Single-cell immune profiling reveals thymus-seeding populations, T cell commitment, and multilineage development in the human thymus

Cordes, M.; Cante-Barrett, K.; Akker, E.B. van den; Moretti, F.A.; Kielbasa, S.M.; Vloemans, S.A.; ... ; Staal, F.J.T.

T CELLS

Single-cell immune profiling reveals thymus-seeding populations, T cell commitment, and multilineage development in the human thymus

Martijn Cordes[1,2]†, Kirsten Canté-Barrett[1,3]†, Erik B. van den Akker[2,4,5], Federico A. Moretti[1], Szymon M. Kiełbasa[6], Sandra A. Vloemans[1], Laura Garcia-Perez[1], Cristina Teodosio[1,7], Jacques J. M. van Dongen[1,7], Karin Pike-Overzet[1], Marcel J. T. Reinders[2,4], Frank J. T. Staal[1,3,8]*

T cell development in the mouse thymus has been studied extensively, but less is known regarding T cell development in the human thymus. We used a combination of single-cell techniques and functional assays to perform deep immune profiling of human T cell development, focusing on the initial stages of prelineage commitment. We identified three thymus-seeding progenitor populations that also have counterparts in the bone marrow. In addition, we found that the human thymus physiologically supports the development of monocytes, dendritic cells, and NK cells, as well as limited development of B cells. These results are an important step toward monitoring and guiding regenerative therapies in patients after hematopoietic stem cell transplantation.

INTRODUCTION

T cell development occurs in the thymus, whereas all other blood cell lineages develop primarily in the bone marrow (BM) and/or spleen. The thymus consists of a mixture of thymic stromal cells and developing thymocytes, and cross-talk between these cell types provides the appropriate developmental signals needed for mature T cells to develop. T cell development follows a series of distinct phenotypic stages characterized by the expression of specific surface receptors, notably CD4 and CD8. In both humans and mice, thymocyte development proceeds through successive stages, starting with $CD4^-CD8^-$ double-negative (DN) cells, followed by $CD4^+CD8^+$ double-positive (DP) cells and ending with $CD4^{+-}CD8^-CD3^+$ or $CD8^+CD4^-CD3^+$ single-positive (SP) cells. The DN subset can be further subdivided into four consecutive stages in mice (DN1 to DN4) and three consecutive stages in humans (DN1 to DN3). In addition to differences in surface marker expression, mouse and human DN subpopulations differ with respect to the rearrangement status of their T cell receptor (TCR) genes and their dependence on cytokines (*1*, *2*). In humans, the three DN stages are traditionally identified on the basis of their expression profile consisting of the stem cell markers CD34, CD38, and CD1a (*3*). CD34 is a marker for hematopoietic stem cells (HSCs) and progenitors of all hematopoietic lineages. In the thymus, the most immature cells express CD34, and this expression declines as the cells mature. Concomitant with this decrease in CD34 expression, thymocytes first express CD38, followed by CD1a, roughly coinciding with full αβ T cell commitment (*4*). Thus, the developing human DN compartment is typically subdivided into the $CD34^{+-}CD38^-CD1a^-$ (DN1) stage (representing the most immature thymocyte subset) followed by the $CD34^+CD38^+CD1a^-$ (DN2) and $CD34^+CD38^+CD1a^+$ (DN3) stages.

Using DNA microarrays, we previously found that the abovementioned stages in human T cell development coincide with precisely regulated patterns of TCR gene rearrangement and up-regulation of T cell lineage–specific genes (*1*, *2*). In addition, our group and others showed that the $CD34^+CD1a^-$ subsets contain progenitors with B lymphoid, myeloid, dendritic cell (DC), natural killer (NK) cell, and even erythroid lineage potential (*3*, *5*–*7*); however, whether any of these cell lineages actually develop within the healthy human thymus is currently unknown.

Another open question is the precise nature of the cells that enter the thymus (*8*). Hematopoietic progenitors from the BM enter the circulation and migrate to the thymus, where they commit to the T cell lineage. Thymic progenitors are continuously outcompeted by newly imported progenitors from the BM, an essential process that maintains healthy T cell development (*9*). However, the precise identity of these so-called thymus-seeding progenitor (TSP) cells is controversial, because some BM progenitor subpopulations may already have T lineage potential (*10*, *11*); moreover, their extremely low numbers hamper their identification. Here, we focused on the most immature DN1 subset, which presumably contains the noncommitted cells that recently seeded the thymus and may already have early specification for the T cell lineage.

Although genetic loss-of-function (LOF) studies in mice have been used to dissect thymocyte developmental trajectories and checkpoints, similar studies in humans are sparse. The closest study of genetic LOF in human T cell development used cells that were obtained from various patients with severe combined immunodeficiency (SCID) and xenotransplanted into nonobese diabetic SCID gamma (NSG) mice, revealing increased heterogeneity and, unexpectedly, previously unknown checkpoints in human T cell development (*12*).

[1]Department of Immunology, Leiden University Medical Center, Leiden, Netherlands. [2]Leiden Computational Biology Center, Leiden University Medical Center, Leiden, Netherlands. [3]Novo Nordisk Foundation Center for Stem Cell Medicine (reNEW), Leiden University Medical Center, Netherlands. [4]Delft Bioinformatics Lab, Delft University of Technology, Delft, Netherlands. [5]Molecular Epidemiology, Leiden University Medical Center, Leiden, Netherlands. [6]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands. [7]Centro de Investigación del Cáncer-Instituto de Biología Molecular y Celular del Cáncer (CIC-IBMCC, USAL-CSIC-FICUS), Department of Medicine, University of Salamanca, Salamanca, Spain. [8]Department of Pediatrics, Leiden University Medical Center, Leiden, Netherlands.
*Corresponding author. Email: f.j.t.staal@lumc.nl
†These authors contributed equally to this work.

Single-cell RNA sequencing (scRNA-seq) can be used to identify and characterize rare thymocyte populations and visualize the transcriptional landscape at the single-cell level, providing insights into T cell development trajectories. Recently, several studies that used scRNA-seq showed that most developmental stages based on surface markers are not linear and homogeneous and have overlapping transcriptional programs, resulting in a heterogeneous continuum (*13*–*15*).

To examine cellular heterogeneity in the human thymus, we performed scRNA-seq and corresponding scRNA-seq for V(D)J-enriched (variable (V), diversity (D) and joining (J)) gene segments in several sorted subsets with overlapping developmental potential. We then performed spectral flow cytometry of developing human thymocytes to perform a detailed phenotypic analysis. Specifically, we characterized the following: (i) the nature and differentiation of TSP subpopulations, (ii) T cell lineage commitment, and (iii) the development of non-T cell lineages in the human thymus. To complement our scRNA-seq analyses, we used functional assays and lineage tracing and identified three distinct TSP subpopulations, several non-T cell developmental pathways in the human thymus, and branching points between αβ and γδ T cells, including a stepwise commitment toward the αβ T cell fate.

RESULTS

Generation of a single-cell atlas of human T cell development

To profile the complete thymocyte population, including its rare subsets, we performed scRNA-seq on sorted thymocytes obtained from six healthy donors (Fig. 1A), focusing on early $CD4^{-}CD8^{-}$ DN thymocytes. We focused specifically on this early stage to gain insight into the nature of the progenitor cells that seed the thymus, the T cell specification and commitment processes, and possible non-T cell lineage differentiation. We therefore used flow cytometry to sort DN1 ($CD34^{+}CD38^{-}CD1a^{-}$), DN2 ($CD34^{+}CD38^{+}CD1a^{-}$), and DN3 ($CD34^{-}CD38^{+}CD1a^{+}$) thymocytes (Fig. 1B and fig. S1A) (*1*) . We also analyzed several additional sorted subsets, including $CD4^{+}$ immature SP (ISP), $CD4^{+}CD8^{+}$ DP, $CD4^{+}$ SP, and $CD8^{+}$ SP thymocytes. The DP subset was further subdivided into $CD4^{+}CD8^{+}CD3^{-}$ and $CD4^{+}CD8^{+}CD3^{+}$ thymocytes (Fig. 1B and fig. S1B). From these eight sorted populations, we generated eight scRNA-seq libraries for gene expression profiling using the 10X Genomics platform. In addition, these eight sequencing libraries were used for targeted amplification of both the αβ and γδ V(D)J genes using TCR-specific primers, resulting in 16 additional libraries for TCR sequencing (*16*). In total, we obtained a dataset for approximately 40,000 cells obtained from six donors (Fig. 1, C and D), with both gene expression data and αβ and γδ TCR rearrangement data for each cell.

Development of γδ T cells and plasmacytoid DCs

After in silico quality control to remove low-quality cells and doublets (*17*), we visualized the overall scRNA-seq data using a two-dimensional and three-dimensional uniform manifold approximation and projection (UMAP) to capture global transcriptional changes in developing thymocytes. We found that our transcriptome data clustered as a continuum spanning the sorted subsets (Fig. 2A and movie S1). Moreover, gene expression profiles were consistent with known changes that occur during human T cell development (Fig. 2B) and with CD34 expression reflecting the most immature thymocytes that appear shortly after entering the thymus. Notch signaling subsequently up-regulates *CD7* expression, marking specification toward the T cell lineage. The first T cell–specific target of the Notch signaling pathway, T cell factor 1 (TCF1; encoded by the *TCF7* gene), is then expressed, followed by *BCL11B*, both of which are critical transcription factors involved in committing early thymocytes to the T cell lineage in mice (*18*–*20*). A subset of early thymocytes with high *BCL11B* expression also expresses *CD1A* and *PTCRA* (Fig. 2B), marking T cell commitment and the initiation of TCR rearrangement. After TCRβ selection in DP cells, the cells differentiate further into SP cells. We noticed that after correcting for cell cycle, the DN cells were either in a state of proliferation (expressing *TOP2A*) or recombination (expressing *RAG1*), confirming that these two processes occur at distinct times during development (Fig. 2C). The timing of TCR rearrangement can be illustrated by aligning the γδ (Fig. 2D) and αβ TCR recombination data (Fig. 2E) to the UMAP plot, showing that γδ rearrangements occur as early as the DN2 stage and through the ISP and even DP stages. The ISP stage that marks the peak of TCRβ rearrangements coincides with the proliferation of *TCRB*-rearranged cells, whereas most of the *TCRG*-expressing cells follow the path with ongoing recombination (Fig. 2, D and E).

Recently, several studies using scRNA-seq reported intrathymic development of plasmacytoid DCs (pDCs) (*21*, *22*). Consistent with these findings, we detected cells with a pDC expression profile, indicated here by expression of *IRF8* (Fig. 2B), *TCF4*, and *SPIB* (fig. S2) within both the DN2 and ISP thymic subpopulations; these cells clustered above the main UMAP plot (Fig. 2B and fig. S2). The ISP-derived pDCs could be distinguished from the DN2-derived pDCs by their expression of *PTCRA*, *CD4*, and *GZMB* (Fig. 2B), suggesting that pDCs either arise from distinct progenitor cells or codevelop with T cells (*23*).

Apoptosis occurs in developing thymocytes after specific checkpoints, and cells undergoing apoptosis can be identified by measuring the percentage of reads mapping to the mitochondrial genome per cell (Fig. 2F) (*17*, *24*). As expected, *TCRB*-rearranged cells that have undergone TCRβ selection followed by rearrangement of the *TCRA* gene (Fig. 2E) subsequently undergo either positive or negative selection (*25*) during the time of *CD3* expression in DP cells (*1*). Negative selection of autoreactive thymocytes accounts for the loss of >90% of thymocytes via apoptosis (*26*). After crossing this tight bottleneck—indicated by the narrow "bridge" connecting the right and left sides of the UMAP plot—positively selected thymocytes continue their development to the $CD3^{+}$ DP stage. Cells expressing *CD27* can be rescued from negative selection and become regulatory T cells, indicated by expression of the transcription factor *FOXP3* (*27*); this is also visible in our dataset (Fig. 2B and fig. S2).

Identification of three TSP cell populations

HSCs that migrate from the BM to the thymus initiate the differentiation process shortly after entering the thymic environment. Thus, naïve TSPs that still resemble HSCs are extremely rare. Moreover, whether partially differentiated lymphoid progenitors from the circulation can initiate T cell development in the thymus is currently unclear (*28*). To identify these early progenitors within the context of early thymic development, we reanalyzed the ~10,000 DN1, DN2, and DN3 cells and generated a new UMAP plot (Fig. 3A and fig.

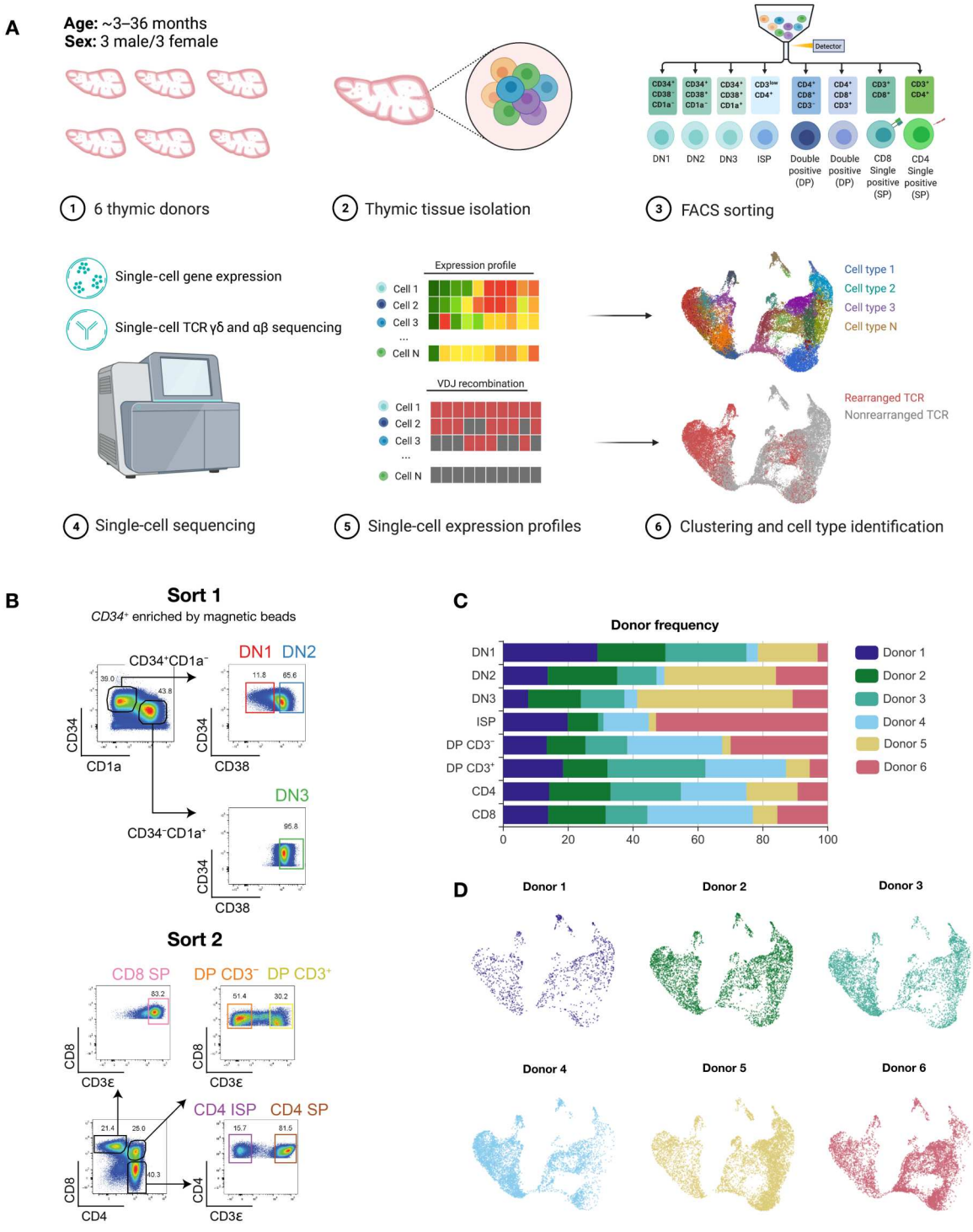

Fig. 1. Single-cell transcriptomics of human thymocytes. (**A**) Experimental scheme: Thymocytes from six different donors were isolated, pooled, and sorted for eight subpopulations representing eight differentiation stages of human T cell development. All eight subpopulations were sequenced for gene expression profiling and TCR analysis. (**B**) Sorting strategies of the three DN populations after CD34$^+$ enrichment (sort 1) and the five DP/SP populations (sort 2). (**C**) Donor proportion per developmental stage. (**D**) Contributions of each individual donor to total UMAP projection (Fig. 2). FACS, fluorescence activated cell sorting.

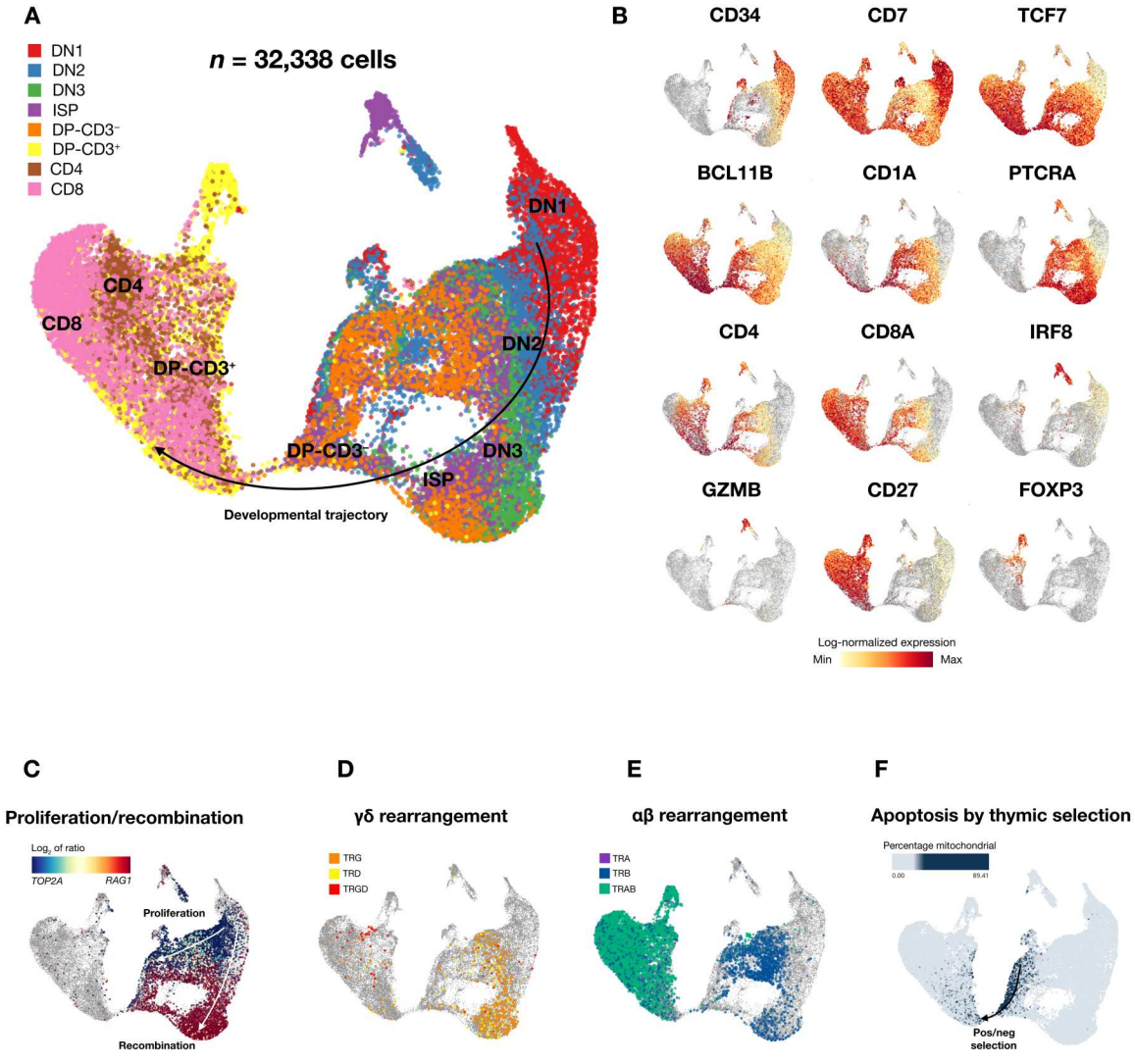

Fig. 2. UMAP projections of all eight sorted populations reveal gene expression related to cellular differentiation and biological processes. (**A**) UMAP projections of complete scRNA-seq datasets colored by sorted populations that reflect T cell developmental stages. (**B**) Gene expression level (red is high expression) of key surface markers and transcription factors projected on the UMAP. (**C**) *TOP2A* (blue) and *RAG1* (red) gene expression indicating proliferation and recombination, respectively. (**D**) Cells with rearranged TCRγ (orange), TCRδ (yellow), or fully rearranged TCRγδ (red). (**E**) Cells with rearranged TCRα (purple), TCRβ (blue), or fully rearranged TCRαβ (green). (**F**) Mitochondrial expression (blue) concomitant with cells in apoptosis during thymocyte selection predominantly just before the major transition among DP, $CD3^-$ to DP, and $CD3^+$ stages.

S3A). We found that the DN1, DN2, and DN3 cells generally followed their expected developmental order (Fig. 3, B and C, and fig. S3B); moreover, cluster marker analysis confirms heterogeneity among the DN cell populations (Fig. 3D). The cells in clusters 1 to 5 have an expression profile that suggests strong multilineage potential on the basis of their expression of the multipotency markers *HOPX* and *ACY3*; the B cell lineage–related genes *IGHM*, *MEF2C*, and *IGLL1*; and the myeloid marker *MPO* (Fig. 3D). The multilineage potential of clusters 1 and 2 was also confirmed by their lack of expression of the key Notch signaling–induced transcription factors *TCF7* and *BCL11B* (*19*).

The earliest DN cells (i.e., in cluster 1) lack *CD7* expression, indicating that the Notch signaling pathway is not yet activated in this stage. The cells in cluster 2 are the first to show *CD7* expression, resulting in low expression of *TCF7*, followed by an up-regulation of both *TCF7* and *BLC11B* in clusters 3 and 4 (Fig. 3E). We then searched for rare thymocyte subpopulations—including TSPs—in our DN dataset. TSPs enter the lobes of the thymus via high endothelial venules (*29*, *30*) and then home directly to the thymic environment by interacting with ligands expressed by thymic epithelial cells (*31*). We therefore searched our dataset for cell populations that express chemokine receptor genes (*CCR7*, *CCR9*, *CXCR3*, and *CXCR4*) (*32*–*35*), integrins (*ITGA4*, *ITGA5*, and *ITGA6*), and other homing factors (*CD62L*/*SELL*, *CD162L*/*SELPG*, and *CD127*/*IL7R*) (*28*, *31*). On the basis of these gene sets, the cells assigned to clusters 1, 2, and 14 seem to have the most thymus-homing activity; we named these thymus-seeding populations TSP1, TSP2, and TSP3, respectively (Fig. 3, F and G). Of note, the cells in TSP1 to TSP3 were derived primarily from the most immature sorted population, namely, the DN1 subpopulation that expresses CD34 but

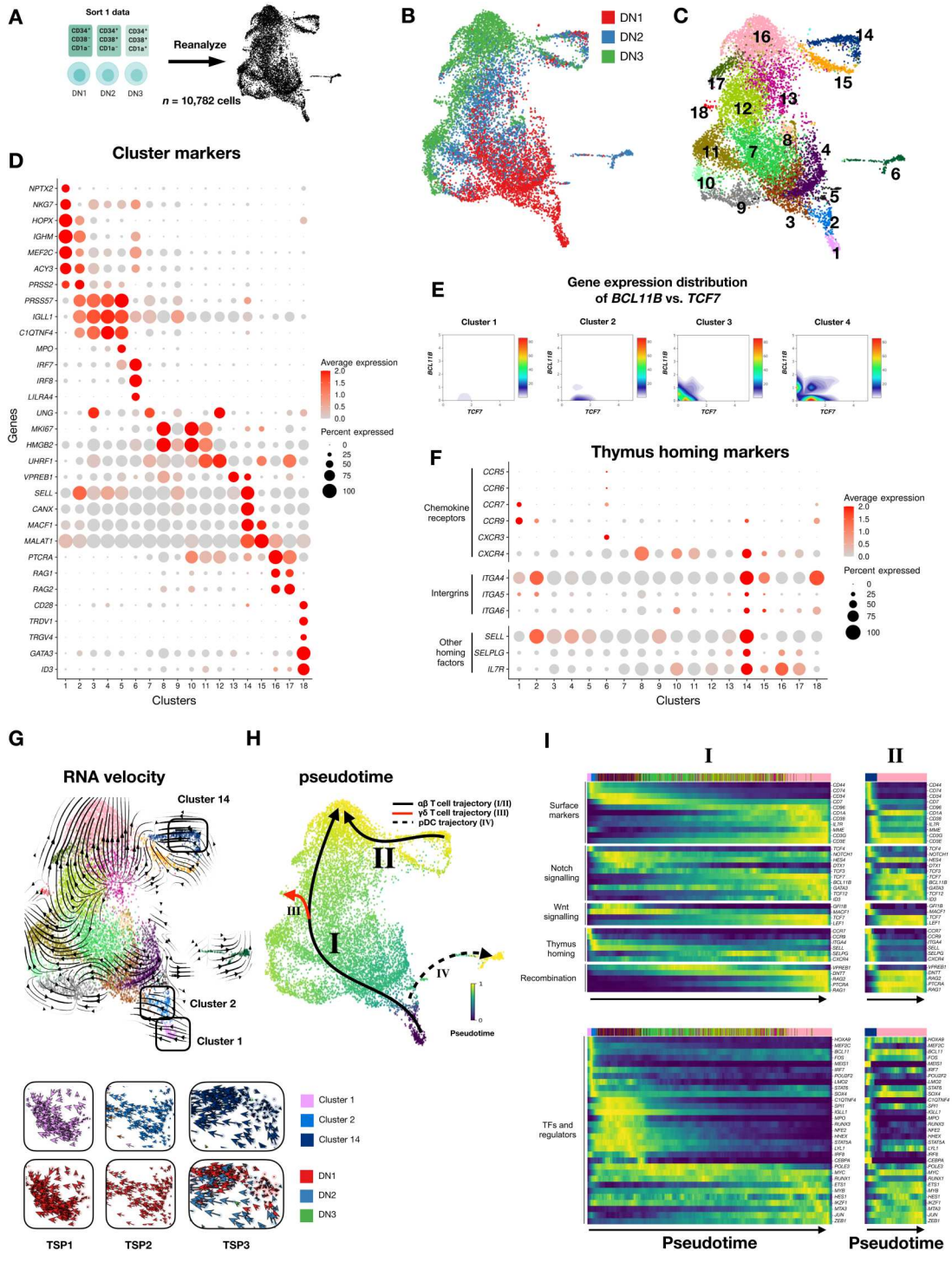

Fig. 3. Reanalysis of CD34⁺ progenitor populations reveals multiple developmental trajectories. (**A**) Reanalysis of DN1, DN2, and DN3 cells. (**B**) UMAP projection of DN1 (red), DN2 (blue), and DN3 (green) cells. (**C**) UMAP projection of 18 clusters identified by cluster analysis. (**D**) Dot plot depicting the genes that identify clusters based on differential gene expression. (**E**) Density plots of *TCF7* and *BCL11B* gene expression in clusters 1 to 4. (**F**) Dotplot indicating relative gene expression of thymus-homing markers per cluster. (**G**) Prediction of the global differentiation of thymocytes by RNA velocity analysis (top). TSP1, TSP2, and TSP3 have been enlarged and indicated for DN1, DN2, or DN3 origin (bottom). Arrows indicate differential splicing activity between TSPs. Global velocity streamlines are visualized by the large arrows, which are a summary of all single-cell velocities (small arrows, bottom inlays). (**H**) Pseudo-temporal ordering of cells colored from dark purple to yellow (0 to 1) according to maturation state. Lines represent differentiation trajectories of two routes of αβ T cell development (black lines), γδ T cell (red line), and pDC (dotted line) development. (**I**) Heatmap showing gene expression of surface markers, signaling pathways, thymus homing markers, recombination genes, and transcription factors involved in the T cell differentiation ordered according to the pseudo-time of the conventional (I) and alternative (II) αβ T cell development trajectories displayed in (H).

not yet CD38 and CD1a (Fig. 3, B and G); therefore, our downstream analyses of the three TSP subsets were performed using data obtained from DN1 thymocytes.

CD34$^+$ progenitor cells follow several developmental trajectories, including an alternative trajectory from progenitors to T cell lineage–committed thymocytes

T cell precursors progress through a series of distinct developmental stages marked by gradual phases of lineage specification followed by a T cell–specific transcriptional program. Before becoming fully committed to the T cell lineage, T cell precursors gradually lose their non-T cell lineage potential (*18*, *36*). To gain new insights into the transcriptional changes that accompany lineage commitment in early progenitor populations, we examined the complete DN dataset. The cells in cluster 6 were classified as pDCs on the basis of their expression of *IRF8* and *LILRA4* (also known as *CD85g*). On the basis of their expression of the proliferation markers *MKI67* and *HMGB2*, clusters 7 to 12 were classified as proliferating cells. We found that cluster 13 had high expression of the precursor B cell (pre-B cell) receptor gene *VPREB1*, whereas cluster 14 (i.e., the TSP3 subset) had high expression of the Wnt pathway–associated gene *MACF1* and the homing marker *SELL* (also known as *CD62L*). The cells in cluster 16 were classified as T lineage–committed thymocytes on the basis of their high expression of the recombination-activating genes *RAG1* and *RAG2* as well as the expression of *PTCRA* (Fig. 3D and fig. S3B) and the presence of *TCRB*-rearranged cells (fig. S3C). Last, we identified a cluster of mature γδ T cells based on the expression of the γδ-associated transcription factors *ID3* and *GATA3* (Fig. 3D), as well as expression of *TRDV* and *TRGV* combined with the presence of *TCRG*- and *TCRD*-rearranged cells (fig. S3D).

Application of RNA velocity (*37*, *38*) and diffusion pseudotime (DPT) (*39*) provided valuable insights into the trajectory of developing thymocytes. We found that overlaying the RNA velocity vectors on the UMAP plot (Fig. 3G) predicted the global differentiation of thymocytes into αβ T cells (cluster 16), γδ T cells (cluster 18), or pDCs (cluster 6). Moreover, ordering of the cells by DPT (Fig. 3H) relative to cluster 1 confirmed that cells in cluster 6 (i.e., pDCs) and cluster 16 (i.e., committed thymocytes) were the most mature cells in our dataset. Together, these analyses suggest the presence of four major differentiation trajectories in our dataset. The first trajectory is the conventional αβ T cell development route (indicated as route I in Fig. 3H) from the most immature, multipotent clusters 1 and 2 (TSP1 and TSP2 cells, respectively) to the rearranged, committed thymocytes in cluster 16. Unexpectedly, however, we identified an alternative T cell trajectory (indicated as route II in Fig. 3H) that also ends at cluster 16 but starts from the relatively more mature cluster 14 (TSP3 cells). A third trajectory leading to γδ T cells splits off from the first trajectory (route III indicated by the red line, exiting route I in Fig. 3H) and ends in cluster 18. Last, as discussed above, we detected two distinct pDC populations (see Fig. 2, A and B), one of which is located in the DN2 population; thus, the fourth trajectory revealed by our DPT analysis is used by cells that differentiate into pDCs (route IV indicated by the dashed line in Fig. 3H).

Transcriptional regulation during the early conventional (route I) and alternative (route II) trajectories of αβ T cell development

To gain additional insight into the transcriptional changes that occur during the conventional and alternative trajectories of αβ T cell development, we ordered the DN1, DN2, and DN3 cells along these two routes from the early progenitors in clusters 1 and 2 to committed thymocytes in cluster 16 (i.e., route I) and from the cells in cluster 14 to committed thymocytes in cluster 16 (i.e., route II), as predicted by the RNA velocity and DPT analysis (Fig. 3, G and H). We also identified differentially activated transcription factors and regulons (transcriptionally coregulated operons) in each cluster using single-cell regulatory network inference and clustering (SCENIC) (*40*), revealing clear changes in regulatory activity in the T cell differentiation trajectories (fig. S3E). For example, the cells in cluster 1 express stem cell–like and progenitor-like genes such as the hematopoietic regulatory genes *HOXA9*, *MEF2C*, *BCL11A*, *MEIS1*, *HOPX*, and *NPTX2* as well as the chemokine receptors CCR7 and CCR9, which are important for homing to the thymus. The stemness genes are the first genes to be down-regulated after Notch expression (Fig. 3, D and I, and fig. S3E). In addition to Notch signaling, Wnt signaling is also active in the DN stages, indicated by the up-regulation of *LEF1* even before *TCF7* expression (Fig. 3I and fig. S3F) (*41*). In these stages, *GFI1B* (growth factor independent 1B transcriptional repressor) is also expressed; this protein regulates Wnt/β-catenin signaling, thus indicating TCF/lymphoid enhancer factor (LEF)–mediated transcription (*42*). Expression of the master hematopoietic regulator gene *SPI1* (encoding the transcription factor PU.1) remains up-regulated through cluster 6 until its expression is finally overridden by the effects of Notch signaling in cluster 7 (Fig. 3I and fig. S3F). During this early window, multilineage transcription factors and regulators are expressed at high levels, including the myeloid-specific genes *MPO*, *CEPBA*, and *LYZ*; the B cell markers *IGLL1*, *VPREB3*, *PAX5*, and *EBF1*; the dendritic markers *IRF8* and *SPIB*; and the erythroid transcription factor *NFE2* (Fig. 3I and fig. S3, F and G). Cells that undergo Notch signaling increase their *HES4* and *DTX1* expression, followed by the expression of *GATA3*, *TCF7*, and *BCL11B* and a rapid down-regulation of *MME* (also known as *CD10*) and *TCF4*. Although the expression of *DNTT* (encoding the enzyme terminal deoxynucleotidyl transferase) starts in cluster 3, its expression increases markedly in cluster 7, coinciding with the increased expression of *TCF7* and *BCL11B* (Fig. 3I). At this point, *PTCRA* and CD1A expression marks the cell′s commitment to the T cell lineage. By comparing the predicted transcription factor activity with our clustering and trajectory analyses, we identified regulators (fig. S3E) with activity during the T cell commitment process (i.e., in clusters 7 to 13). These regulators include *POLE3*, which is consistent with the recent finding in *Pole3*-deficient mice that T cell differentiation is blocked at the DN3 stage (*43*); *MTA3*, a cell fate regulator involved in B cell differentiation (*44*); and *ZEB1*, which was recently described as playing an essential role in the transition of mouse thymocytes from the DN2 stage to the DP stage (*45*).

Of note, we also found that the pre-B cell receptor gene *VPREB1* is expressed in the earliest cells and continues until the cells begin to express *CD1A* (Fig. 3I and fig. S3G). In addition, expression of the *RAG* genes together with the expression of *TCF12* [which encodes transcription factor 12, also known as hela E-box binding factor

(HEB)] initiates TCR rearrangement. The cells in cluster 8 are the first to contain a rearranged *TCRG* locus (fig. S3D), followed by the completely rearranged *TCRD* locus in clusters 12 and 16, marking the first *TCRB*-rearranged thymocytes (fig. S3C).

Fully rearranged γδ T cells in developing DP thymocytes

We found that *TCRB* rearrangement occurs during the ISP and CD3$^-$ DP stages of T cell development, followed by *TCRA* rearrangement in the CD3$^+$ DP stage (fig. S4, A to E). In addition, we found that cells either undergo or have completed their *TCRG* and *TCRD* rearrangement in the ISP and DP stages, including the mature γδ T cells in cluster 16 (fig. S4, C and F), thus after the initial γδ T cells split off in cluster 18 (Fig. 3, B and C); CD4$^+$CD8$^+$ γδ T cells were found in the DP population. This finding suggests that cells can differentiate into γδ T cells at various developmental stages and follow the αβ path for longer than previously anticipated.

TSP1, TSP2, and TSP3 subsets generally represent HSC-like, MPP-like, and CLP-like TSPs, respectively

Because TSPs originate from the BM, we sought to determine the BM progenitors that give rise to the three TSP subsets that we identified. To this end, we mapped the TSP1, TSP2, and TSP3 cells in the DN1 dataset to a large multimodal dataset that includes all major BM-derived blood cell types and progenitors (*46*) (Fig. 4, A and B, and fig. S5, A to C). We also subannotated the CD34$^+$ progenitor population in this dataset using the recently reported BM progenitor markers (*47*). This transcriptomics-based mapping approach revealed general correspondence between the three TSP clusters and distinct BM progenitor populations. Specifically, most TSP1 cells match the HSC annotation, most TSP2 cells match the multipotent progenitor (MPP) annotation, and most TSP3 cells match the lympho-myeloid primed progenitor (LMPP) and common lymphoid progenitor (CLP) annotations (Fig. 4, C to E); our TSP3 subset resembles CD34$^+$CD7$^+$ CLPs previously reported in humans (*48*). However, when we integrated our dataset with two thymus-seeding populations identified by Lavaert *et al.* (*21*), we observed that only the first population has immature gene expression and resembles our TSP1 and TSP2 clusters due to proximity in the integrated UMAP (fig. S5D). In contrast, their second cluster shows overlap with our pDC and pDC progenitor cluster and lacks the expression of stemness marker CD34 and thymus-homing marker CCR9.

Next, we analyzed the differential gene expression between the three TSP subsets and identified subset-specific markers (Fig. 5A). TSP1 is a relatively rare CD34$^+$CD7$^-$ HSC-like population that lacks *CD7* expression but expresses several immature stemness-like genes, including *HOXA9*, *HOPX*, *MEF2C*, and *NPTX2* (Fig. 5A; see also cluster 1 in Fig. 3D). Strikingly, this population seems to be quiescent based on the lack of expression of proliferation genes (fig.

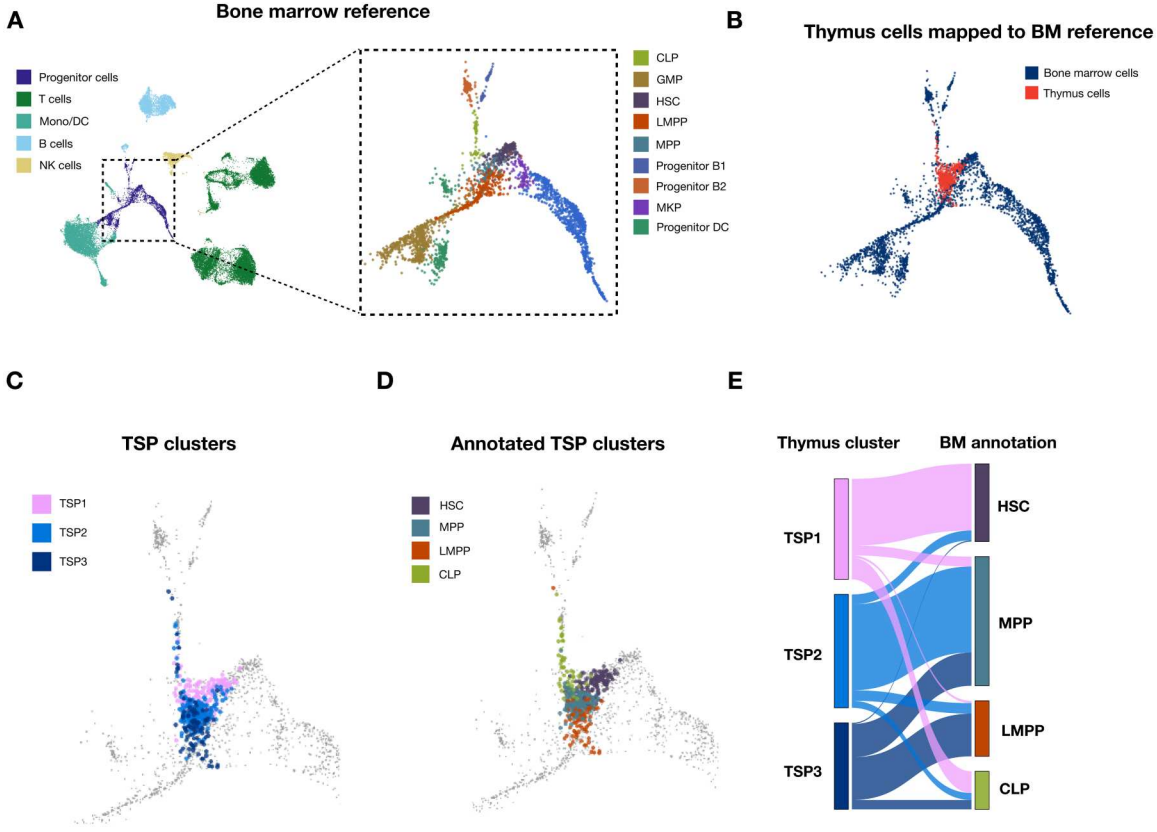

Fig. 4. Identification of three TSP populations. (**A**) Annotated multimodal BM reference UMAP (*46*) with a focus on progenitor cell populations (outlined by the dotted square). (**B**) Three thymocyte TSP populations (red) were mapped onto the BM progenitor cell (blue) reference UMAP. (**C**) Thymocyte TSP1 (pink), TSP2 (blue), and TSP3 (dark blue) were mapped separately. (**D**) The same TSPs colored for BM progenitor label after reference mapping. (**E**) TSPs linked to their BM annotation after reference mapping in (D).

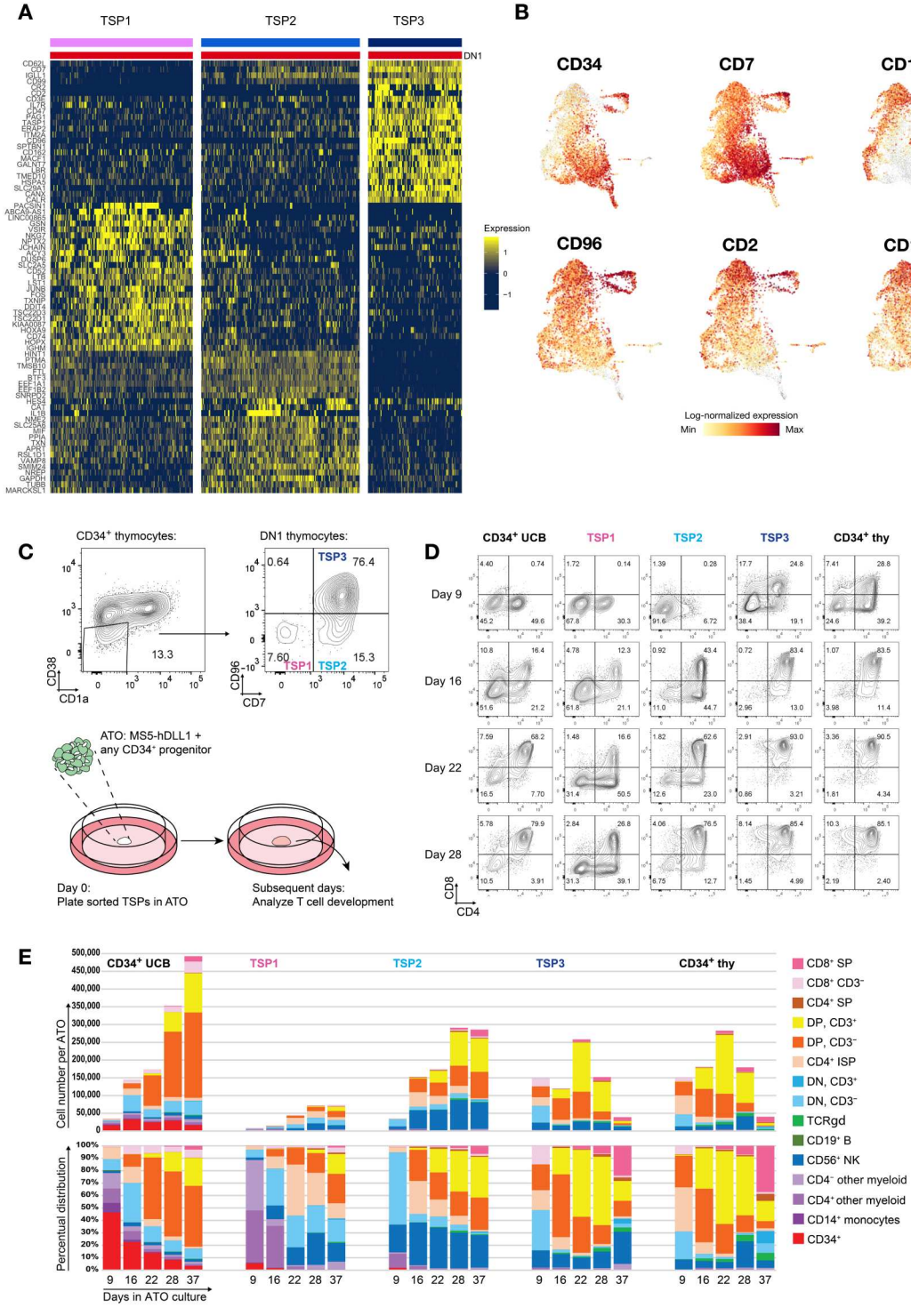

Fig. 5. TSP1, TSP2, and TSP3 differ in T cell differentiation kinetics. (**A**) Heatmap of differential gene expression between the three identified TSPs (within DN1 cells). (**B**) Examples of gene expression of phenotypic markers of TSPs. (**C**) Flow sorting strategy of TSPs (within $CD34^+CD38^-CD1a^-$ DN1 thymocytes) using CD2, CD7, and CD96 as distinguishing markers. Sorted TSPs were cultured in an ATO. (**D**) Flow cytometry analysis of TSP cells and $CD34^+$-enriched UCB cells or thymocytes (thy) after 9, 16, 22, and 28 days of culture in an ATO. CD4/CD8 DP plots were gated $CD45^+CD34^-CD33^-CD14^-CD56^-CD19^-CD7^+TCR\gamma\delta^-$. (**E**) Bar graphs indicating various cell type numbers (top) and proportions (bottom) of sorted populations after 9, 16, 22, 28, and 37 days of ATO culture. All populations were gated $CD45^+$. $CD14^+$ monocytes: $CD33^+CD14^+$, $CD4^+$ and $CD4^-$ other myeloid, $CD56^+$ NK, and $CD19^+$ B cells were $CD33^+CD14^-$. The TCRgd ($TCR\gamma\delta^+CD3^+$) population was $CD33^-CD14^-CD56^-CD19^-CD7^+$. All subsequent T cell populations were gated negative for TCRγδ and divided into $CD3^+$ and $CD3^-$ populations. DN: $CD4^-CD8^-$; $CD4^+$ ISP: $CD4^+CD8^-CD3^-$; DP: $CD4^+CD8^+$; $CD4^+$ SP: $CD4^+CD8^-CD3^+$; and $CD8^+$ SP: $CD4^-CD8^+CD3^+$. A representative of two biological experiments is shown.

S3B) and is therefore not a rapidly expanding TSP population. TSP2 is a $CD34^+CD7^+$ MPP-like population that expresses *CD7* but no other T cell–specific genes; in contrast, these cells express genes in non-T cell lineages, including *IGHM*, *IGLL1*, *NFE2*, *MPO*, *IRF8*, *LY6E*, *MME*, and *SPI1*, suggesting that they retain their multipotency (Fig. 5A; see also cluster 2 in Fig. 3D); although expression of these genes continues through clusters 3 to 5, these clusters also begin to express T cell genes such as *NOTCH1*, *HES4*, and *LEF1*. Last, TSP3 is a $CD34^+CD7^+$ CLP-like population that also expresses the homing marker genes *CD62L* and *CXCR4* and Wnt pathway genes such as *MACF1* (*49*) [see also cluster 14 in Fig. 3 (D and F)]. Given the high expression of *IL7R* and other lymphoid marker genes and the low expression of non-T cell markers, in combination with the fact that they are still immature due to the presence of CD34 and absence of CD1a, TSP3 cells appear to be immature while primed to develop rapidly into T cells upon entering the thymus.

TSP1, TSP2, and TSP3 cells have distinct T cell differentiation kinetics based on in silico data

On the basis of our in silico DPT analyses (Fig. 3, H and I), we predicted that the two trajectories to reaching αβ T cells in cluster 16 (i.e., trajectory I from clusters 1 and 2 and trajectory II from cluster 14) follow distinct time courses. When we focused on the RNA velocity data for the various TSP subsets, we found differences in splicing activity (reflected by differences in the length of the arrows) (Fig. 3G, insets). Compared with the cells in TSP1, the cells in TSP2 have relatively high splicing activity (i.e., longer arrows); moreover, the arrows in TSP2 have clear directionality. In contrast, the cells in TSP3 have virtually no splicing activity at their putative starting point in the topmost right corner of cluster 14 (Fig. 3G, bottom insets), whereas the neighboring cells in cluster 14 have high splicing activity, suggesting the onset of differentiation. We then ordered the cells along the two predicted trajectories (with TSP1 and TSP3 cells as the starting points for trajectories I and II, respectively) and plotted the gene expression data on a pseudotime axis (Fig. 3I). We found that the two trajectories follow a similar gene expression pattern but differed considerably with respect to their total duration; specifically, the TSP1 cells first differentiated into several consecutive intermediate cell types (identified as clusters 3 to 8) before becoming T cell lineage–committed thymocytes, whereas the TSP3 cells reached this stage far more rapidly.

Artificial thymic organoids seeded with TSP1, TSP2, and TSP3 support and confirm in silico findings regarding the differences in proliferation and differentiation

To confirm our in silico results suggesting that TSP3 cells are primed to develop more rapidly into T cells upon entering the thymus compared with the less mature TSP1 and TSP2 subsets (which take longer to reach cluster 16), we isolated and cultured these subpopulations. Using flow cytometry, we investigated all of the cell surface markers that were differentially expressed between the three TSP subsets (Fig. 5A). We found that the $CD34^+CD38^{-/-}CD1a^-$ DN1 thymocytes contained a distinct $CD2^-CD96^-$ subpopulation, part of which was $CD7^-$ and $CD162^+$, consistent with TSP1 and TSP2 cells; in contrast, the $CD2^+CD96^+$ subpopulation was $CD7^+$, $CD162^+$, and interleukin-7R positive ($IL\text{-}7R^+$), consistent with TSP3 cells (Fig. 5, B and C). We then tracked the development of flow-sorted $CD2^-CD96^-CD7^-$ DN1 (i.e., TSP1), $CD2^-CD96^-CD7^+$ DN1 (i.e., TSP2), and $CD2^+CD96^+CD7^+$ DN1 (i.e., TSP3) cells in artificial thymic organoids (ATOs) (*50*) cultured for up to 37 days (Fig. 5C and fig. S6A). Of note, these TSP1, TSP2, and TSP3 cells can also be detected in peripheral blood mononuclear cells from umbilical cord blood (UCB) (fig. S6B). As a control for the ATOs, we found that when starting with $CD34^+$ cells isolated from UCB, only 16% of the cells were $CD4^+CD8^+$ by day 16, whereas TSP1 does not proliferate as much as the other populations. In contrast, virtually all of the TSP3 and total $CD34^+$ thymocytes (comprising not only DN1 but also DN2 and DN3) cells were $CD4^+CD8^+$ by that same time point (Fig. 5, D and E). $CD34^+$ progenitors appear to represent a steady pool in UCB-derived ATOs but not in thymocyte/TSP-derived ATOs, which explains the relatively rapid exhaustion of thymocyte/TSP ATOs. Last, non-T cells do not readily develop in these ATOs except for NK cells that mainly arise from the multipotent TSP2. (Fig. 5E). Thus, these in vitro data support our notion that TSP1 is a quiescent, nonproliferating population and that TSP3 cells are primed to develop much more rapidly into T cells compared with TSP1 and TSP2 cells.

Development of other, non-T cell types in the thymus

To gain more insight into the multilineage potential of thymocytes, we integrated our complete thymus dataset with two different human BM datasets; one dataset contains 40,000 mature BM cells from the immune census dataset (*51*), and the other dataset contains 25,000 early hematopoietic progenitor cells from a CD34-enriched BM dataset (Fig. 6A and fig. S7A) (*52*). All three datasets were processed, integrated, and annotated (for details, see methods), resulting in a combined dataset containing a total of about 77,500 cells. We then generated a three-dimensional UMAP plot using Seurat; the plot was then loaded into the BioTuring browser (*53*), which made it possible to rotate the UMAP plot and visualize the cells at different angles (Fig. 6, fig. S7, and movie S2). This integrated single-cell dataset fully recapitulated the differentiation of HSCs via progenitor populations into a wide range of mature immune cell subsets in the BM (Fig. 6B and movies S2 to S4). Of note, the thymocytes that were integrated into the BM setting still followed the overall T cell differentiation trajectory from DN cells to ISP, DP, and lastly SP cells (Fig. 6C and movie S3).

Most of the mature BM cell types overlapped with small populations of thymic cells. The most immature thymocytes, namely, DN1 and—to a lesser extent—DN2 cells, overlapped with these mature BM cell types, suggesting a link between immature and mature non-T cells in the thymus (Fig. 6D), including B cells, NK cells, pDCs, neutrophils, monocytes, and macrophages (Fig. 6, C to E; fig. S7, B to D; and movies S5 and S6).

To confirm these findings using a different technique, we used 28-color spectral flow cytometry of human thymocytes to identify mature non-T cell types and their progenitors. Using a principal components analysis (PCA) to visualize this data in lower dimensions, we again observed a continuum of developing cell types, ranging from immature $CD34^+$ progenitors to mature monocytes, DCs, and B cells, in both the BM and thymus datasets (Fig. 6F). Moreover, this data confirms the recently reported development of pDCs in the thymus (*21*, *22*). With respect to NK cells, although we identified mature NK cells in both the BM and thymus (Fig. 6, E and F), we were unable to identify developing NK cells by flow cytometry due to the lack of suitable surface markers. In addition, although mature erythrocytes were lost during the thawing

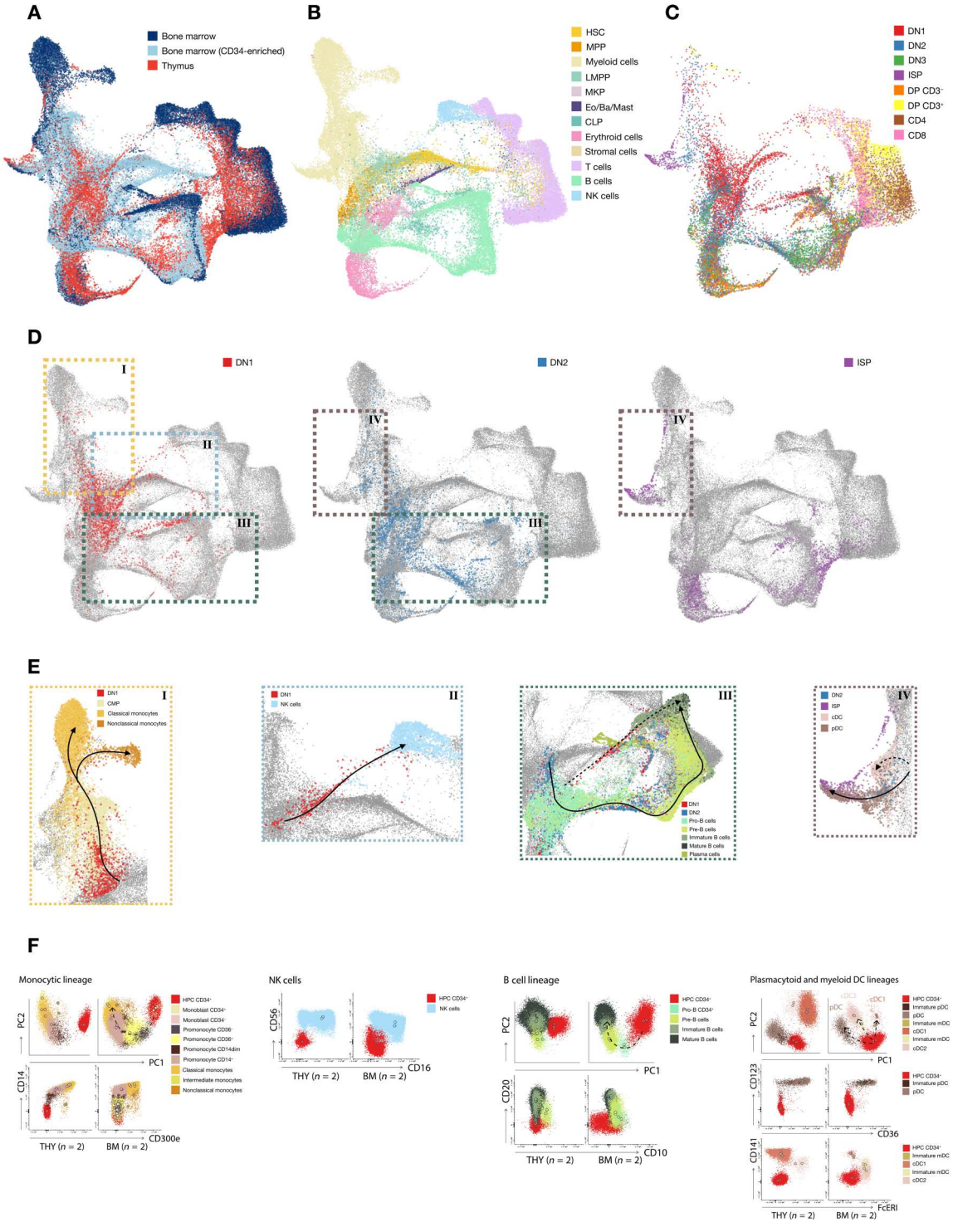

Fig. 6. Intrathymic development of alternative lineages. (**A**) UMAP projection after integration of mature BM cells (*51*) with CD34$^+$ BM (*52*) and our complete thymus dataset. (**B**) UMAP projection of annotated BM cells. (**C**) All eight sorted thymocyte subsets were projected on the integrated UMAP. (**D**) Projection of DN1 (red), DN2 (blue), and ISP cells (purple) on the integrated UMAP. (**E**) Insets from (D): DN1, DN2, and ISP cell differentiation toward monocytes, NK cells, B cells, cDCs, and pDCs. (**F**) Principal components (PC) analyses and example flow plots (bottom rows) of 28-color spectral flow cytometry of human thymocytes (THY) (n = 2) and human bone marrow (BM) (n = 2) to indicate mature non-T cell types (monocytes, NK cells, B cells, cDCs, and pDCs) and their progenitors. mDC, myeloid DC (which includes cDC1 and cDC2).

procedure, we found developing erythroblasts in the BM but not in the thymus. Classic DCs from the myeloid lineage (i.e., cDC1 and cDC2) are present in both the BM and thymus (albeit with more cDC1 cells than cDC2 cells in the thymus), but the differentiation steps for these DCs are unclear due to the lack of suitable markers for their corresponding progenitor cells. With respect to neutrophils, their lineage is clearly visible in the BM, but these cells are virtually absent in the thymus samples. However, mature plasma B cells, mast cells, and eosinophils are all present in the thymus (fig. S7E).

Genomic lineage tracing of non-T cells reveals their thymic origin

In the thymus, the first TCR rearrangements occur in the *TCRD* locus before T cell commitment. The first delta rearrangement occurs between the Dδ2 and Dδ3 segments and is nonfunctional, followed by the first functional rearrangement between Dδ2 and Jδ1 (Fig. 7A) (*1*, *2*). Evidence of Dδ2-Jδ1 rearrangements in any given cell indicates that the cell's differentiation originated in the thymus. Therefore, we sorted B cells, NK cells, monocytes, cDC1 cells, and pDCs from both BM and thymus samples and looked for the presence of these early rearrangement steps (fig. S7F); $CD34^+$-enriched cord blood cells were included as a negative control, and $CD34^+$-enriched thymocytes were included as a positive control (*1*, *2*). The presence of the Dδ2-Dδ3 rearrangement—and particularly the Dδ2-Jδ1 rearrangement—in the NK cells, monocytes, cDC1 cells, and pDCs in the thymus samples indicates that these cells can develop in the thymus; in contrast, these same cell types isolated from the BM lack these rearrangements (Fig. 7B).

A small but measurable percentage of B cells in the thymus samples contain the nonfunctional Dδ2-Dδ3 rearrangement (but not the functional Dδ2-Jδ1 rearrangement); however, the Dδ2-Dδ3 rearrangement is also present at the same frequency in B cells obtained from the BM samples, suggesting that the Dδ2-Dδ3 rearrangement may not be specific to T cells but also can occur in B cells, although Dδ2-Jδ1 rearrangements are extremely rare in B cells. A large number of precursor B cell acute lymphoblastic leukemia cells contain Dδ2-Dδ3 rearrangements but never contain Dδ2-Jδ1 rearrangements, supporting the notion that the Dδ2-Dδ3 rearrangement marks a step toward the lymphoid lineage that may be common to both T cells and B cells (*54*, *55*). Thus, the presence of this rearrangement does not necessarily indicate that the B cells in the thymus developed from thymic progenitor cells. Nevertheless, our combined single-cell transcriptomics, flow cytometry, and *TCRD* rearrangement data provide compelling evidence that non-T cells lineages including—but not necessarily limited to—NK cells, cDC1 cells, pDCs, monocytes, and possibly B cells can develop within the human thymus.

DISCUSSION

Recent advances in single-cell techniques such as RNA-seq and spectral flow cytometry have led to important new insights into the heterogeneity of the immune system. Here, we used these techniques together with quantitative polymerase chain reaction (qPCR) analysis and functional assays to better define the road map of T cell development in the human thymus. In our study, we focused on the earliest stages of T cell development, including thymic seeding by progenitor cells and T cell lineage commitment. Because we focused on these early stages, they are overrepresented in our data. Nevertheless, we observed a clear continuum in

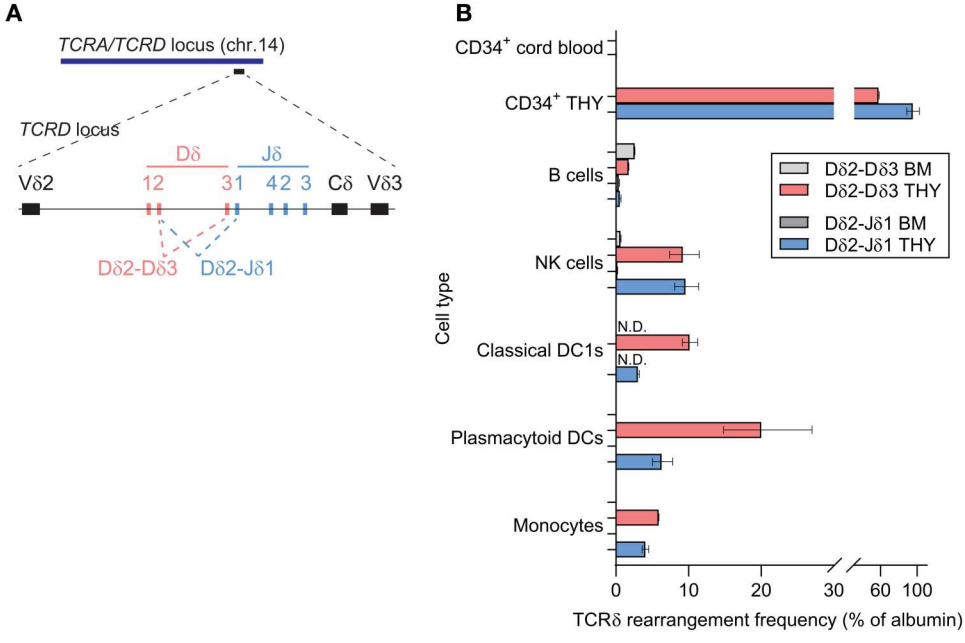

Fig. 7. qPCR analysis of *TCRD* gene rearrangements in non-T cells isolated from thymus and BM. (**A**) Schematic diagram of the human *TCRD* gene complex. (**B**) Alleles with Dδ2-Dδ3 (pink) and Dδ2-Jδ1 (blue) rearrangements expressed as a percentage of albumin in B cells, NK cells, cDC1s, pDCs, and monocytes isolated from thymus and from BM as controls (light and dark gray, respectively). $CD34^+$ cord blood represents a second negative control, whereas immature $CD34^+$ thymocytes serve as a positive control. One representative of two biological replicates with triplicate samples for qPCR is shown.

thymocyte development, with no major gaps in the developmental trajectory.

Using mice with targeted mutations in key transcription factors that regulate T cell development, our group and others previously showed that Tcf7, Gata3, Runx1, and Bcl11b play important roles in this process (*18*–*20*, *56*). For example, we showed that a transcription factor cascade drives T cell development in which Notch signaling induces its first T cell–specific target gene, *Tcf7*, which then induces (in combination with Notch signaling) *Bcl11b* and *Gata3*; *Bcl11b* primarily induces T cell–specific genes, whereas *Gata3* primarily represses non-T cell–specific genes (*20*, *57*). We observed *TCF7* expression before *BCL11B* expression, suggesting that a similar transcription factor cascade also exists in the human thymus, although carefully designed LOF experiments similar to experiments involving BCL11B (*58*) are needed to address this question in the human thymus.

When sorting mature thymocyte subsets, we excluded non-T cell types using lineage markers (except for pDCs, for which sorting markers were not included). Although this strategy allowed us to clearly define the T cell specification and commitment steps, it also made it more difficult to define non-T cell development pathways. However, our $CD34^+$-enriched dataset containing DN1, DN2, and DN3 cells did not deplete these alternate lineages and provided important insights into the development of non-T cell types. Moreover, our complementary spectral flow cytometry approach revealed a continuum of cells that develop into several non-T cell lineages in the total thymocyte samples. We also sorted several populations of mature non-T cells from the thymus and confirmed their thymic origin using the earliest *TCRD* locus rearrangements as a lineage-tracing marker. Collectively, our data show that monocytes, NK cells, myeloid DCs, and pDCs can develop in the human thymus. Recently, Lavaert and colleagues (*21*, *22*) reported that pDCs are the only non-T cells that develop in the human thymus; in addition to confirming this result using lineage tracing, we also observed the development of other non-T cells. We therefore believe that excluding most non-T cells using lineage surface markers during sorting may lead to the false conclusion that only pDCs and T cells develop in the thymus, whereas our detailed scRNA-seq analysis (including immature $CD34^+CD1a^-$ sorted cells) combined with deep flow cytometry profiling and lineage tracing clearly demonstrate that monocytes, NK cells, myeloid DCs, and—albeit to a lesser extent—B cells also develop in the human thymus. Our finding of mast cells in the human thymus is particularly interesting given that mast cell development in the thymus was reported previously in mice overexpressing GATA3 (*59*) but not in wild-type mice.

Previous studies suggest that an extremely small fraction of thymocytes consists of progenitor cells for NK cells, B cells, DCs, monocytic cells, and even erythroid cells in addition to the overwhelming majority of T cell progenitors. $CD34^+$ thymocyte progenitors can be redirected to differentiate down these non-T cell lineages in vitro under the appropriate conditions (*3*, *60*, *61*) and in vivo in humanized NSG mice (*62*). In addition, mouse thymic progenitors were shown recently to develop in vivo into B cells and myeloid cells in *Tcf7* knockout mice, which lack robust T cell development (*20*). Unlike mouse thymocytes, the most immature human thymocytes have an extremely broad differentiation capacity, including the potential to develop into erythrocytes (*3*). However, the fact that progenitors have this capacity does not necessarily mean that this developmental path actually occurs in the human thymus; despite this inherent capacity, we found no indication that erythrocytes develop in the human thymus.

The possibility that B cells can develop in the human thymus has long intrigued immunologists. In mice, clear evidence exists that most B cells in the thymus develop from thymic precursor cells (*63*, *64*), although a relatively small percentage of B cells can develop from distinct progenitors, and some mature B cells may seed the thymus (*65*). Using flow cytometry, we previously showed that the full range of B cell developmental stages can be measured in the thymus (*66*); however, this finding does not exclude the possibility that distinct precursor B cells may have entered the thymus. Our genetic lineage tracing studies using the initial *TCRD* rearrangements as a marker of T cell progenitors indicate that only an extremely small percentage of thymic B cells are derived from this thymic progenitor. However, it is possible that the TSP3 subpopulation could represent a B/T cell progenitor that follows different developmental trajectories. Nevertheless, the developmental origin of B cells in the human thymus is clearly diverse. With respect to their function, evidence in mice suggests that thymic B cells may play an antigen-presenting role during negative selection (*67*). In addition, evidence suggests that antibody-producing plasma B cells may be present in the human thymus (*68*), a finding supported by our own results.

We characterized the nature of thymus-seeding cells using RNA velocity and RNA-seq data and identified three distinct TSP subpopulations, namely, an HSC-like subset (TSP1), an MPP-like subset (TSP2), and a CLP-like subset (TSP3). The HSC-like TSP1 subpopulation appears to be quiescent and may represent a stem cell–like reservoir for thymocytes (*69*). On the other hand, the CLP-like TSP3 subpopulation is particularly interesting, because T cells develop much more rapidly from this subset compared with the other two subsets, consistent with previous studies in mice (*70*). On the basis of marker expression, the TSP3 subset identified here is similar to a previously described CLP subset in human BM (*48*). The TSP1 subset is a population with much slower kinetics toward developing T cells and, based on the RNA velocity analysis, remains largely quiescent. Given the transcriptomics data that relate the TSP1 subset much more closely to quiescent HSCs than the TSP2 population, it is tempting to speculate that TSP1 represents an intrathymic stem cell population. Needless to say, much more future work will be needed to validate this hypothesis.

Recently, Lavaert and colleagues (*21*) reported two human TSP subsets, the first of which likely corresponds to our TSP1- and TSP2-combined subset due to the expression of *CD34*, stemness genes, and multipotency genes. Our results do not confirm that their second TSP subset is an early TSP because it lacks *CD34* expression but expresses *CD3E*, indicating a T cell lineage–committed thymocyte subset. Moreover, their second subset likely has robust pDC progenitor potential, because it clustered in our dataset with pDCs and their immediate progenitors; instead, we identified the previously unrecognized CLP-like TSP subset (which we call TSP3) that develops into committed thymocytes much more rapidly than our HSC-like TSP1 and MPP-like TSP2 subsets.

In addition to characterizing these distinct thymus-seeding populations, we also investigated the T cell lineage commitment for these TSP subsets. We found that compared with the TSP3 subset, the HSC-like TSP1 subset and the MPP-like TSP2 subset have a considerably longer trajectory to becoming committed T cells, characterized by a fully rearranged TCRB (for the αβ lineages)

and the expression of *CD3*, *pTA*, and other T cell–specific genes such as *LCK*, *LAT*, and *MAL*. The branch into the γδ lineage occurs relatively late—from late DN2 to DN3 or even from cells in the ISP or DP stage—in contrast with our previous findings (*1*). Although this study reports on previously unrecognized developmental trajectories and thymic seeding populations, they are not confirmed by genetic LOF studies that have been very insightful in the mouse. For human T cell development, such studies completely rely on "experiments of nature" using patient cells from severe combined immune deficiencies as LOF models (*12*). Recent technological advances with gene editing techniques should make such studies possible in human thymocytes. Intra-individual variation invariably presents a confounding factor in human studies, as opposed to inbred mouse studies, but we minimized this source of variation by using expression data from six individual human thymi that all contributed to the full dataset (Fig. 1, C and D). In summary, we report data on the nature of cells seeding the thymus, the potential for development of non-T cell lineages, and the establishment of T cell lineage commitment. These findings are important for therapeutic efforts to restore thymic function (*71*), for instance, through groundbreaking gene therapy efforts to cure inherited disorders of T cell development, which have recently become clinical reality (*72*) but still need further development to become robust therapies.

MATERIALS AND METHODS

Study design

In this study, we performed scRNA-seq on human postnatal thymocytes obtained from six donors. Thymocytes from the six donors were pooled and flow-sorted in eight consecutive stages of T cell development that were sequenced separately to allow in-depth analysis per stage, including the earliest and most rare subpopulations. High-dimensional 29-color spectral flow cytometry and lineage tracing using early TCR delta rearrangements were performed to reveal development of alternative immune lineages in the thymus. Last, functional characterization of the three identified TSP populations in in vitro ATOs validated the results on TSPs.

Samples

Thymic tissues were obtained as surgical tissue discards from children 7 weeks to 3 years of age (median: 6 months of age) who underwent cardiac surgery; written informed consent was obtained from the parents. The children did not have immunological abnormalities. Thymocytes were isolated from the tissues by cutting the thymic lobes into small pieces and then squeezing the pieces through a metal mesh; the cells were then frozen and stored in liquid nitrogen until further use. Healthy surplus BM mononuclear cell suspensions were obtained from donor BM samples and stored in liquid nitrogen until further use.

Isolation of thymocyte subsets

To isolate the thymocyte subsets, thymocytes were thawed and pooled from six samples prepared as described above. To isolate DN1, DN2, and DN3 thymocytes, we used CD34 magnetic beads (CD34 MicroBead Kit UltraPure, Miltenyi Biotec) to enrich for $CD34^+$ thymocytes.

Flow cytometry and cell sorting

Spectral flow cytometry and cell sorting were performed at the Leiden University Medical Center Flow Cytometry Core Facility (www.lumc.nl/research/facilities/fcf) using a Cytek Aurora 5L (Cytek Biosciences, Fremont, CA) and a BD FACSAria III 4L (BD Biosciences, San Jose, CA), respectively. To sort the more mature thymocytes, including ISP, DP, and SP cells, we excluded non-T lineage cells using antibodies against the following surface markers: CD26, CD56, CD13, CD33, CD19, and CD34. The purity of the sorted cells is shown in figs. S1 and S7F, and the antibodies used for flow cytometry are listed in table S1.

Single-cell RNA sequencing

Cell suspensions were barcoded (10X Chromium Single Cell platform, 10X Genomics) using the Chromium Single-Cell 5′ Library (10X Genomics) from $Lin^-CD34^+CD38^-CD1a^-$ (DN1), $Lin^-CD34^+CD38^+CD1a^-$ (DN2), $Lin^-CD34^+CD38^+CD1a^+$ (DN3), $CD34^-CD3^{low}CD4^+$ (ISP), $CD3^-CD4^+CD8^-$ (CD4SP), and $CD3^+CD4^-CD8^+$ (CD8SP) and used to generate eight single-cell 5′ gene expression libraries. The loaded cell numbers ranged from 300 to 500,000, aiming for 5000 cells per reaction. All libraries were sequenced using an Illumina HiSeq 4000 [paired-end 75–base pair (bp) reads] to an average depth of 50,000 reads per cell, resulting in eight datasets (one for each developmental thymocyte subset).

Paired single-cell TCRαβ/γδ sequencing

Single-cell TCRαβ sequencing libraries were generated using the Chromium Single-Cell 5′ Library Construction Kit (PN-1000020, 10X Genomics) and the Chromium Single-Cell V(D)J Enrichment Kit for human T cells (PN-100005, 10X Genomics). Single-cell TCRγδ sequencing libraries were generated using the Chromium Single-Cell 5′ Library Construction Kit (PN-1000020, 10X Genomics) and the following custom primer sets: first PCR:forward primer (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC-3′), outer reverse primers for T cell receptor delta constant (*TRDC*) (5′-GGCAGAAAACCATCAATGCC-3′ and 5′-GTCTCATTTCTGTTCCTCCC-3′), and inner reverse primers for *TRDC* (5′-CCCAGGACTTTTGTCTTCCC-3′ and 5′-GAGTGTAGCTTCCTCATGCC-3′); second PCR: forward primer (5′-AATGATACGGCGACCACCGAGATCT-3′), outer reverse primers for T cell receptor gamma constant (*TRGC)* (5′-CTCCATTGCAGCAGAAAGCC-3′ and 5′-AGACAGCAGGTGATGATGGC-3′), and inner reverse primers for *TRGC* (5′-TTCTGGCACCGTTAACCAGC-3′ and 5′-TAGTCTTCATGGTGTTCCCC-3′). Pooled libraries were sequenced using an Illumina HiSeq4000 (paired-end 150-bp reads), aiming for an average depth of 5000 read pairs per cell.

Demultiplexing cells to donors

To determine the identity of the individual donor of every cell, we used the Bayesian demultiplexing tool Vireo (v0.4.2, R version). In brief, we first generated a list of single-nucleotide polymorphism (SNP) positions by aligning all expressed reads from all cells and by selecting positions with a minimum allele frequency of 0.1 and minimum total coverage of 20. Next, in each cell and at each position, we identified overlapping SNPs and counted them in two disjoint groups corresponding to the reference and nonreference alleles. The allelic count matrices were then used to fit a Vireo

model that either identified the most likely donor for each cell or classified the cell as a doublet.

Processing of scRNA-seq data

Sequencing data were demultiplexed and mapped against the GRCh38 genome using Cell Ranger 3.0.0. The resulting gene count matrices were loaded into R using Seurat v3. On average, gene expression was detected for 1000 to 4000 genes per cell. After investigating the overall gene expression distributions for every sample, we excluded DN1 cells with <100 genes; DN2, DN3, and $CD3^-$ DP cells with <200 genes; ISP cells with <150 genes; $CD3^+$ DP cells with <85 genes; and $CD4^+$ and $CD8^+$ SP cells with <75 genes. Low-quality genes were defined as being expressed in fewer than three cells. We used the Vireo method based on genotyping to identify the donors in silico and used the genotyping profiles to detect doublets. We then corrected the data for low-quality cells and removed cells without a genotype, cells with low gene expression, and doublets. We also corrected for donor effects using canonical correlation analysis (CCA), an integrative method in Seurat version 3.2. Before performing the PCA, we performed cell cycle regression.

After filtration, we merged the datasets from all eight thymocyte subsets using the MergeSeurat function. We then split this merged dataset into six subsets on the basis of the six identified genotypes. We used CCA in Seurat to identify common sources of variation between the donors. After identifying anchors using 40 principal components, the dataset was integrated, and the integrated dataset was used for dimensionality reduction. Before running the PCA dimensionality reduction, cell cycle scores were calculated and used for cell cycle regression based on the expression of genes involved in the cell cycle. The residuals were used to scale the dataset. To maintain the global structure within the dataset as much as possible, we then visualized cells using a two-dimensional UMAP plot generated using the RunUMAP function in Seurat package with 10 principal components. We used the PercentageFeatureSet function to calculate the percentage of mitochondrial gene expression per cell.

Reanalysis was performed on the immature thymocyte populations ("sort 1": DN1, DN2, and DN3) and on the more mature thymocyte populations ("sort 2": ISP, DP_$CD3^-$, DP_$CD3^+$, CD4, and CD8). Both datasets started with the raw data and were reanalyzed using the same steps mentioned above. In addition, we removed ribosomal genes, mitochondrial genes, histone genes, and *XIST* (X-inactive specific transcript) from the original count matrix. Low-quality cells were defined as <1000 unique molecular identifier (UMI) counts and >30% mitochondrial reads.

After PCA of the sort 1 dataset, 35 principal components were used for UMAP generation and clustering analysis with the FindNeighbours (dims 1:35) and FindClusters function (resolution = 1), which identified 19 clusters. After scaling the RNA data, we used the FindAllMarkers function in Seurat was used to identify differentially expressed genes between the 19 clusters. Cluster 19 was identified as a small cluster of stromal cells and therefore excluded from subsequent analysis. Next, FindAllMarkers was again run on the remaining 18 clusters. Reanalysis of the sort 2 dataset was performed in Seurat using the same parameters as sort 1, identifying 18 clusters that were used for cluster marker analysis using FindAllMarkers in Seurat.

Single-cell TCR repertoire analysis

Raw reads from the paired V(D)J sequencing runs were processed using cellranger_vdj in Cell Ranger (v.3.0.0) with a custom reference provided by the manufacturer (version 2.0.0 GRCh38 VDJ-alts-ensembl). V(D)J sequence information was then extracted from the all_contig_annotations.csv output file.

Chains that contained the full-length recombinant sequence and were supported by >2 UMI counts were selected and linked to the cellular transcriptome data based on the cell barcodes. These chains were considered productive if a functional open reading frame covering the CDR3 region could be identified.

Integrated analysis of the thymus and BM datasets

We downloaded the Human Cell Atlas (HCA) BM immune census dataset from https://data.humancellatlas.org/explore/projects/cc95ff89-2e68-4a08-a234-480eca21ce79; this dataset consists of approximately 300,000 cells from eight donors. We removed ribosomal genes, mitochondrial genes, histone genes, and *XIST* from the original count matrix and randomly down-sampled the dataset to 40,000 cells (5000 cells per donor). In addition, a CD34-enriched BM dataset was downloaded from https://data.humancellatlas.org/explore/projects/091cf39b-01bc-42e5-9437-f419a66c8a45; this dataset consists of approximately 30,000 cells from three donors. We also removed ribosomal genes, mitochondrial genes, histone genes, and *XIST* from the original count matrix from this dataset. Both datasets were further processed in Seurat using the same steps used to analyze the thymus data. In addition to the cell annotations provided in the original meta-data, we annotated the cells from both datasets by comparing the differential markers between clusters against cell markers for 35 transcriptionally coherent cell populations (*73*) in the HCA census dataset, and for further subclassification of cell types, we performed differential expression of the annotated cells within the proteo-genomic BM reference maps (*47*).

After cell annotation, we merged both BM datasets with our complete thymus dataset to generate a single combined Seurat object. We then performed CCA between our dataset and public datasets using the donor information to identify common sources of variation. We aligned the data using 30 CCA dimensions, regressed out the cell cycle effects, performed PCA, and then generated an integrated three-dimensional UMAP with 30 principal components. We used BBrowser (version 2.9.23) (*53*) to visualize and analyze the data.

Integrative analysis of published TSP populations

We downloaded the raw data from the publication of immature human thymocytes deposited in Gene Expression Omnibus (GEO) with accession number GSE144870. Data were processed using the same steps used to analyze the dataset generated for this publication. We then selected the TSP populations in the analyzed dataset based on the markers as shown in the heatmap in Fig. 5J from the publication (*21*) combined with the positioning of the populations in the UMAP as shown in Fig. 6E and fig. S6 from the same publication. We merged and integrated the cells of the two TSP populations with our DN123 dataset using CCA and generated a new UMAP based on 30 principal components.

Automated annotation of TSP subtypes

Version 4 of Seurat introduced weighted nearest neighbor (WNN) analysis as a strategy to integrate multimodal single-cell sequencing

data and included a large multimodal BM dataset of 25 cell surface markers with simultaneous whole-transcriptome measurements (*74*). Using WNN analysis to learn the relative utility of each data modality in each cell, this reference dataset contains highly robust cell type annotations. The FindTransferAnchors and MapQuery functions were used to map our DN1/2/3 dataset to the annotated multimodal reference (*74*). In brief, anchors between the query and the reference dataset were identified using a precomputed supervised PCA on the reference dataset to maximally capture the structure of the WNN graph. Next, cell type labels from the reference dataset, as well as imputations of all measured protein markers, were transferred to each cell of the query using the previously identified anchors. Last, the query dataset was projected onto the UMAP structure of the reference dataset. Using the differential expressed gene markers between the $CD34^+$ progenitor populations (*47*), we subannotated the $CD34^+$ cell types in the BM multimodal reference set (*74*).

Pseudotime trajectory analysis

For velocity analysis, we generated loom files for each of the eight datasets using velocyto (version 0.17.17). Using the ReadVelocity function in the SeuratWrappers library, we added the splicing information from the cell stored in the loom files as new assays to the individual SeuratObjects. We then used the anndata2ri package in Python to convert the Seurat object of sort 1 (DN1, DN2, and DN3) to an AnnData object. This converted all of the results generated in Seurat (including the velocity, PCA, UMAP, and clustering data) to a format that can be used for trajectory analysis with the Scanpy toolkit. Using Scanpy, the count matrix was normalized for each cell and then log-normalized. We used scVelo (https://github.com/theislab/scvelo) to calculate the first- and second-order moments for each cell across its nearest neighbors [scvelo.pp.moments (n_pcs = 30 and n_neighbors = 50)]. Next, the velocities were estimated, and the velocity graph was constructed using the scvelo.tl.velocity function in the dynamic mode and the scvelo.tl.velocity_graph functions. Velocities were visualized using the scvelo.tl.velocity_embedding function on top of the UMAP coordinates that were calculated in Seurat.

For pseudotime ordering of the cells, we used a DPT that infers progression of the cells through a geodesic distance along a diffusion map that was calculated from 10 principal components. From this diffusion map, the DPT was computed using the tl.dpt function from Scanpy using 10 diffusion components and by setting the root at cluster 1. Together with scVelo, we defined four developmental paths within the sort 1 (DN1/2/3) dataset. We then reordered the cells for each developmental path before generating gene expression heatmaps.

Analysis of transcriptional regulation

We used SCENIC v.1.1.2 in R (*40*) to infer the Gene Regulatory Networks in our dataset. SCENIC uses the raw count matrix as the input and filters out genes with fewer than three UMIs in 1% of the cells and genes that are detected in at least 1% of the cells. This resulted in a matrix of 220 potential transcriptional regulators from which a gene regulatory network was created using GENIE3/GRNBoost, and potential regulons were selected using DNA-motif analysis. We then analyzed the activity of regulons across the cells using AUCell and subsequently visualized the results by plotting the average regulon activity in a heatmap for each cluster in our DN1/2/3 dataset.

Genomic DNA isolation and qPCR

Genomic DNA was extracted from sorted cell suspensions using the QIAAmp DNA Micro Kit (QIAGEN), and DNA concentration was measured using a NanoDrop (Thermo Fisher Scientific). qPCR was performed using the TaqMan Universal Master Mix II with uracil-N-glycoslyase (Thermo Fisher Scientific) in combination with the primers and probes listed in table S2 (*1*). All PCR reactions were run on a QuantStudio 3 real-time PCR system (Thermo Fisher Scientific).

Cultured ATOs

ATOs were initiated with 3750 sorted TSPs or $CD34^+$-enriched hematopoietic progenitor cells from cord blood and mixed with 75,000 MS5-hDLL1 cells [murine stromal cell line (MS5) transduced with human Delta-like ligand 1, Millipore, SCC167]. The ATOs were placed in 0.4-μm cell culture inserts (Millipore) in six-well plates and cultured in RPMI-1640 medium (1 ml per well; Thermo Fisher Scientific) containing 4% B-27 (Life Technologies), 30 μM ascorbic acid (Sigma-Aldrich), 1× penicillin/streptomycin (Gibco), 1× GlutaMAX (Thermo Fisher Scientific), human IL-7 (5 ng/ml; Miltenyi Biotec), and human FMS-related tyrosine kinase 3 ligand (FLT3L) (5 ng/ml; Miltenyi Biotec) as described previously (*50*).

Supplementary Materials

This PDF file includes:
Tables S1 and S2
Figs. S1 to S7

Other Supplementary Material for this manuscript includes the following:
Movies S1 to S6
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. W. A. Dik, K. Pike-Overzet, F. Weerkamp, D. de Ridder, E. F. de Haas, M. R. Baert, P. van der Spek, E. E. Koster, M. J. Reinders, J. J. van Dongen, A. W. Langerak, F. J. 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).
2. F. Weerkamp, J. J. van Dongen, F. J. Staal, Notch and Wnt signaling in T-lymphocyte development and acute lymphoblastic leukemia. *Leukemia* **20**, 1197–1205 (2006).
3. F. Weerkamp, M. R. Baert, M. H. Brugman, W. A. Dik, E. F. de Haas, T. P. Visser, C. J. de Groot, G. Wagemaker, J. J. van Dongen, F. J. Staal, Human thymus contains multipotent progenitors with T/B lymphoid, myeloid, and erythroid lineage potential. *Blood* **107**, 3131–3137 (2006).
4. F. Famili, A. S. Wiekmeijer, F. J. Staal, The development of T cells from stem cells in mice and humans. *Future Sci. OA* **3**, FSO186 (2017).
5. H. Spits, B. Blom, A. C. Jaleco, K. Weijer, M. C. Verschuren, J. J. van Dongen, M. H. Heemskerk, P. C. Res, Early stages in the development of human T, natural killer and thymic dendritic cells. *Immunol. Rev.* **165**, 75–86 (1998).
6. R. Haddad, P. Guardiola, B. Izac, C. Thibault, J. Radich, A. L. Delezoide, C. Baillou, F. M. Lemoine, J. C. Gluckman, F. Pflumio, B. Canque, Molecular characterization of early human T/NK and B-lymphoid progenitor cells in umbilical cord blood. *Blood* **104**, 3918–3926 (2004).
7. K. Cante-Barrett, R. D. Mendes, Y. Li, E. Vroegindeweij, K. Pike-Overzet, T. Wabeke, A. W. Langerak, R. Pieters, F. J. Staal, J. P. Meijerink, Loss of $CD44^{dim}$ expression from early

progenitor cells marks T-cell lineage commitment in the human thymus. *Front. Immunol.* **8**, 32 (2017).
8. H. T. Petrie, P. W. Kincade, Many roads, one destination for T cell progenitors. *J. Exp. Med.* **202**, 11–13 (2005).
9. V. C. Martins, K. Busch, D. Juraeva, C. Blum, C. Ludwig, V. Rasche, F. Lasitschka, S. E. Mastitsky, B. Brors, T. Hielscher, H. J. Fehling, H. R. Rodewald, Cell competition is a tumour suppressor mechanism in the thymus. *Nature* **509**, 465–470 (2014).
10. H. R. Rodewald, K. Kretzschmar, S. Takeda, C. Hohl, M. Dessing, Identification of pro-thymocytes in murine fetal blood: T lineage commitment can precede thymus colonization. *EMBO J.* **13**, 4229–4240 (1994).
11. A. Krueger, H. von Boehmer, Identification of a T lineage-committed progenitor in adult blood. *Immunity* **26**, 105–116 (2007).
12. A. S. Wiekmeijer, K. Pike-Overzet, H. IJspeert, M. H. Brugman, I. L. Wolvers-Tettero, A. C. Lankester, R. G. Bredius, J. J. van Dongen, W. E. Fibbe, A. W. Langerak, M. van der Burg, F. J. Staal, Identification of checkpoints in human T-cell development using severe combined immunodeficiency stem cells. *J. Allergy Clin. Immunol.* **137**, 517–526.e3 (2016).
13. 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).
14. 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).
15. J. E. Park, R. A. Botting, C. Dominguez Conde, D. M. Popescu, M. Lavaert, D. J. Kunz, I. Goh, E. Stephenson, R. Ragazzini, E. Tuck, A. Wilbrey-Clark, K. Roberts, V. R. Kedlian, J. R. Ferdinand, X. He, S. Webb, D. Maunder, N. Vandamme, K. T. Mahbubani, K. Polanski, L. Mamanova, L. Bolt, D. Crossland, F. de Rita, A. Fuller, A. Filby, G. Reynolds, D. Dixon, K. Saeb-Parsy, S. Lisgo, D. Henderson, R. Vento-Tormo, O. A. Bayraktar, R. A. Barker, K. B. Meyer, Y. Saeys, P. Bonfanti, S. Behjati, M. R. Clatworthy, T. Taghon, M. Haniffa, S. A. Teichmann, A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020).
16. M. Lee, E. Lee, S. K. Han, Y. H. Choi, D. I. Kwon, H. Choi, K. Lee, E. S. Park, M. S. Rha, D. J. Joo, E. C. Shin, S. Kim, J. K. Kim, Y. J. Lee, Single-cell RNA sequencing identifies shared differentiation paths of mouse thymic innate T cells. *Nat. Commun.* **11**, 4367 (2020).
17. T. Ilicic, J. K. Kim, A. A. Kolodziejczyk, F. O. Bagger, D. J. McCarthy, J. C. Marioni, S. A. Teichmann, Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (2016).
18. L. Li, M. Leid, E. V. Rothenberg, An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* **329**, 89–93 (2010).
19. B. N. Weber, A. W. Chi, A. Chavez, Y. Yashiro-Ohtani, Q. Yang, O. Shestova, A. Bhandoola, A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* **476**, 63–68 (2011).
20. L. Garcia-Perez, F. Famili, M. Cordes, M. Brugman, M. van Eggermond, H. Wu, J. Chouaref, D. S. L. Granado, M. M. Tiemessen, K. Pike-Overzet, L. Daxinger, F. J. T. Staal, Functional definition of a transcription factor hierarchy regulating T cell lineage commitment. *Sci. Adv.* **6**, eaaw7313 (2020).
21. M. Lavaert, K. L. Liang, N. Vandamme, J. E. Park, J. Roels, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, B. Vandekerckhove, G. Leclercq, A. Regev, P. Van Vlierberghe, M. Guilliams, S. A. Teichmann, Y. Saeys, T. Taghon, Integrated scRNA-seq identifies human postnatal thymus seeding progenitors and regulatory dynamics of Differentiating Immature Thymocytes. *Immunity* **52**, 1088–1104.e6 (2020).
22. J. Le, J. E. Park, V. L. Ha, A. Luong, S. Branciamore, A. S. Rodin, G. Gogoshin, F. Li, Y. E. Loh, V. Camacho, S. B. Patel, R. S. Welner, C. Parekh, Single-cell RNA-seq mapping of human thymopoiesis reveals lineage specification trajectories and a commitment spectrum in T cell development. *Immunity* **52**, 1105–1118.e9 (2020).
23. H. Shigematsu, B. Reizis, H. Iwasaki, S. Mizuno, D. Hu, D. Traver, P. Leder, N. Sakaguchi, K. Akashi, Plasmacytoid dendritic cells activate lymphoid-specific genetic programs irrespective of their cellular origin. *Immunity* **21**, 43–53 (2004).
24. L. Galluzzi, O. Kepp, G. Kroemer, Mitochondria: Master regulators of danger signalling. *Nat. Rev. Mol. Cell Biol.* **13**, 780–788 (2012).
25. P. Kisielow, H. von Boehmer, Development and selection of T cells: Facts and puzzles. *Adv. Immunol.* **58**, 87–209 (1995).
26. C. D. Surh, J. Sprent, T-cell apoptosis detected in situ during positive and negative selection in the thymus. *Nature* **372**, 100–103 (1994).
27. J. M. Coquet, J. C. Ribot, N. Babala, S. Middendorp, G. van der Horst, Y. Xiao, J. F. Neves, D. Fonseca-Pereira, H. Jacobs, D. J. Pennington, B. Silva-Santos, J. Borst, Epithelial and dendritic cells in the thymic medulla promote CD4+Foxp3+ regulatory T cell development via the CD27-CD70 pathway. *J. Exp. Med.* **210**, 715–728 (2013).
28. L. A. Kohn, Q. L. Hao, R. Sasidharan, C. Parekh, S. Ge, Y. Zhu, H. K. Mikkola, G. M. Crooks, Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin. *Nat. Immunol.* **13**, 963–971 (2012).
29. E. F. Lind, S. E. Prockop, H. E. Porritt, H. T. Petrie, Mapping precursor movement through the postnatal thymus reveals specific microenvironments supporting defined stages of early lymphoid development. *J. Exp. Med.* **194**, 127–134 (2001).
30. K. Mori, M. Itoi, N. Tsukamoto, H. Kubo, T. Amagai, The perivascular space as a path of hematopoietic progenitor cells and mature T cells between the blood circulation and the thymic parenchyma. *Int. Immunol.* **19**, 745–753 (2007).
31. A. Bhandoola, A. Sambandam, From stem cell to T cell: One route or many? *Nat. Rev. Immunol.* **6**, 117–126 (2006).
32. P. Robertson, T. K. Means, A. D. Luster, D. T. Scadden, CXCR4 and CCR5 mediate homing of primitive bone marrow-derived hematopoietic cells to the postnatal thymus. *Exp. Hematol.* **34**, 308–319 (2006).
33. B. A. Schwarz, A. Sambandam, I. Maillard, B. C. Harman, P. E. Love, A. Bhandoola, Selective thymus settling regulated by cytokine and chemokine receptors. *J. Immunol.* **178**, 2008–2017 (2007).
34. A. Krueger, S. Willenzon, M. Lyszkiewicz, E. Kremmer, R. Forster, CC chemokine receptor 7 and 9 double-deficient hematopoietic progenitors are severely impaired in seeding the adult thymus. *Blood* **115**, 1906–1912 (2010).
35. D. A. Zlotoff, A. Sambandam, T. D. Logan, J. J. Bell, B. A. Schwarz, A. Bhandoola, CCR7 and CCR9 together recruit hematopoietic progenitors to the adult thymus. *Blood* **115**, 1897–1905 (2010).
36. A. C. Carpenter, R. Bosselut, Decision checkpoints in the thymus. *Nat. Immunol.* **11**, 666–673 (2010).
37. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lonnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundstrom, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
38. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
39. L. Haghverdi, M. Buttner, F. A. Wolf, F. Buettner, F. J. Theis, Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
40. S. Aibar, C. B. Gonzalez-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J. C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
41. F. Weerkamp, M. R. Baert, B. A. Naber, E. E. Koster, E. F. de Haas, K. R. Atkuri, J. J. van Dongen, L. A. Herzenberg, F. J. Staal, Wnt signaling in the thymus is regulated by differential expression of intracellular signaling molecules. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3322–3326 (2006).
42. P. Shooshtarizadeh, A. Helness, C. Vadnais, N. Brouwer, H. Beauchemin, R. Chen, H. Bagci, F. J. T. Staal, J. F. Cote, T. Moroy, Gfi1b regulates the level of Wnt/β-catenin signaling in hematopoietic stem cells and megakaryocytes. *Nat. Commun.* **10**, 1270 (2019).
43. I. Siamishi, N. Iwanami, T. Clapes, E. Trompouki, C. P. O'Meara, T. Boehm, Lymphocyte-specific function of the DNA polymerase epsilon subunit pole3 revealed by neomorphic alleles. *Cell Rep.* **31**, 107756 (2020).
44. N. Fujita, D. L. Jaye, C. Geigerman, A. Akyildiz, M. R. Mooney, J. M. Boss, P. A. Wade, MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. *Cell* **119**, 75–86 (2004).
45. J. Zhang, M. Wencker, Q. Marliac, A. Berton, U. Hasan, R. Schneider, D. Laubreton, D. E. Cherrier, A.-L. Mathieu, A. Rey, W. Jiang, J. Caramel, L. Genestier, A. Marçais, J. Marvel, Y. Ghavi-Helm, T. Walzer, Zeb1 represses TCR signaling, promotes the proliferation of T cell progenitors and is essential for NK1.1$^+$ T cell development. *Cell. Mol. Immunol.* **18**, 2140–2152 (2021).
46. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck III, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
47. S. Triana, D. Vonficht, L. Jopp-Saile, S. Raffel, R. Lutz, D. Leonce, M. Antes, P. Hernandez-Malmierca, D. Ordonez-Rueda, B. Ramasz, T. Boch, J. C. Jann, D. Nowak, W. K. Hofmann, C. Muller-Tidow, D. Hubschmann, T. Alexandrov, V. Benes, A. Trumpp, M. Paulsen, L. Velten, S. Haas, Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* **22**, 1577–1589 (2021).
48. I. Hoebeke, M. De Smedt, F. Stolz, K. Pike-Overzet, F. J. Staal, J. Plum, G. Leclercq, T-, B- and NK-lymphoid, but not myeloid cells arise from human CD34$^+$CD38$^-$CD7$^+$ common lymphoid progenitors expressing lymphoid-specific genes. *Leukemia* **21**, 311–319 (2007).

49. H. J. Chen, C. M. Lin, C. S. Lin, R. Perez-Olle, C. L. Leung, R. K. Liem, The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. *Genes Dev.* **20**, 1933–1945 (2006).
50. C. S. Seet, C. He, M. T. Bethune, S. Li, B. Chick, E. H. Gschweng, Y. Zhu, K. Kim, D. B. Kohn, D. Baltimore, G. M. Crooks, A. Montel-Hagen, Generation of mature T cells from human hematopoietic stem and progenitor cells in artificial thymic organoids. *Nat. Methods* **14**, 521–530 (2017).
51. A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Eils, W. Enard, A. Farmer, L. Fugger, B. Gottgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klenerman, A. Kriegstein, E. Lein, S. Linnarsson, E. Lundberg, J. Lundeberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijn, M. Netea, G. Nolan, D. Pe'er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T. N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. T. Stubbington, F. J. Theis, M. Uhlen, A. van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, N. Yosef; Human Cell Atlas Meeting Participants, The human cell atlas. *eLife* **6**, e27041 (2017).
52. M. Setty, V. Kiseliovas, J. Levine, A. Gayoso, L. Mazutis, D. Pe'er, Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).
53. T. Le, T. Phan, M. Pham, D. Tran, L. Lam, T. Nguyen, T. Truong, H. Vuong, T. Luu, N. Phung, N. Pham, T. Nguyen, O. Pham, A. Nguyen, H. Nguyen, H. Tran, L. Tran, H. A. Nguyen, T. Tran, N. Nguyen, N. Tran, C. Boysen, U. Nguyen, V. Pham, T. Kim, N. Pham, T. Gill, S. Pham, BBrowser: Making single-cell data easily accessible. *bioRxiv* 414136. 11 December 2020. https://doi.org/10.1101/2020.12.11.414136.
54. T. M. Breit, I. L. Wolvers-Tettero, A. Beishuizen, M. A. Verhoeven, E. R. van Wering, J. J. van Dongen, Southern blot patterns, frequencies, and junctional diversity of T-cell receptor-delta gene rearrangements in acute lymphoblastic leukemia. *Blood* **82**, 3063–3074 (1993).
55. T. Szczepański, A. W. Langerak, I. L. Wolvers-Tettero, G. J. Ossenkoppele, G. Verhoef, M. Stul, E. J. Petersen, M. A. de Bruijn, M. B. van't Veer, J. J. van Dongen, Immunoglobulin and T cell receptor gene rearrangement patterns in acute lymphoblastic leukemia are less mature in adults than in children: Implications for selection of PCR targets for detection of minimal residual disease. *Leukemia* **12**, 1081–1088 (1998).
56. H. Hosokawa, M. Romero-Wolf, M. A. Yui, J. Ungerback, M. L. G. Quiloan, M. Matsumoto, K. I. Nakayama, T. Tanaka, E. V. Rothenberg, Bcl11b sets pro-T cell fate by site-specific cofactor recruitment and by repressing Id2 and Zbtb16. *Nat. Immunol.* **19**, 1427–1440 (2018).
57. M. E. Garcia-Ojeda, R. G. Klein Wolterink, F. Lemaitre, O. Richard-Le Goff, M. Hasan, R. W. Hendriks, A. Cumano, J. P. Di Santo, GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice. *Blood* **121**, 1749–1759 (2013).
58. V. L. Ha, A. Luong, F. Li, D. Casero, J. Malvar, Y. M. Kim, R. Bhatia, G. M. Crooks, C. Parekh, The T-ALL related gene BCL11B regulates the initial stages of human T-cell differentiation. *Leukemia* **31**, 2503–2514 (2017).
59. T. Taghon, M. A. Yui, E. V. Rothenberg, Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nat. Immunol.* **8**, 845–855 (2007).
60. P. Res, E. Martínez-Cáceres, A. Cristina Jaleco, F. Staal, E. Noteboom, K. Weijer, H. Spits, CD34+CD38dim cells in the human thymus can differentiate into T, natural killer, and dendritic cells but are distinct from pluripotent stem cells. *Blood* **87**, 5196–5206 (1996).
61. M. H. Heemskerk, B. Blom, G. Nolan, A. P. Stegmann, A. Q. Bakker, K. Weijer, P. C. Res, H. Spits, Inhibition of T cell and promotion of natural killer cell development by the dominant negative helix loop helix factor Id3. *J. Exp. Med.* **186**, 1597–1602 (1997).
62. K. Weijer, C. H. Uittenbogaart, A. Voordouw, F. Couwenberg, J. Seppen, B. Blom, F. A. Vyth-Dreese, H. Spits, Intrathymic and extrathymic development of human plasmacytoid dendritic cell precursors in vivo. *Blood* **99**, 2752–2759 (2002).
63. S. Luc, T. C. Luis, H. Boukarabila, I. C. Macaulay, N. Buza-Vidas, T. Bouriez-Jones, M. Lutteropp, P. S. Woll, S. J. Loughran, A. J. Mead, A. Hultquist, J. Brown, T. Mizukami, S. Matsuoka, H. Ferry, K. Anderson, S. Duarte, D. Atkinson, S. Soneji, A. Domanski, A. Farley, A. Sanjuan-Pla, C. Carella, R. Patient, M. de Bruijn, T. Enver, C. Nerlov, C. Blackburn, I. Godin, S. E. Jacobsen, The earliest thymic T cell progenitors sustain B cell and myeloid lineage potential. *Nat. Immunol.* **13**, 412–419 (2012).
64. J. Castaneda, Y. Hidalgo, D. Sauma, M. Rosemblatt, M. R. Bono, S. Nunez, The multifaceted roles of B cells in the thymus: From immune tolerance to autoimmunity. *Front. Immunol.* **12**, 766698 (2021).
65. K. Akashi, L. I. Richie, T. Miyamoto, W. H. Carr, I. L. Weissman, B lymphopoiesis in the thymus. *J. Immunol.* **164**, 5221–5226 (2000).
66. F. Weerkamp, E. F. de Haas, B. A. Naber, W. M. Comans-Bitter, A. J. Bogers, J. J. van Dongen, F. J. Staal, Age-related changes in the cellular composition of the thymus in children. *J. Allergy Clin. Immunol.* **115**, 834–840 (2005).
67. I. Ferrero, F. Anjuère, P. Martín, G. Martínez del Hoyo, M. L. Fraga, N. Wright, R. Varona, G. Márquez, C. Ardavín, Functional and phenotypic analysis of thymic B cells: Role in the induction of T cell negative selection. *Eur. J. Immunol.* **29**, 1598–1609 (1999).
68. S. Nuñez, C. Moore, B. Gao, K. Rogers, Y. Hidalgo, P. J. Del Nido, S. Restaino, Y. Naka, G. Bhagat, J. C. Madsen, M. R. Bono, E. Zorn, The human thymus perivascular space is a functional niche for viral-specific plasma cells. *Sci. Immunol.* **1**, (2016).
69. V. C. Martins, E. Ruggiero, S. M. Schlenner, V. Madan, M. Schmidt, P. J. Fink, C. von Kalle, H. R. Rodewald, Thymus-autonomous T cell development in the absence of progenitor import. *J. Exp. Med.* **209**, 1409–1417 (2012).
70. M. Kondo, I. L. Weissman, K. Akashi, Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
71. A. Chhatta, H. M. M. Mikkers, F. J. T. Staal, Strategies for thymus regeneration and generating thymic organoids. *J. Immunol. Regen. Med.* **14**, 100052 (2021).
72. F. J. T. Staal, A. Aiuti, M. Cavazzana, Autologous stem-cell-based gene therapy for inherited disorders: State of the art and perspectives. *Front. Pediatr.* **7**, 443 (2019).
73. S. B. Hay, K. Ferchen, K. Chetal, H. L. Grimes, N. Salomonis, The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp. Hematol.* **68**, 51–61 (2018).
74. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

Acknowledgments: We are grateful to the Leiden University Medical Center Flow Cytometry Core Facility for assistance with cell sorting, the Leiden Genome Technology Center, and J. Melsen for helpful discussions regarding the data analyses. **Funding:** K.P.O. and L.G.-P. are supported by H22020 grant RECOMB (755170-2), which has received funding from the European Union Horizon 2020 research and innovation program. K.C.B. and F.J.T.S. are supported in part by Novo Nordisk Foundation grants (NNF21CC0073729) that supports the Novo Nordisk Foundation Center for Stem Cell Medicine at LUMC. E.B.A. was funded by a personal grant from the Dutch Research Council (NWO; VENI: 09150161810095). **Author contributions:** M.C., K.C.-B., F.A.M., E.B.v.d.A, K.P-O., M.J.T.R., and F.J.T.S designed experiments. M.C., K.C.-B., F.A.M., S.M.K., S.A.V., and L.G.-P. performed experiments. M.C., K.C.-B., S.M.K., and C.T. analyzed the data. M.C., K.C.-B., E.B.v.d.A., K.P.-O., M.J.T.R., and F.J.T.S. interpreted the data. M.C., K.C.-B., E.B.v.d.A., J.J.M.v.D., K.P.-O., M.J.T.R., and F.J.T.S. consulted and discussed the manuscript. K.C.-B., E.B.v.d.A., K.P.O., M.J.T.R., and F.J.T.S. supervised the project. M.C., K.C.-B., E.B.v.d.A., K.P.-O., M.J.T.R., and F.J.T.S. wrote the manuscript. All authors reviewed and approved the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. All scRNA-seq data generated during this study were deposited in GEO with accession number GSE195812. This study did not use any unique codes, and all analyses were performed in R and Python using standard protocols from previously published packages. All primers and other nucleic acid sequences are provided in Materials and Methods. Requests for materials should be directed to F.J.T.S.

Submitted 19 July 2022
Accepted 18 October 2022
Published 11 November 2022
10.1126/sciimmunol.ade0182