Supplementary Materials for

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

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

*Corresponding author. Email: m.a.haniffa@newcastle.ac.uk (M.H.); tom.taghon@ugent.be (T.T.); st9@sanger.ac.uk (S.A.T.)

Published 21 February 2020, *Science* **367**, eaay3224 (2020)
DOI: 10.1126/science.aay3224

This PDF file includes:

Other supplementary material for this manuscript includes the following:

Tables S1 to S5 (Excel format)
Data S1 (HTML format)

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.

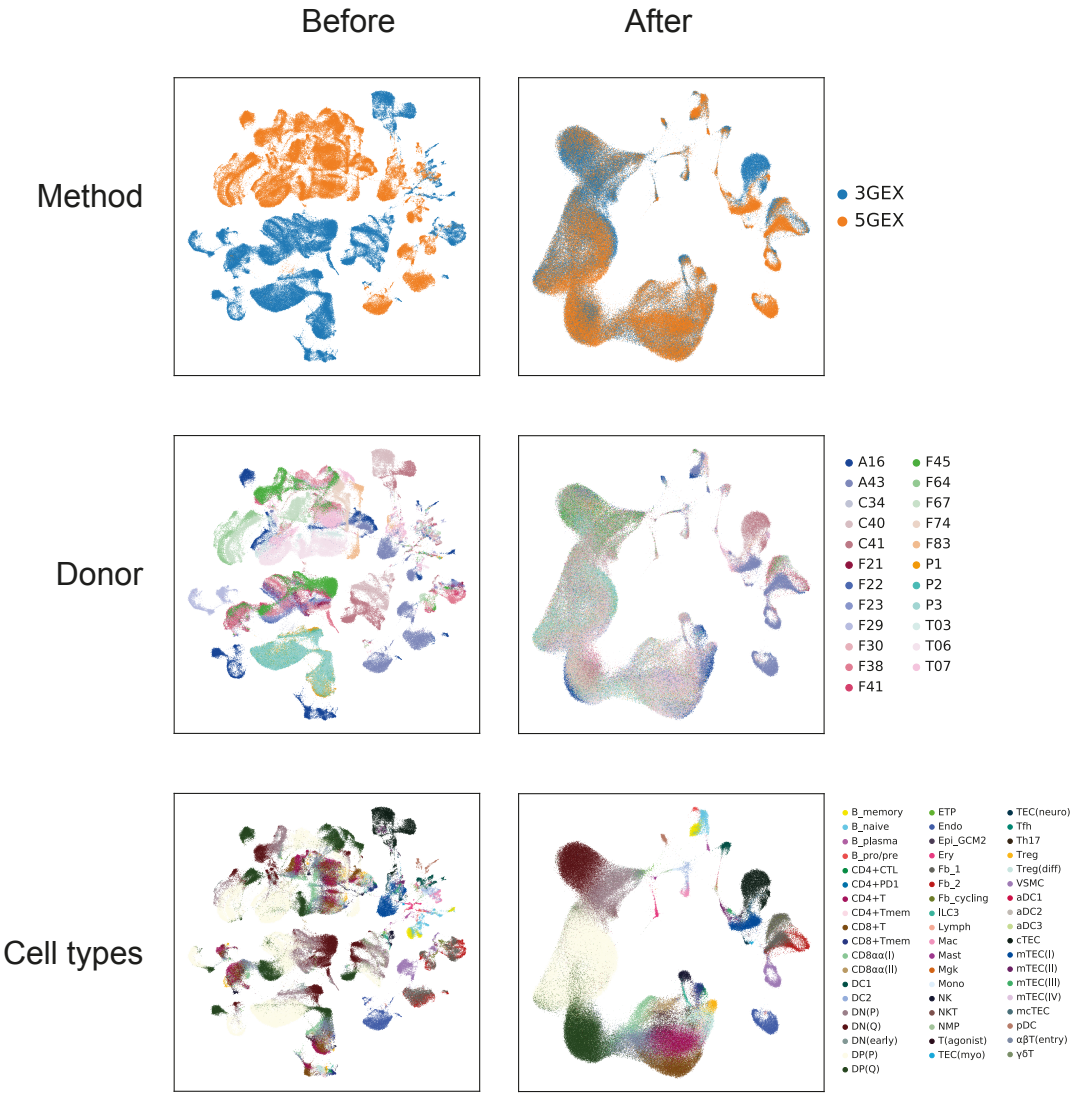


Fig. S1.

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).

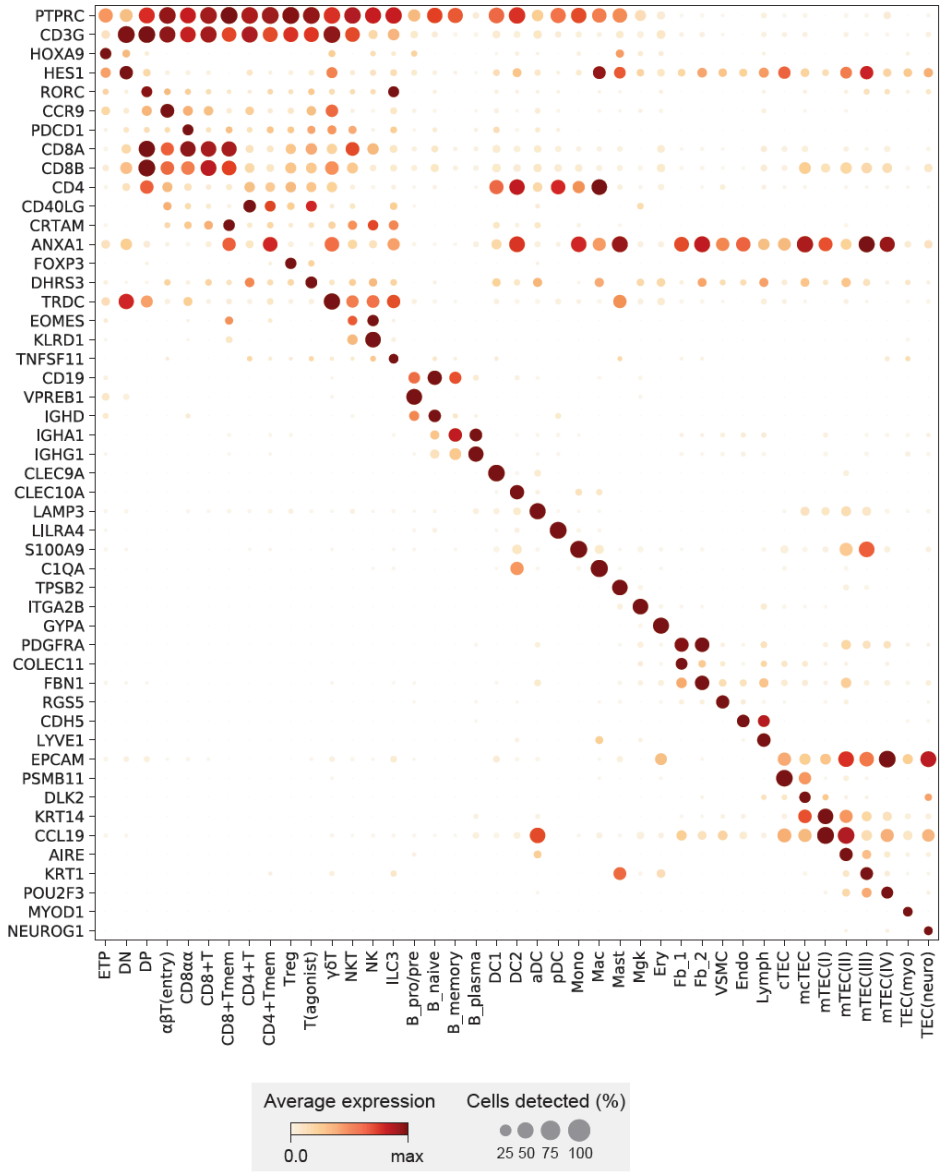


Fig. S2.

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

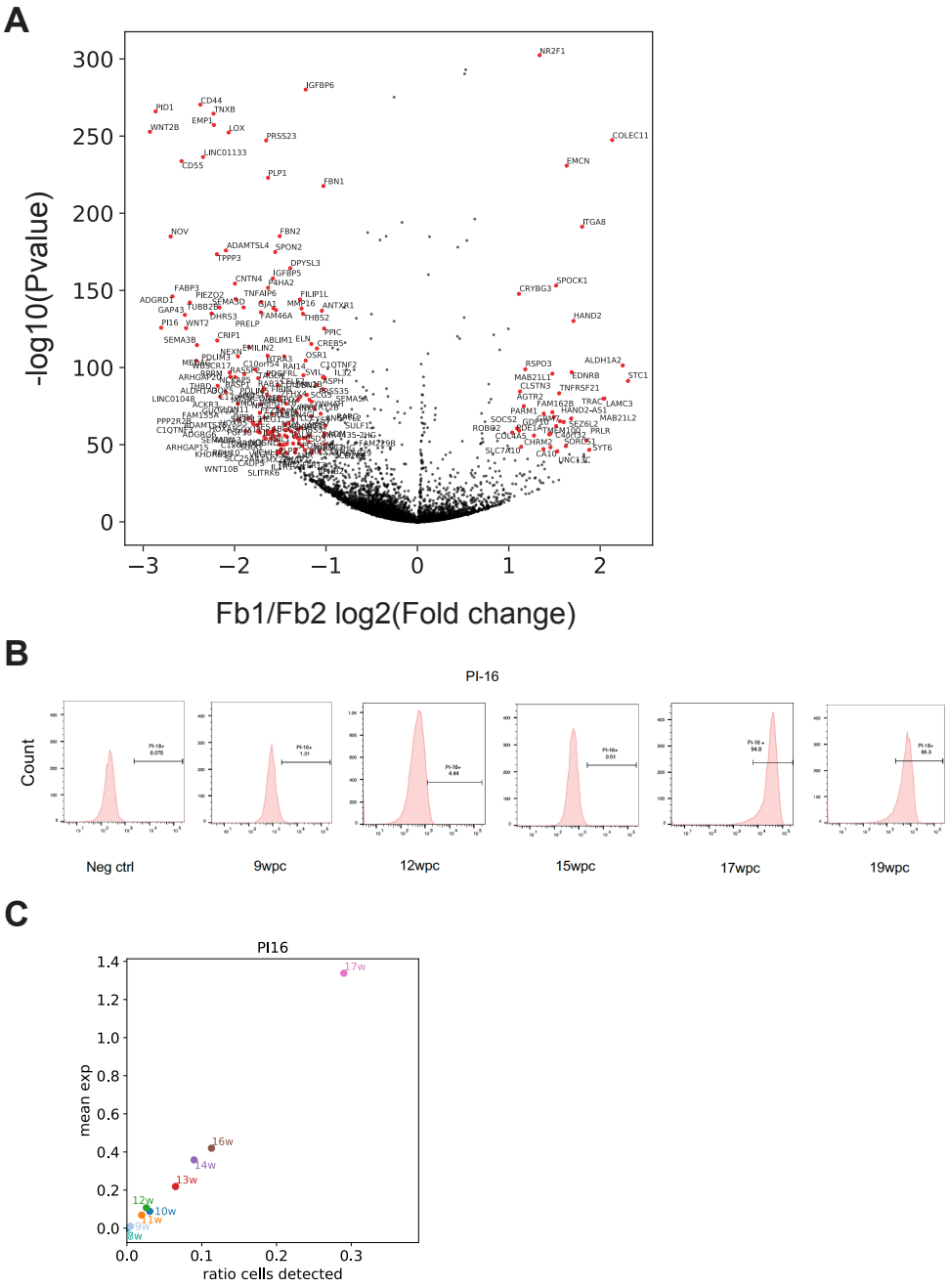


Fig. S3.

(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

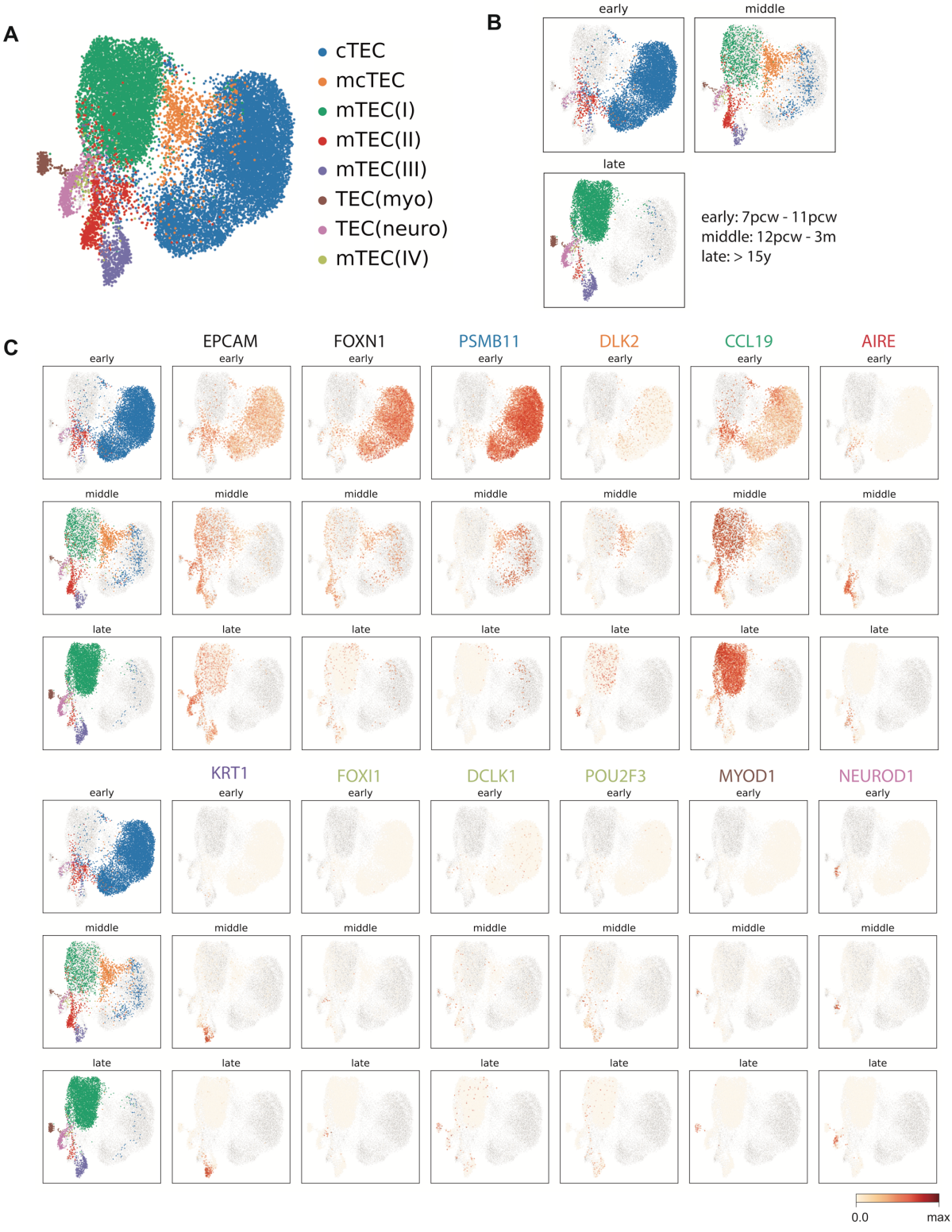

A
cTEC
mcTEC
mTEC(I)
mTEC(II)
mTEC(III)
TEC(myo)
TEC(neuro)
mTEC(IV)
B
early
middle
late
early: 7pcw - 11pcw
middle: 12pcw - 3m
late: > 15y
C
EPCAM
FOXN1
PSMB11
DLK2
CCL19
AIRE
KRT1
FOXI1
DCLK1
POU2F3
MYOD1
NEUROD1
0.0
max

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.

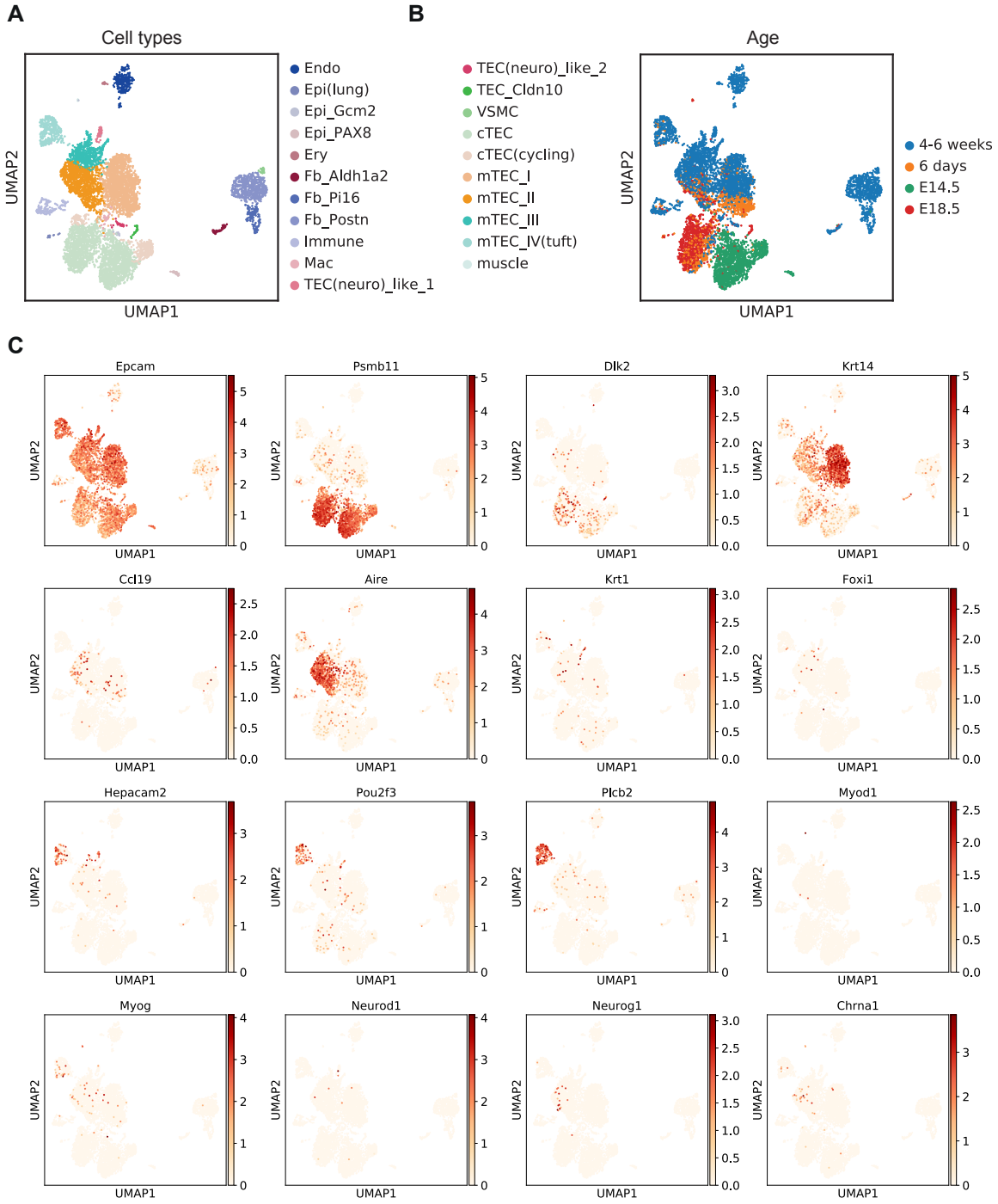


Fig. S5.

(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.

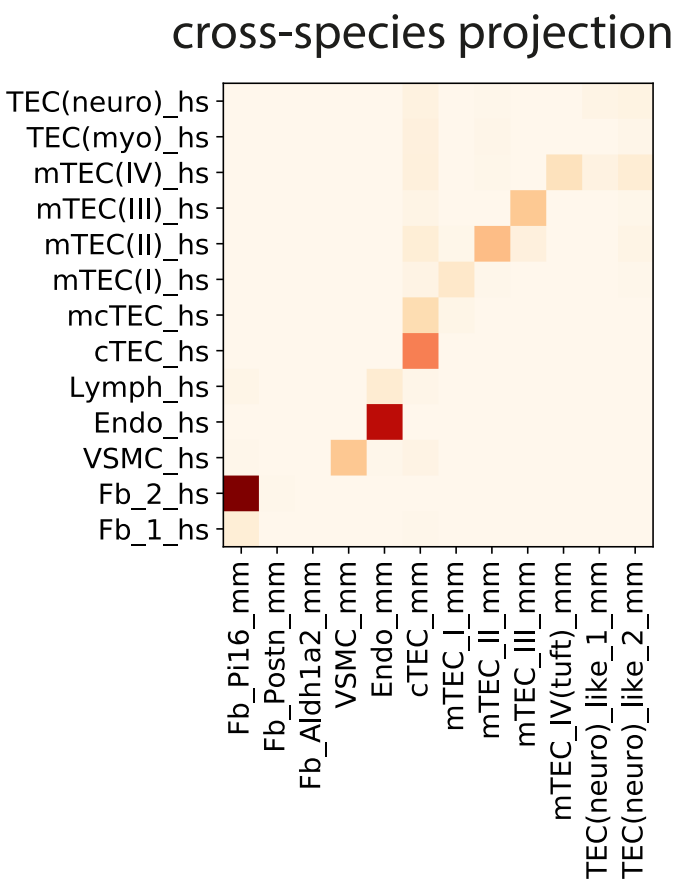


Fig. S6.

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.

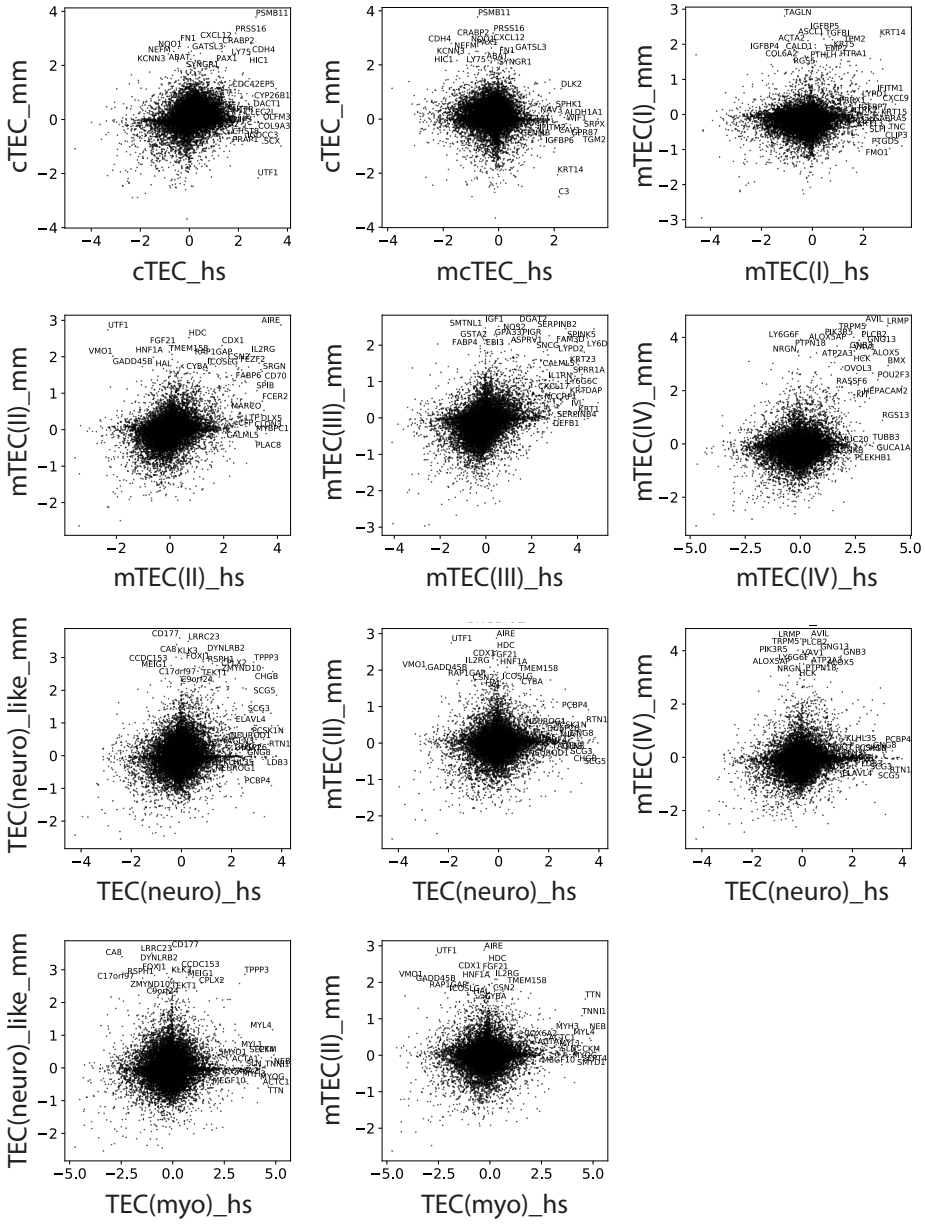


Fig. S7.

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)

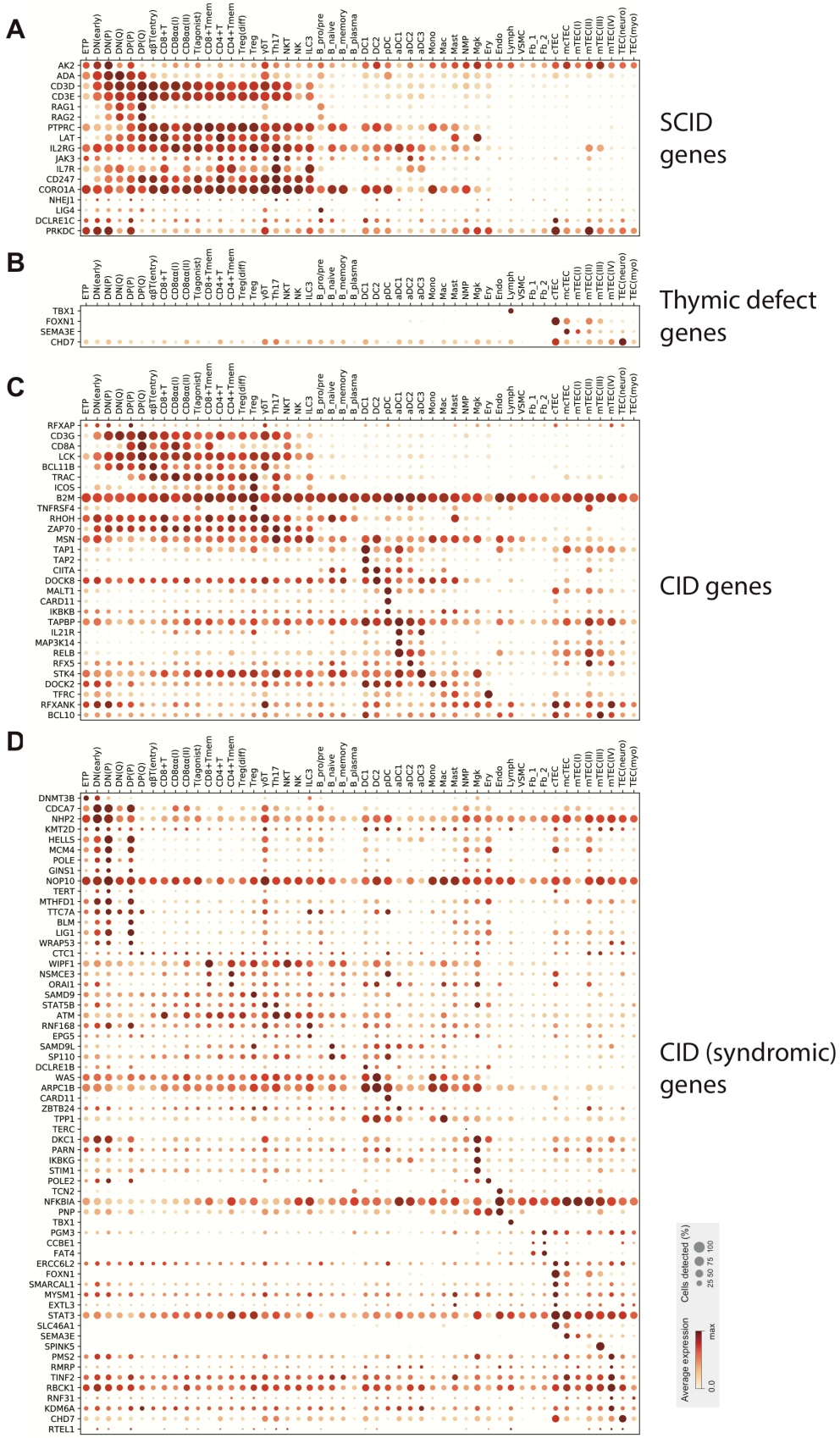

A
ETP
DN(early)
DN(T)
DN(Q)
DP(P)
DP(Q)
αβT(entry)
CD8+T
CD8αα(I)
CD8αα(II)
T(agonist)
CD8+Tmem
CD4+T
CD4+Tmem
Treg(diff)
Treg
γδT
Th17
NKT
NK
ILC3
B_pro/pre
B_naive
B_memory
B_plasma
DC1
DC2
pDC
aDC1
aDC2
aDC3
Mono
Mac
Mast
NMP
Mgk
Ery
Endo
Lymph
VSMC
Fb_1
Fb_2
cTEC
mcTEC
mTEC(I)
mTEC(II)
mTEC(III)
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
TBX1
FOXN1
SEMA3E
CHD7
Thymic defect genes
C
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
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.

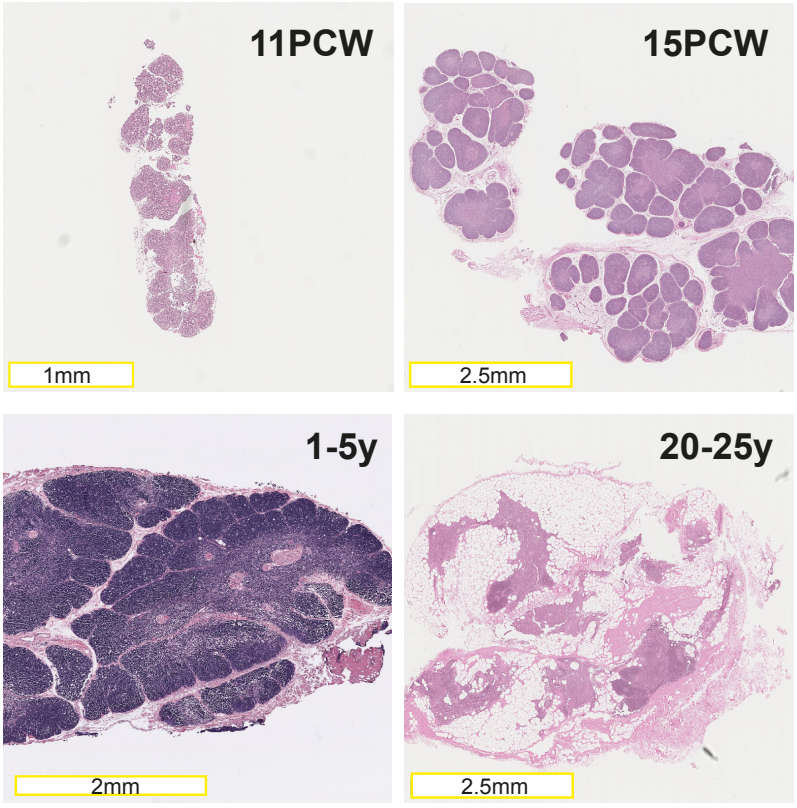


Fig. S9.
H&E staining of cross-sectioned thymic tissue at different developmental and postnatal ages.

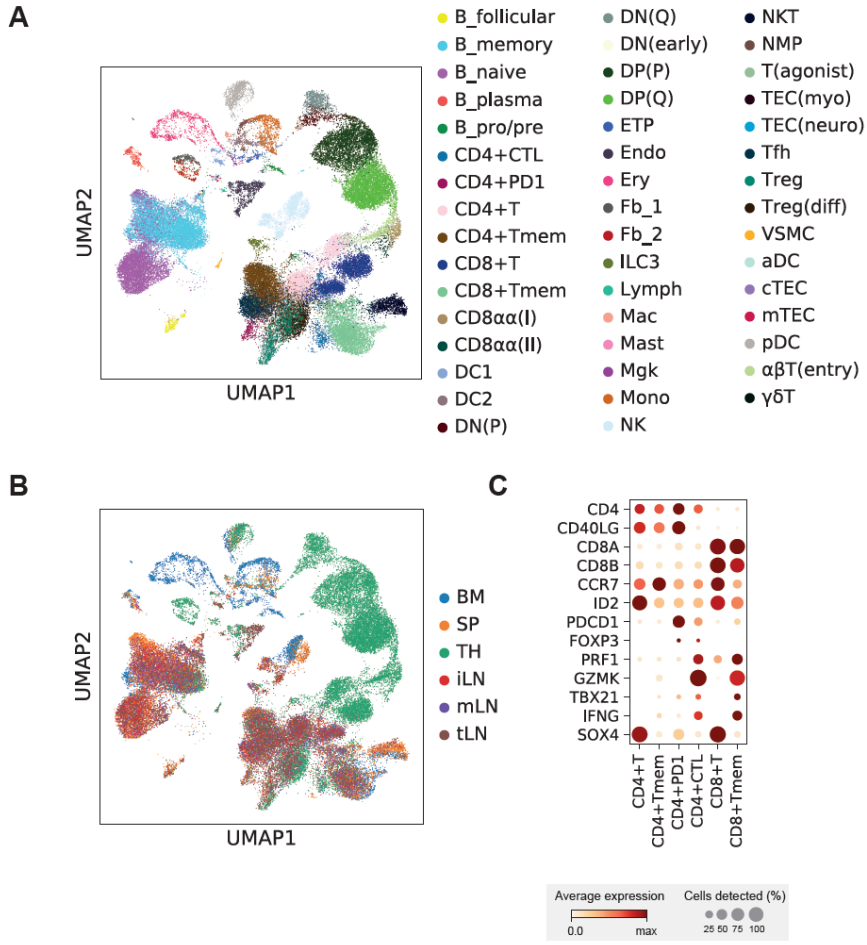


Fig. S10.

(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.

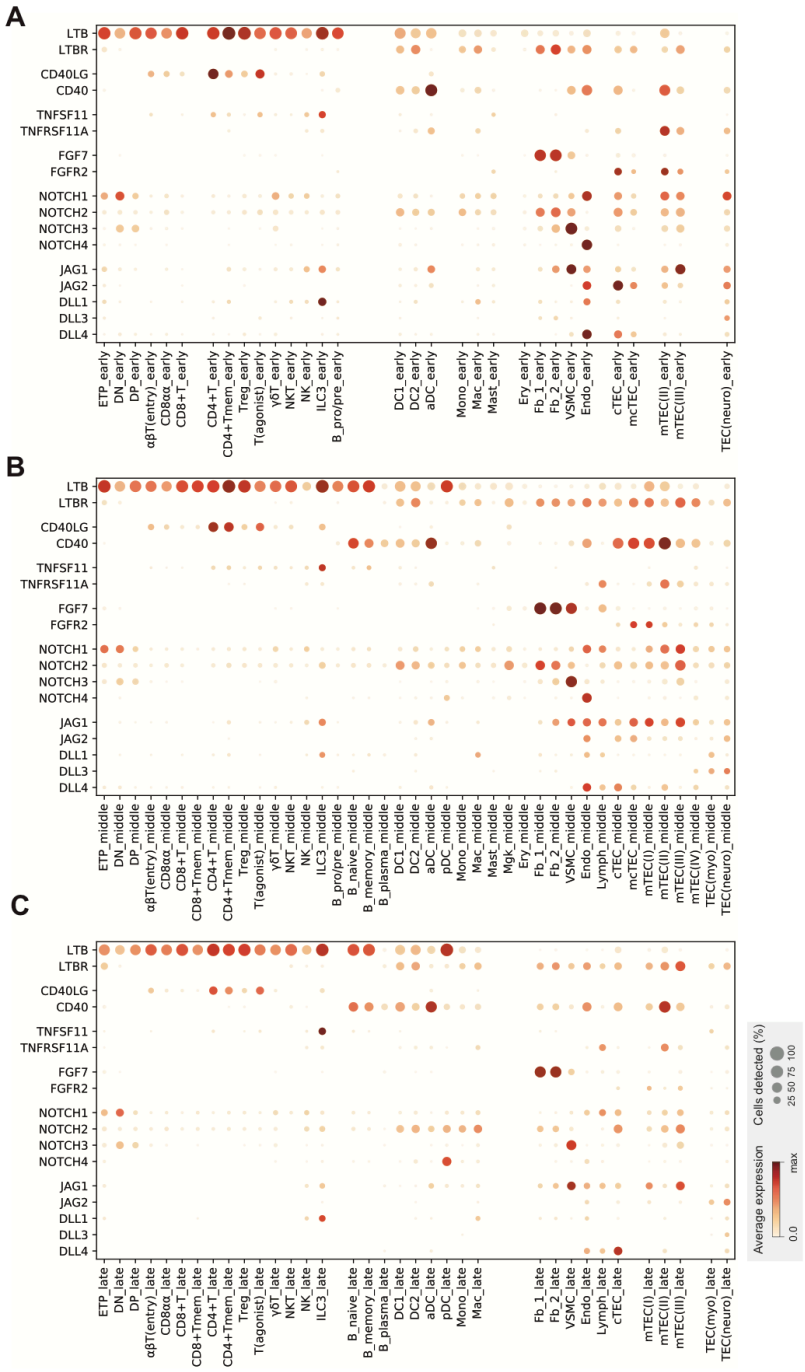


Fig. S11.

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.

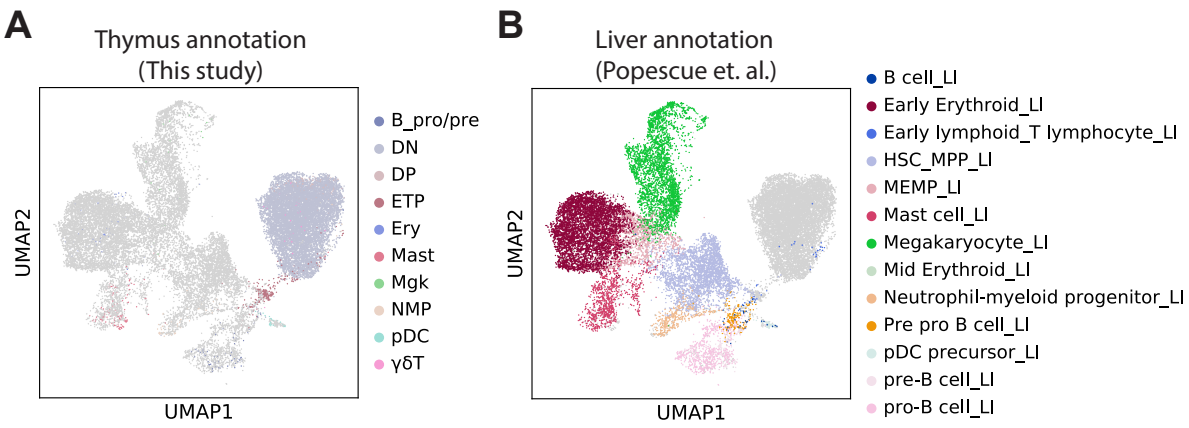


Fig. S12.

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)

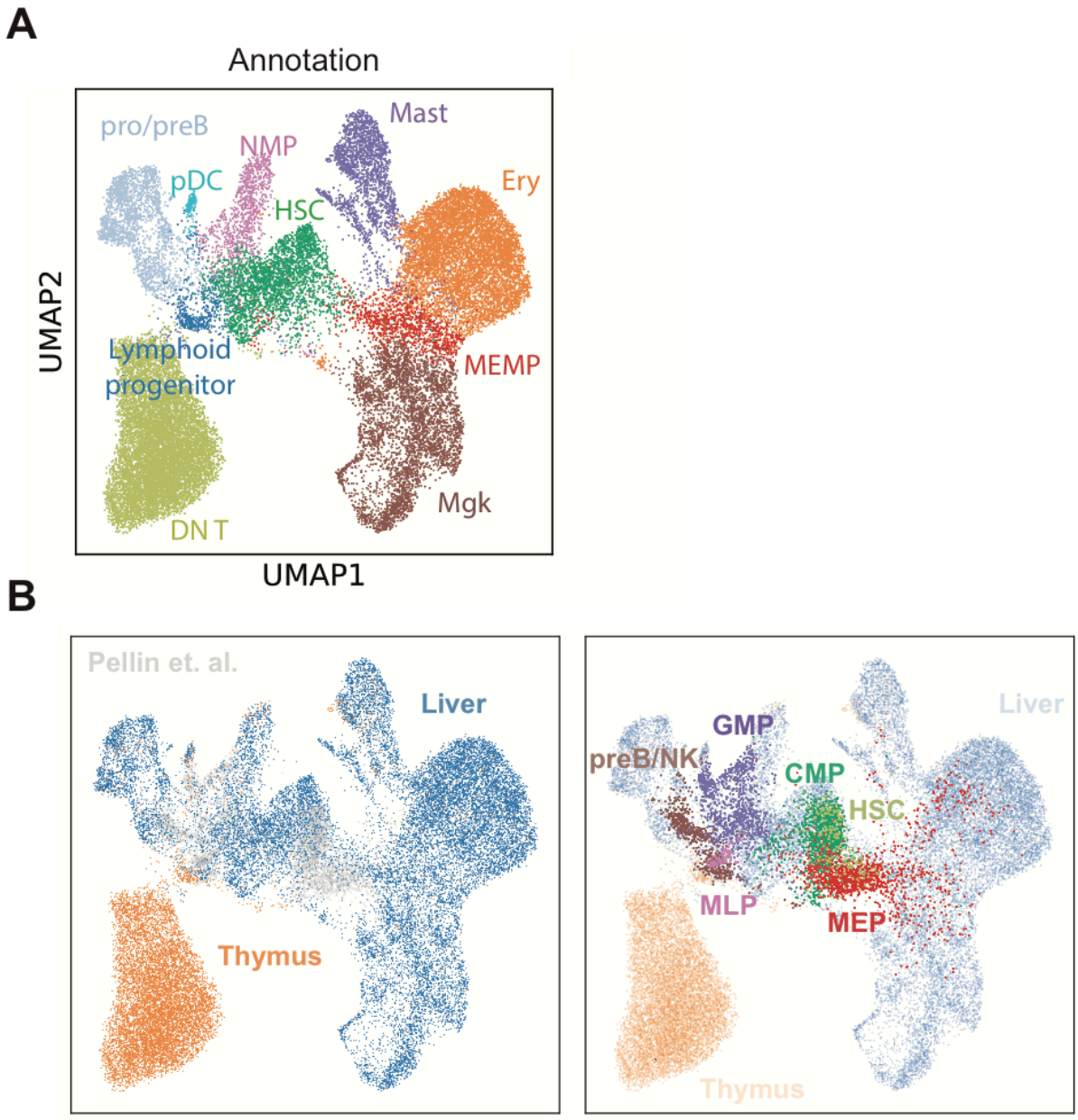


Fig. S13.

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.

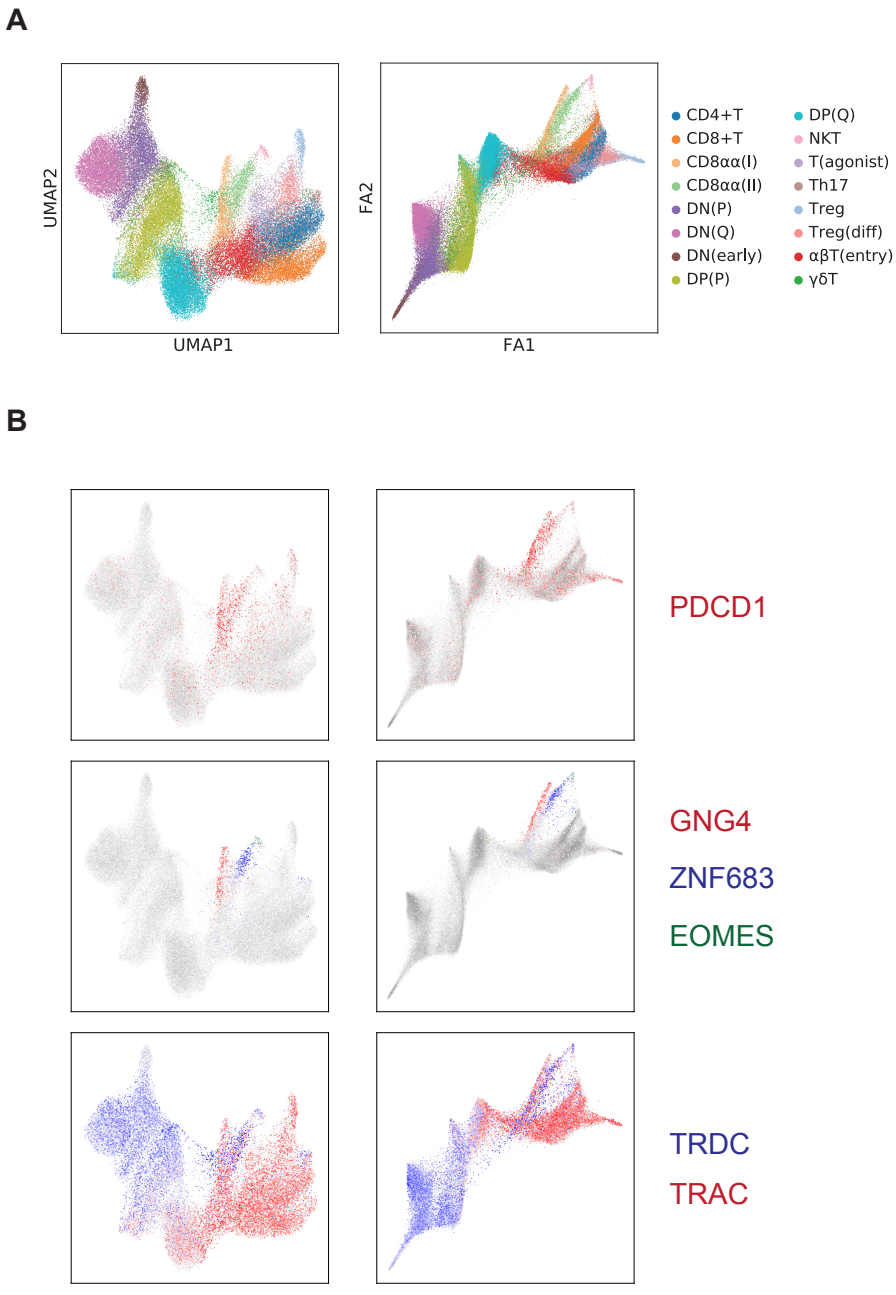


Fig. S14.

(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.

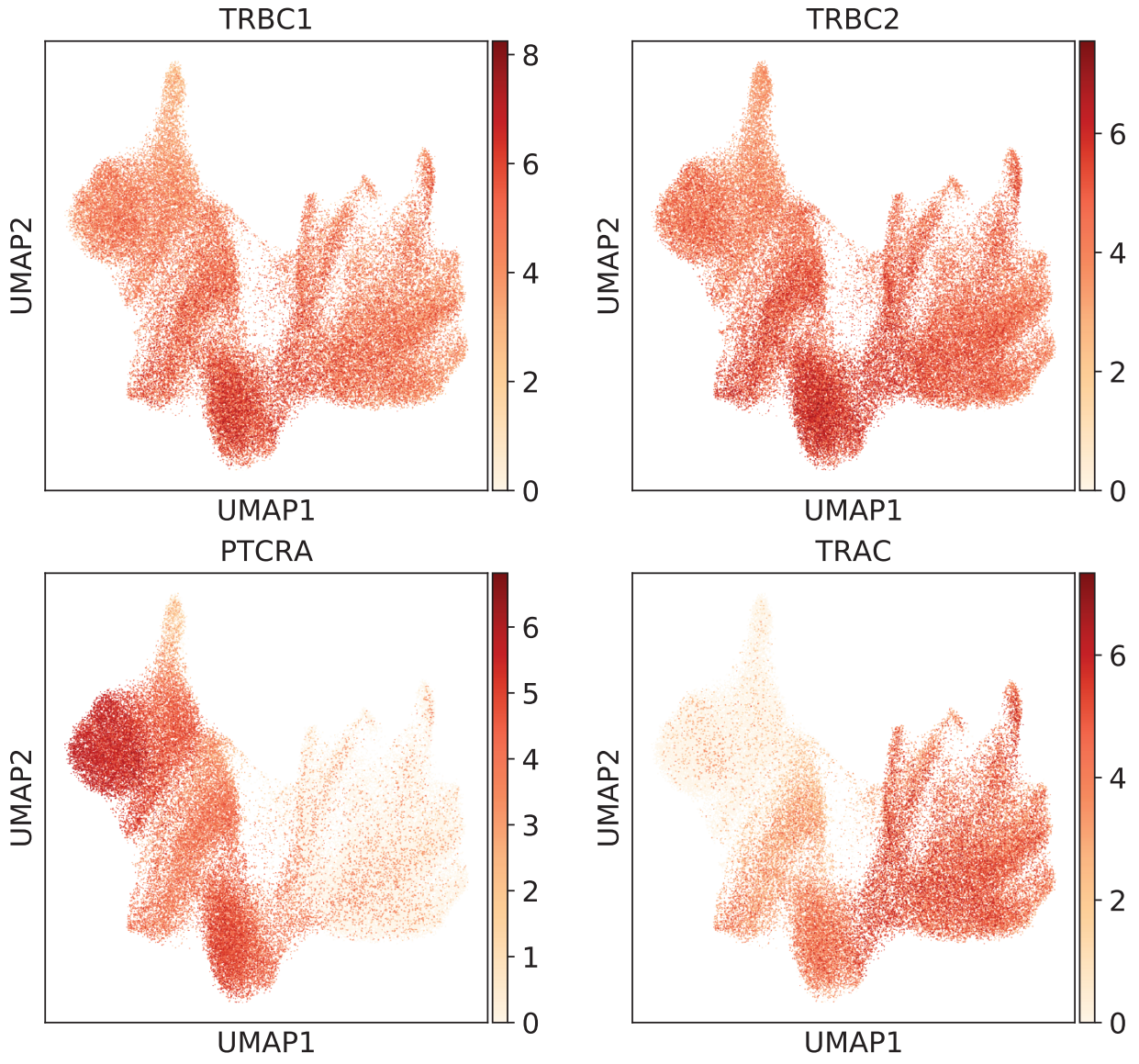


Fig. S15.

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.

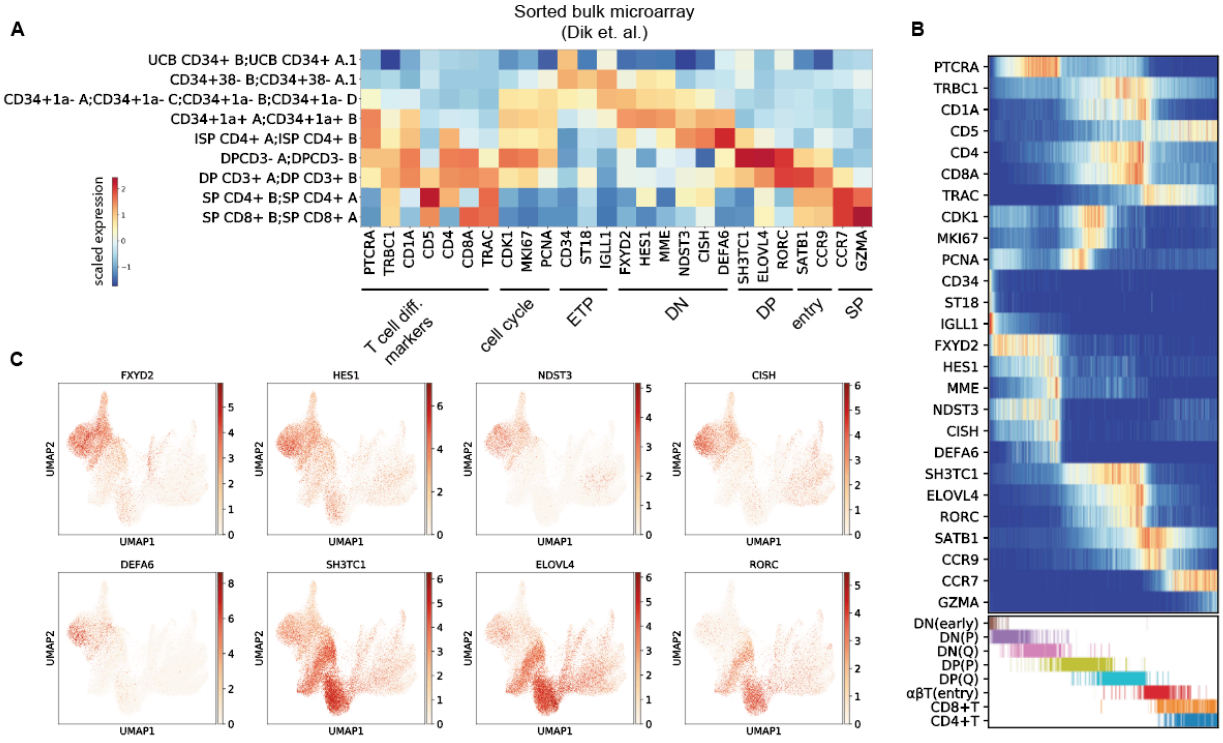


Fig. S16.

(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).

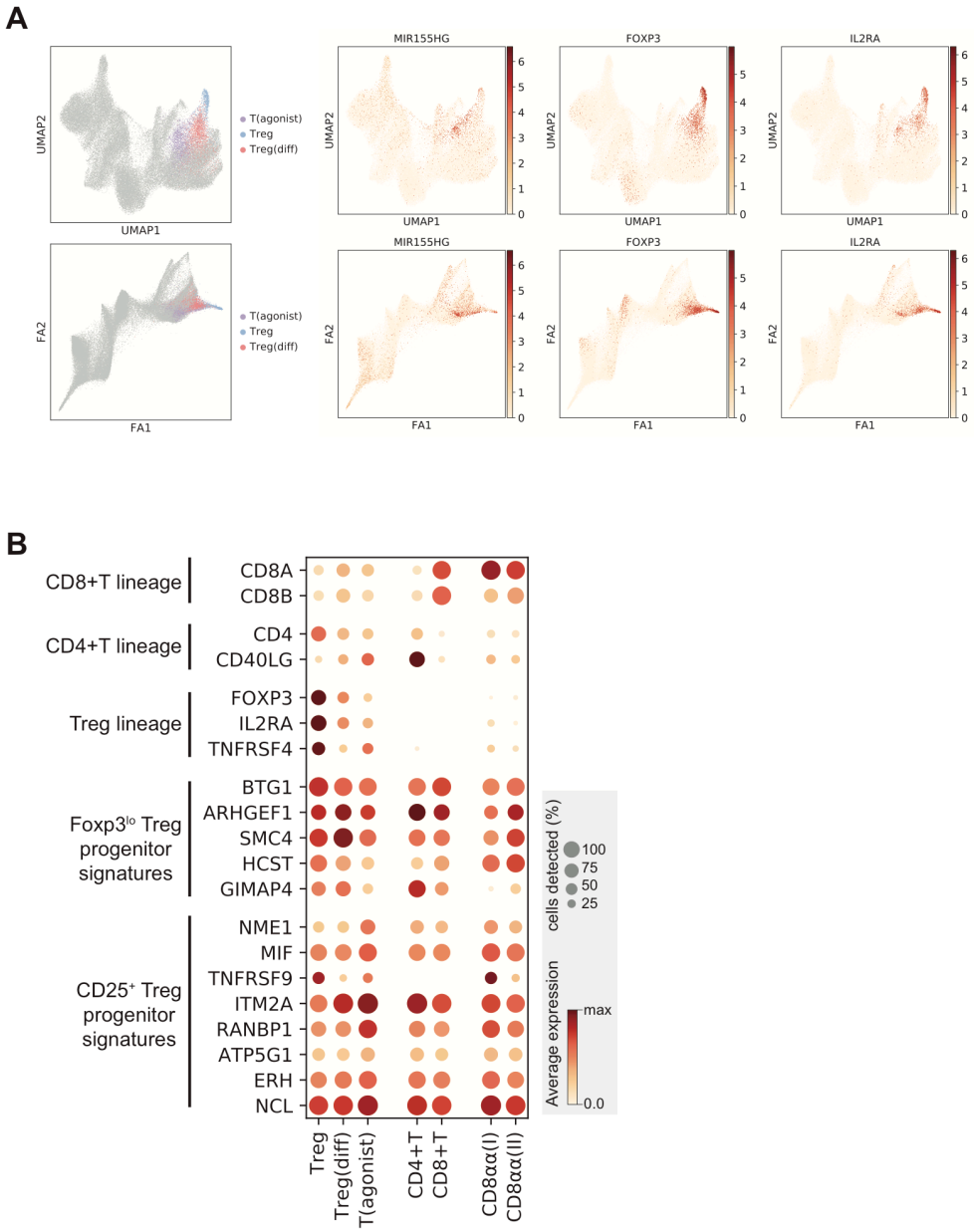


Fig. S17.

(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*).

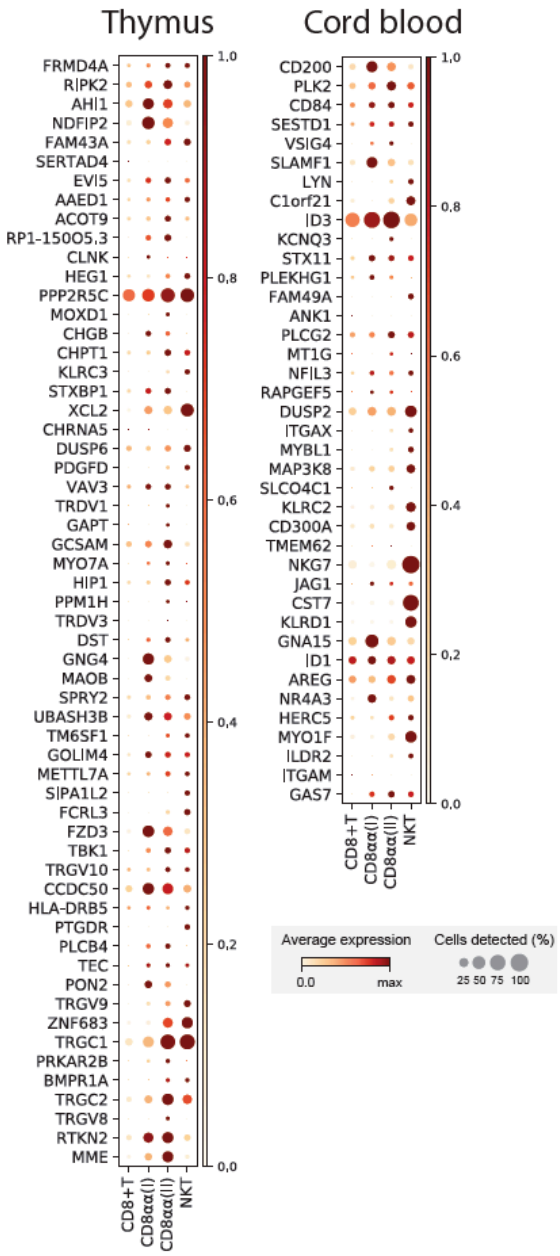


Fig. S18.

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.

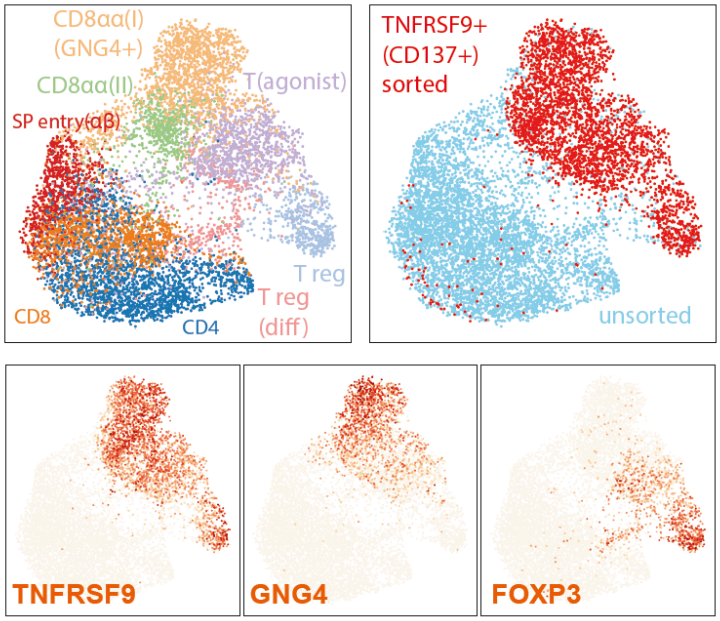


Fig. S19.

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.

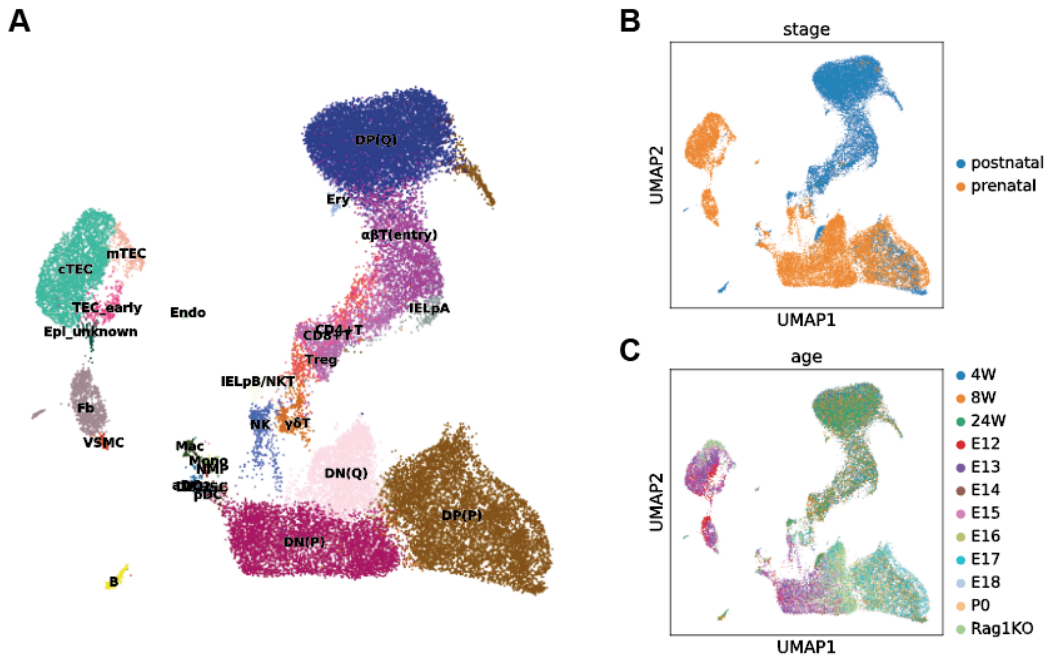


Fig. S20.

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.

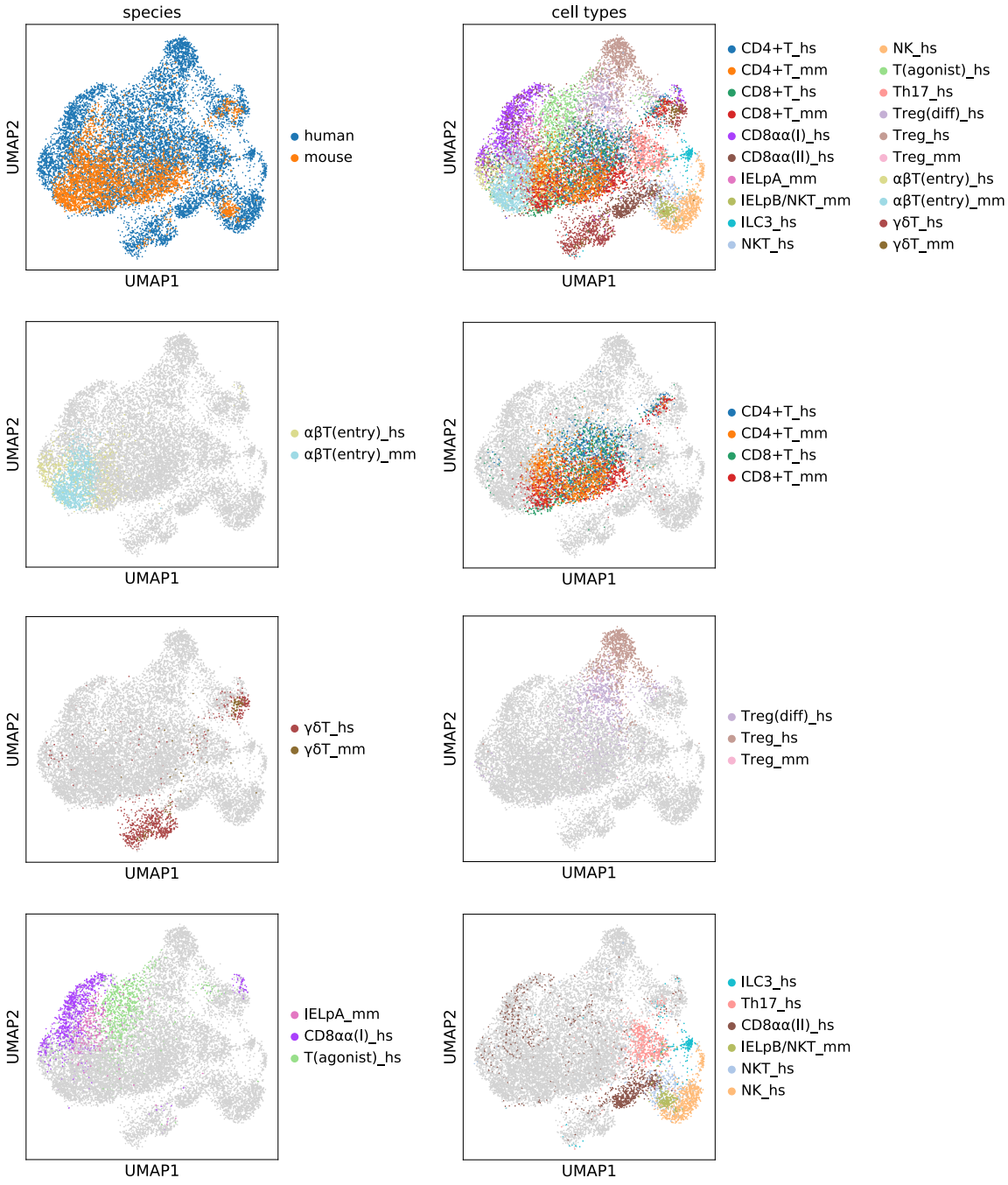
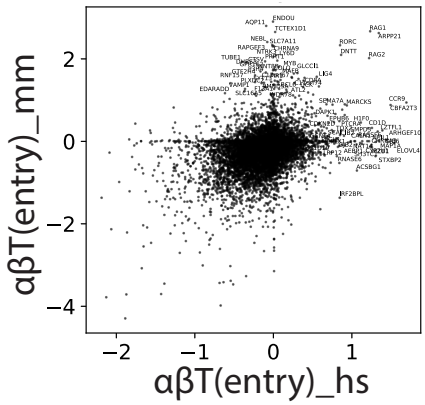
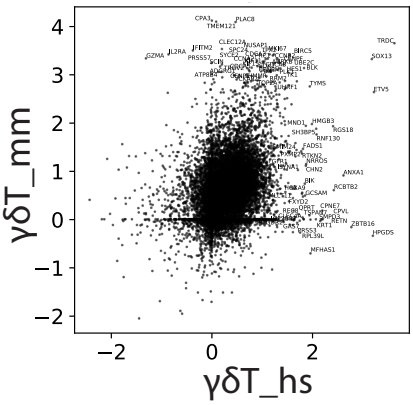
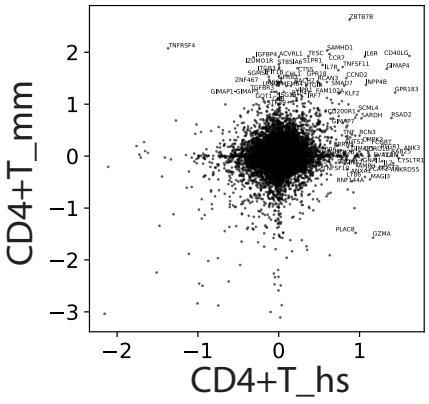


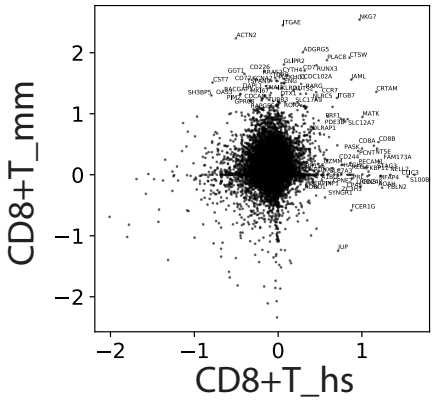
Fig. S21.

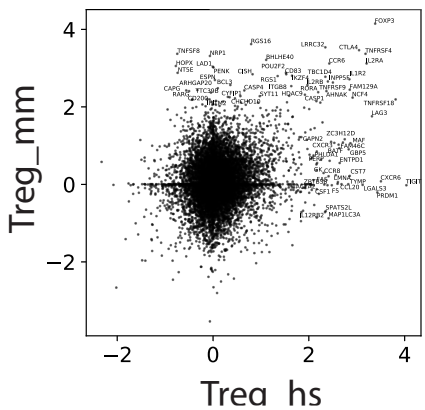
UMAP plot showing the integrative data analysis of mature T cell populations from human and mouse. Human cell types are annotated by ‘hs’ and mouse cell types are marked with ‘mm’. Matching cell groups are shown together.

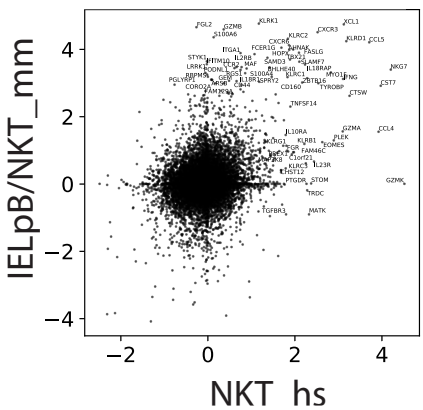

αβT(entry)_mm
αβT(entry)_hs

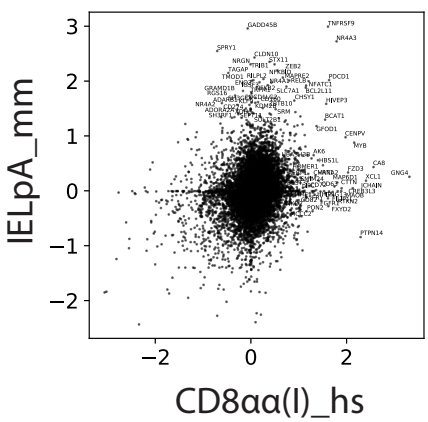

γδT_mm
γδT_hs


CD4+T_mm
CD4+T_hs


CD8+T_mm
CD8+T_hs


Treg_mm
Trea_hs


IELpB/NKT_mm
NKT_hs


IELpA_mm
CD8αα(I)_hs

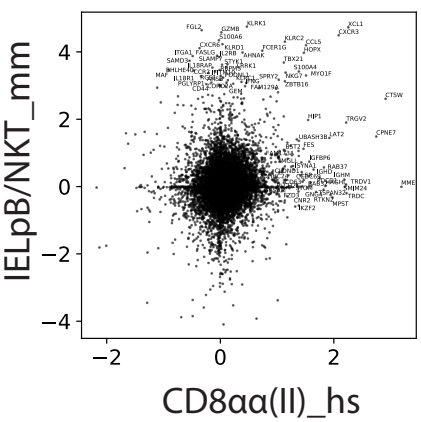

IELpB/NKT_mm
CD8αα(II)_hs

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)

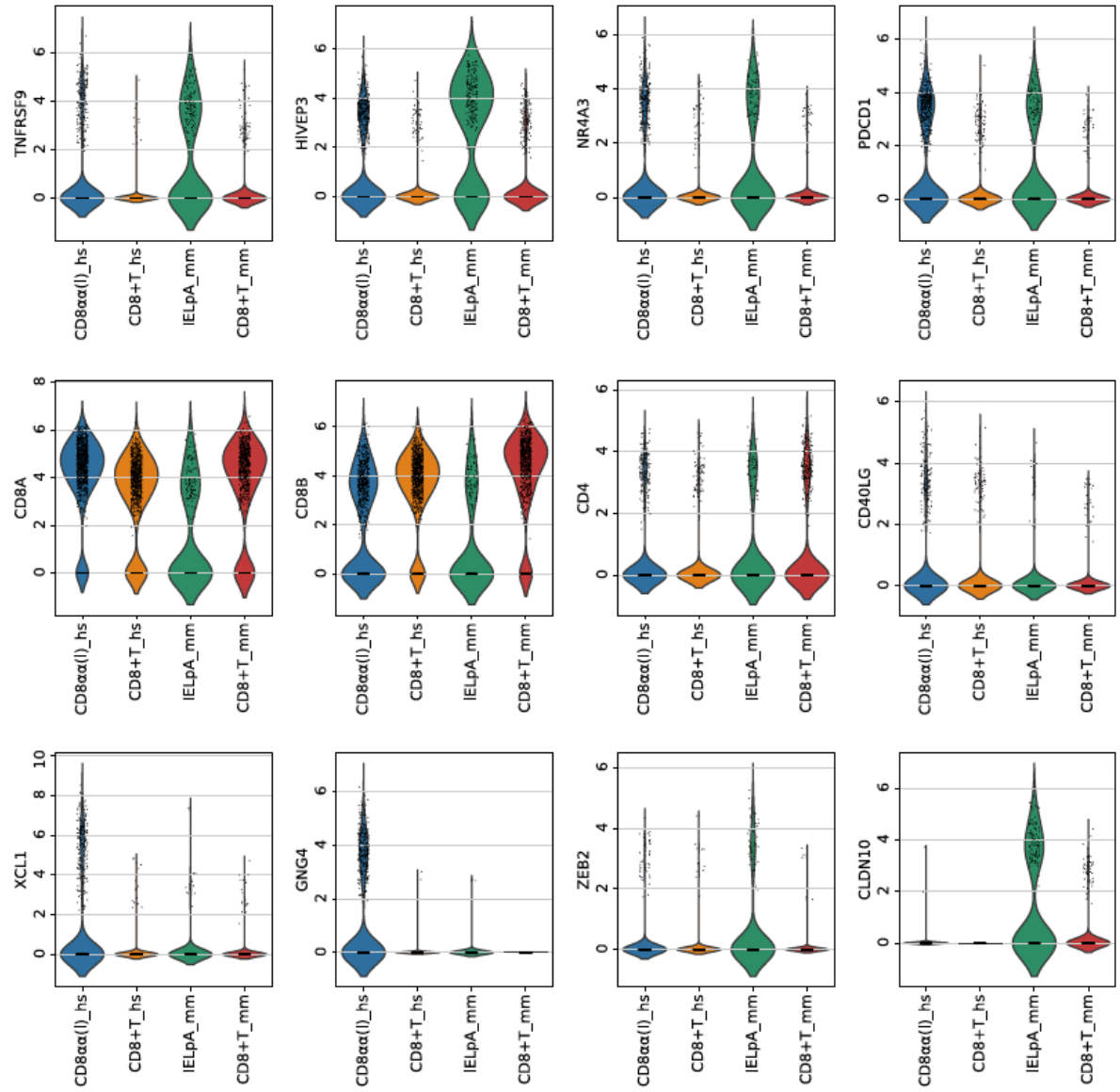


Fig. S23.

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.

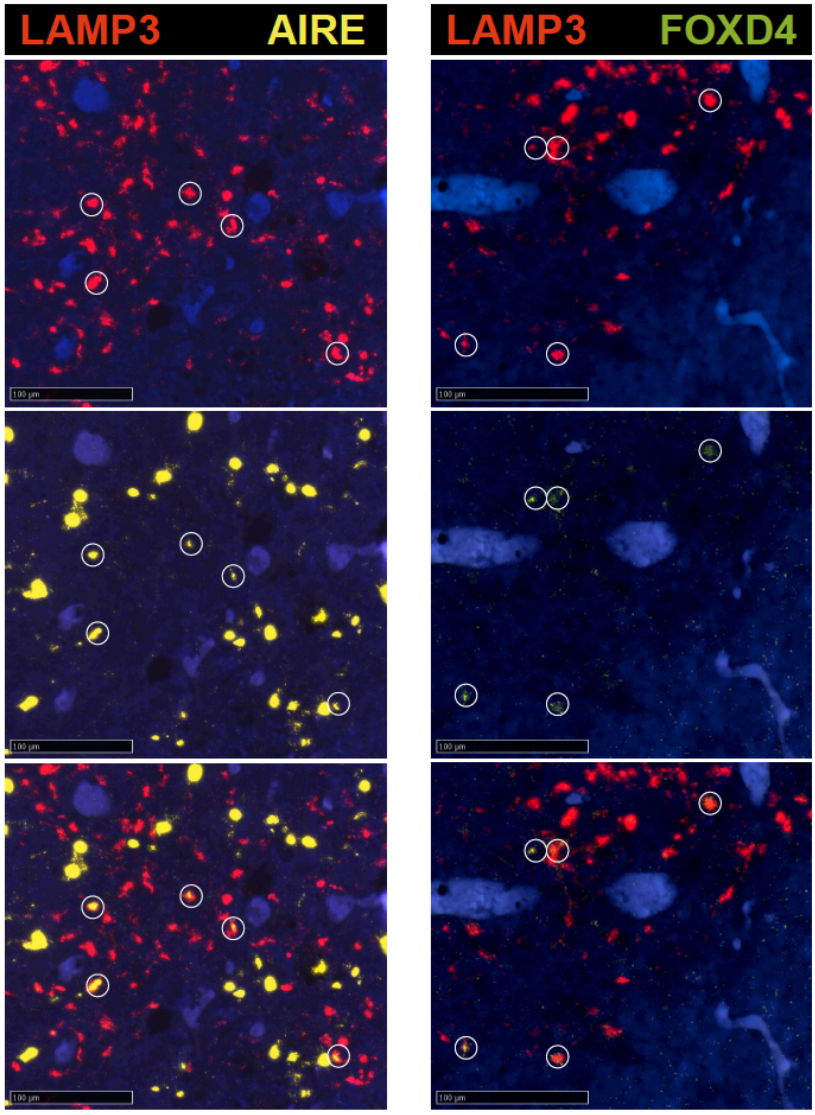


Fig. S24.

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.

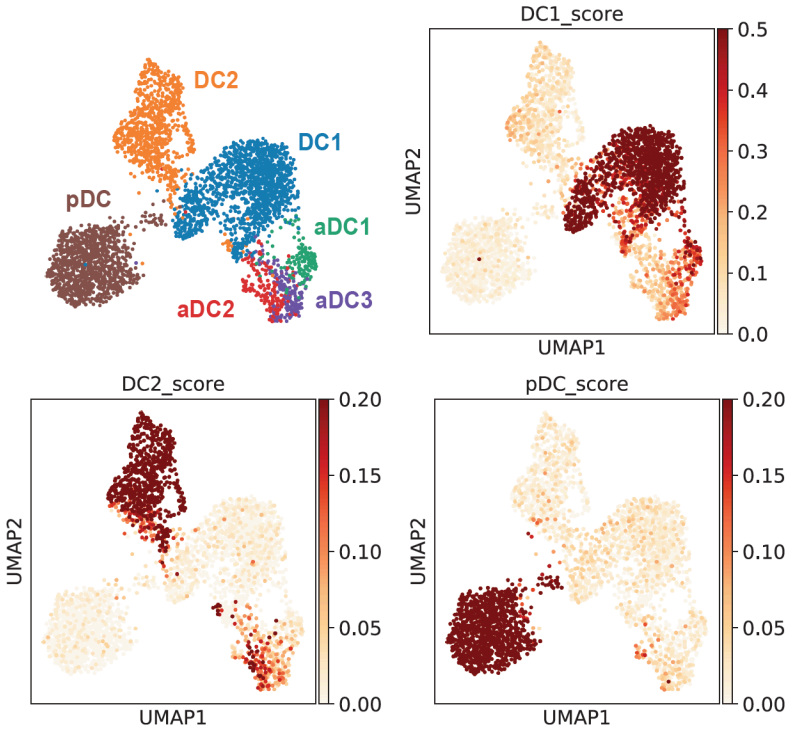


Fig. S25.

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.

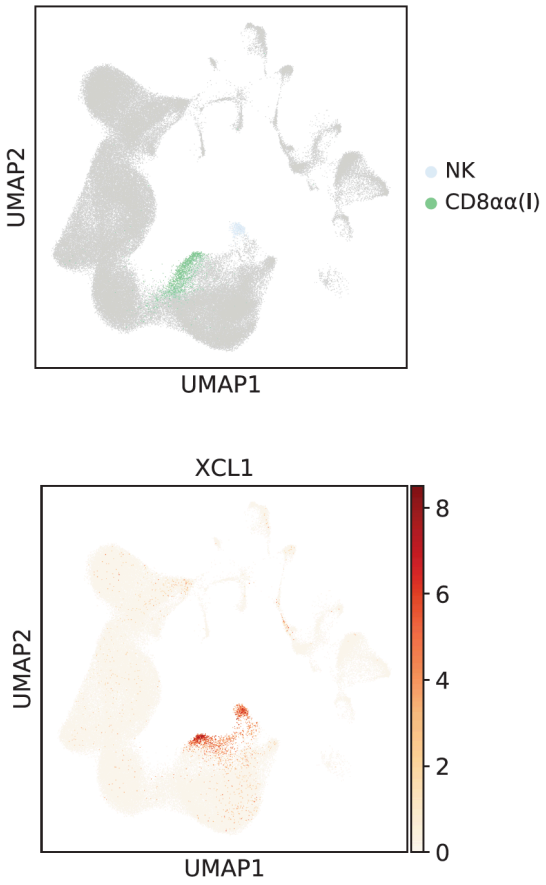
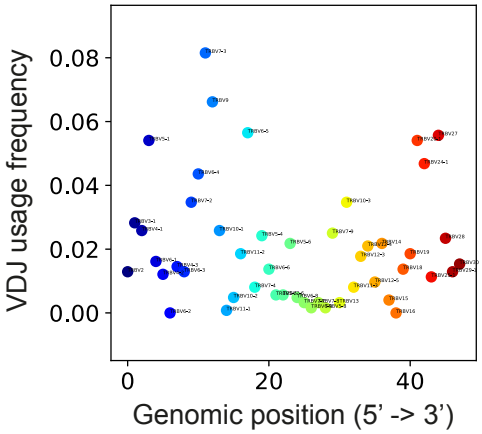
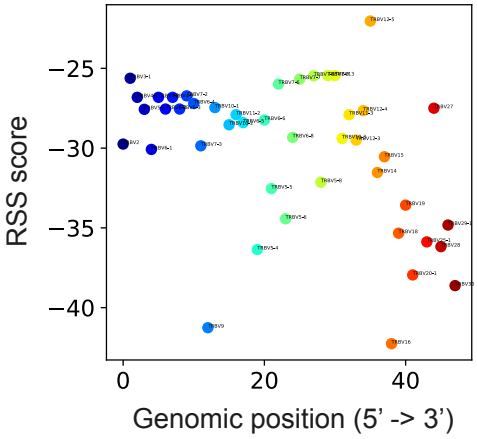


Fig. S26.
UMAP plot showing cell types expressing XCL1 (top) and XCL1 expression level in fetal thymus (bottom)

A


VDJ usage frequency
0.08
0.06
0.04
0.02
0.00
0
20
40
Genomic position (5' -> 3')

B


RSS score
−25
−30
−35
−40
0
20
40
Genomic position (5' -> 3')

C

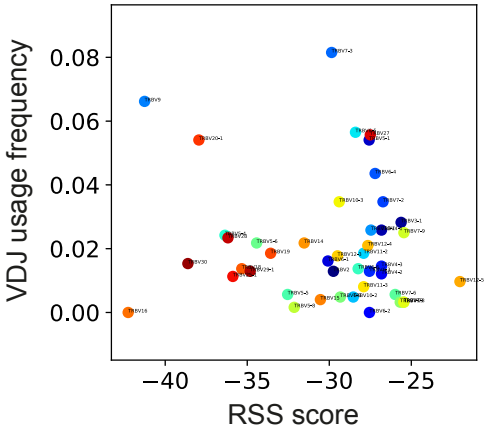

VDJ usage frequency
0.08
0.06
0.04
0.02
0.00
−40
−35
−30
−25
RSS score

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.

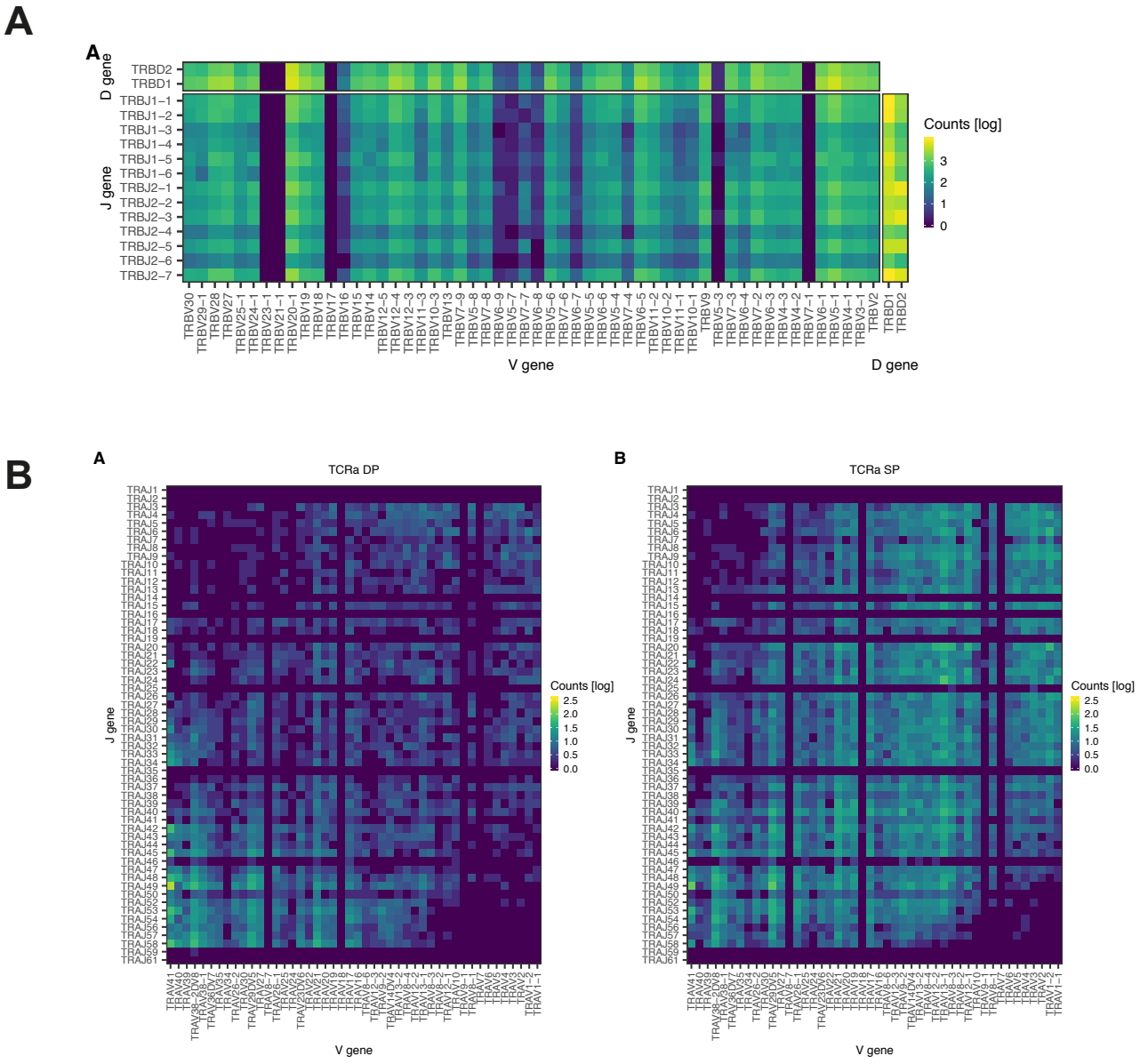


Fig. S28.

(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.

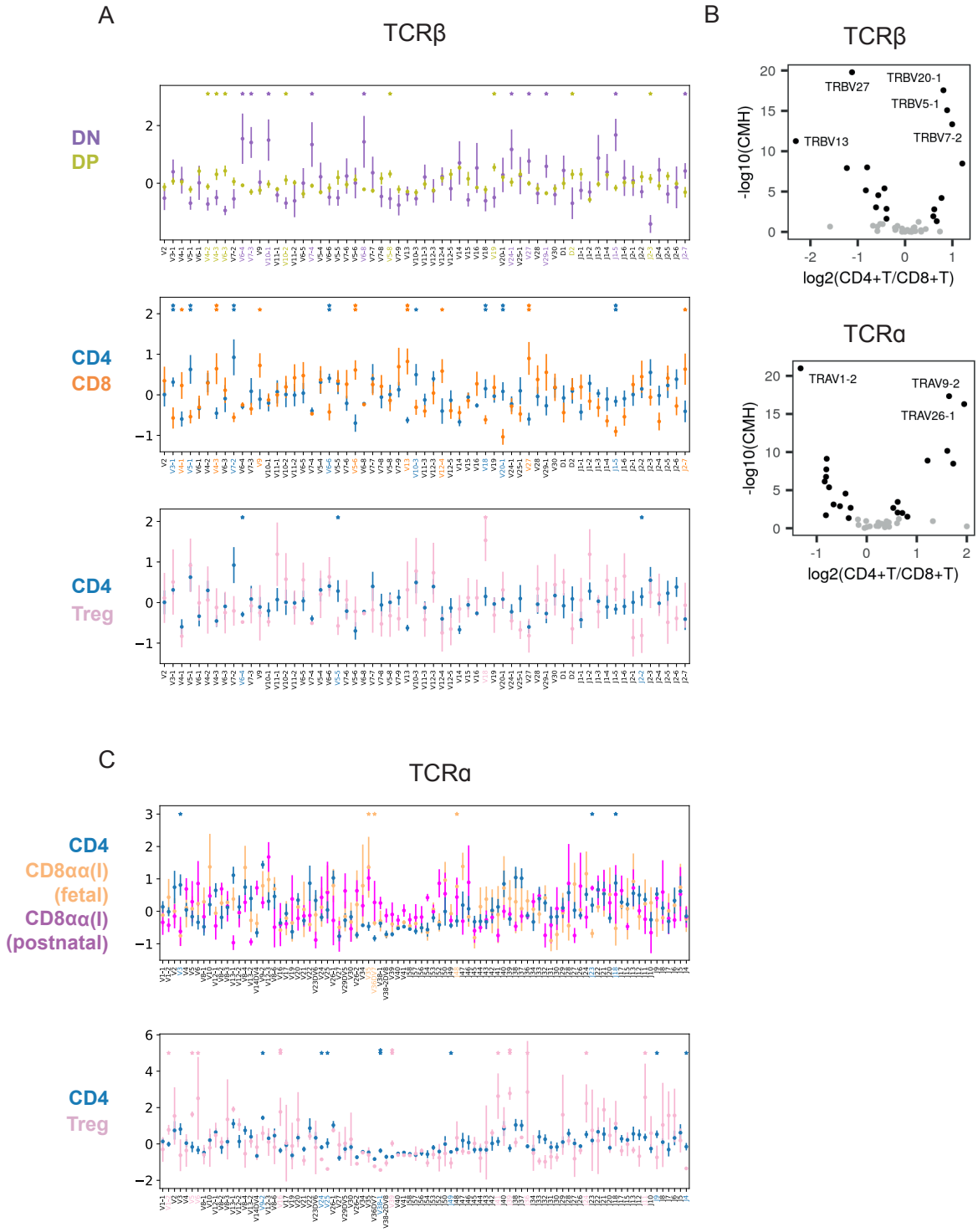

A
TCRβ
DN
DP
CD4
CD8
CD4
Treg
B
TCRβ
TRBV27
TRBV20-1
TRBV5-1
TRBV7-2
TRBV13
-log10(CMH)
log2(CD4+T/CD8+T)
TCRα
TRAV1-2
TRAV9-2
TRAV26-1
-log10(CMH)
log2(CD4+T/CD8+T)
C
TCRα
CD4
CD8αα(I)
(fetal)
CD8αα(I)
(postnatal)
CD4
Treg

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. **(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.

References

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). doi:10.1073/pnas.1714478115 Medline

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). doi:10.1016/j.it.2009.04.003 Medline

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). doi:10.1146/annurev-immunol-020711-075035 Medline

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). doi:10.1084/jem.177.1.19 Medline

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). doi:10.1038/srep14871 Medline

6. L. Wu, K. Shortman, Heterogeneity of thymic dendritic cells. *Semin. Immunol.* **17**, 304–312 (2005). doi:10.1016/j.smim.2005.05.001 Medline

7. G. Anderson, Y. Takahama, Thymic epithelial cells: Working class heroes for T cell development and repertoire selection. *Trends Immunol.* **33**, 256–263 (2012). doi:10.1016/j.it.2012.03.005 Medline

8. Y.-J. Liu, A unified theory of central tolerance in the thymus. *Trends Immunol.* **27**, 215–221 (2006). doi:10.1016/j.it.2006.03.004 Medline

9. J. Gameiro, P. Nagib, L. Verinaud, The thymus microenvironment in regulating thymocyte differentiation. *Cell Adhes. Migr.* **4**, 382–390 (2010). doi:10.4161/cam.4.3.11789 Medline

10. W. E. Jenkinson, S. W. Rossi, S. M. Parnell, E. J. Jenkinson, G. Anderson, PDGFRα-expressing mesenchyme regulates thymus growth and the availability of intrathymic niches. *Blood* **109**, 954–960 (2007). doi:10.1182/blood-2006-05-023143 Medline

11. S. Inglesfield, E. J. Cosway, W. E. Jenkinson, G. Anderson, Rethinking Thymic Tolerance: Lessons from Mice. *Trends Immunol.* **40**, 279–291 (2019). doi:10.1016/j.it.2019.01.011 Medline

12. J. F. A. P. Miller, The golden anniversary of the thymus. *Nat. Rev. Immunol.* **11**, 489–495 (2011). doi:10.1038/nri2993 Medline

13. J. Gordon, N. R. Manley, Mechanisms of thymus organogenesis and morphogenesis. *Development* **138**, 3865–3878 (2011). doi:10.1242/dev.059998 Medline

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). doi:10.1016/j.immuni.2018.04.015 Medline

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). doi:10.1038/s41586-018-0346-1 Medline

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). doi:10.1038/s41586-018-0345-2 Medline

17. J. Mestas, C. C. W. Hughes, Of mice and not men: Differences between mouse and human immunology. *J. Immunol.* **172**, 2731–2738 (2004). doi:10.4049/jimmunol.172.5.2731 Medline

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). doi:10.1242/dev.087320 Medline

19. B. V. Kumar, T. J. Connors, D. L. Farber, Human T Cell Development, Localization, and Function throughout Life. *Immunity* **48**, 202–213 (2018). doi:10.1016/j.immuni.2018.01.007 Medline

20. J. Nikolich-Žugich, M. K. Slifka, I. Messaoudi, The many important facets of T cell repertoire diversity. *Nat. Rev. Immunol.* **4**, 123–132 (2004). doi:10.1038/nri1292 Medline

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

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 293852 [preprint]. 3 April 2018.

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). doi:10.1371/journal.pone.0140815 Medline

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). doi:10.1101/gr.115428.110 Medline

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–E3215 (2013). doi:10.1073/pnas.1304048110 Medline

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* btz625 (2019). doi:10.1093/bioinformatics/btz625 Medline

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). doi:10.4049/jimmunol.1002185 Medline

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). doi:10.4049/jimmunol.1200358 Medline

29. C. Gu, E. Giraudo, The role of semaphorins and their receptors in vascular development and cancer. *Exp. Cell Res.* **319**, 1306–1316 (2013). doi:10.1016/j.yexcr.2013.02.003 Medline

30. H. J. Garchon, F. Djabiri, J. P. Viard, P. Gajdos, J. F. Bach, Involvement of human muscle acetylcholine receptor α-subunit gene (CHRNA) in susceptibility to myasthenia gravis. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4668–4672 (1994). doi:10.1073/pnas.91.11.4668 Medline

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). doi:10.1016/j.jneuroim.2003.11.013 Medline

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). Medline

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**, eaan8664 (2018). doi:10.1126/sciimmunol.aan8664 Medline

34. N. Lopes, A. Sergé, P. Ferrier, M. Irla, Thymic Crosstalk Coordinates Medulla Organization and T cell Tolerance Induction. *Front. Immunol.* **6**, 365 (2015). doi:10.3389/fimmu.2015.00365 Medline

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). doi:10.1038/s41586-018-0698-6 Medline

36. D. Elewaut, C. F. Ware, The unconventional role of LTαβ in T cell differentiation. *Trends Immunol.* **28**, 169–175 (2007). doi:10.1016/j.it.2007.02.005 Medline

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). doi:10.1084/jem.20062497 Medline

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). doi:10.4049/jimmunol.167.4.1954 Medline

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**, dev165597 (2018). doi:10.1242/dev.165597 Medline

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). doi:10.4049/jimmunol.1201815 Medline

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). doi:10.1084/jem.20171000 Medline

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). doi:10.1182/blood-2010-06-290049 Medline

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). doi:10.1084/jem.20121798 Medline

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). doi:10.1038/s41586-019-1652-y Medline

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). doi:10.1016/j.immuni.2019.09.008 Medline

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). doi:10.1038/s41467-019-10291-0 Medline

47. D. K. Shah, J. C. Zúñiga-Pflücker, An overview of the intrathymic intricacies of T cell development. *J. Immunol.* **192**, 4017–4023 (2014). doi:10.4049/jimmunol.1302259 Medline

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). doi:10.4049/jimmunol.165.6.3094 Medline

49. M. R. Tourigny, S. Mazel, D. B. Burtrum, H. T. Petrie, T cell receptor (TCR)-β gene recombination: Dissociation from cell cycle regulation and developmental progression

during T cell ontogeny. *J. Exp. Med.* **185**, 1549–1556 (1997). doi:10.1084/jem.185.9.1549 Medline

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). doi:10.1084/jem.20042524 Medline

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). doi:10.1016/j.cell.2018.10.022 Medline

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). doi:10.1038/s41590-018-0289-6 Medline

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**, eaah4232 (2017). doi:10.1126/sciimmunol.aah4232 Medline

54. J. R. Fergusson, V. M. Fleming, P. Klenerman, CD161-expressing human T cells. *Front. Immunol.* **2**, 36 (2011). doi:10.3389/fimmu.2011.00036 Medline

55. E. S. Alonzo, D. B. Sant'Angelo, Development of PLZF-expressing innate T cells. *Curr. Opin. Immunol.* **23**, 220–227 (2011). doi:10.1016/j.coi.2010.12.016 Medline

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). doi:10.1038/ni.3751 Medline

57. J. Oh, J.-S. Shin, The Role of Dendritic Cells in Central Tolerance. *Immune Netw.* **15**, 111–120 (2015). doi:10.4110/in.2015.15.3.111 Medline

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**, eaah4573 (2017). doi:10.1126/science.aah4573 Medline

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). doi:10.1038/nature03886 Medline

60. P. J. Fairchild, J. M. Austyn, Thymic dendritic cells: Phenotype and function. *Int. Rev. Immunol.* **6**, 187–196 (1990). doi:10.3109/08830189009056629 Medline

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 (2019). doi:10.3389/fimmu.2018.02902 Medline

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). doi:10.3389/fimmu.2015.00398 Medline

63. Y. Lei, A. M. Ripen, N. Ishimaru, I. Ohigashi, T. Nagasawa, L. T. Jeker, M. R. Bösl, G. A. Holländer, Y. Hayashi, Rde. 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). doi:10.1084/jem.20102327 Medline

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). doi:10.1038/ni1448 Medline

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). doi:10.1073/pnas.1504971112 Medline

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). doi:10.1016/j.celrep.2017.05.045 Medline

67. P. Bullen, D. I. Wilson, in *Molecular Genetics of Early Human Development*, T. Strachan, S. Lindsay, D. I. Wilson, Eds. (BIOS, 1997), pp. 27–35.

68. W. M. Hern, Correlation of fetal age and measurements between 10 and 26 weeks of gestation. *Obstet. Gynecol.* **63**, 26–32 (1984). Medline

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). doi:10.1016/j.cels.2018.11.005 Medline

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). doi:10.1038/nmeth.3971 Medline

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 397042 [preprint]. 22 August 2018.

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). doi:10.1093/nar/gky822 Medline