TRBV
genes
5′
3′
D1
J1
D2
J2
D
TRAV genes
TRAJ genes
5′
3′
DP(P)
DP(Q)
CD8αα,
CD4+T,
Treg
CD8+T
E
TRBV
CD8
DN
DP
CD4
PC1
PC2
TRAV+TRAJ
CD8
DP
CD4
PC1
PC2
DN(P)
DN(Q)
DP(P)
DP(Q)
CD4+T
CD8+T
CD8αα(I)
Treg
F
TRAV genes
TRAJ genes
relative usage (z-score)
2
0
DP(P)
CD4+T
DP(Q)
CD8+T
V1-1 V1-2 V2 V3 V4 V5 V6 V8-1 V10 V12-1 V8-2 V8-3 V13-1 V12-2 V8-4 V13-2 V14DV4 V9-2 V12-3 V8-6 V16 V17 V19 V20 V21 V22 V23DV6 V24 V25 V26-1 V27 V29DV5 V30 V26-2 V34 V35 V36DV7 V38-1 V38-2DV8 V39 V40 V41 J58 J57 J56 J54 J53 J52 J50 J49 J48 J47 J46 J45 J44 J43 J42 J41 J40 J39 J38 J37 J36 J34 J33 J32 J31 J30 J29 J28 J27 J26 J24 J23 J22 J21 J20 J18 J17 J15 J13 J12 J11 J10 J9 J8 J7 J6 J5 J4
5′
genomic position
3′

Fig. 5. Intrinsic bias in human TCR repertoire formation and selection

(A) Heatmap showing the proportion of each TCRβ V, D, J gene segment present at progressive stages of T cell development. Gene segments are positioned according to genomic location.
(B) Same scheme as in (A) applied to TCRα V and J gene segments. While there is a usage bias of segments at the beginning of development, segments are evenly used by the late developmental stages, indicating progressive recombination leading to even usage of segments.
(C-D) Schematics illustrating a hypothetical chromatin loop that may explain genomic location bias in recombination of TCRβ locus (C) and the mechanism of progressive recombination of TCRα locus leading to even usage of segments (D).
(E) PCA plots showing TRBV or TRAV and TRAJ gene usage pattern in different T cell types. Arrows depict T cell developmental order. For TRBV, there is a strong effect from beta selection, after which point the CD4+ and CD8+ repertoires diverge. The development for TRAV+TRAJ is more progressive, with stepwise divergence into the CD4+ and CD8+ repertoires.
(F) Relative usage of TCRα V and J gene segments according to cell type. The *Z*-score for each segment is calculated from the distribution of normalised proportions stratified by the cell type and sample. *P*-value is calculated by comparing z-scores in CD4+T and CD8+T cells using t-test, and FDR is calculated using Benjamini-Hochberg correction. (*: *p*-value < 0.05, **: FDR < 10%). Gene names on the x-axis and asterisks are coloured by significant enrichment in CD4+T cells (blue) or CD8+T cells (orange).

Supplementary Materials for

A cell atlas of human thymic development defines T cell repertoire formation

Jong-Eun Park[1], Rachel A. Botting[2], Cecilia Domínguez Conde[1], Dorin-Mirel Popescu[2], Marieke Lavaert[3,4], Daniel J. Kunz[1,19,20], Issac Goh[2], Emily Stephenson[2], Roberta Ragazzini[9], Elizabeth Tuck[1], Anna Wilbrey-Clark[1], Kenny Roberts[1], Veronika R. Kedlian[1], John R. Ferdinand[5], Xiaoling He[25], Simone Webb[2], Daniel Maunder[2], Niels Vandamme[6,21], Krishnaa Mahbubani[7], Krzysztof Polanski[1], Lira Mamanova[1], Liam Bolt[1], David Crossland[8,24], Fabrizio de Rita[24], Andrew Fuller[2], Andrew Filby[2], Gary Reynolds[2], David Dixon[2], Kourosh Saeb-Parsy[7], Steven Lisgo[8], Deborah Henderson[8], Roser Vento-Tormo[1], Omer A. Bayraktar[1], Roger A. Barker[25], Kerstin B. Meyer[1], Yvan Saeys[6,21], Paola Bonfanti[9,10,11], Sam Behjati[1,22,23], Menna R. Clatworthy[1,5,12], Tom Taghon[3,4,*], Muzlifah Haniffa[1,2,13,*], Sarah A. Teichmann[1,19,*]

Correspondence to: st9@sanger.ac.uk

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S29
Captions for Tables S1 to S5
Tables S6 to S8
Captions for Data S1

Other Supplementary Materials for this manuscript include the following:

Table S1 to S5
Data S1

Materials and Methods

Tissue Acquisition

All tissue samples used for this study were obtained with written informed consent from all participants in accordance with the guidelines in The Declaration of Helsinki 2000 from multiple centres. Human fetal tissues were obtained from the MRC/Wellcome Trust-funded Human Developmental Biology Resource (HDBR, http://www.hdbr.org) with appropriate maternal written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee (08/H0906/21+5). HDBR is regulated by the UK Human Tissue Authority (HTA; www.hta.gov.uk) and operates in accordance with the relevant HTA Codes of Practice. Some human embryonic thymic tissues were also obtained from Wellcome-MRC Cambridge Stem Cell Institute and Department of Clinical Neurosciences with appropriate maternal written consent and approval from Research Ethics Committee (REC No: 96/085). Human paediatric samples were obtained from Ghent University Hospital and Newcastle Hospitals NHS Trust with appropriate written consent and approval from the Ghent University Hospital Ethics Committee (B670201319452) and the East Midlands-Derby Research Ethics Committee (REC No: 18/EM/0314) respectively. The human adult deceased donor sample was obtained from the Cambridge Biorepository for Translational Medicine (CBTM) with appropriate written consent and approval from the Cambridge University Ethics Committee (reference 15/EE/0152, East of England Cambridge South Research Ethics Committee).

Tissue Processing

All tissues were processed immediately after isolation using consistent protocols with variation in enzymatic digestion strength. Tissue was transferred to a sterile 10mm^2 tissue culture dish and cut into <1mm^3 segments before being transferred to a 50mL conical tube. For mild digestion, tissues were digested with 1.6mg/mL collagenase type IV (Worthington) in RPMI (Sigma-Aldrich) supplemented with 10%(v/v) heat-inactivated fetal bovine serum (FBS; Gibco), 100U/mL penicillin (Sigma-Aldrich), 0.1mg/mL streptomycin (Sigma-Aldrich), and 2mM L-glutamine (Sigma-Aldrich) for 30 minutes at 37°C with intermittent shaking. For stringent digestion, tissue was digested with 0.2 mg/ml Liberase ™ (Roche)/0.125 KU DNase1 (Sigma-Aldrich)/10mM HEPES in RPMI for 30 minutes at 37°C with intermittent shaking. The dissociated cells were separated and remaining undigested tissue were digested again with fresh media. This procedure was repeated until the tissue was completely dissociated. Digested tissue was passed through a 100μm filter, and cells collected by centrifugation (500g for 5 minutes at 4°C). Cells were treated with 1X red blood cell (RBC lysis buffer (eBioscience) for 5 minutes at room temperature and washed once with flow buffer (PBS containing 5%(v/v) FBS and 2mM EDTA) prior to cell counting.

Fetal developmental stage assignment and chromosomal assessment

Embryos up to 8 post conception weeks (PCW) were staged using the Carnegie staging method (*67*). After 8 PCW, developmental age was estimated from measurements of foot length and heel to knee length and compared against a standard growth chart (*68*). A piece of skin, or where this was not possible, chorionic villi tissue, was collected from every sample for Quantitative Fluorescence-Polymerase Chain Reaction analysis using markers for the sex chromosomes and the following autosomes: 13, 15, 16, 18, 21, 22. All samples analysed were of normal karyotype.

Flow cytometry and FACS for Single-cell RNA Sequencing

Isolated thymus cells were stained with a panel of antibodies prior to sorting based on CD45 or CD3 expression gate. The anti-human monoclonal antibodies used for flow cytometry based immunophenotyping and FAC sorting are listed in Table S6. An antibody cocktail was freshly prepared by adding 3μL of each antibody in 50μL Brilliant Stain Buffer (BD) per tissue. Cells ($\leq 10x10^6$) were resuspended in 50-100μL flow buffer and an equal volume of antibody mix was added to cells from each tissue. Cells were stained for 30 minutes on ice, washed with flow buffer and resuspended at $10x10^6$ cells/mL. Immediately prior to sorting, DAPI (Sigma-Aldrich) was added to a final concentration of 3μM and cells strained through a 35µm filter. Flow sorting was performed on a BD FACSAria™ Fusion instrument using DIVA V8, and data analysed using FlowJo V10.4.1. Cells were gated to remove dead cells and doublets, and then sorted for 10X or SS2 scRNAseq analysis. For 10X droplet microfluidic analysis, cells were sorted into chilled FACS tubes coated with FBS and prefilled with 500µL sterile PBS. Paediatric samples were sorted into 50% FCS and 50% IMDM medium (supplemented with 1% L-glutamine, 1% Penicillin/Streptomycin and 10% FCS). For SS2 scRNAseq analysis, single cells were index-sorted into 96-well lo-bind plates (Eppendorf) containing 10µL lysis buffer (TCL 858 (Qiagen) + 1% (v/v) 2-mercaptoethanol) per well.

MACS for Single-cell RNA Sequencing

Enrichment of EPCAM positive cells were performed using CD326 (EPCAM) microbeads (Miltenyi Biotec., 130-061-101) according to manufacturer's protocol. CD45 depleted cells were obtained using CD45 microbeads (Miltenyi Biotec., 130-045-801) according to manufacturer's protocol. Cell number and viability were checked after the enrichment to ensure that no significant cell death has been caused by the process.

Coverage of cells per sample

From each sample, we obtained 1,000-20,000 cells which varies due to the size of the tissue/sample obtained. If roughly estimated by comparing this number to the total number of cells obtained after dissociation (**Data S1**), we have profiled 1 out of 10 cells for 7-8 wks fetus, 1 out of 100 cells for 9-11 wks fetus, 1 out of 5,000 cells for 12-13 wks fetus, 1 out of 10,000 cells for 16-17 wks fetus, 1 out of 500,000 cells for paediatric thymus and 1 out of 10,000 cells from adult thymus. The difference in sampling depth is caused by the rapid increase in thymic size throughout development, and decrease in cellularity in the aging process. The CD45+ population accounts for 90% of cells in thymus, most of them being thymocytes. To increase the coverage of CD45- stromal cells, we sampled the same number of cells from both

CD45+/CD45- sorting gate from fetal and adult samples, which increased the coverage of stromal cells by ~10 fold. We also specifically enriched for EPCAM+ epithelial cells from one fetal, paediatric and adult samples, to ensure higher coverage of epithelial cells. Thus, our sampling strategy was most extensive on CD45- stromal cells from fetus and adult, thymocytes and epithelial cells.

Single molecule RNA FISH

Samples were fixed in 10% NBF, dehydrated through an ethanol series and embedded in paraffin wax. Five-micrometre samples were cut, baked at 60 °C for 1 h and processed using standard pre-treatment conditions, as per the RNAscope multiplex fluorescent reagent kit version 2 assay protocol (manual) or the RNAscope 2.5 LS fluorescent multiplex assay (automated). The RNAscope probes used for this study are listed in Table S7. TSA-plus fluorescein, Cy3 and Cy5 fluorophores were used at 1:1500 dilution for the manual assay, or 1:300 dilution for the automated assay. Slides were imaged on different microscopes: Hamamatsu Nanozoomer S60 or 3DHistech Pannoramic MIDI. Filter details were as follows: DAPI: excitation 370–400, BS 394, emission 460–500; FITC: excitation 450–488, BS 490, emission 500–550; Cy3: excitation 540–570, BS 573, emission 540–570; Cy5: excitation 615–648, BS 691, emission 662–756. Stained sections were also imaged with a Perkin Elmer Opera® Phenix™ High-Content Screening System, in confocal mode with 1 μm z-step size, using 20× (NA 0.16, 0.299 μm/pixel) and 40× (NA 1.1, 0.149 μm/pixel) water-immersion objectives

Thymic fibroblasts culture derivation and phenotypic characterisation

Thymic explants were derived from foetal biopsies at different thymic stages (HDBR Newcastle University - Newcastle Upon Tyne, REC reference: 19/NE/0290 and HDBR University College of London - London, REC reference: 18/LO/0822) and cultured on a precoated Matrigel (Corning) 6mm dish in DMEM (Life Technologies) supplemented with 15% heat-inactivated FBS (Life Technologies) + 1% Penicillin/Streptomycin (Sigma-Aldrich), 1% L-glutamine (Life Technologies), 1% Non-Essential

Aminoacids (Life Technologies) and 100mM beta-Mercaptoethanol (Life Technologies). Fibroblast cells come out of explants at around 7 days of culture and are left on the plate until outgrowths are confluent enough to pass. The culture is therefore kept for 5-6 passages and phenotypic analysis was performed at multiple passages. Fibroblasts were detached with trypsin 1X (Sigma-Aldrich) for 3 minutes at 37°C and subsequently resuspended in completed media before collection. Cells are harvested and phenotypic analysis is performed on 500,000 cells per sample. Cells were stained at 4°C for 30 min in Hanks Balanced Salt Solution-2% FBS with the following markers: anti-THY1 AF700 1:100 (Biolegend), anti-PDGFRalpha PE 1:100 (Biolegend) and PI-16 (BD) 1:50. Cells are washed in an excess of HBSS + 2% FBS and are resuspended in HBSS + 2% FBS with DAPI (Sigma-Aldrich) to discriminate live from dead cells.

Library Preparation and Sequencing

For the droplet-encapsulation scRNA-seq experiments, 8000 live, single, CD45^{+} or CD45^{-} FACS-isolated cells or MACS-enriched cells were loaded on to each of the Chromium Controller (10x Genomics). Single cell cDNA synthesis, amplification and sequencing libraries were generated using the Single Cell 3' and 5' Reagent Kit following the manufacturer's instructions. The libraries from up to eight loaded channels were multiplexed together and sequenced on an Illumina HiSeq 4000. The libraries were distributed over eight lanes per flow cell and sequenced using the following parameters: Read1: 26 cycles, i7: 8 cycles, i5: 0 cycles; Read2: 98 cycles to generate 75bp paired end reads.

For the plate-based scRNA-seq experiments, a slightly modified Smart-Seq2 protocol was used as previously described (56). After cDNA generation, libraries were prepared (384 cells per library) using the Illumina Nextera XT kit. Index v2 sets A, B, C and D were used per library to barcode each cell for multiplexing. Each library was sequenced (384 cells) per lane at a sequencing depth of 1-2 million reads per cell on HiSeq 4000 using v4 SBS chemistry to create 75bp paired end reads.

Alignment, quantification and quality control of single cell RNA sequencing data

Droplet-based sequencing data was aligned and quantified using the Cell Ranger Single-Cell Software Suite (version 2.0.2 for 3' chemistry and version 2.1.0 for 5' chemistry, 10x Genomics Inc) using the GRCh38 human reference genome (official Cell Ranger reference, version 1.2.0). Cells with fewer than 2000 UMI counts and 500 detected genes were considered as empty droplets and removed from the dataset. Cells with more than 7000 detected genes were considered as potential doublets and and removed from the dataset. Smart-seq2 sequencing data was aligned with *STAR* (version 2.5.1b), using the STAR index and annotation from the same reference as the 10x data. Gene-specific read counts were calculated using *htseq-count* (version 0.10.0). Scanpy (version 1.3.4) python package was used to load the cell-gene count matrix and perform downstream analysis.

Doublet detection

To exclude doublets from single-cell RNA sequencing data, we applied scrublet (https://github.com/AllonKleinLab/scrublet, (*69*)) algorithm per sample to calculate scrublet-predicted doublet score per cell with following parameters: sim_doublet_ratio = 2; n_neighbors=30; expected_doublet_rate= 0.1. Any cell with scrublet score > 0.7 was flagged as doublet. To propagate the doublet detection into potential false-negatives from scrublet analysis, we over-clustered the dataset (*sc.tl.louvain* function from scanpy package version 1.3.4; resolution = 20), and calculated the average doublet score within each cluster. Any cluster with averaged scrublet score > 0.6 was flagged as a doublet cluster. All remaining cell clusters were further examined to detect potential false-negatives from scrublet analysis according to the following criteria: (1) Expression of marker genes from two distinct cell types which are unlikely according to prior knowledge (i.e. *CD3* for T cells and *CD19* for B cells), (2) Higher number of UMI counts and (3) Lack of unique marker gene defining the cluster. All clusters flagged as doublets were removed from further downstream biological analysis.

Defining contaminating populations from other tissues

We noticed that embryonic thymus can be contaminated with thyroid or parathyroid derived tissue, which is annotated as Epi_PAX8 (marked by PAX8, HHEX, TG, NKX2.1) and Epi_GCM2 (marked by PTH, GCM2, GATA3, CHGA). We removed cell clusters defined by these markers and removed entire dataset if it has larger cell cluster belonging to these contaminating populations compared to thymic epithelial cells.

Clustering and annotation of scRNA-seq data

Downstream analysis included data normalisation (*scanpy.api.pp.normalize_per_cell* method, scaling factor 10000), log-transformation (*scanpy.api.pp.log1p),* variable gene detection (*scanpy.api.pp.filter_gene_dispersion*), data feature scaling (*scanpy.api.pp.scale*), PCA analysis (*scanpy.api.pp.pca*, from variable genes), batch-balanced neighbourhood graph building (*scanpy.api.pp.bbknn*) and Louvain graph-based clustering (*scanpy.api.tl.louvain*, clustering resolution manually tuned) performed using the python package scanpy (version 1.3.4). Custom defined cell cycle gene sets (**Table S8**) were removed from the list of variable genes to remove cell-cycle associated variation. Cluster cell identity was assigned by manual annotation using known marker genes as well as computed differentially expressed genes (DEGs) using custom python function. Clusters with clear and uniform identity were annotated first, and a logistic regression model was trained based on this annotation. This model was used to predict the identity of cells in a cluster with a mixture of different cell types, which can be computationally clustered together due to transcriptional similarity. To achieve a high-resolution annotation, we separated broadly annotated cells (e.g. Epithelial cells, single positive T cells) and repeated the procedure of variable gene selection, which allowed the annotation of smaller and fine-grained cell subsets (e.g. mTECs, regulatory T cells).

Alignment of data across different batches

Batches for batch alignment can come from different chemistries used on the same set of cells, e.g. 10X chemistry (5' and 3'), or from cells from different donors analysed using the same chemistry. In other words, there can be technical or biological differences between batches. We performed iterative batch correction, first by roughly aligning batches across similar samples (e.g. all foetal samples or paediatric samples) using *scanpy.api.pp.bbknn* function. We used this batch-aligned manifold to annotate cell types. After achieving a coarse-grained cell type annotation, we fitted a L2-regularised linear model using batches (e.g. 10X chemistry, donors) or cell type annotation as a categorical variable. Then we regressed out variations explained by batch variables, and kept residuals, which contain biological information. After this, we aligned batches again using the *scanpy.api.pp.bbknn* function to achieve a high-resolution and batch-mixed manifold, which is used for refining annotation, visualisation and trajectory analysis.

Estimating cellular composition per sample

To estimate the relative proportion of each cell type in different samples, we defined broad categories of cell types (e.g. lymphocytes, myeloid cells, total cells), and calculated the proportions of each cell type within selected group of cells. If all cell types used for a comparison come from the same sorting gate, we simply calculated the proportion as: number of cells in specific cell type / total number of cells in comparison set. When cell types used for comparison are derived from multiple sorting gates, we calculated a normalisation factor for each sorting gate as: number of cells sorted in a specific sorting gate / total number of sorted cells across multiple sorting gates, and multiplied this normalisation factor to the number of cells in each sorting gate. These normalised numbers are used to calculate proportions, which eliminates bias caused by sorting different number of cells into different gates. The significance of changes in cellular proportions are tested by t-test on cell proportions.

Trajectory analysis

To model differentiation trajectories, a combination of linear regression and batch-alignment algorithms were applied as described above to generate a neighbourhood graph. The robustness and accuracy of batch-alignment was tested by comparing multiple batch-alignment methods. Among the resulting manifolds, we selected the one with the best fit to well-known sequential events in T-cell differentiation such as TCR recombination. We then calculated diffusion pseudotime (*70*) using the *scanpy.api.tl.dpt* function in scanpy, which starts from the manually selected progenitor cell. The progenitor cell is selected from the extremities of diffusion components. Cells are binned based on the pseudotime ordering, and differentially expressed genes are identified as genes whose expression is significantly different from the randomly permuted background in any of the bins.

Visualisation of the transcription factor network

Transcription factor network analysis was performed as previously described (*71*). First, gene expression levels were imputed by taking an average of 30-nearest neighbors in three-dimensional UMAP space. An annotation score for each cell type was calculated by measuring the frequency of cell types amongst the 30-nearest neighbors which are used for imputation. To remove redundant information, cells were randomly sampled from each unit voxel from the three-dimensional UMAP space. The human transcription factors were selected from AnimalTFDB3 (*72*). Only highly-variable transcription factors were subject to calculation of the correlation matrix, which was subsequently used for graph building and visualisation using the force-directed graph function implemented in the scanpy package.

TCR VDJ sequence analysis

10X TCR-enriched libraries are mapped with the Cell Ranger Single-Cell Software Suite (version 2.1.0, 10x Genomics Inc) to the custom reference provided by the manufacturer (version 2.0.0 GRCh38 VDJ reference). VDJ sequence information was extracted from the output file “filtered_contig_annotations.csv.” The merged VDJ output dataset is available in our data repository (see Data and materials availability).

Chains which contained full-length recombinant sequence and supported by more than 2 UMI counts were selected, and linked to the cellular transcriptome data based on cell barcodes. These chains were considered as productive if a functional ORF covering the CDR3 region could be found. To compare V, D, J gene usage per cell type, each V, D, J gene count in each specific cell type was normalised by the sum of counts within that cell type, and then converted to a z-score per gene. Student's t-test was used to compare the z-scores between different cell types. Cochran–Mantel–Haenszel test was also used to compare profiles between CD4+T and CD8+T cells, which yielded comparable results.

Comparison to published dataset

Human liver dataset collected from the same donors which has been described in (*44*). were processed in the same way as the thymic cells. The human bone marrow-derived hematopoietic progenitors dataset has been downloaded from Gene Expression Omnibus (GSE117498) as processed count matrix. The dataset has been processed through the same pipeline and combined with human liver and thymus dataset. Batch alignment was performed across thymus, liver and bone marrow datasets using the BBKNN algorithm assisted by linear regression. The human sorted thymocytes microarray dataset has been downloaded from ArrayExpress (E-MEXP-337) as a processed expression matrix.

Mouse thymus cell atlas

The mouse stromal dataset has been collected from Gene Expression Omnibus (GSE103967) and mouse fetal thymus dataset has been downloaded from Gene Expression Omnibus (GSE107910). Mouse postnatal thymus dataset has been generated for C57BL/6J mice (4, 8, 24 weeks old). Dissected thymi were dissociated with Liberase TH protocol and two 10X 3' v3 lanes were loaded for each sample. All data has been processed in the same way as the human thymic cells. The mouse stromal dataset has been re-annotated following the original description by authors. We noted some minor cell populations which were not defined in the original study. Mouse fetal and postnatal cells are integrated into the same dataset and annotated altogether. A logistic model trained from the annotated human data was applied to assist the annotation process to achieve coherent annotation between human and mouse.

Cross-species comparison

The alignment of the mouse dataset to the human dataset has been achieved by two methods: (1) Bi-directional prediction based on logistic models trained from each dataset. The prediction probability from human to mouse and mouse to human cell pairs are multiplied to derive the final similarity score. (2) Batch alignment using BBKNN algorithm assisted by linear regression to remove species-specific variations while keeping the biological structure. For this, an initial round of BBKNN integration has been performed across all samples to produce a graph structure with connections between nearest neighbors across batches. Low-resolution graph-based clustering was performed on this to obtain a clustering structure based on biological variation. Then L2-regularised linear regression was performed using this cluster structure as biological variables and species/sample structure as batch variable. The variation explained by batch structure was regressed out from the data, and this corrected matrix is used for the second round of BBKNN integration. This resulted in a manifold that is well-mixed across species. Of note, we confined this approach to subsets of cells (e.g. mature T cells), to achieve better alignment by reducing complexity.

Cell-cell interaction analysis

Specific interactions between cells are modeled using CellPhoneDB (www.CellPhoneDB.org) as previously described (*35*). To minimise computational burden and equally represent different cell types, we downsampled the dataset by randomly sampling 1000 cells from each cell type. We modified cell-cell interaction scores by multiplying average expression level of each ligand and receptor gene within cell-cell pairs, and maximum-normalising this score. The list of chemokines was retrieved from the HUGO Gene Nomenclature Committee. To visualise the interactions, we first selected interaction pairs based on significance of specificity from CellPhoneDB and calculated normalised interaction score for each cell pairs.

This normalised interaction score has been calculated by multiplying the average expression level of ligand and receptors for all cell pairs, and maximum normalising these values.

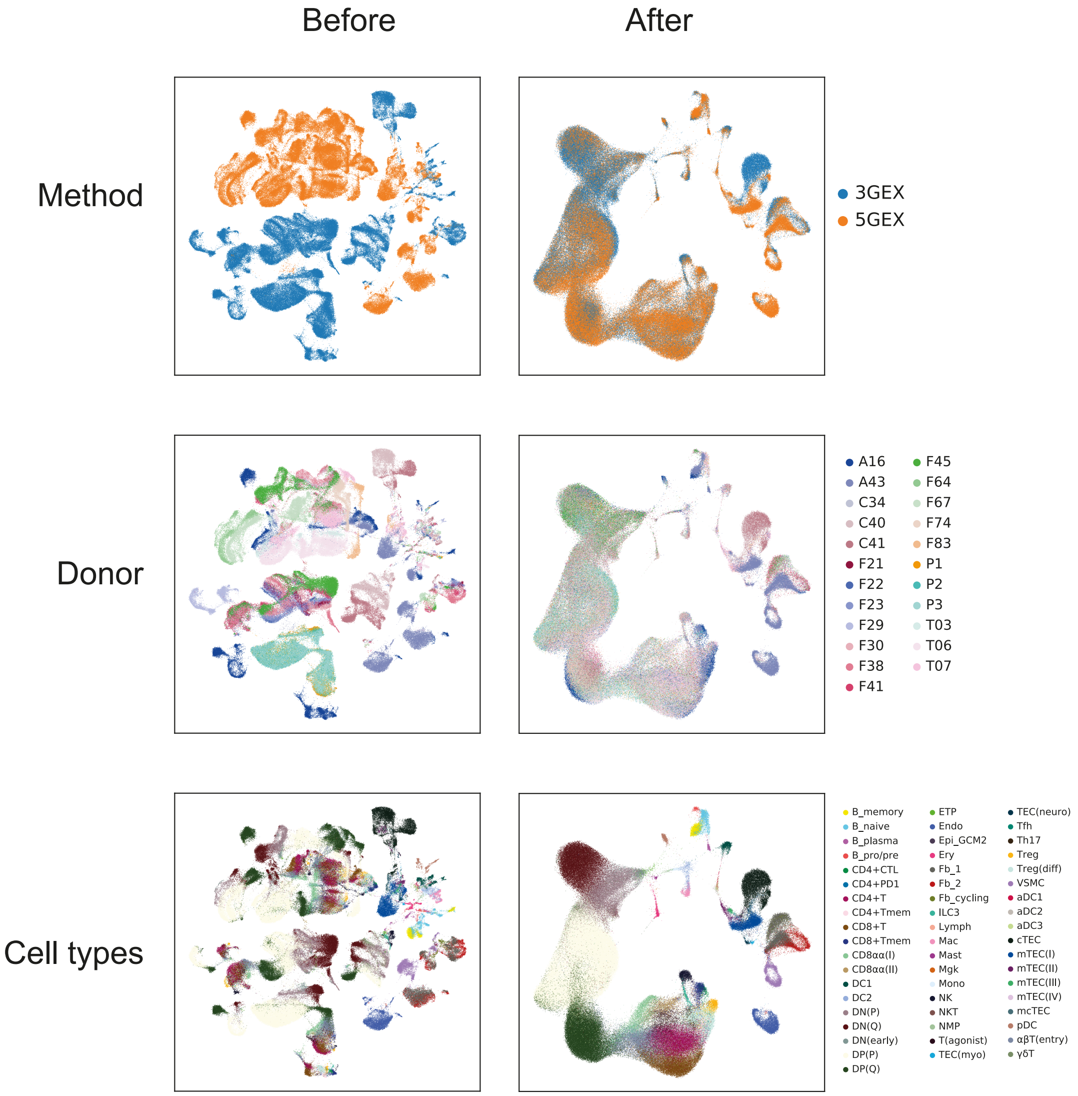

Fig. S1.

UMAP visualisation of the entire dataset before (left) and after (right) batch alignment. Cells are coloured by methods (top), donors (middle) and cell types (bottom).

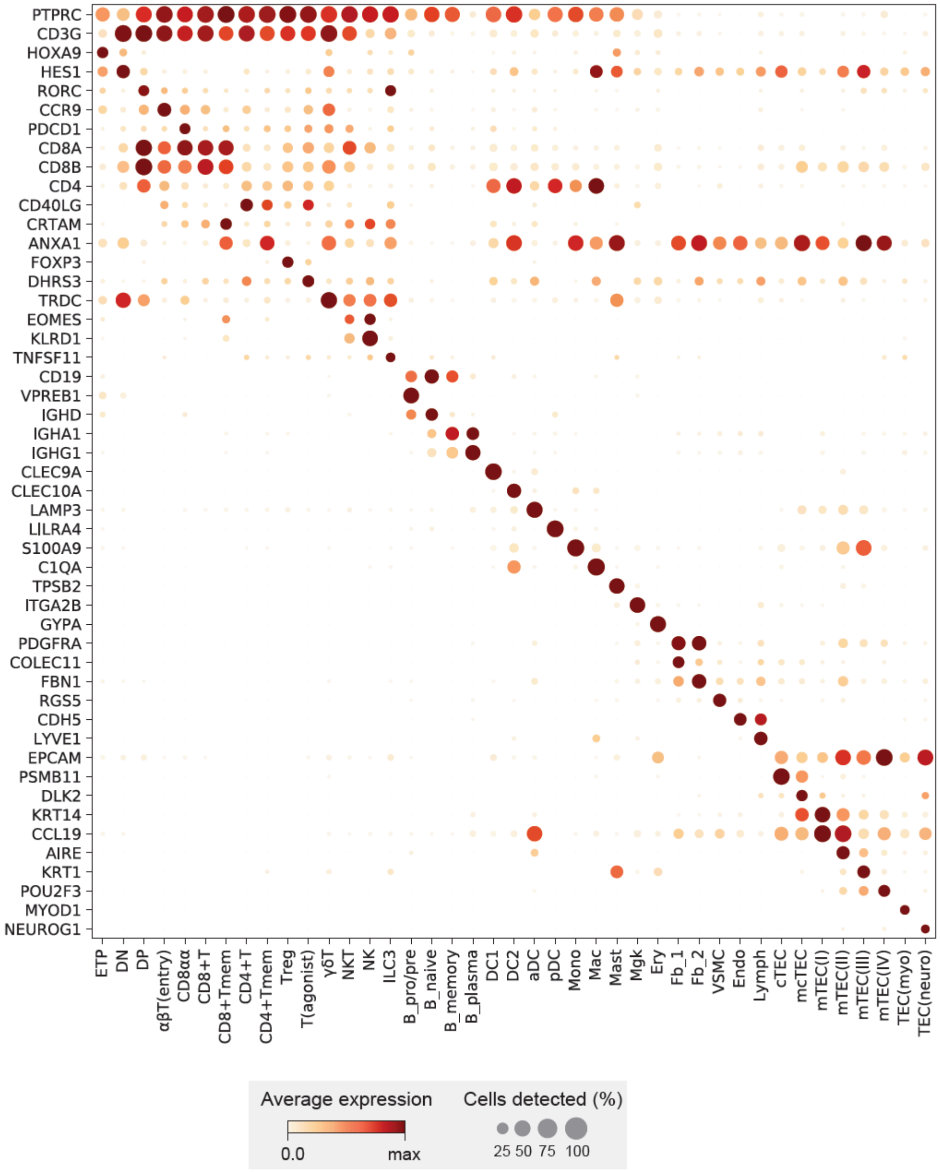

Fig. S2.

Dot plot showing marker gene expression for annotated cell types. Color represents maximum-normalised mean expression of marker genes in each cell group, and size indicates the proportion of cells expressing marker gene. ETP: early thymic progenitors, DN: double negative T cells, DP: double positive T cells, Treg: regulatory T cells, ILC3: innate lymphoid cell type 3, B_pro/pre: pro-B cells and pre-B cells, DC1: conventional dendritic cell type 1, DC2: conventional dendritic cell type 2, aDC: activated dendritic cells, pDC: plasmacytoid dendritic cells, Mono: monocytes, Mac: macrophage, Mast: mast cells, Mgk: megakaryocytes, Ery: erythrocytes, Endo: endothelial cells, VSMC: vesicular smooth muscle cells, Fb_1, Fb_2: fibroblasts type 1 and 2

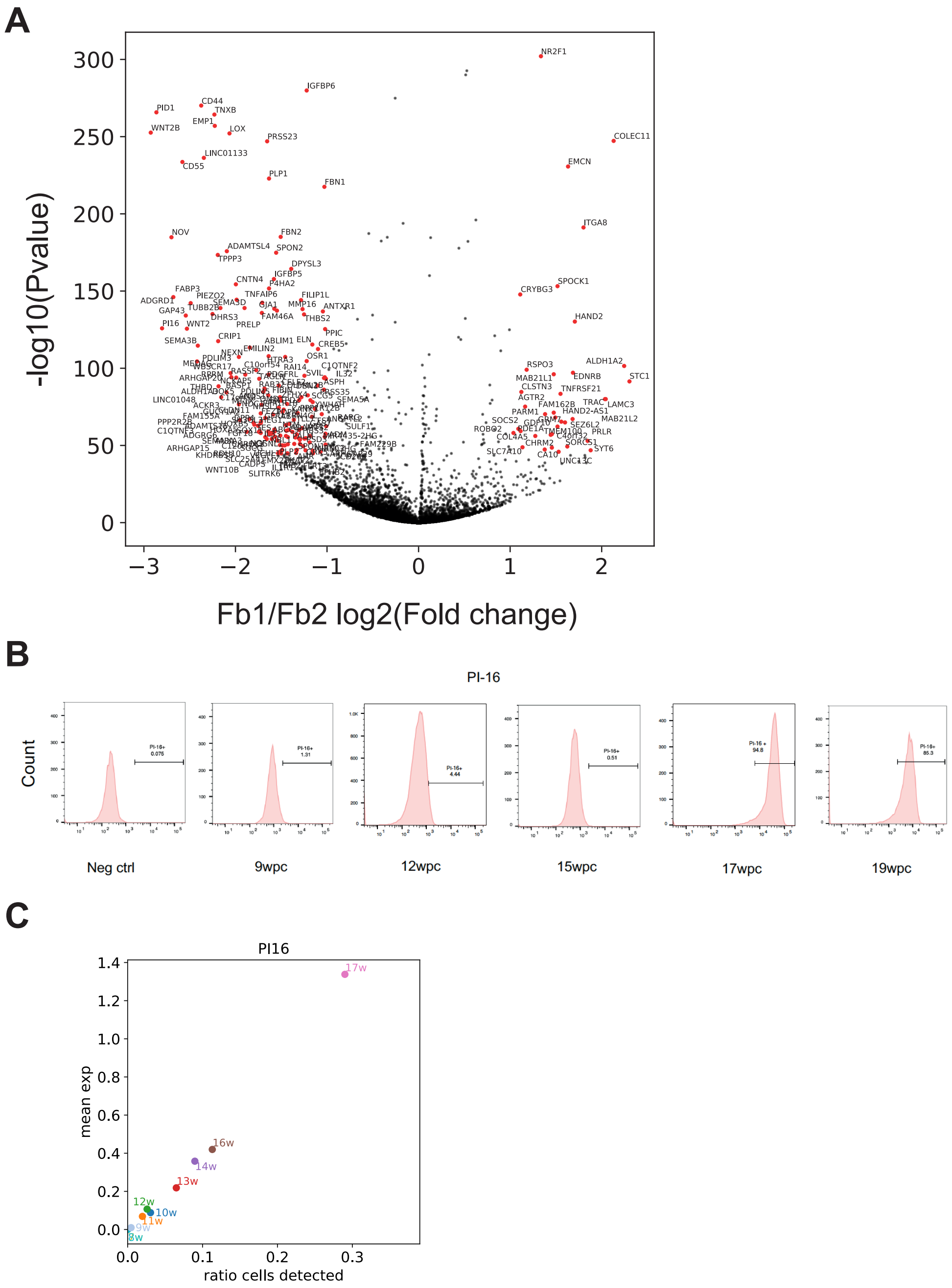

Fig. S3.

(A) Volcano plot showing differentially expressed genes between fb1 and fb2 type of thymic fibroblasts. X-axis and y-axis represent log2(fold change) and -log10(p-value) respectively. **(B)** FACS analysis of PI16 protein level in thymic fibroblast explant culture from different stages of human fetal thymus. **(C)** Expression level of PI16 mRNA level in single-cell RNA sequencing data from different stages of human fetal thymus

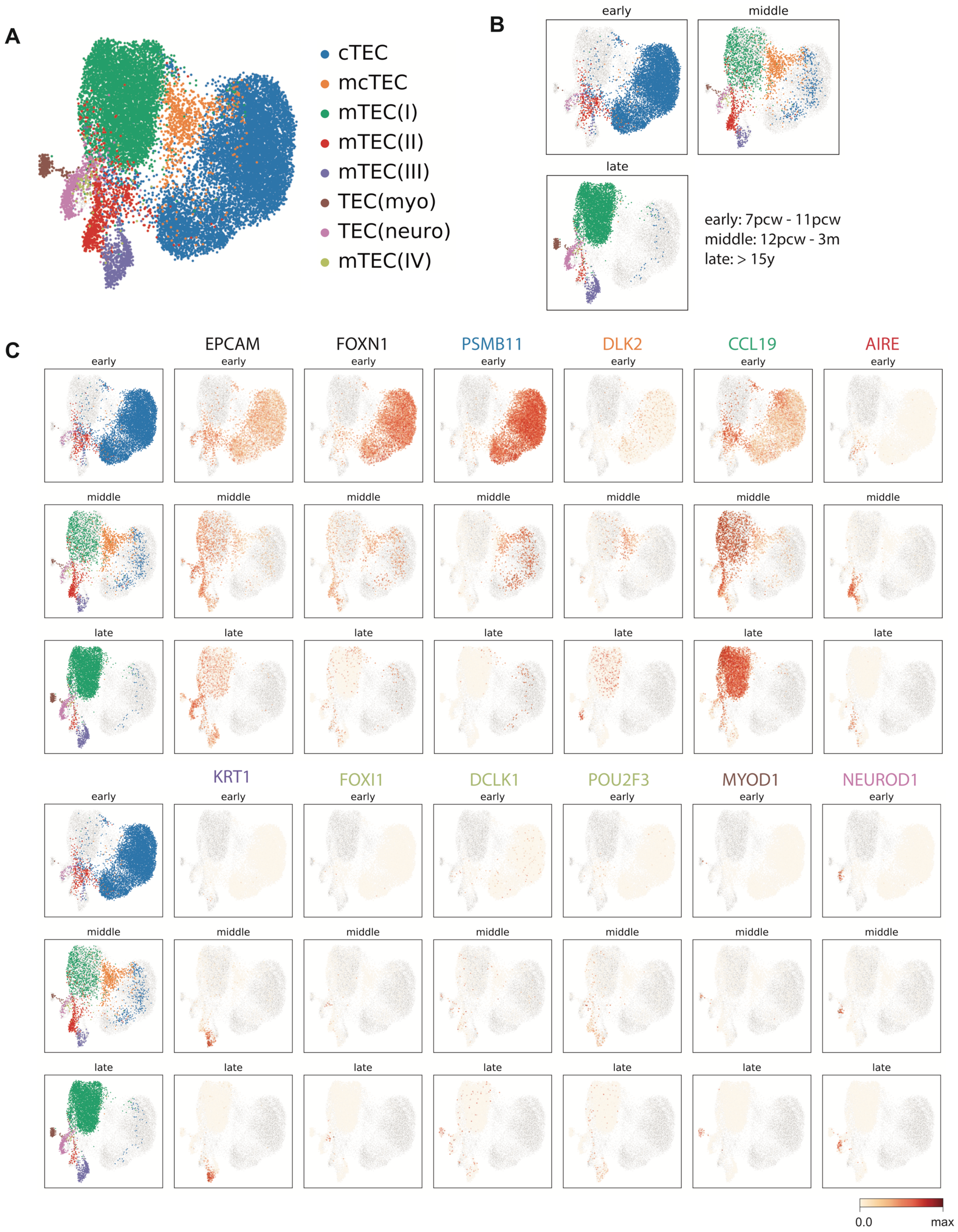
A
● cTEC
● mcTEC
● mTEC(I)
● mTEC(II)
● mTEC(III)
● TEC(myo)
● TEC(neuro)
● mTEC(IV)
B
early
middle
late
early: 7pcw - 11pcw
middle: 12pcw - 3m
late: > 15y
C
EPCAM
FOXN1
PSMB11
DLK2
CCL19
AIRE
KRT1
FOXI1
DCLK1
POU2F3
MYOD1
NEUROD1
early
middle
late
0.0
max

Fig. S4.

(A) UMAP plot showing human thymic epithelial cell (TEC) populations **(B)** UMAP plot showing the human TECs stratified into three different stages (early: 7 pcw -11 pcw fetal, middle: 12 wks pcw – 3 months postnatal, late: more than 15 years old). The grey shaded data points correspond to the other two stages. **(C)** UMAP plot showing the expression pattern of marker genes for each epithelial cell type (same colour scheme as applied in (A)) stratified into early, middle, late stages. The shaded data points correspond to the other two stages.

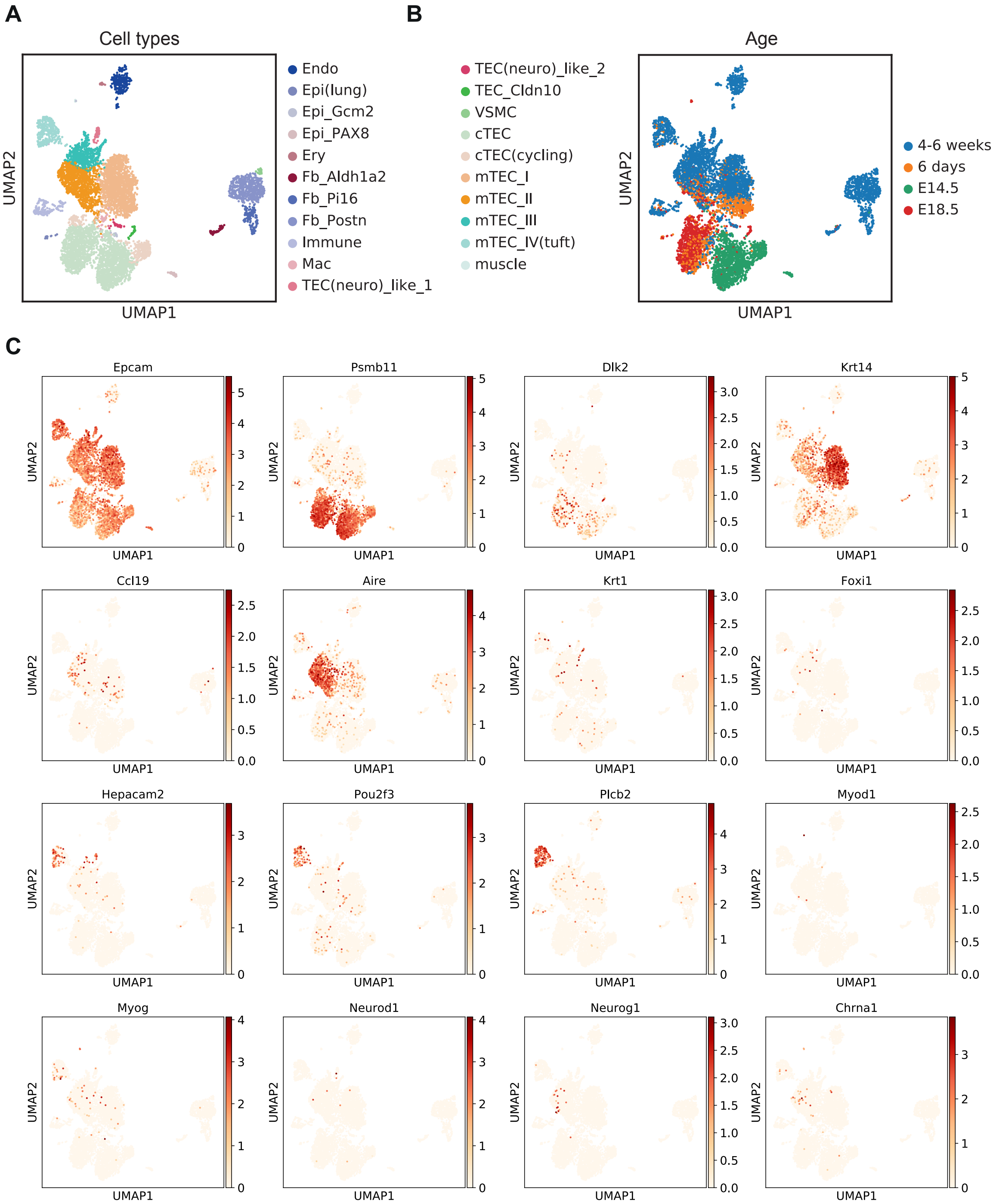

Fig. S5.
(A) UMAP plot showing mouse thymic stromal cell populations (*15*) stratified by cell type and **(B)** by age. **(C)** UMAP plots showing the expression pattern of marker genes for each epithelial cell type in mouse thymic stromal cell population.

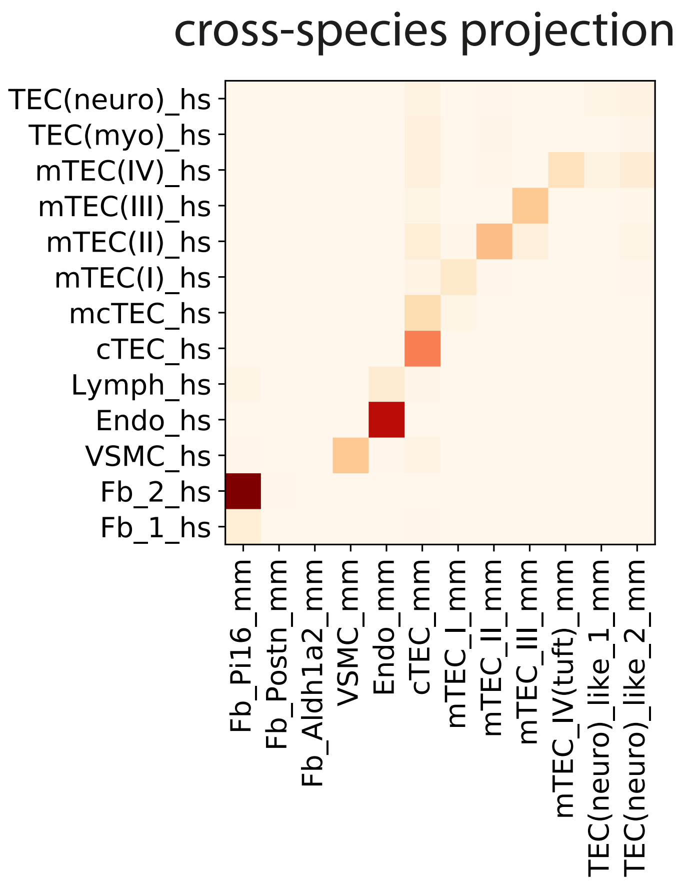

Fig. S6.

Projection score between human and mouse thymic epithelial cell types calculated by multiplying the cross-species predicted projection probability of logistic models trained on each species data.

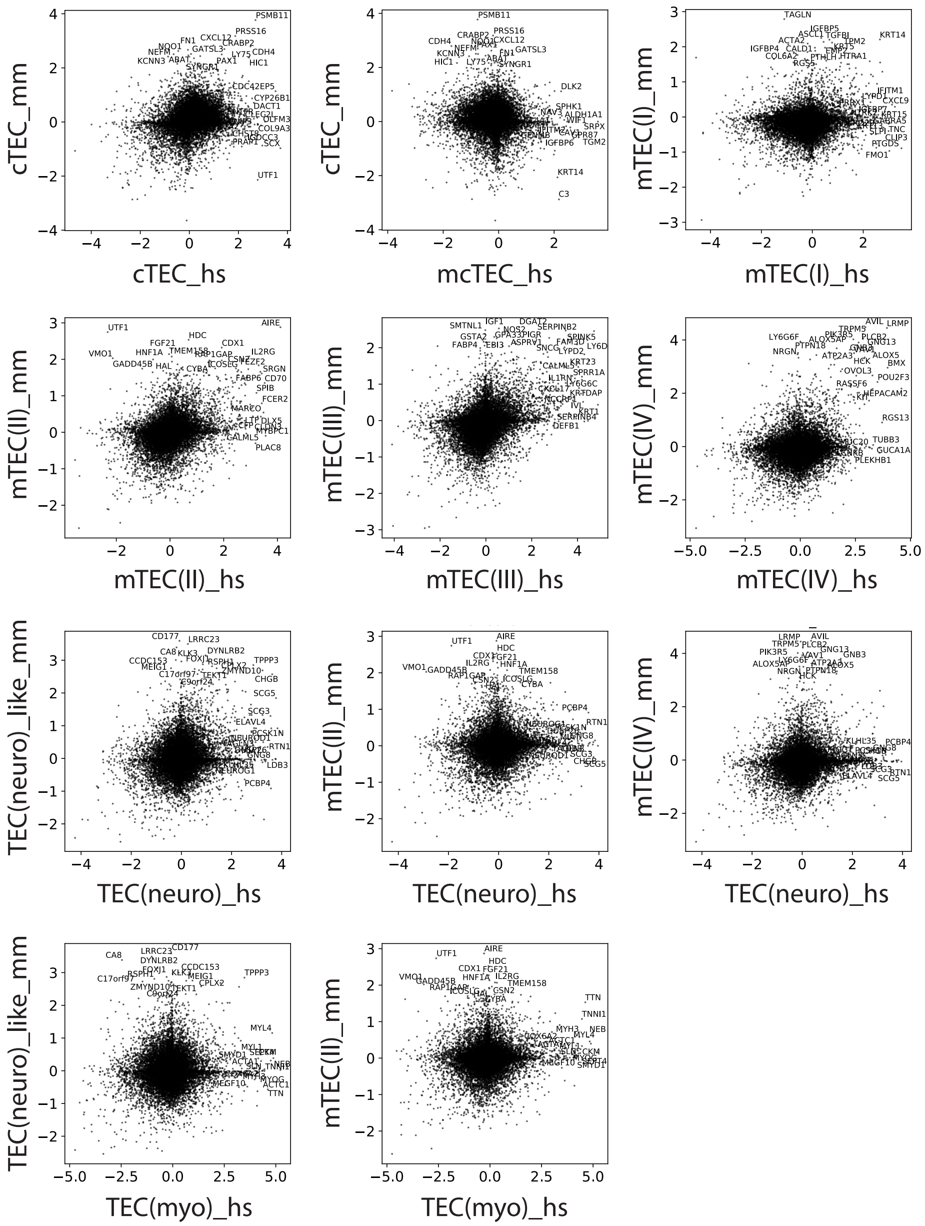

Fig. S7.

Comparison of markers for thymic epithelial cell types from human and mouse. X-axis is gene expression log2-fold change for the designated human cell types against all other human thymic epithelial cells. Y-axis is gene expression log2-fold change for the designated mouse cell types against all mouse thymic epithelial cells. Comparison sets are determined based on the cross-species projection score (fig. S6)

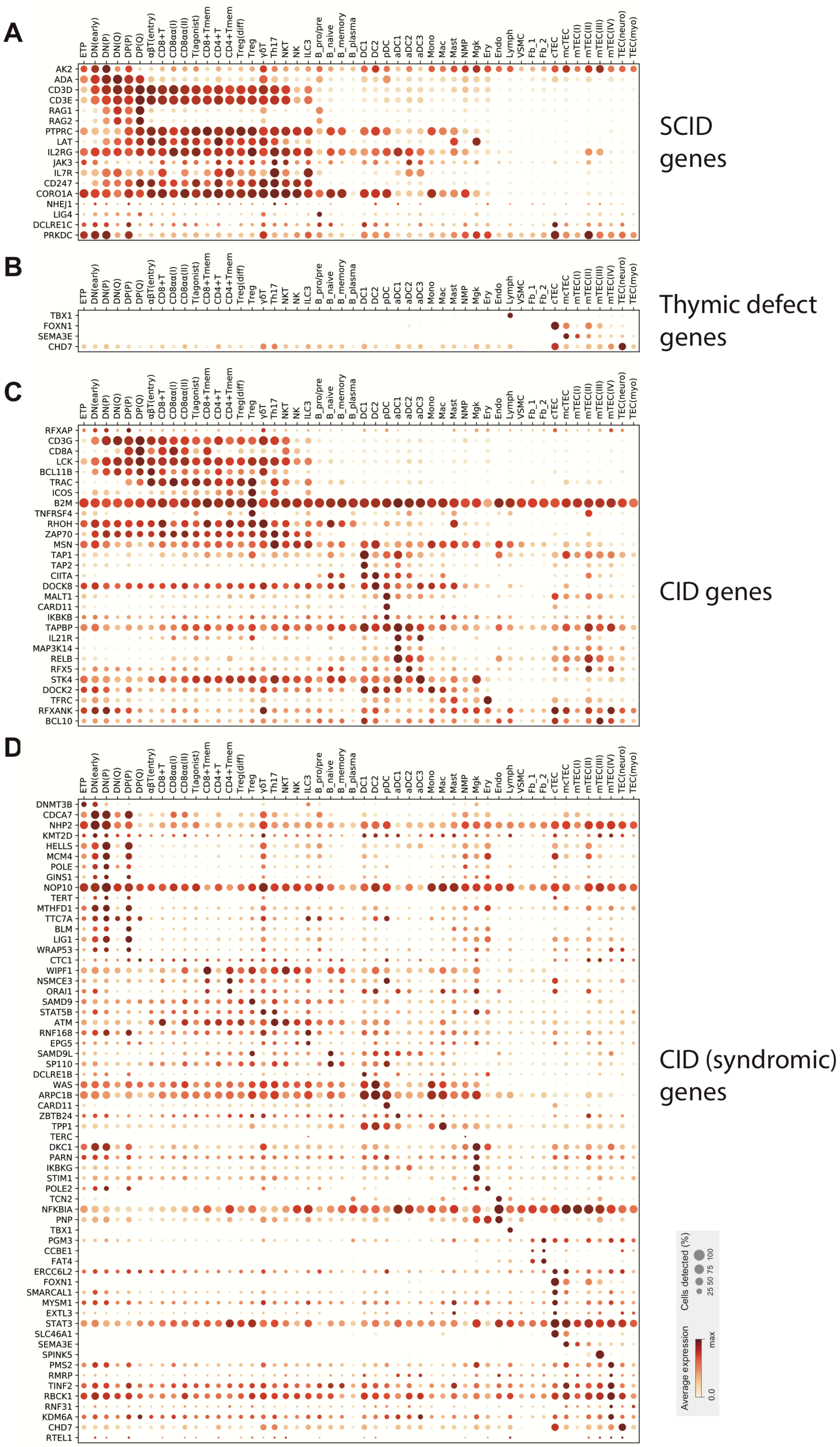
A
ETP
DN(early)
DN(P)
DN(Q)
DP(P)
DP(Q)
αβT(entry)
CD8+T
CD8αα(I)
CD8αα(II)
T(agonist)
CD8+Tmem
CD4+T
CD4+Tmem
Treg(diff)
Treg
γδT
Th17
NKT
NK
ILC3
B_pro/pre
B_naive
B_memory
B_plasma
DC1
DC2
pDC
aDC1
aDC2
aDC3
Mono
Mac
Mast
NMP
Mgk
Ery
Endo
Lymph
VSMC
Fb_1
Fb_2
cTEC
mcTEC
mTEC(I)
mTEC(II)
mTEC(III)
mTEC(IV)
TEC(neuro)
TEC(myo)
AK2
ADA
CD3D
CD3E
RAG1
RAG2
PTPRC
LAT
IL2RG
JAK3
IL7R
CD247
CORO1A
NHEJ1
LIG4
DCLRE1C
PRKDC
SCID genes
B
ETP
DN(early)
DN(P)
DN(Q)
DP(P)
DP(Q)
αβT(entry)
CD8+T
CD8αα(I)
CD8αα(II)
T(agonist)
CD8+Tmem
CD4+T
CD4+Tmem
Treg(diff)
Treg
γδT
Th17
NKT
NK
ILC3
B_pro/pre
B_naive
B_memory
B_plasma
DC1
DC2
pDC
aDC1
aDC2
aDC3
Mono
Mac
Mast
NMP
Mgk
Ery
Endo
Lymph
VSMC
Fb_1
Fb_2
cTEC
mcTEC
mTEC(I)
mTEC(II)
mTEC(III)
mTEC(IV)
TEC(neuro)
TEC(myo)
TBX1
FOXN1
SEMA3E
CHD7
Thymic defect genes
C
ETP
DN(early)
DN(P)
DN(Q)
DP(P)
DP(Q)
αβT(entry)
CD8+T
CD8αα(I)
CD8αα(II)
T(agonist)
CD8+Tmem
CD4+T
CD4+Tmem
Treg(diff)
Treg
γδT
Th17
NKT
NK
ILC3
B_pro/pre
B_naive
B_memory
B_plasma
DC1
DC2
pDC
aDC1
aDC2
aDC3
Mono
Mac
Mast
NMP
Mgk
Ery
Endo
Lymph
VSMC
Fb_1
Fb_2
cTEC
mcTEC
mTEC(I)
mTEC(II)
mTEC(III)
mTEC(IV)
TEC(neuro)
TEC(myo)
RFXAP
CD3G
CD8A
LCK
BCL11B
TRAC
ICOS
B2M
TNFRSF4
RHOH
ZAP70
MSN
TAP1
TAP2
CIITA
DOCK8
MALT1
CARD11
IKBKB
TAPBP
IL21R
MAP3K14
RELB
RFX5
STK4
DOCK2
TFRC
RFXANK
BCL10
CID genes
D
ETP
DN(early)
DN(P)
DN(Q)
DP(P)
DP(Q)
αβT(entry)
CD8+T
CD8αα(I)
CD8αα(II)
T(agonist)
CD8+Tmem
CD4+T
CD4+Tmem
Treg(diff)
Treg
γδT
Th17
NKT
NK
ILC3
B_pro/pre
B_naive
B_memory
B_plasma
DC1
DC2
pDC
aDC1
aDC2
aDC3
Mono
Mac
Mast
NMP
Mgk
Ery
Endo
Lymph
VSMC
Fb_1
Fb_2
cTEC
mcTEC
mTEC(I)
mTEC(II)
mTEC(III)
mTEC(IV)
TEC(neuro)
TEC(myo)
DNMT3B
CDCA7
NHP2
KMT2D
HELLS
MCM4
POLE
GINS1
NOP10
TERT
MTHFD1
TTC7A
BLM
LIG1
WRAP53
CTC1
WIPF1
NSMCE3
ORAI1
SAMD9
STAT5B
ATM
RNF168
EPG5
SAMD9L
SP110
DCLRE1B
WAS
ARPC1B
CARD11
ZBTB24
TPP1
TERC
DKC1
PARN
IKBKG
STIM1
POLE2
TCN2
NFKBIA
PNP
TBX1
PGM3
CCBE1
FAT4
ERCC6L2
FOXN1
SMARCAL1
MYSM1
EXTL3
STAT3
SLC46A1
SEMA3E
SPINK5
PMS2
RMRP
TINF2
RBCK1
RNF31
KDM6A
CHD7
RTEL1
CID (syndromic) genes
Cells detected (%)
25 50 75 100
Average expression
0.0
max

Fig. S8.

Dot plot showing the expression of genes causing Severe Combined Immunodeficiency (SCID) (A), thymic defects (B), Combined Immunodeficiency (CID) (C), and syndromic CID (D). Genes are taken from the IUIS Classification of Inborn Errors of immunity (February 2018). Color represents maximum-normalised mean expression of marker genes in each cell group, and size indicates the proportion of cells expressing marker gene.

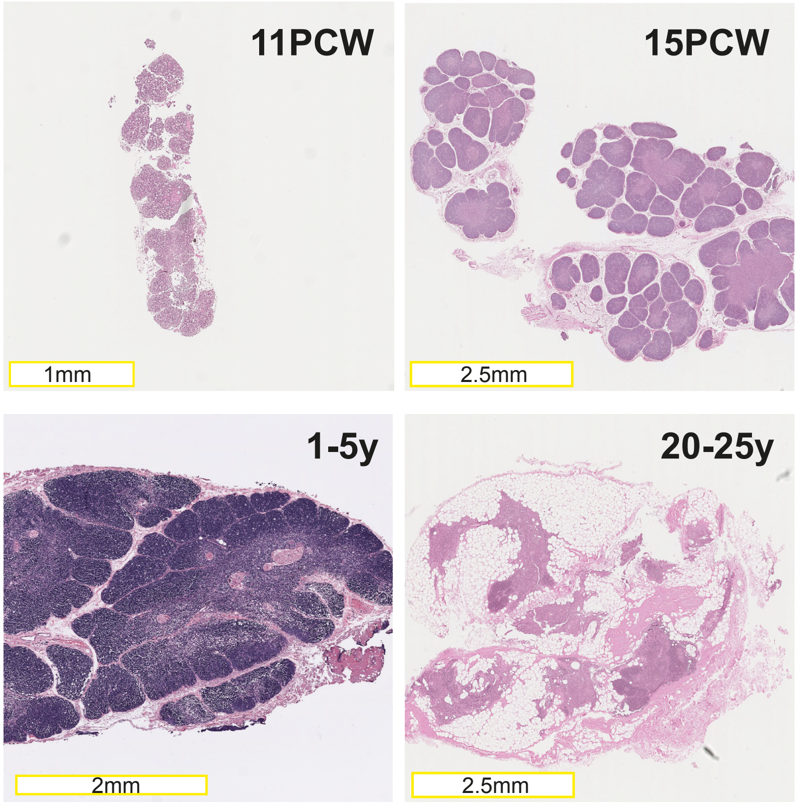

Fig. S9.
H&E staining of cross-sectioned thymic tissue at different developmental and postnatal ages.

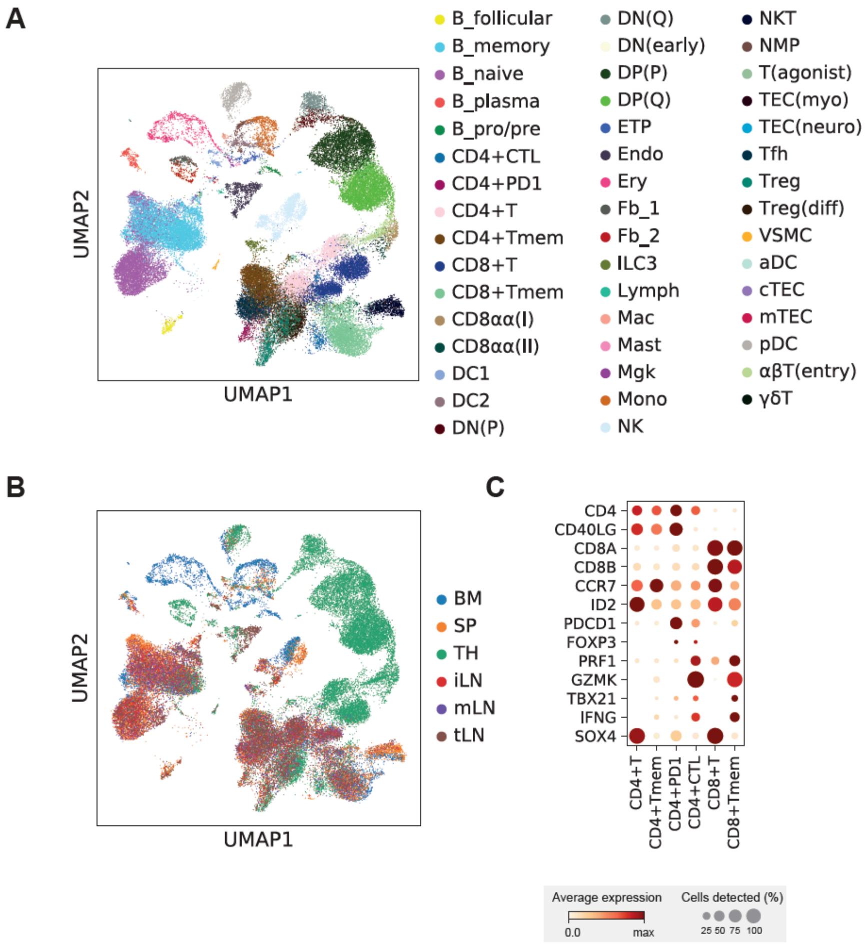

Fig. S10.

(A) UMAP plot showing cell type annotations for the young adult sample (20-25 years old). **(B)** Organ composition for UMAP plot shown in (A). **(C)** Dot plot showing marker gene expression for mature T cells found in young adult sample. Abbreviations are as defined from Fig. S2. BM: bone marrow; SP: spleen; TH: thymus; iLN: inguinal lymph node; mLN: mesenteric lymph node; tLN: thoracic lymph node.

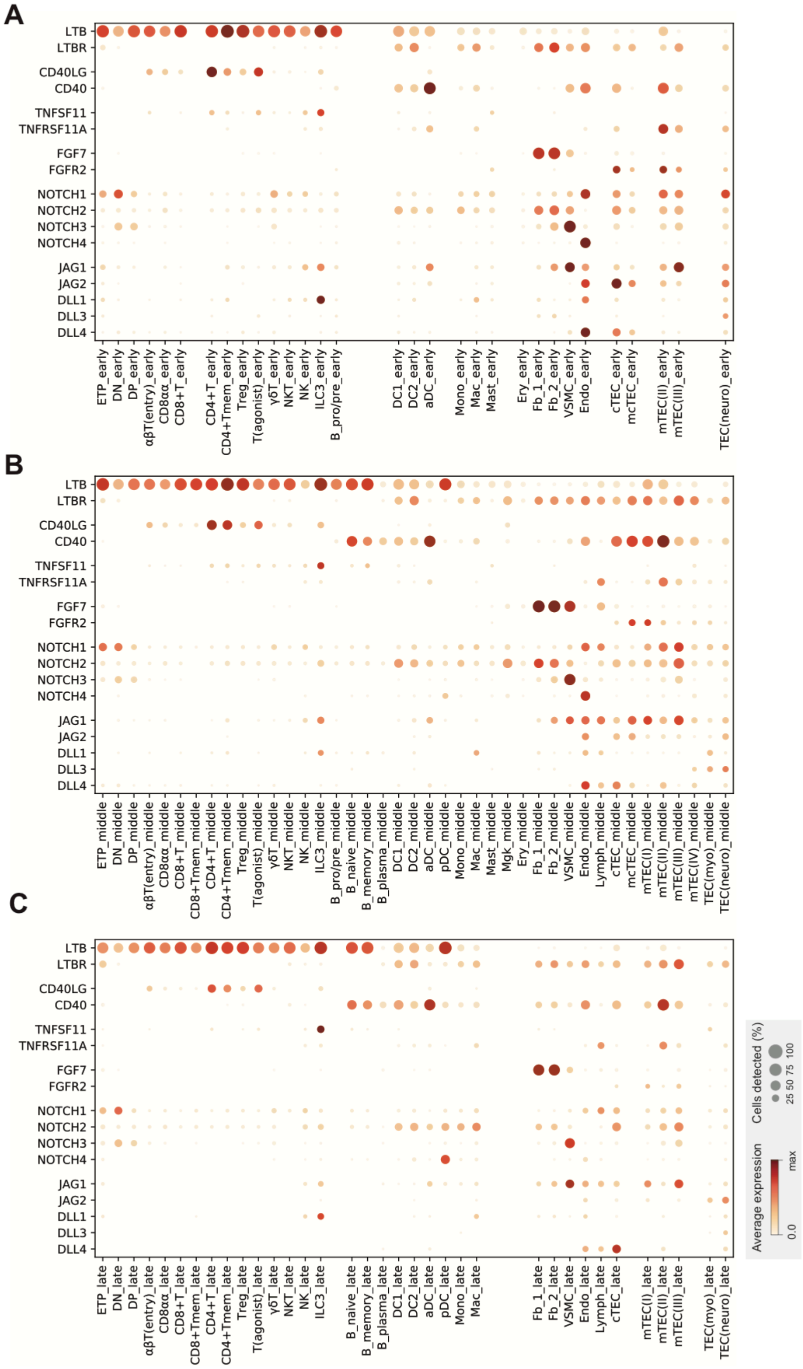

Fig. S11.
Dot plot showing cell type specific expression of signalling pathways which are known to regulate thymic development. Dataset is separated according to developmental stages as indicated in Fig. S4B. Cell types with less than 50 cells detected per stages are omitted from the plot.

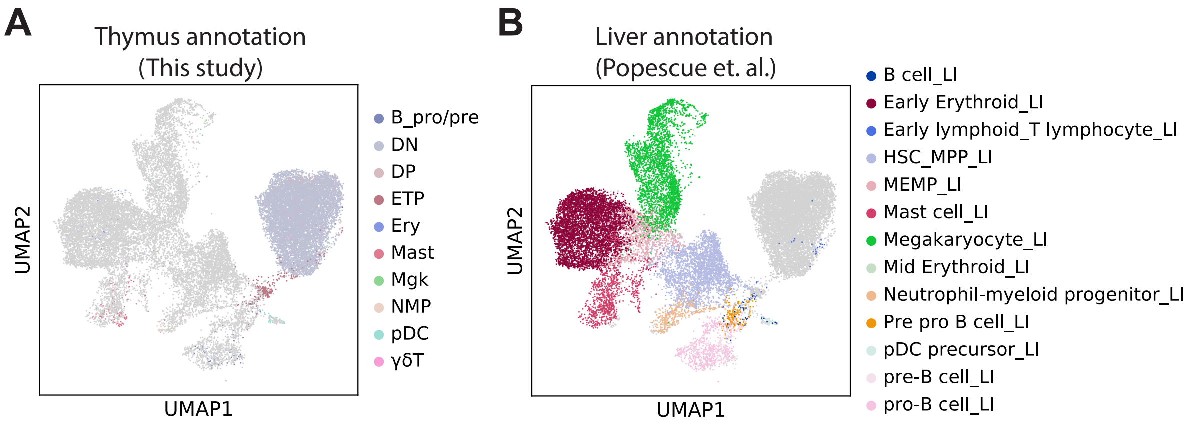

Fig. S12.
UMAP plot displayed in Figs. 2A and B coloured according to the original annotation on thymus cells from this study (left panel) and liver cells sampled from same donors (*44*). (right panel)

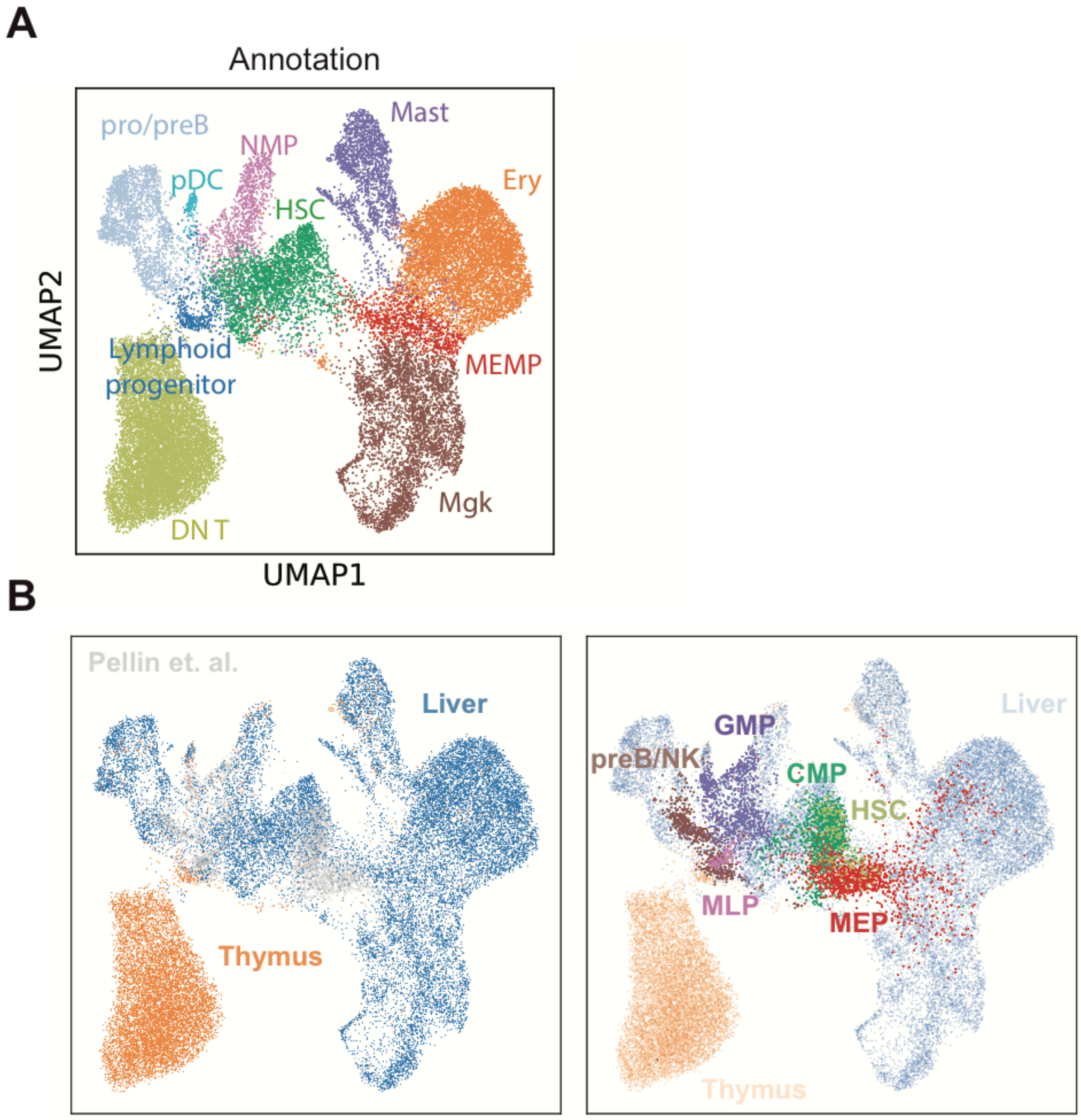

Fig. S13.

UMAP plot showing the integrative analysis between hematopoietic progenitors from thymus and liver and sorted human hematopoietic progenitors from bone marrow (*46*). Cells are labeled based on **(A)** annotation described in Fig 2A, **(B)** derived organs (left panel, shown for thymus and liver cells) or sorting scheme (right panel). MLP: multi-lymphoid progenitors. MEP: megakaryocyte-erythrocyte progenitors. HSC: hematopoietic stem cells. GMP: granulocyte-macrophage progenitors. CMP: common-myeloid progenitors.

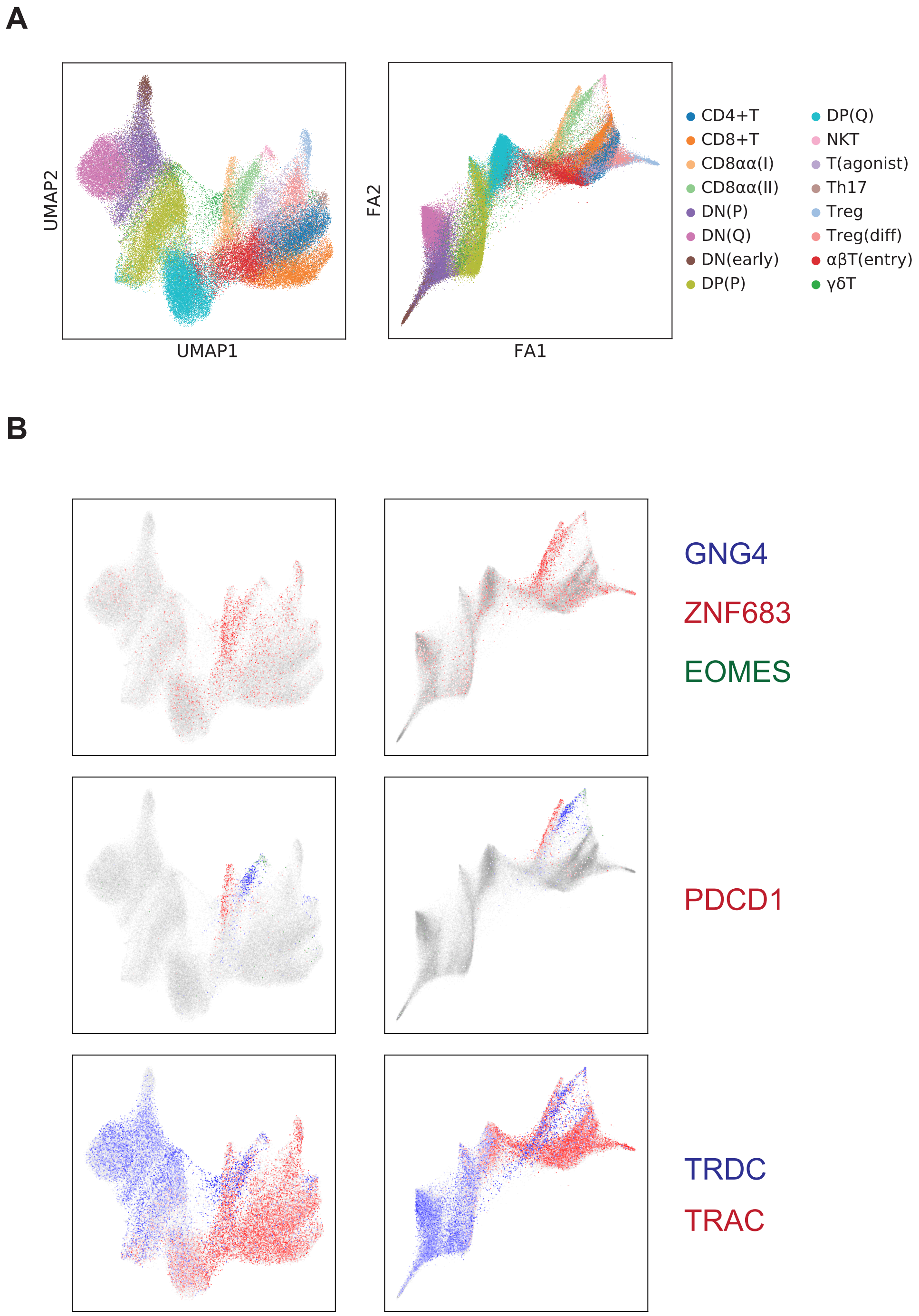

Fig. S14.

(A) UMAP plot (left, same one as shown in Fig. 2C) and force directed graph plot (right) showing T cell development trajectory. **(B)** UMAP plot (left) and force directed graph plot (right) showing marker gene expression for CD8αα^{+} T subtypes found in human thymus.

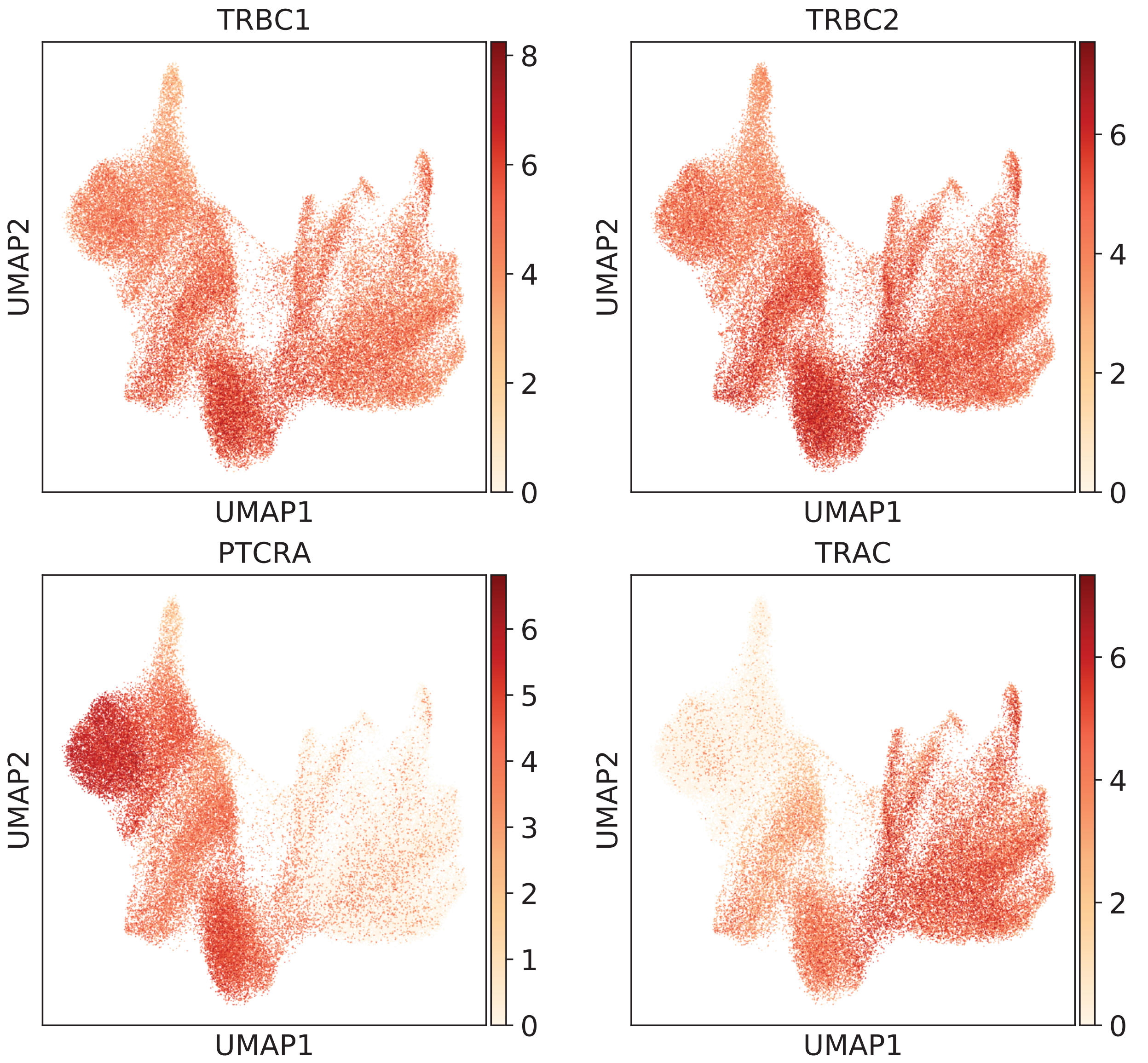

Fig. S15.

UMAP plot (same as used in Fig. 2C) displaying expression pattern of TRBC1, TRBC2, PTCRA (pre-TCR complex) and TRAC genes. Peak expression of PTCRA is found in DN(Q) cells.

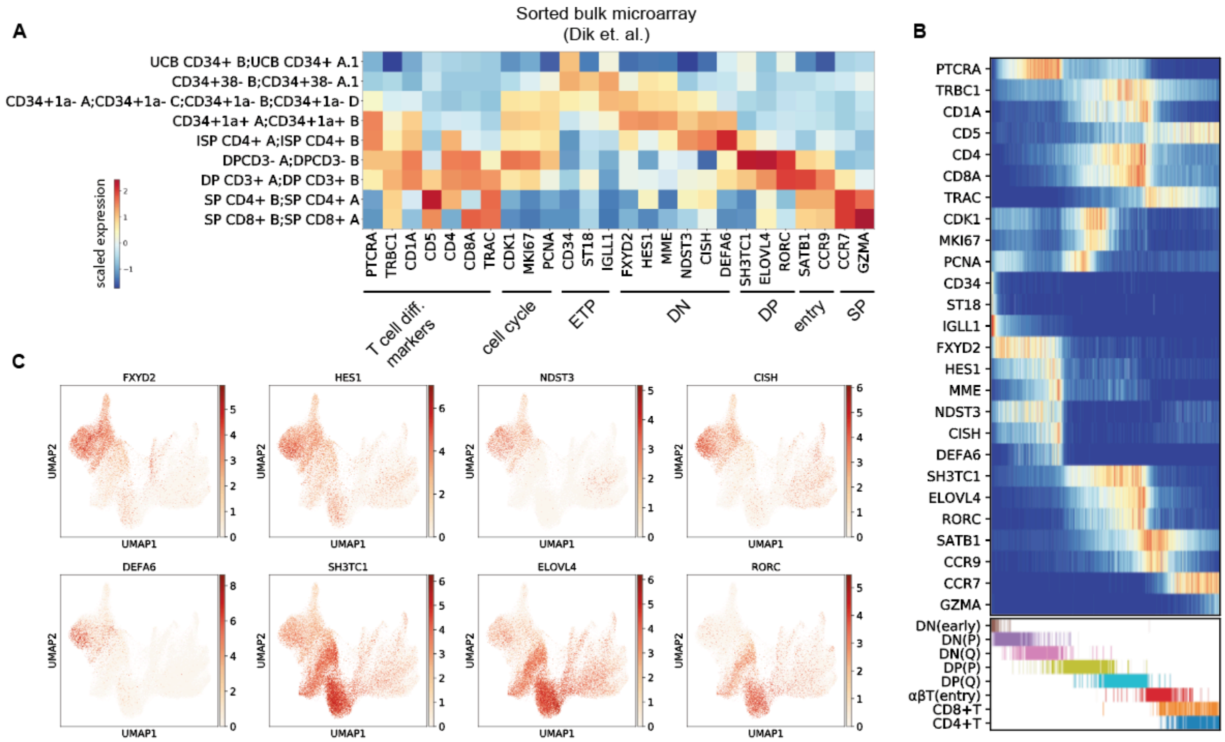

Fig. S16.

(A) Heatmap showing expression pattern of T cell differentiation marker genes (x-axis) from sorted cell populations (y-axis) (*50*). **(B)** Heatmap showing the expression pattern of T cell differentiation marker genes (same set used in (A)) across modelled pseudotime. Distribution of cell types are depicted in the lower panel. **(C)** Umap plot (same as Fig. 2C) showing expressing pattern of selected marker genes. *DEFA6* is marker gene for ISP CD4+ population (Fig. S16A), which overlaps largely with DN(Q) cells (Fig. 2C).

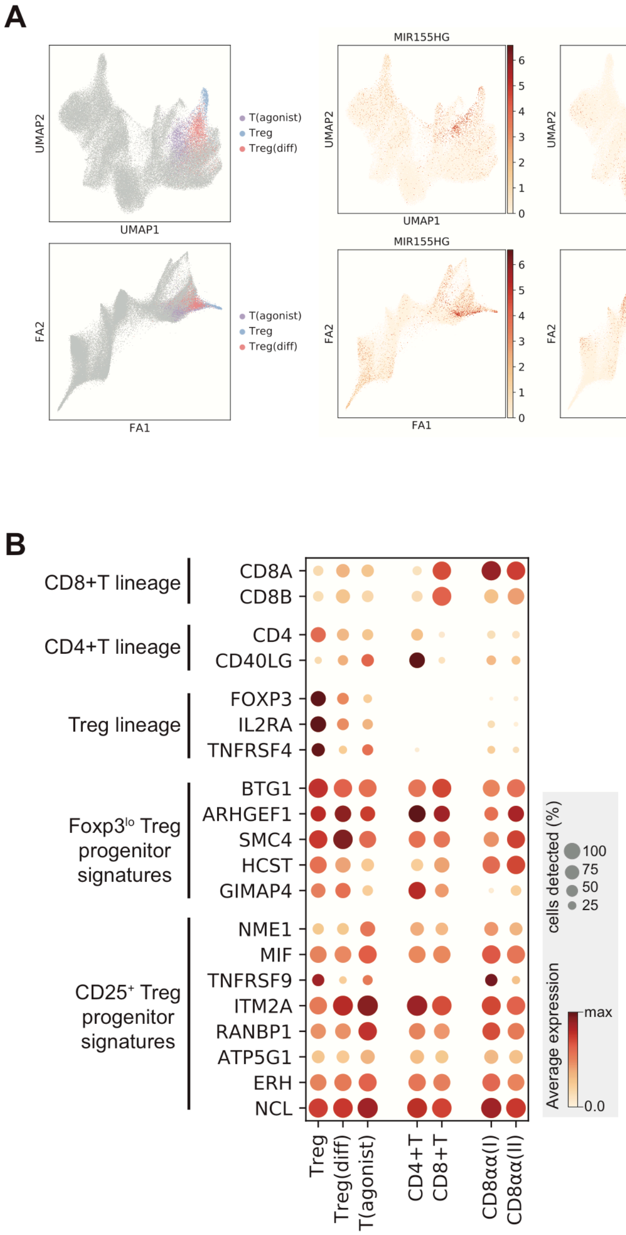

Fig. S17.

(A) UMAP and force directed graph for T cell trajectory (Same as displayed in fig. S14), highlighting Treg lineage cell types. (left panel). Right panel shows marker for T(agonist) cells (MIR155HG), Tregs and Treg(diif) cells (FOXP3), and IL2RA expression is shared among all three cell types. **(B)** Dot plot showing the expression of lineage markers and signatures for two Treg progenitors defined in (*52*).

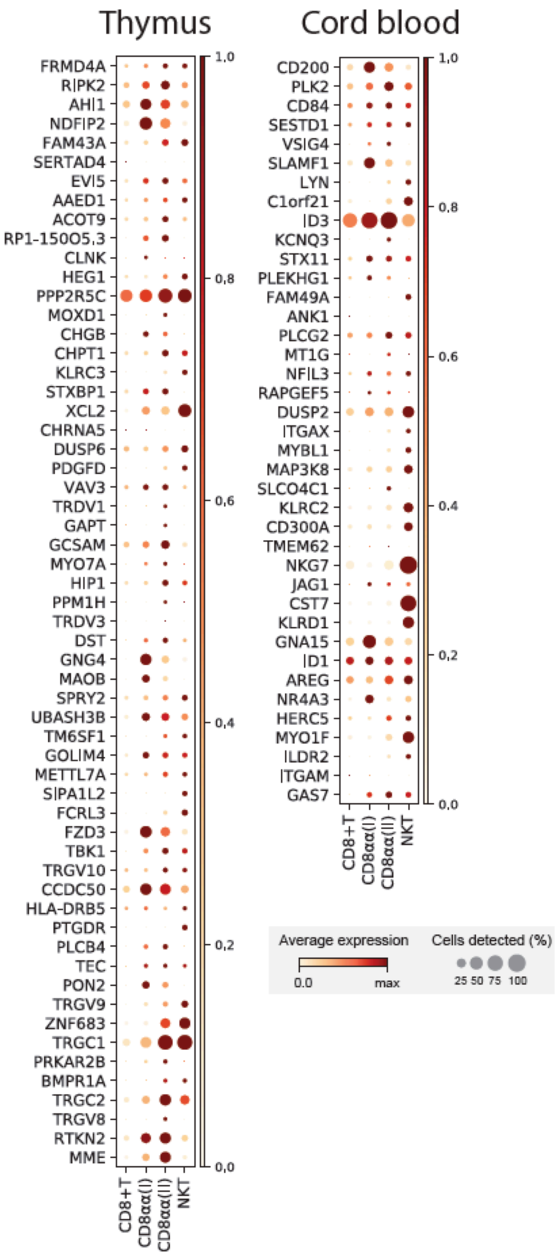

Fig. S18.

Dot plot showing the expression level of CD8αα$^+$ T marker genes enriched in thymus (left) or cord blood (right) across conventional CD8$^+$ T cells and three CD8αα$^+$ T types found in thymus.

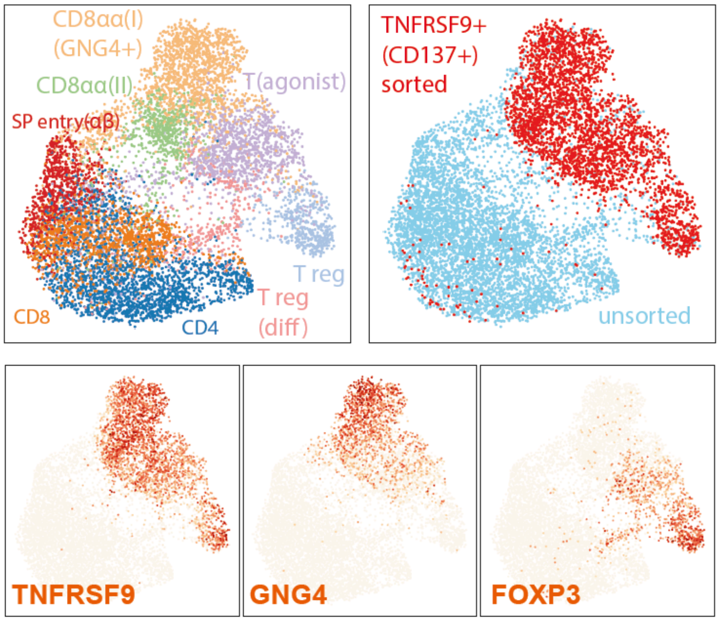

Fig. S19.
UMAP plot showing CD3+CD137+ sorted population from 12 PCW fetal thymus. Sorted cells (top right, red) were compared to unsorted mature T cells (top right, skyblue) from the same sample. Gene expression of CD8αα+T(I) marker (*GNG4*), Treg marker (*FOXP3*) and marker shared between these two groups (*TNFRSF9/CD137*) are shown.

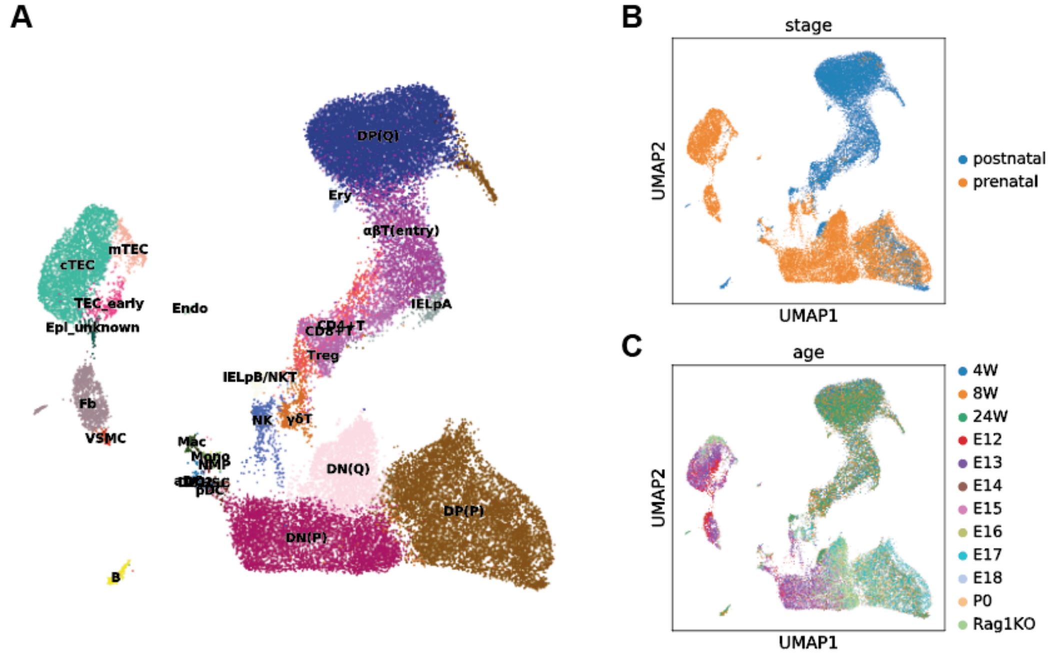

Fig. S20.
UMAP plot showing single-cell atlas of mouse thymic cells coloured by **(A)** cell types, **(B)** developmental stages and **(C)** age. E: embryonic day, W: weeks of postnatal age, Rag1KO: Rag1 knockout mouse.

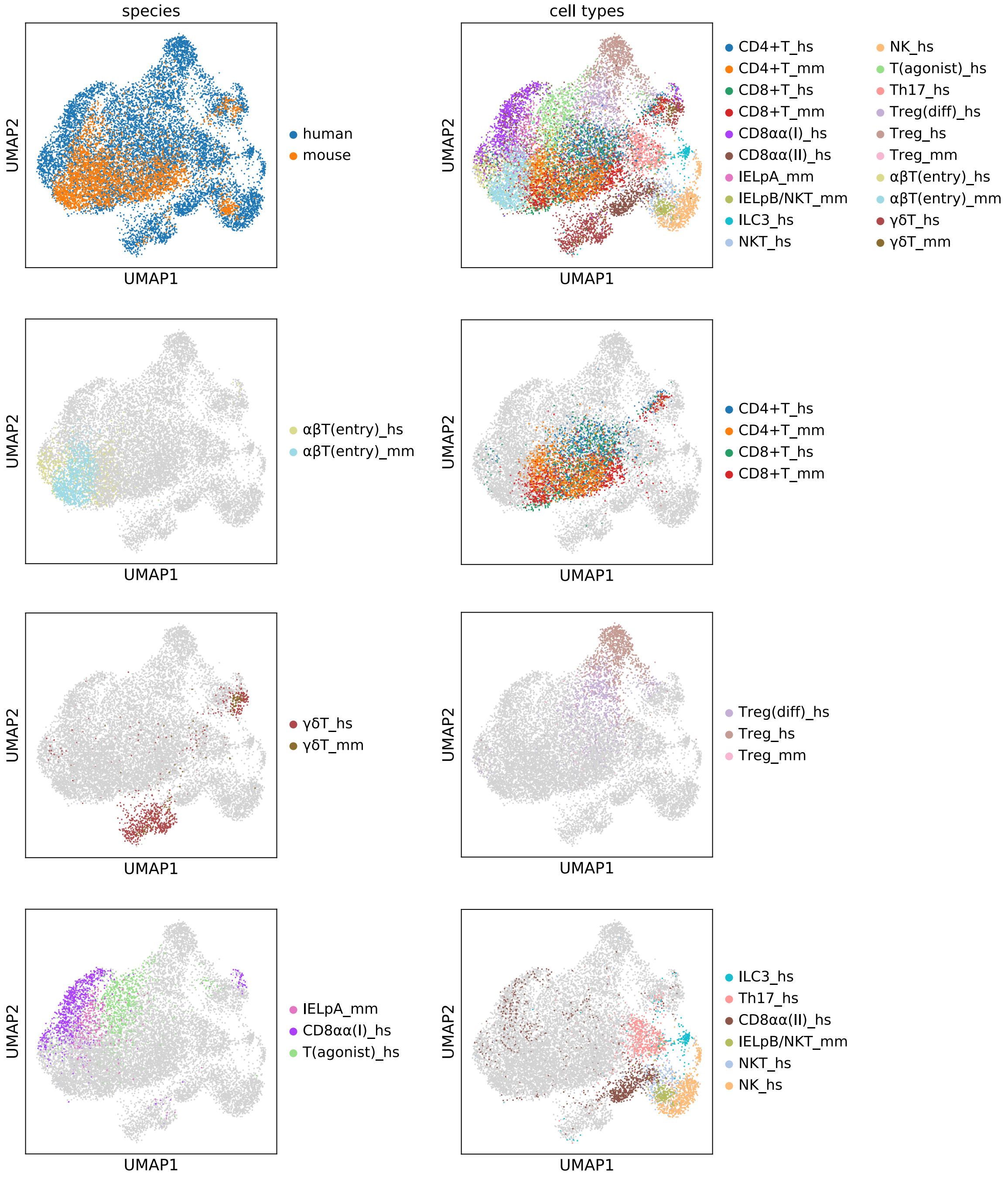

Fig. S21.

UMAP plot showing the integrative data analysis of mature T cell populations from human and mouse. Human cell types are annotated by ‘hs’ and mouse cell types are marked with ‘mm’. Matching cell groups are shown together.

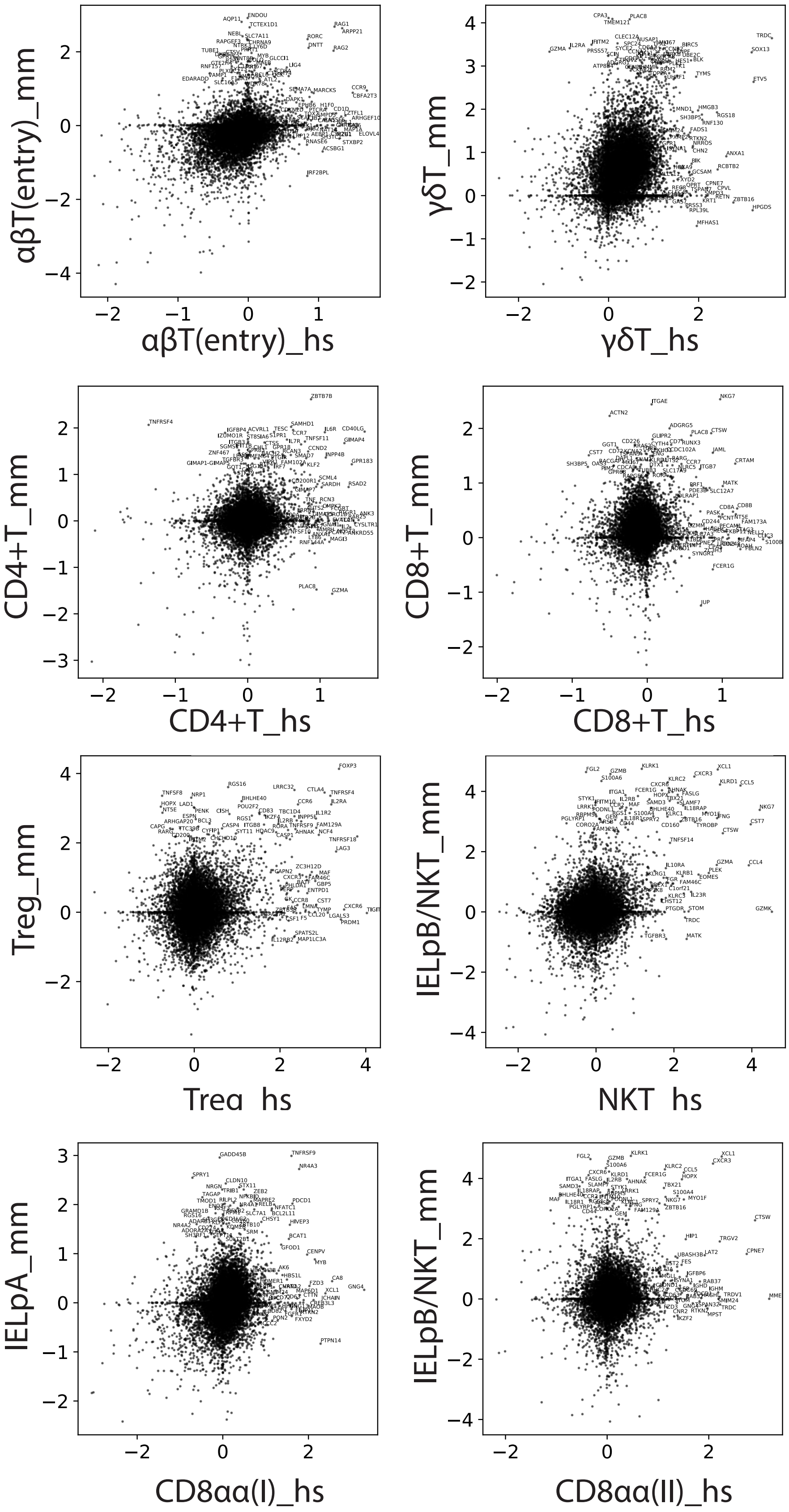
αβT(entry)_mm
αβT(entry)_hs
2
0
−2
−4
−2
−1
0
1
ENDOU
AQP11
TCTEX1D1
RAG1
ARPP21
RORC
DNTT
RAG2
CCR9
CBFA2T3
ELOVL4
STXBP2
ACSBG1
IRF2BPL
γδT_mm
γδT_hs
4
3
2
1
0
−1
−2
−2
0
2
CPA3
PLAC8
TMEM121
GZMA
IL2RA
TRDC
SOX13
ETV5
HMGB3
RGS18
ANXA1
RCBTB2
ZBTB16
HPGDS
MFHAS1
CD4+T_mm
CD4+T_hs
2
1
0
−1
−2
−3
−2
−1
0
1
ZBTB7B
TNFRSF4
CD40LG
GIMAP4
GPR183
RSAD2
PLAC8
GZMA
CD8+T_mm
CD8+T_hs
2
1
0
−1
−2
−2
−1
0
1
NKG7
ITGAE
ACTN2
CTSW
JAML
CRTAM
MATK
CD8B
S100B
FCER1G
JUP
Treg_mm
Trea hs
4
2
0
−2
−2
0
2
4
FOXP3
RGS16
LRRC32
CTLA4
TNFRSF4
IL2RA
TNFRSF18
LAG3
CXCR6
TIGIT
PRDM1
IL12RB2
MAP1LC3A
IELpB/NKT_mm
NKT hs
4
2
0
−2
−4
−2
0
2
4
FGL2
GZMB
KLRK1
XCL1
CXCR3
KLRD1
CCL5
NKG7
CST7
CTSW
GZMA
CCL4
GZMK
IELpA_mm
CD8aa(I)_hs
3
2
1
0
−1
−2
−2
0
2
GADD45B
TNFRSF9
NR4A3
SPRY1
PDCD1
GNG4
PTPN14
IELpB/NKT_mm
CD8aa(II)_hs
4
2
0
−2
−4
−2
0
2
FGL2
KLRK1
XCL1
CXCR3
CTSW
CPNE7
TRGV2
MME

Fig. S22.

Comparison of markers for mature T cell types from human and mouse thymus. X-axis is gene expression log2-fold change for the designated human cell types against all other human epithelial cells. Y-axis is gene expression log2-fold change for the designated mouse cell types against all mouse epithelial cells. Comparison sets are determined based on the data integration (Fig. S21)

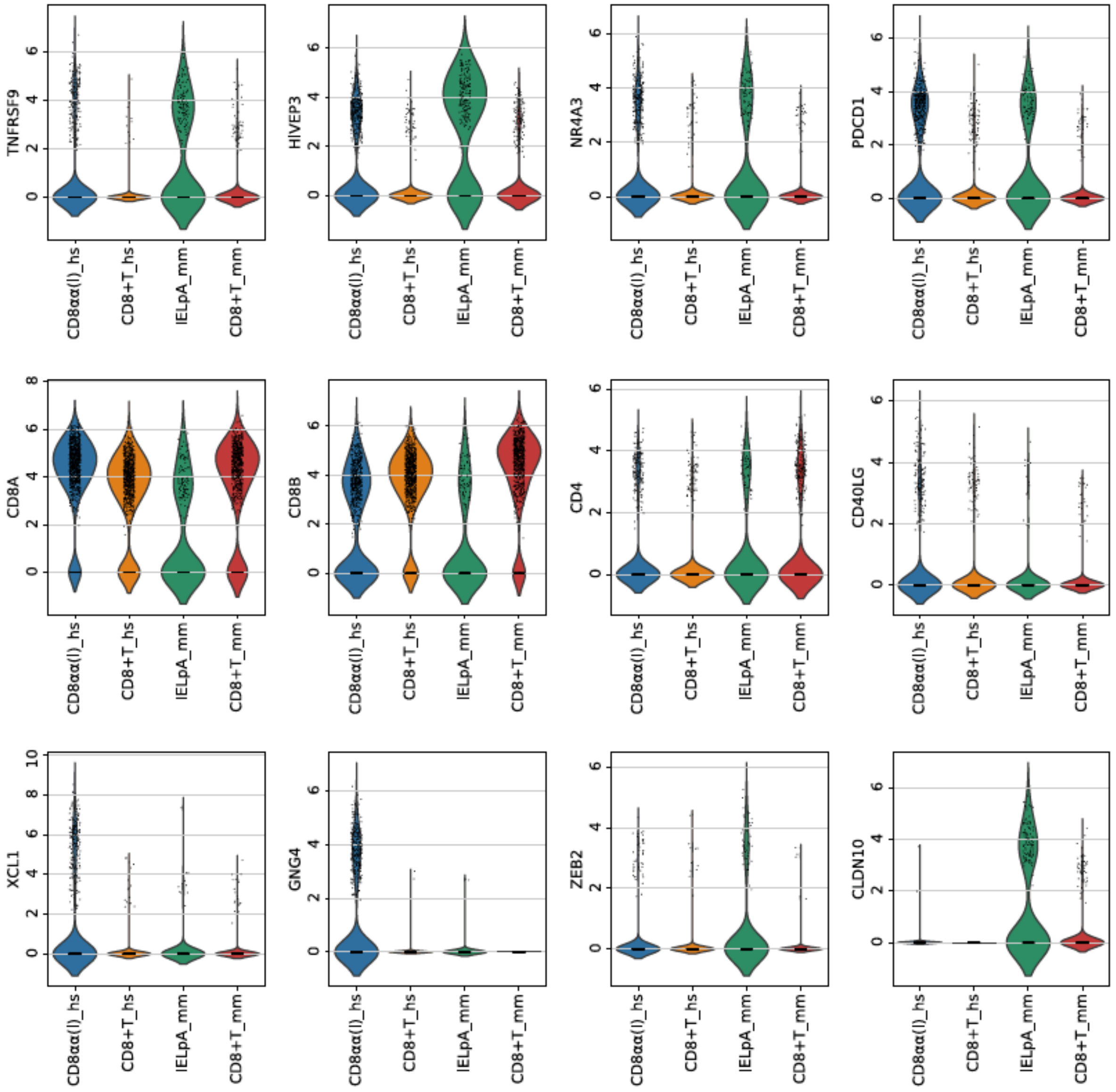

Fig. S23.
Violin plots showing the gene expression level (normalised to total reads per cell, log-transformed) across CD8αα+T(I) (human), IELpA (mouse) and CD8+T cells from both species. ‘hs’ and ‘mm’ suffix is used to identify cells from human and mouse, respectively. Gene names are designated in the y-axis.

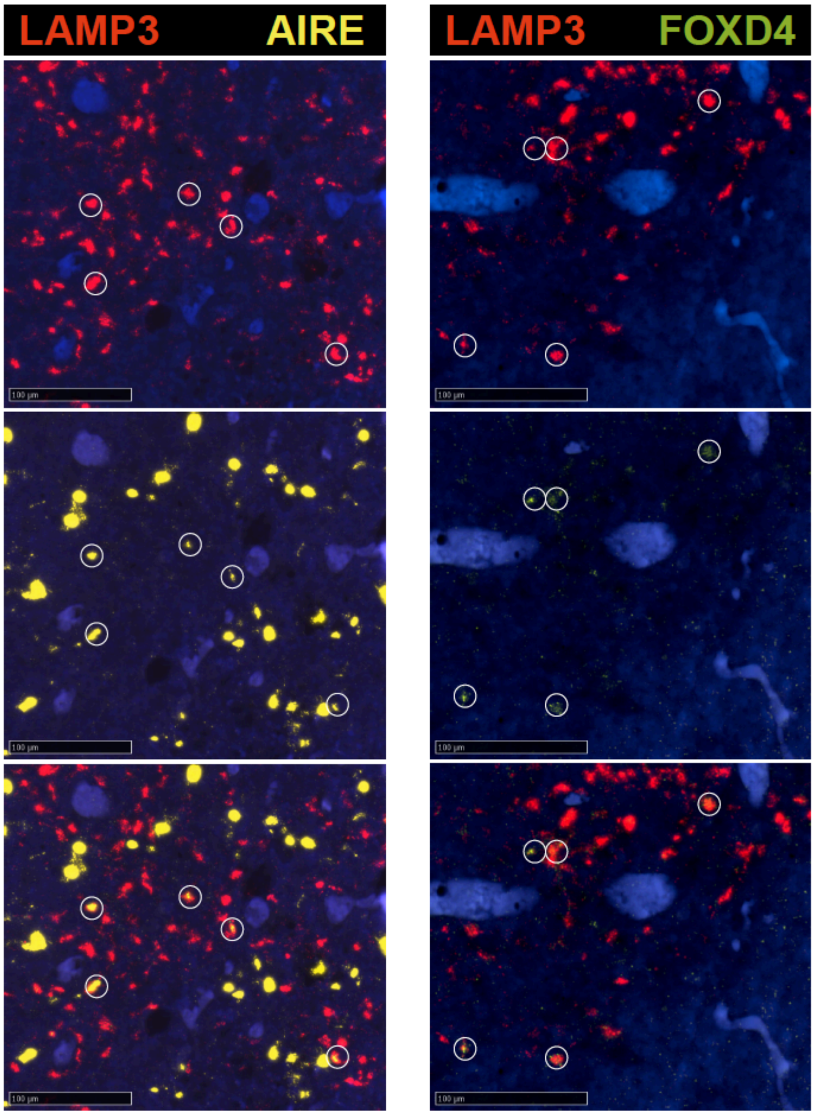

Fig. S24.
RNA single-molecule FISH detection of various genes expressed in aDCs (*LAMP3*, *AIRE*, *FOXD4*) on 15 PCW fetal thymus tissue section. Cells with expression of both genes are marked with a circle.

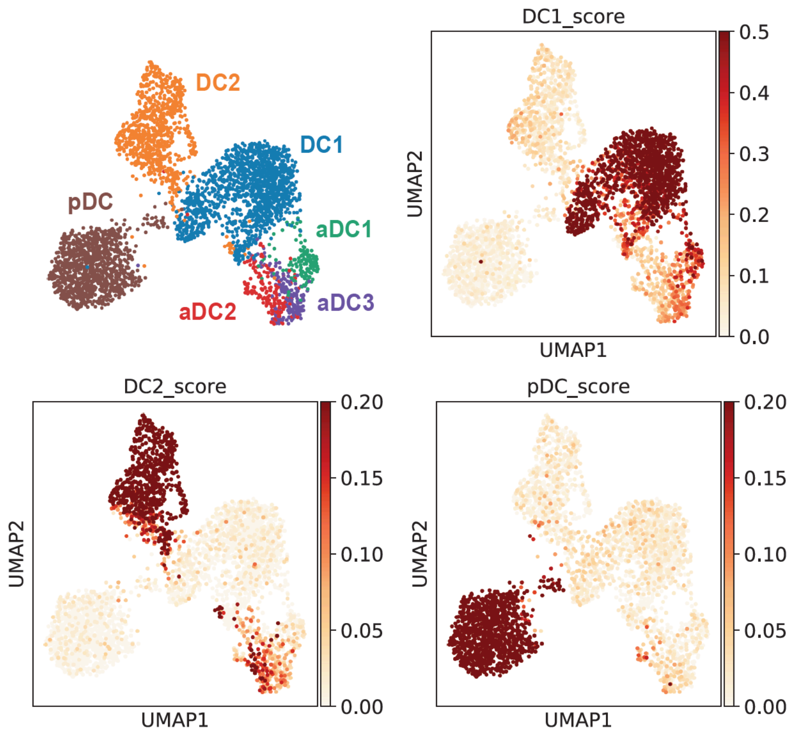

Fig. S25.

UMAP plot showing DC subtypes found in human thymus (top left). The same UMAP plot is used to show the cells with high DC1, DC2 and pDC scores, which are calculated by taking the average of expression level for lineage-specific genes.

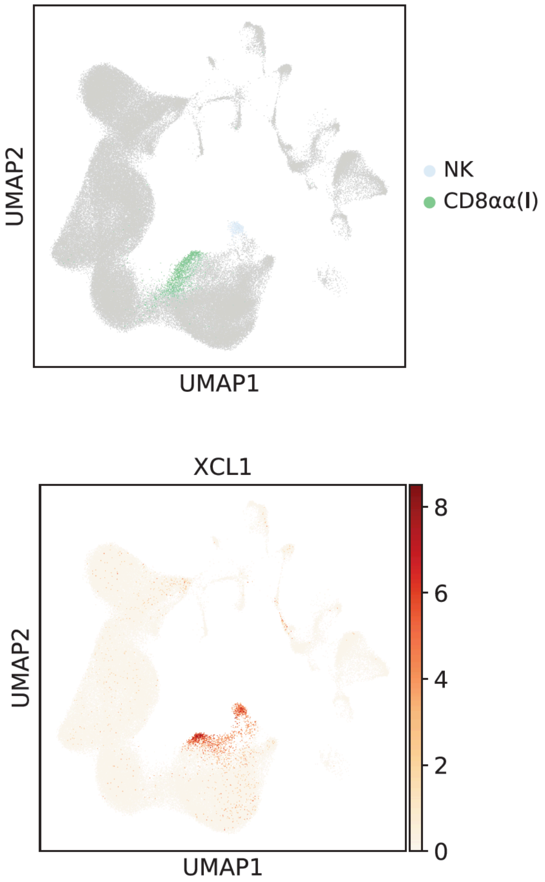

Fig. S26.
UMAP plot showing cell types expressing XCL1 (top) and XCL1 expression level in fetal thymus (bottom)

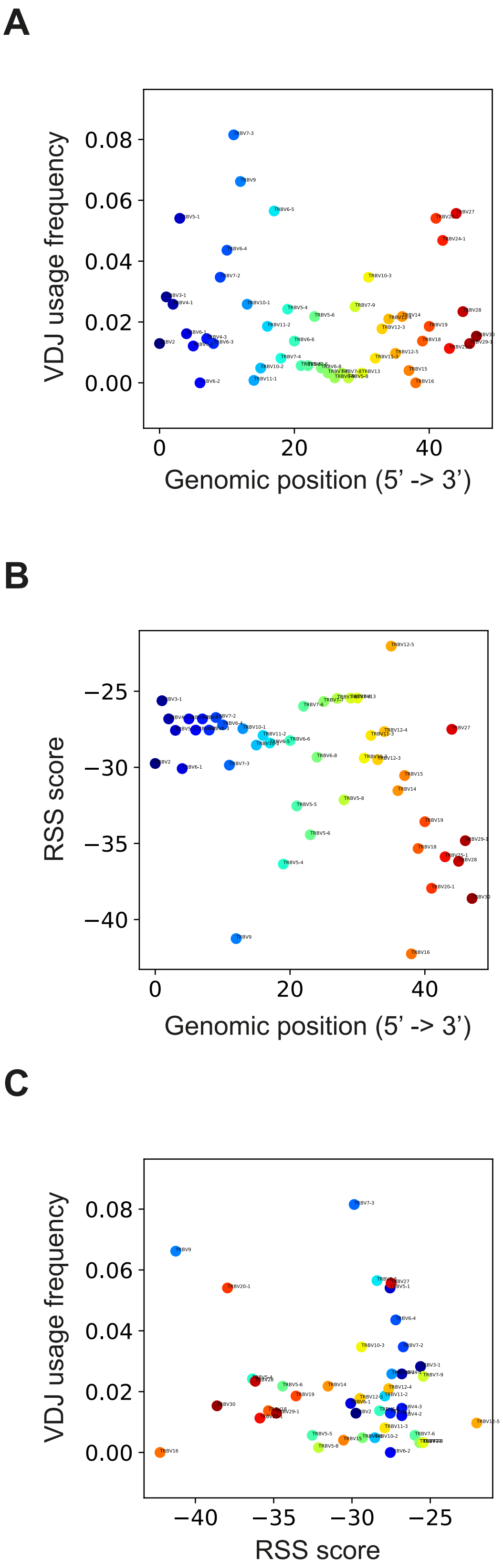
A
VDJ usage frequency
0.08
0.06
0.04
0.02
0.00
0
20
40
Genomic position (5' -> 3')
TRBV7-3
TRBV9
TRBV6-5
TRBV5-1
TRBV6-4
TRBV7-2
TRBV3-1
TRBV4-1
TRBV10-1
TRBV5-4
TRBV5-6
TRBV11-2
TRBV6-6
TRBV6-1
TRBV4-3
TRBV6-3
TRBV2
TRBV7-4
TRBV10-2
TRBV11-1
TRBV6-2
TRBV10-3
TRBV7-9
TRBV14
TRBV12-3
TRBV19
TRBV18
TRBV12-5
TRBV11-3
TRBV15
TRBV16
TRBV13
TRBV5-8
TRBV6-8
TRBV27
TRBV24-1
TRBV28
TRBV30
TRBV29-1
B
RSS score
−25
−30
−35
−40
0
20
40
Genomic position (5' -> 3')
TRBV12-5
TRBV3-1
TRBV7-2
TRBV6-4
TRBV10-1
TRBV11-2
TRBV6-6
TRBV7-6
TRBV7-7
TRBV12-4
TRBV12-3
TRBV13
TRBV27
TRBV6-8
TRBV2
TRBV6-1
TRBV7-3
TRBV15
TRBV14
TRBV5-5
TRBV5-8
TRBV19
TRBV29-1
TRBV18
TRBV5-6
TRBV25-1
TRBV28
TRBV5-4
TRBV20-1
TRBV30
TRBV9
TRBV16
C
VDJ usage frequency
0.08
0.06
0.04
0.02
0.00
−40
−35
−30
−25
RSS score
TRBV7-3
TRBV9
TRBV20-1
TRBV27
TRBV5-1
TRBV6-4
TRBV10-3
TRBV7-2
TRBV3-1
TRBV7-9
TRBV5-6
TRBV14
TRBV19
TRBV12-4
TRBV11-2
TRBV30
TRBV29-1
TRBV6-1
TRBV2
TRBV4-3
TRBV4-2
TRBV5-5
TRBV15
TRBV10-2
TRBV11-3
TRBV7-6
TRBV5-8
TRBV12-5
TRBV6-2
TRBV16

Fig. S27.

(A) Scatter plot comparing genomic position (x-axis) and relative usage (y-axis) for TCRβ V genes. Genes are coloured based on genomic position. The same colour scheme is applied for following figures. **(B)** Scatter plot comparing genomic position (x-axis) and RSS score (y-axis) for TCRβ V genes. **(C)** Scatter plot comparing RSS score (x-axis) and relative usage (y-axis) for TCRβ V genes.

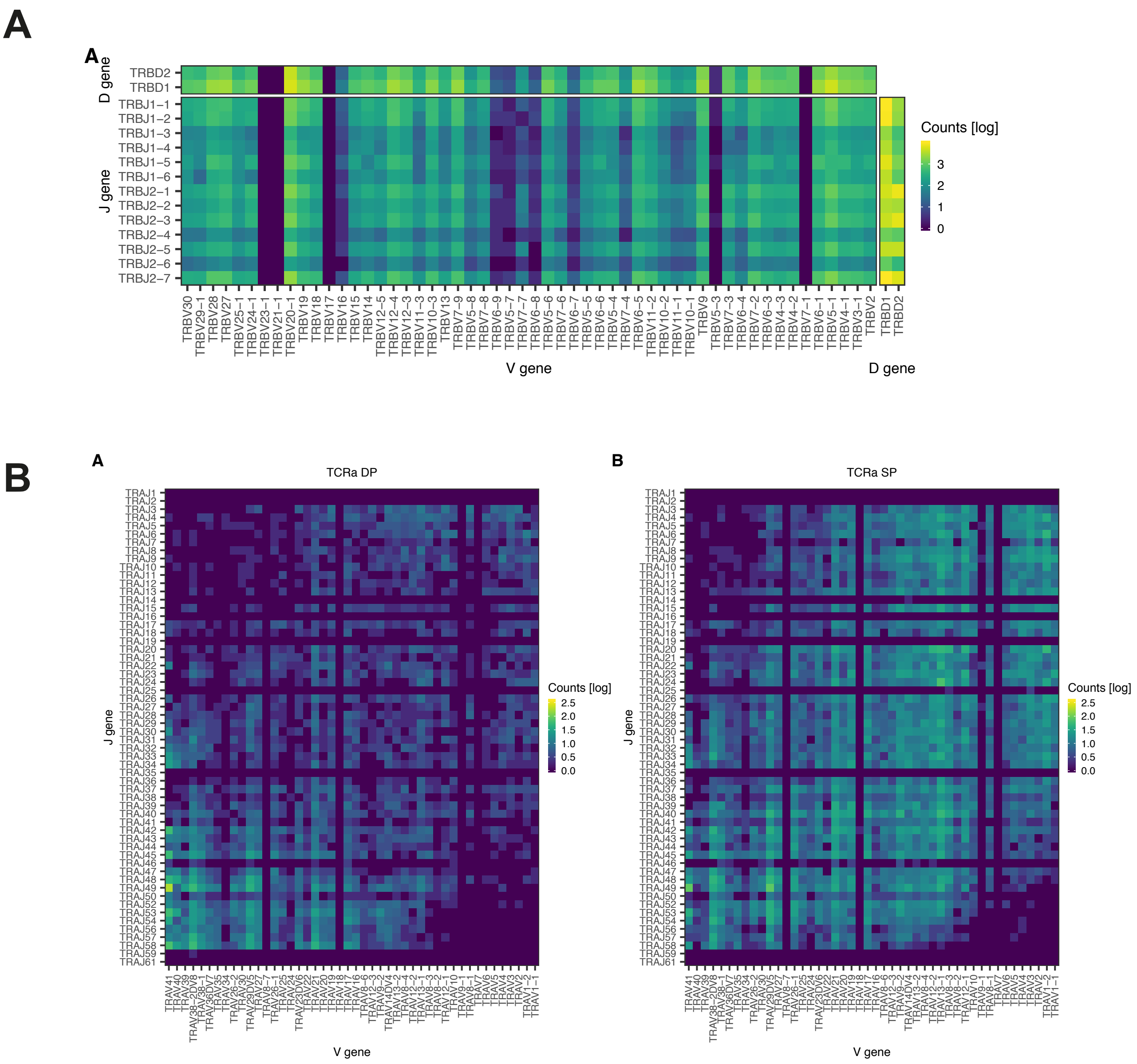

Fig. S28.

(A) Relative frequency (log scale) of V-J, V-D, J-D gene pairs in TCRβ locus. **(B)** Relative frequency (log scale) of V-J gene pairs in TCRα locus. Dataset is divided into DP and SP stages to highlight the enrichment of proximal pairs in DP stage.

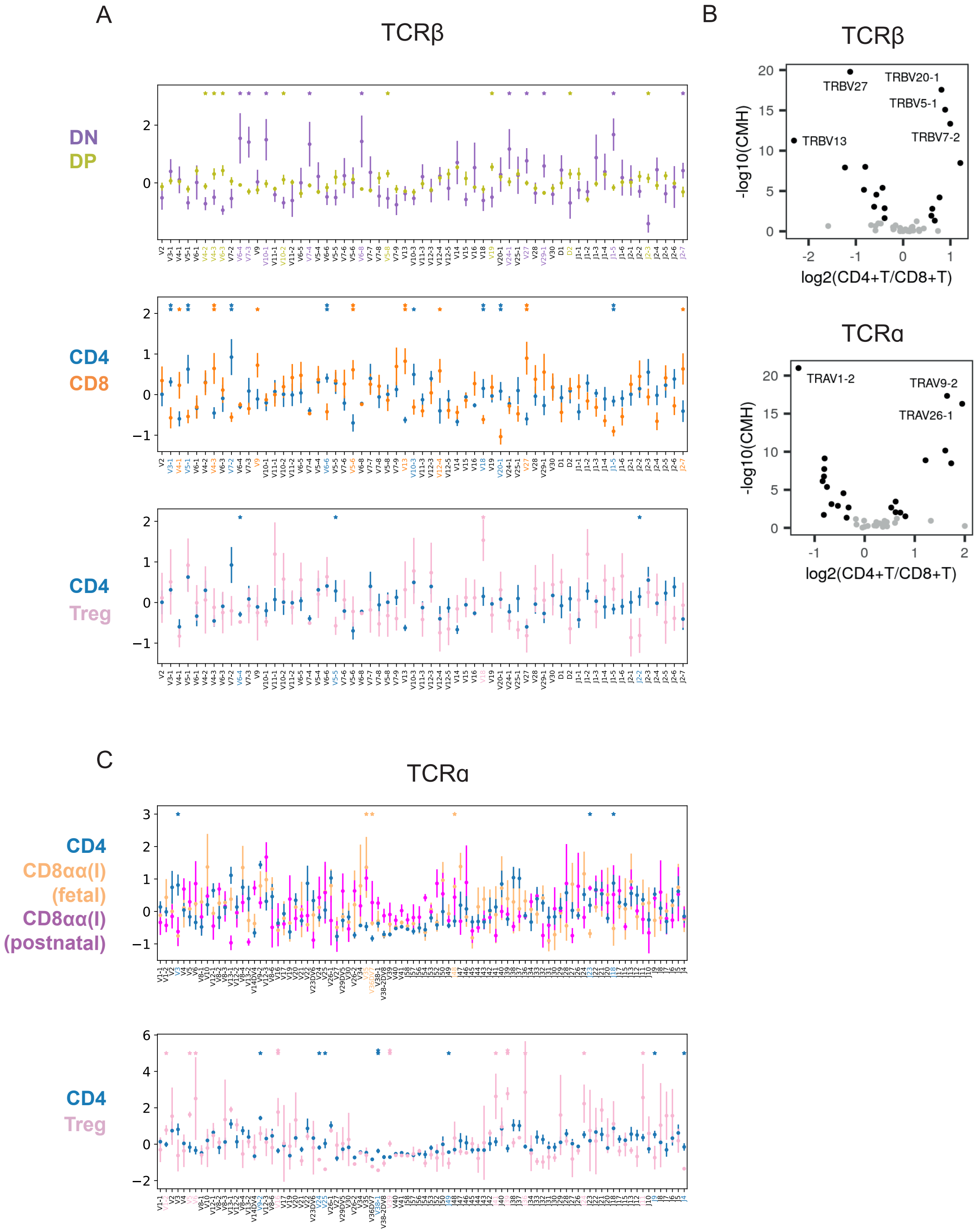
A
TCRβ
DN
DP
CD4
CD8
CD4
Treg
B
TCRβ
TRBV27
TRBV20-1
TRBV5-1
TRBV13
TRBV7-2
-log10(CMH)
log2(CD4+T/CD8+T)
TCRα
TRAV1-2
TRAV9-2
TRAV26-1
-log10(CMH)
log2(CD4+T/CD8+T)
C
TCRα
CD4
CD8αα(I)
(fetal)
CD8αα(I)
(postnatal)
CD4
Treg

Fig. S29.

(A, B) Relative usage of V, D and J gene segments according to cell types for TCRβ locus (A) or TCRα locus (B). *Z*-score for each segment is calculated from the distribution of normalised proportions stratified by the cell type and sample. *P*-value is calculated by comparing z-scores using t-test, and FDR is calculated using Benjamini-Hochberg correction. (*: *p*-value < 0.05, **: FDR < 10%) Gene names on the x-axis and asterisks are coloured by significant enrichment. For CD4 vs CD8αα+T (I) comparison, CD8αα+T (I) data points are separated into fetal samples (n=4) and post-natal sample (n=1, young adult) to highlight differences between fetal sampels and young adult sample. All other comparisons are inclusive of both fetal and post-natal samples. Consistency between fetal and post-natal samples are separately confirmed (data not shown).
(C) Volcano plot showing log2(fold change) of V, D, J gene frequencies between CD4+T and CD8+T cells (x-axis) and -log10(p-value) calculated by Cochran–Mantel–Haenszel test. Genes with most significant changes are annotated.

Table S1. Table_S1.xlsx (separate file)

Excel file containing metadata for sequenced samples.

[FileName] Prefix for raw sequencing files

[SampleID] Sample description (DonorID-Organ-Sort-Method)

[Organ] TH: thymus, SP: spleen, LI: liver, BM: bone marrow, iLN: iliac lymph node, tLN: thoracic lymph node, mLN: mesenteric lymph node

[DonorID] Unique ID assigned for each donor

[Sort] Sorting scheme used for each sample. 45P: CD45+, 45N: CD45-, CD3P: CD3+, CD3N: CD3-, CD137: CD137, 45NM: CD45 depletion by MACS, EPCAM: EPCAM enrichment by MACS

[Method] 3GEX: 10X 3' chemistry gene expression profiling, 5GEX: 10X 5' chemistry gene expression profiling

[VDJ_file] File names for TCR enrichment sequencing if available

[Enzyme] Protocols used for dissociation

Table S2. Table_S2.xlsx (separate file)

Excel file containing absolute cell numbers for each cell type in each sample.

Table S3. Table_S3.xlsx (separate file)

Excel file containing hierarchy of cell type annotations used in the study

Table S4. Table_S4.csv (separate file)

CSV file containing top 20 marker genes for each cell type.

Table S5. Table_S5_cpdb_means.csv (separate file)

CellPhoneDB analysis output file containing means calculated for each ligand-receptor pair within each cell-cell pair. The output has been selected for the ligand-receptor pairs which are specific to at least one cell-cell pair. Please refer to CellPhoneDB manual for details.

Table S6.

Antibodies used for FACS staining

Marker	Fluorochrome	Clone	Isotype	Supplier
CD123	BUV395	7G3	Mouse IgG2a κ	BD Biosciences
CD11c	APC-Cy7	Bu15	Mouse IgG1 κ	Biolegend
CD14	PE CF594	MφP9	Mouse IgG2b κ	BD Biosciences
CD137	PE-Cy5	4B4-1	Mouse IgG1 κ	Biolegend
CD141	PerCP-Cy5.5	M80	Mouse IgG1 κ	Biolegend
CD19	FITC	4G7	Mouse IgG1 κ	BD Biosciences
CD20	FITC	L27	Mouse IgG1 κ	BD Biosciences
CD3	BV605	SK7	Mouse IgG1 κ	Biolegend
CD4	BV711	RPA-T4	Mouse IgG1 κ	Biolegend
CD8A	AF700	HIT8a	Mouse IgG1 κ	Biolegend
CD8B	FITC	REA715	Human IgG1	Miltenyi Biotec
HLA-DR	BV785	L243	Mouse IgG2a κ	Biolegend
EpCAM	Vioblue	HEA125	Mouse IgG1 κ	Miltenyi Biotec
CD45	APC	HI30	Mouse IgG1 κ	BD Biosciences
CCR7	PerCP-Cy5.5	G043H7	Mouse IgG2a κ	Biolegend
CD56	PE	NCAM16.2	IgG2b, k	BD Biosciences
CD34	PE-Cy7	581	Mouse IgG1 κ	Biolegend
CD3	APC	SK7	Mouse IgG1 κ	Biolegend
THY1	Af700	5E10	Mouse IgG1 κ	Biolegend
PEGFRa	PE	16A1	Mouse IgG1 κ	Biolegend
PI16	BV605	RUO	Mouse IgG1 κ	BD Biosciences

Table S7.
Probes used for smRNA FISH

Gene ID	Cat. Number	Channel
CSF2RA	409341	C1
CCR7	410721	C1
LAMP3	468761-C2	C2
CD80	421471-C3	C3
CD8A	560391-C3	C3
FOXP3	418471	C1
TNFRSF9	415171	C1
ITGAX	419151	C1
FBN1	482478-C2	C2
COLEC11	542438	C1
ACTA2	311818-C3	C3
PDGFRA	604488	C1
CDH5	437458-C3	C3
XCR1	custom	C3
FOXD4	custom	C3
GNG4	custom	C2
AIRE	custom	C1
GDF10	506168	
NEUROG1	444398-C2	C2
MYOD1	562728-C2	C2
EPCAM	310288-C4	C4
ALDH1A2	528748	

Table S8.

List of cell cycle genes (559 genes) defined and used in this study

AC004381.6, ACAT2, ACOT7, ACSL3, ACTL6A, ACYP1, ADK, AIFM1, ALYREF, ANKRD36C, ANLN, ANP32B, ANP32E, AP000251.3, ARHGAP11A, ARHGAP11B, ARHGAP33, ARHGEF39, ASF1B, ASPM, ASRGL1, ATAD2, ATAD5, ATP5G1, ATP8B3, AURKA, AURKB, BAG2, BARD1, BAZ1B, BCL2L12, BIRC5, BLM, BLMH, BOP1, BORA, BRCA1, BRCA2, BRD7, BRD8, BRIP1, BUB1, BUB1B, BUB3, BUD13, C16orf91, C19orf48, C1QBP, C1orf112, C1orf35, C21orf58, C4orf27, C4orf46, C5orf34, C8orf88, C9orf40, CARHSP1, CASC5, CASP8AP2, CBX2, CBX5, CCDC14, CCDC15, CCDC167, CCDC18, CCDC34, CCDC58, CCDC86, CCNA2, CCNB1, CCNB2, CCNE2, CCNF, CCP110, CCSAP, CDC20, CDC25A, CDC25B, CDC25C, CDC27, CDC45, CDC6, CDC7, CDCA2, CDCA3, CDCA4, CDCA5, CDCA7, CDCA8, CDK1, CDK2, CDK5RAP2, CDKN2AIP, CDKN3, CDT1, CENPA, CENPE, CENPF, CENPH, CENPJ, CENPK, CENPL, CENPM, CENPN, CENPO, CENPP, CENPQ, CENPU, CENPV, CENPW, CEP152, CEP55, CEP57L1, CEP76, CEP78, CEP97, CHAC2, CHAF1A, CHAF1B, CHEK1, CISD1, CIT, CKAP2, CKAP2L, CKAP5, CKLF, CKS1B, CKS2, CLGN, CLSPN, CMSS1, CNP, CRNDE, CSE1L, CTC-260E6.6, CTCF, CTDSPL2, CTNNAL1, CTPS1, DAZAP1, DBF4, DCAF12, DDB2, DDX11, DDX39A, DEK, DEPDC1, DEPDC1B, DHCR24, DHFR, DIAPH3, DLEU2, DLGAP5, DNA2, DNAJC9, DNMT1, DSCC1, DSG2, DSN1, DTL, DTYMK, DUT, E2F2, E2F7, E2F8, EBNA1BP2, ECT2, EIF1AY, ELP5, EMC9, ENO2, ENOSF1, EPCAM, ERCC6L, ERH, ERI2, ESCO2, ESPL1, EXO1, EXOC5, EXOSC5, EXOSC8, EXOSC9, EZH2, FAIM, FAM111A, FAM111B, FAM122B, FAM221A, FAM64A, FAM72B, FAM76B, FAM83D, FANCA, FANCD2, FANCG, FANCI, FBXO5, FEN1, FH, FHL2, FKBP5, FOXM1, G2E3, GALK1, GAPDH, GAR1, GARS, GEN1, GGH, GINS1, GINS2, GINS4, GKAP1, GLRX5, GMCL1, GMNN, GMPPB, GOT2, GPANK1, GPATCH4, GPN3, GPSM2, GSG2, GTF3A, GTF3C5, GTSE1, H1FX, H2AFV, H2AFX, H2AFY, H2AFZ, HADH, HAT1, HAUS6, HELLS, HIRIP3, HIST1H1A, HIST1H1B, HIST1H1D, HIST1H1E, HIST1H2AH, HIST1H2AM, HIST1H3G, HIST1H4C, HIST2H2AC, HIST3H2A, HJURP, HLTF, HMGA2, HMGB1, HMGB2, HMGB3, HMGCS1, HMGN2, HMGN5, HMGXB4, HMMR, HN1, HNRNPLL, HNRNPR, HPRT1, HSPA14, HSPB11, IARS, IDH2, IFRD2, IGF2BP1, ILF2, IMMP1L, INCENP, ING2, ITGB3BP, JAM3, KCTD9, KDM1A, KIAA0101, KIAA1524, KIF11, KIF14, KIF15, KIF18A, KIF18B, KIF20A, KIF20B, KIF22, KIF23, KIF2C, KIF4A, KIFC1, KLHL23, KMT5A, KNSTRN, KNTC1, KPNA2, LDHA, LDLR, LEO1, LIG1, LIN9, LMNB1, LMNB2, LRR1, LRRC42, LRRCC1, LSM4, MAD2L1, MAD2L2, MAGOHB, MASTL, MCM10, MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MCM8, MELK, MGME1, MIS18A, MIS18BP1, MKI67, MLH1, MMS22L, MND1, MNS1, MRPS2, MRPS23, MRTO4, MSH2, MSH6, MTFR2, MTHFD1, MTHFD2, MXD3, MYBL2, MYEF2, MZT1, NAE1, NASP, NCAPD2, NCAPD3, NCAPG, NCAPG2, NCAPH, NCAPH2, NCBP2-AS2, NDC80, NEDD1, NEIL3, NEK2, NFYB, NOP14, NOP16, NRM, NTPCR, NUCKS1, NUDT1, NUDT15, NUDT8, NUF2, NUP107, NUP155, NUP37, NUP50, NUP93, NUSAP1, ODF2, OIP5, ORC1, ORC6, OXCT1, PAICS, PARPBP, PAWR, PBK, PCNA, PDCD2, PGAM1, PGP, PHF19, PHGDH, PIDD1, PIF1, PKMYT1, PLCB4, PLK1, PLK4, PM20D2, POC1A, POLA1, POLA2, POLD1, POLD3, POLE, POLE2, POLQ, POLR2D, POLR3K, POP7, PPA1, PPIL1, PRC1, PRDX2, PRIM1, PRIM2, PRKDC, PRPS1, PRR11, PRSS21, PSIP1, PSMC3IP, PSMG1, PSMG3, PSRC1, PTMA, PTTG1, PUM3, PXMP2, RACGAP1, RAD18, RAD21, RAD51, RAD51AP1, RAD51C, RAD54L, RAN, RANBP1, RANGAP1, RBBP8, RBL1, RCC1, RDM1, RFC2, RFC3, RFC4, RFC5, RFWD3, RHEB, RHNO1, RMI1, RMI2, RNASEH2A, RNF168, RP11-196G18.23, RPA1, RPA3, RPL39L, RPS4Y1, RRM1, RRM2, RTKN2, RUVBL1, SAAL1, SAC3D1, SAE1, SAMD1, SASS6, SEH1L, SFXN4, SGOL1, SGOL2, SGTA, SHCBP1, SHMT1, SIVA1, SKA1, SKA2, SKA3, SKP2, SLC16A1, SLC2A1, SLC39A8, SLC43A3, SLC7A3, SLF1, SLFN13, SMC1A, SMC2, SMC3, SMC4, SNRNP48, SNRPD1, SPAG5, SPC24, SPC25, SPDL1, SRD5A3, SRM, SSRP1, STIL, STMN1, SUV39H2, SVIP, TACC3, TCF19, TCOF1, TDP1, TEX30, TFDP1, THOC3, THOC6, THOP1, TICRR, TIMELESS, TK1, TM7SF3, TMEM106C, TMEM237, TMEM97, TMPO, TOMM40, TOMM5, TOP2A, TOPBP1, TPGS2, TPX2, TRAIP, TRAP1, TRIP13, TROAP, TTF2, TTK, TUBA1B, TUBB, TUBB4B, TUBG1, TXN, TXNRD1, TYMS, UBE2C, UBE2S, UBE2T, UBR7, UCHL5, UCK2, UHRF1, UNG, USP1, USP39, VRK1, WDHD1, WDR34, WDR43, WDR62, WDR76, WDR77, WEE1, WHSC1, XRCC6BP1, YBX1, YDJC, YEATS4, ZGRF1, ZNF714, ZNF738, ZWILCH, ZWINT

Data S1. T_cell_development_3D_umap.html (separate file)

HTML file containing 3D umap structure for T cell developmental trajectory.