

TECHNICAL ADVANCES AND RESOURCES

# Single-cell analyses reveal early thymic progenitors and pre-B cells in zebrafish

Sara A. Rubin<sup>1,2,3</sup>, Chloé S. Baron<sup>1,3</sup>, Cecilia Pessoa Rodrigues<sup>1,3</sup>, Madeleine Duran<sup>4</sup>, Alexandra F. Corbin<sup>1</sup>, Song P. Yang<sup>1</sup>, Cole Trapnell<sup>4</sup>, and Leonard I. Zon<sup>1,2,3,5</sup>

The zebrafish has proven to be a valuable model organism for studying hematopoiesis, but relatively little is known about zebrafish immune cell development and functional diversity. Elucidating key aspects of zebrafish lymphocyte development and exploring the breadth of effector functions would provide valuable insight into the evolution of adaptive immunity. We performed single-cell RNA sequencing on ~70,000 cells from the zebrafish marrow and thymus to establish a gene expression map of zebrafish immune cell development. We uncovered rich cellular diversity in the juvenile and adult zebrafish thymus, elucidated B- and T-cell developmental trajectories, and transcriptionally characterized subsets of hematopoietic stem and progenitor cells and early thymic progenitors. Our analysis permitted the identification of two dendritic-like cell populations and provided evidence in support of the existence of a pre-B cell state. Our results provide critical insights into the landscape of zebrafish immunology and offer a foundation for cellular and genetic studies.

## Introduction

The zebrafish is a useful immunological model, but much remains unknown about zebrafish immune cell development (Trede et al., 2004). Although work has been done to characterize zebrafish TCR and B cell receptor loci (Haire et al., 2000; Meeker et al., 2010; Seelye et al., 2016) and track the seeding and development of the thymus (Bertrand et al., 2008; Kissa et al., 2008; Lam et al., 2002; Murayama et al., 2006; Willett et al., 1997b), literature on the intricacies of T- and B-cell development is sparse and these processes are often inferred from other teleosts (Bajoghli et al., 2019; Barraza et al., 2020; Liu et al., 2017; Page et al., 2013). Mutagenesis studies have proven fruitful for identifying conserved aspects of lymphocyte development (Schorpp et al., 2006), but functional assays in teleosts often lack specificity as few monoclonal antibodies exist for cell type purification and targeted stimulation or inhibition (Lugo-Villarino et al., 2010; Mitra et al., 2010). Creative use of existing technology in zebrafish can overcome these shortcomings to more precisely characterize immune cell populations and suggest more specific targets for downstream functional studies.

Single-cell RNA sequencing (scRNA-seq) has emerged as a powerful tool to reveal underappreciated cellular diversity and developmental trajectories. Great technological advances have enabled the capture and profiling of thousands of individual cells

from varied tissues across different organisms (Svensson et al., 2018). Many computational structures and algorithms have been developed for extracting new biological insights from scRNA-seq data. This has afforded researchers opportunities to go beyond basic questions of cellular identification and tissue-specific population quantification to generate comprehensive organism-level cell atlases, discover novel subpopulations, and explore developmental processes. Trajectory-inferring algorithms, including graph-based methods like PAGA (Wolf et al., 2019), tree-based methods like Monocle (Cao et al., 2019; Qiu et al., 2017; Trapnell et al., 2014), and multifurcation methods like FateID (Herman et al., 2018), have further expanded our capability to explore questions relating to cellular ancestry. Although these methods cannot, on their own, provide definitive orderings of cellular development, their pseudotime predictions are powerful for identifying gene targets for future study. Overall, scRNA-seq provides vast opportunities to advance the field of zebrafish immunology.

Building upon the groundwork of transcriptional analyses of adult zebrafish hematopoiesis (Carmona et al., 2017; Hernandez et al., 2018; Macaulay et al., 2016; Moore et al., 2016; Tang et al., 2017), we capitalized on recent technological advances to take a deep dive into zebrafish immune cell development. We

<sup>1</sup>Stem Cell Program and Division of Hematology/Oncology, Boston Children's Hospital and Dana-Farber Cancer Institute, Boston, MA; <sup>2</sup>Department of Immunology, Blavatnik Institute, Harvard Medical School, Boston, MA; <sup>3</sup>Stem Cell and Regenerative Biology Department, Harvard University, Cambridge, MA; <sup>4</sup>Department of Genome Sciences, University of Washington, Seattle, WA; <sup>5</sup>Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA.

Correspondence to Leonard I. Zon: [zon@enders.tch.harvard.edu](mailto:zon@enders.tch.harvard.edu).

© 2022 Rubin et al. This article is distributed under the terms of an Attribution-Noncommercial-Share Alike-No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution-Noncommercial-Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

transcriptionally profiled the zebrafish kidney marrow and thymus, which uncovered new insights into zebrafish T- and B-cell development. Specifically, we observed evolutionarily conserved transcriptional regulation and developmental states, including populations consistent with early thymic progenitors (ETPs) and pre-B cells. We also identified previously uncharacterized immune cell populations in the zebrafish, including those resembling conventional type 1 dendritic cells (DCs) and plasmacytoid DCs. This knowledge builds upon our evolutionary understanding of T-cell development (Bajoghli et al., 2009), clarifies some discrepancies in the field, and better situates the zebrafish as an immunological model moving forward. Our findings reveal new genetic targets and populations prime for investigation to characterize zebrafish immunology.

## Results

### The adult zebrafish thymus exhibits rich cellular heterogeneity akin to the mammalian thymus

To characterize the transcriptional landscape of the adult zebrafish thymus, we performed droplet-based scRNA-seq (10x Genomics) on live cells sorted from whole thymi derived from four zebrafish at 3–4 mo post-fertilization (mpf). SoupX was used to mitigate the impact of ambient mRNA contamination before quality-control filtering and sample integration in Seurat (Stuart et al., 2019; Young and Behjati, 2020). A total of 14,394 cells meeting the inclusion criteria were normalized by scTransform, integrated using reciprocal principal component analysis (PCA), clustered with the nearest neighbor approach, and visualized by uniform manifold approximation and projection (UMAP; Fig. 1 A and Table S1). Multiple stages of T cell development were evident: (1) *rag1/2*<sup>+</sup> T cells, both cycling and non-cycling; (2) *rag1/2*<sup>-</sup> cycling T cells; (3) *ipcefl*<sup>+</sup> maturing T cells, with a distinct *ccr7*<sup>+</sup> subpopulation; and (4) *il2rb*<sup>+</sup> mature T cells (Fig. 1 B). Distinct subsets of T or T-like populations were also present: (1) T helper 2 (Th2) cells and/or type 2 innate lymphoid cells (ILC2s) expressing *gata3*, *il4*, and *il13*; (2) *gzmk*<sup>+</sup>, *gzma*<sup>+</sup>, *eomesa/b*<sup>+</sup> cytotoxic T cells; and (3)  $\gamma\delta$  T cells expressing *sox13*, *trdc*, and *tcr $\gamma$*  (*si:dkeyp-13d12.23*). Additionally, many non-T cell populations were identified, including thymic epithelial cells (TECs; *epcam*<sup>+</sup>, *cdh1*<sup>+</sup>), dendritic-like (DC-like) cells (DC1, *spock3*<sup>+</sup>, *spila*<sup>+</sup>; DC2, *spock3*<sup>+</sup>, *ctsbb*<sup>+</sup>), macrophages (*havcr2*<sup>+</sup>, *mfap4*<sup>+</sup>), a mixed population composed largely of granulocytes (*mpx*<sup>+</sup>, *cpa5*<sup>+</sup>), natural killer-like (NK-like) cells (*eomesa*<sup>+</sup>, *fcer1g1*<sup>+</sup>), erythrocytes (*hbaa1*<sup>+</sup>, *hbaa2*<sup>+</sup>), and B cells (*pax5*<sup>+</sup>, *cd79a*<sup>+</sup>; Fig. S1 A). B cells were found to be predominately of the IgT isotype (previously named IgZ; Dornburg et al., 2021): 63% *ighz* cluster vs. 37% *ighd* cluster. Finally, we identified an ETP population expressing genes consistent with progenitor identity, including *si:dkey-261h17.1*, a CD34 ortholog, and *csflrb*, previously reported to be expressed in monocytes, macrophages, and some granulocytes in addition to hematopoietic stem and progenitor cells (HSPCs; Hason et al., 2022). Overall, these results reveal great cell type diversity within the zebrafish thymus, on par with mammalian thymic composition.

### Juvenile and adult thymi are largely composed of the same cell types but exhibit developmental differences in cellular composition and states

To assess how thymic composition changes over time, we analyzed juvenile thymi at 4 wk post-fertilization (wpf). Live cells were sorted from a pool of dissected thymi from 21 zebrafish and four technical replicates were sequenced (Fig. 1 C, Fig. S1 B, and Table S2). By increasing cell number by integrating cells from juvenile and adult thymi, we improved resolution and identified the following: (1) thrombocytes (*itga2b*<sup>+</sup>, *thbs1b*<sup>+</sup>); (2) fibroblast-like cells (*pdgfra/b*<sup>+</sup>, *col2a1a*<sup>+</sup>); and (3) *sox13*<sup>-</sup>  $\gamma\delta$  T cells (*sox4a1*<sup>+</sup>, *gata3*<sup>+</sup>, *etv5a*<sup>-</sup>; Fig. 1 D). Greater clarity was gained on the mixed granulocyte-rich population from the adult thymus composed of neutrophil-like granulocytes (*mpx*<sup>+</sup>, *npsn*<sup>+</sup>) and eosinophils (*gata2a*<sup>+</sup>, *viml*<sup>+</sup>; Fig. 1 D and Fig. S1 C). Seeing great overlap in the cell types present in the adult and juvenile thymi, we next asked whether cell type abundances were similar at both time points. We found the juvenile thymus was enriched in *rag1/2*<sup>+</sup> T cells, *rag1/2*<sup>+</sup> cycling T cells, and *rag1/2*<sup>-</sup> cycling T cells (beta-binomial test; Fig. 1 E and Table S3). The juvenile thymus displayed greater joint expression of TCR  $\alpha/\beta$  and  $\gamma/\delta$  (Fig. S1 D). We also noted juvenile enrichments of *sox13*<sup>+</sup>  $\gamma\delta$  T cells, *sox13*<sup>-</sup>  $\gamma\delta$  T cells, and fibroblasts. Conversely, the adult thymus was enriched in maturing *ccr7*<sup>-</sup> and *ccr7*<sup>+</sup> T cells, TECs, and other non-T cell populations, including B cells, DC-like cells (1), mixed granulocytes, and macrophages. Examining global and cell-type specific gene expression between timepoints, we observed a striking difference in normalized expression of ribosomal subunit genes within adult and juvenile thymus erythrocytes, including that of *rps29*, *rps15a*, and *rplp1*, detected in <3–10% of juvenile erythrocytes vs. >85–91% of adult erythrocytes, with an average expression 12 to 20-fold higher in adult thymus erythrocytes (Fig. S1, E–G; and Table S4). These data show that although cell populations are similar between the adult and juvenile thymi, the cellular composition and transcriptional landscape of the zebrafish thymus does change over time, with adult thymi enriched in later-stage T cells and non-T cell populations and relatively fewer *rag1/2*<sup>+</sup> T cells and  $\gamma\delta$  T cells.

### Zebrafish TECs share mammalian gene expression programs

To characterize the epithelial and stromal cells in greater detail, we re-integrated, clustered, and visualized the subset of *epcam*<sup>+</sup> TECs and fibroblast-like cells (Fig. 1 F and Fig. S2, A and B; and Table S5). Analysis revealed 10 clusters, which we named in accordance with the mammalian literature (Baran-Gale et al., 2020). These populations included the following: (1) cortical TECs (cTECs; *psmb11a/b*<sup>+</sup>, *ccl25a*<sup>+</sup>); (2) medullary TECs (mTECs; *aire*<sup>+</sup>, *pvalb5*<sup>+</sup>); (3) structural TECs (*pros1*<sup>+</sup>, *col4a3*<sup>+</sup>); (4) neural TECs (*sstr5*<sup>+</sup>, *mal2*<sup>+</sup>); and (5) tuft-like mTECs (*prrr15la*<sup>+</sup>, *fybb*<sup>+</sup>). We also noted fatty acid TECs (*elovl7a*<sup>+</sup>, *elovl6l*<sup>+</sup>) and a cluster consistent with corticomedullary junction TECs (*notch3*<sup>+</sup>, *krt5*<sup>high</sup>), expressing both medullary (*ccl19b*) and cortical chemokines (*ccl25a*, *cxcl12a*; Kadouri et al., 2020). Fibroblasts were distinguished from TECs by the lack of *epcam* expression and enrichment in *pdgfra/b*, *twist1a*, and *prrx1a/b*. By beta-binomial test, we observed an enrichment of fibroblasts in the juvenile thymus, consistent with our unsubsetted results, in addition to a

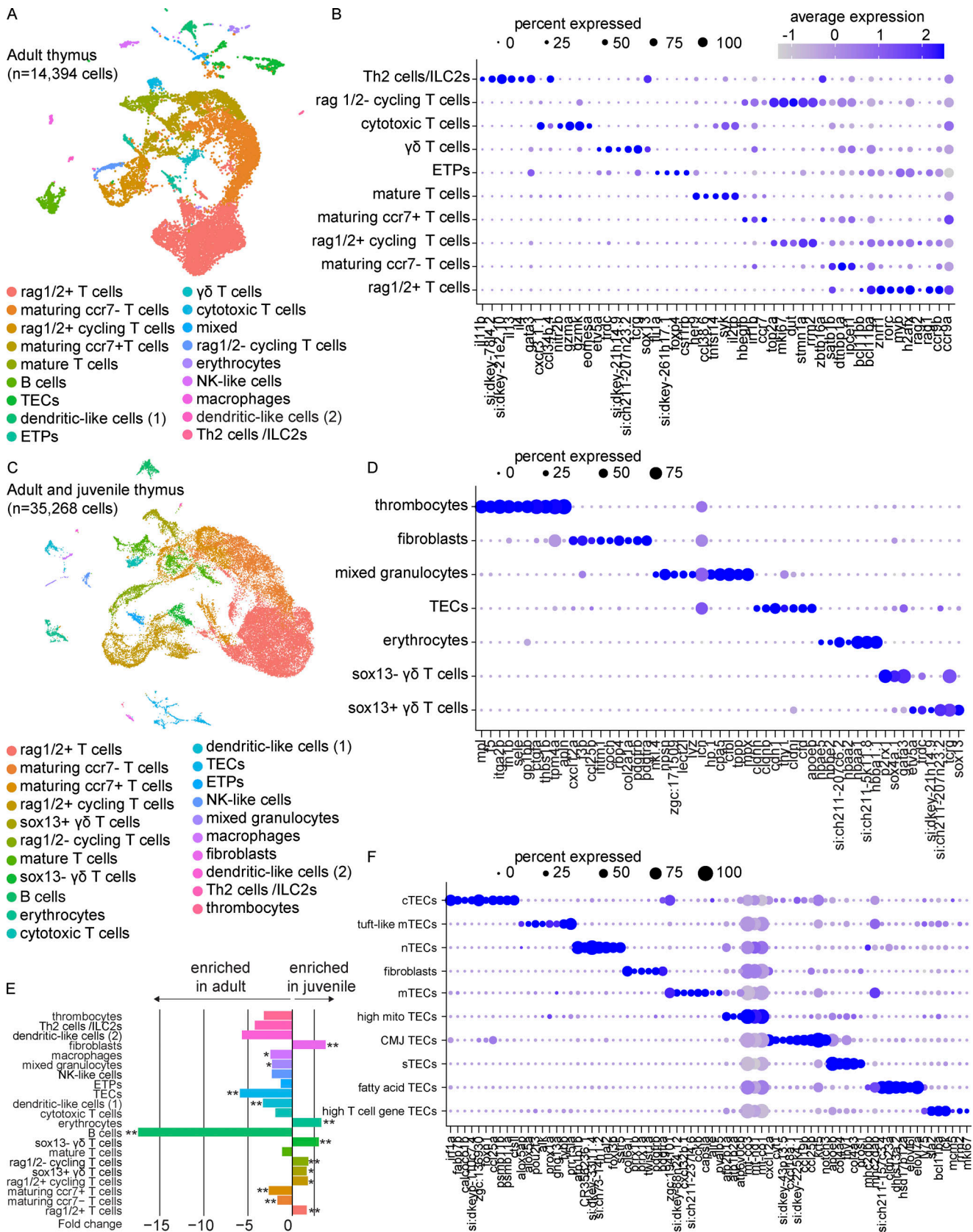


Figure 1. **Transcriptional characterization of the zebrafish thymus.** (A) UMAP visualization of 14,394 cells obtained from the thymi of four adult zebrafish (3–4 mpf). All thymi were dissected and processed within the same experiment. Cell type annotations were based on enriched markers (Table S1) that were identified by Wilcoxon rank sum test. (B) Dot plot of selected markers of the adult thymus showing expression across T cell populations. See Fig. S1 A for full

dot plot. Dot size reflects the percentage of cells within a population expressing a given marker and dot color shows the average expression within that population. **(C)** UMAP visualization of 35,268 cells obtained from the thymi of four adult zebrafish and four technical replicates of juvenile zebrafish (4 wpf; pool of cells from 21 zebrafish). Cell type annotations were based on enriched markers (Table S2) that were identified by Wilcoxon rank sum test and nomenclature consistent with (A) was used when appropriate. The adult and juvenile thymi were dissected and processed in independent experiments. **(D)** Dot plot of selected markers of the adult and juvenile thymi highlighting the expression of newly/better-resolved clusters following integration. **(E)** Bar plot of the beta-binomial test comparing cell type composition of the juvenile and adult thymi (Table S3). The average fold change for each cell type is depicted by bar length and direction, with negative values enriched in the adult thymus and positive values enriched in the juvenile thymus. Statistical significance is shown for Benjamini-Hochberg adjusted P values: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ . **(F)** Dot plot of selected markers distinguishing the epithelial and fibroblast-like populations from the subset analysis visualized by UMAP in Fig. S2 A and identified by the Wilcoxon rank sum test (Table S5). nTEC, neural TEC; mito, mitochondrial-gene expressing; CMJ, corticomedullary junction; sTEC, structural TEC.

smaller but significant enrichment of cTECs in the juvenile thymus (Fig. S2 C and Table S6). Our analysis highlights comparable diversity between zebrafish and mammalian TECs, with many putatively orthologous epithelial and stromal cell subtypes recognizable by conserved marker genes.

### Zebrafish T-cell development involves the equivalent of a CD4-CD8 double-positive state

To investigate T-cell developmental trajectories, we performed FateID analysis (Herman et al., 2018) on a RaceID3 clustering of the adult thymus T cell populations in addition to Monocle 3 trajectory analysis (Cao et al., 2019). FateID analysis resolved three trajectories: trajectory 1 of the mature T cells, enriched in *il2rb* and *gata3* expression; trajectory 19 of the cytotoxic T cells, enriched in *gzma* and *eomesa* expression; and trajectory 5 of the  $\gamma\delta$  T cells (Fig. 2, A-E; and Fig. S2, D and E). We focused our downstream analysis on the more robustly inferred mature T cell (t1) and cytotoxic T cell (t19) trajectories. Self-organizing maps (SOMs) for t1 and t19 along with pseudotime plots of notable genes in T cell development revealed that the peak detection of CD4 and CD8 orthologs occurred within the *rag1/2*<sup>+</sup> T cells with reduced levels of expression in both mature and cytotoxic T cells (Fig. 2, D and E, middle and right panels; and Table S7). Both trajectories demonstrated relatively high levels of *csflrb* expression in the cluster assigned earliest in pseudotime (cluster 7, ETPs) and elevated levels of *il2rb* in the most mature populations (Fig. 2, D and E, right panels). These results were consistent with an independent trajectory analysis of the mature and cytotoxic T cells in Monocle 3 (Fig. 3, A-E). Furthermore, we identified cells that expressed at least one ortholog of CD4 concomitantly with at least one ortholog of CD8, reminiscent of the CD4-CD8 double-positive state in mammals (Fig. 3, D and E). These “double-positive” cells were identified predominantly within the *rag1/2*<sup>+</sup> T cells, *rag1/2*<sup>+</sup> cycling T cells, and maturing *ccr7*<sup>-</sup> T cells. To our knowledge, these results provide the first demonstration of CD4-CD8 ortholog co-expression in zebrafish thymus at the single-cell level.

### ETPs retain myeloid gene expression but are enriched in T cell commitment genes

To better understand the earliest stages of T-cell development from HSPCs in the kidney marrow, we performed a large-scale scRNA-seq characterization of sorted populations from marrows of five adult (3–4 mpf) zebrafish. Consistent with previous work, from the 34,492 cells profiled, we identified cells from all major

hematopoietic lineages, including erythrocytes, thrombocytes, granulocytes (neutrophil-like and eosinophil-like), monocytes, macrophages, T cells, B cells, and NK-like cells (Fig. 4, A-C; and Table S8). In addition to the canonical markers and cell populations previously reported, we identified *si:ch211-207c6.2* as an erythrocyte-specific pre-microRNA (pre-miRNA), *miRNA144/451* (Dore et al., 2008; Fig. 4 D) and recovered two populations of DC-like cells. Focusing specifically on characterizing the thymic ETPs and their relationship to marrow HSPCs, we next integrated cells from adult thymi and progenitor and lymphoid cells from adult marrows ( $n = 35,464$  cells; Fig. 5, A and B). Thymic ETPs clustered in cluster 4, which was identified as HSPCs for its expression of *csflrb*, *flila*, and a CD34 ortholog (Fig. 5 C and Fig. S2, F and G). We additionally note that these cells characterized as ETPs in our Seurat analysis share great consistency with our RaceID3 ETP assignment (cluster 7; Fig. 5 D). StemID2 analysis revealed that cluster 7 had one of the highest StemID scores, greatest in clusters with high transcriptome entropy (i.e., uncertainty in differentiation) and connectedness to other clusters, consistent with the progenitor identity (Grun et al., 2016; Herman et al., 2018; Fig. 5 E). Cluster 7 was unbiased toward the three trajectories and was enriched in pathways related to hematopoietic stem cell differentiation but not T-cell differentiation like cluster 11 (*rag1/2*<sup>+</sup> T cells), the other high StemID scoring cluster (Fig. 5, F and G). Further support for the successful identification of zebrafish ETPs came from the analysis of gene modules. Specifically, two independently determined HSPC modules from our marrow investigation (see Materials and methods) and a literature-based ETP gene list from mice and humans all had the greatest module scores in the cells we called ETPs (Fig. 5, H-J; and Table S9; Cante-Barrett et al., 2017; Cordes et al., 2022 Preprint; Porritt et al., 2004; Rothenberg et al., 2008; Zeng et al., 2019). This evidence strongly suggests the successful identification of zebrafish ETPs in our analysis.

Having identified zebrafish ETPs, we next sought to compare the gene expression of these ETPs with marrow HSPCs. While analyzing and comparing the progenitor cells in these two hematopoietic organs is insufficient to inform us of the transitory states between marrow HSPCs and thymic ETPs, this comparison does allow us to identify changes in gene expression that occur at some point between these two states. This analysis identified 10 genes as being highly conserved (Fig. 5 K, Fig. S3, and Table S10). These genes included those whose mammalian orthologs have known stem cell functions like *plk2b*, which is highly expressed in aged hematopoietic stem cells and involved

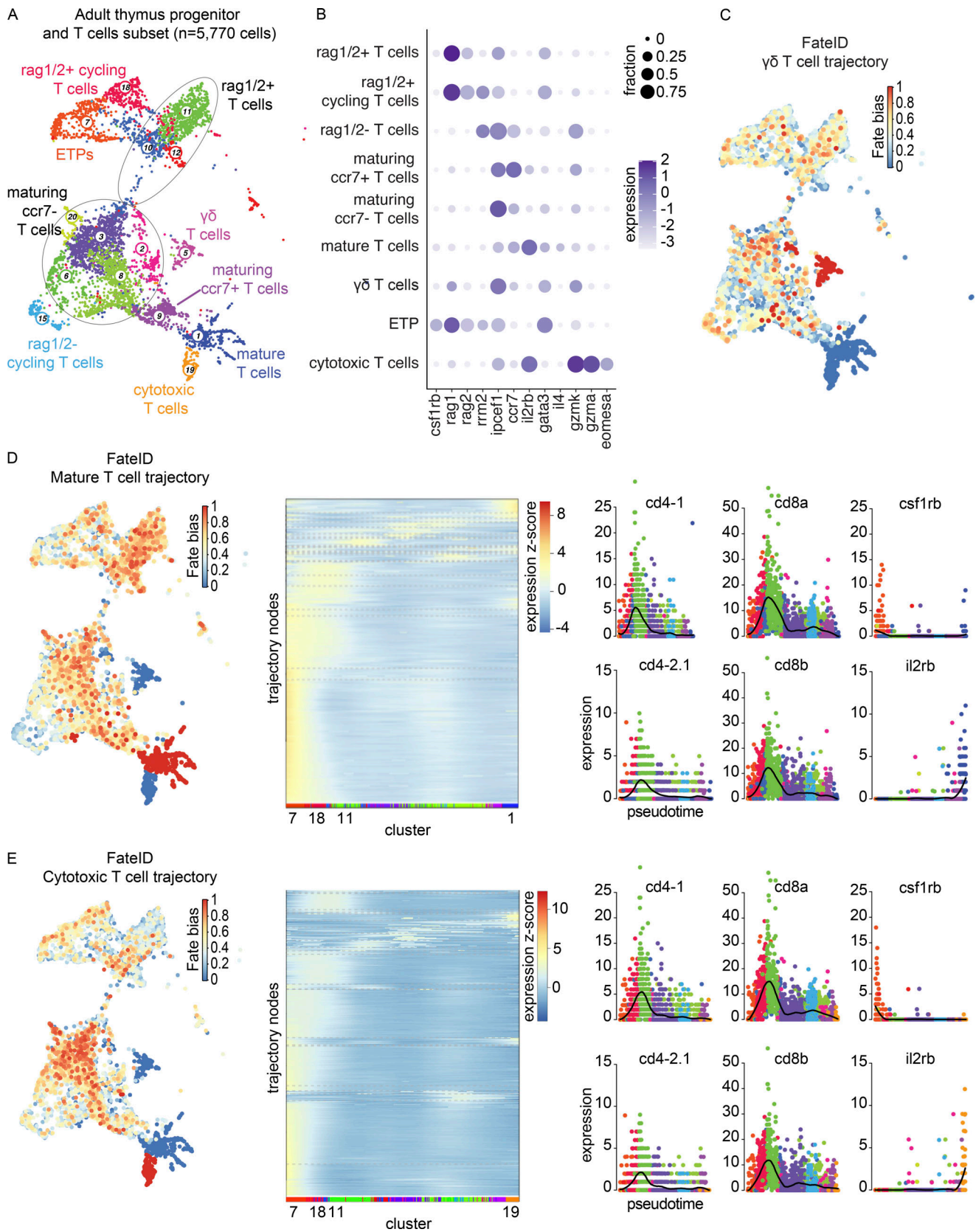


Figure 2. **Trajectory analysis of the adult thymus using FateID.** (A) RaceID3 UMAP visualization of the adult thymus progenitor and T cell subset. This subset of cells was derived from four adult zebrafish (3–4 mpf) dissected and processed within the same experiment. Cell type annotations were based on enriched markers and nomenclature consistent with Seurat clustering was used when appropriate. Circle outlines group multiple RaceID3 clusters falling within one Seurat-assigned cell type. (B) Dot plot of selected markers of the adult thymus progenitor and T cell subset showing expression across T cell populations.

Dot size reflects the percentage of cells within a population expressing a given marker and dot color shows the log<sub>2</sub> transformation of expression within that population. **(C)** Fate bias visualization depicts in color (red = high, blue = low) the probability that a cell will be assigned to the γδ T cell trajectory. Fate bias is 1 for the γδ T cell lineage-defining cluster, cluster 5, whereas clusters 1 and 19, lineage-defining clusters for the mature T cell and cytotoxic T cell trajectories, respectively, have a fate bias of 0. **(D)** Left: Fate bias visualization depicts in color (red = high, blue = low) the probability that a cell will be assigned to the mature T cell trajectory. Fate bias is 1 for the mature T cell lineage-defining cluster, cluster 1, whereas clusters 5 and 19, lineage-defining clusters for the γδ T cell and cytotoxic T cell trajectories, respectively, have a fate bias of 0. Middle: SOM for the mature T cell trajectory demonstrating pseudotime expression, grouping genes into nodes with RaceID3 cluster labeling along the x axis. Expression z-score is reflected by the color bar, with red being the highest and blue being the lowest. See also Table S7. Right: Pseudotime plots of the expression of *cd4-1*, *cd4-2.1*, *cd8a*, *cd8b*, *csf1rb*, and *il2rb* in the mature T cell trajectory. **(E)** Left: Fate bias visualization depicting in color (red = high, blue = low) the probability that a cell will be assigned to the cytotoxic T cell trajectory. Fate bias is 1 for the cytotoxic T cell lineage-defining cluster, cluster 19, whereas clusters 1 and 5, lineage-defining clusters for the mature T cell and γδ T cell trajectories, respectively, have a fate bias of 0. Middle: SOM for the cytotoxic T cell trajectory demonstrating pseudotime expression, grouping genes into nodes with RaceID3 cluster labeling along the x axis. Expression z-score is reflected by the color bar, with red being the highest and blue being the lowest. See also Table S7. Right: Pseudotime plots of the expression of *csf1rb*, *cd4-1*, *cd4-2.1*, *cd8a*, *cd8b*, *csf1rb*, and *il2rb* in the cytotoxic T cell trajectory.

in regulating cell cycle G1 to S phase progression (Kowalczyk et al., 2015), *dnmt3bb.1* (Challen et al., 2014), *hmgb3a* (Nemeth et al., 2003), and the dual HSPC and myeloid lineage markers *cebpa* and *csf1rb* (Smith et al., 1996). Genes without a known connection to HSPC biology included *maptb*, whose human ortholog *MAPT* encodes microtubule-associated protein τ; *foxp4*;

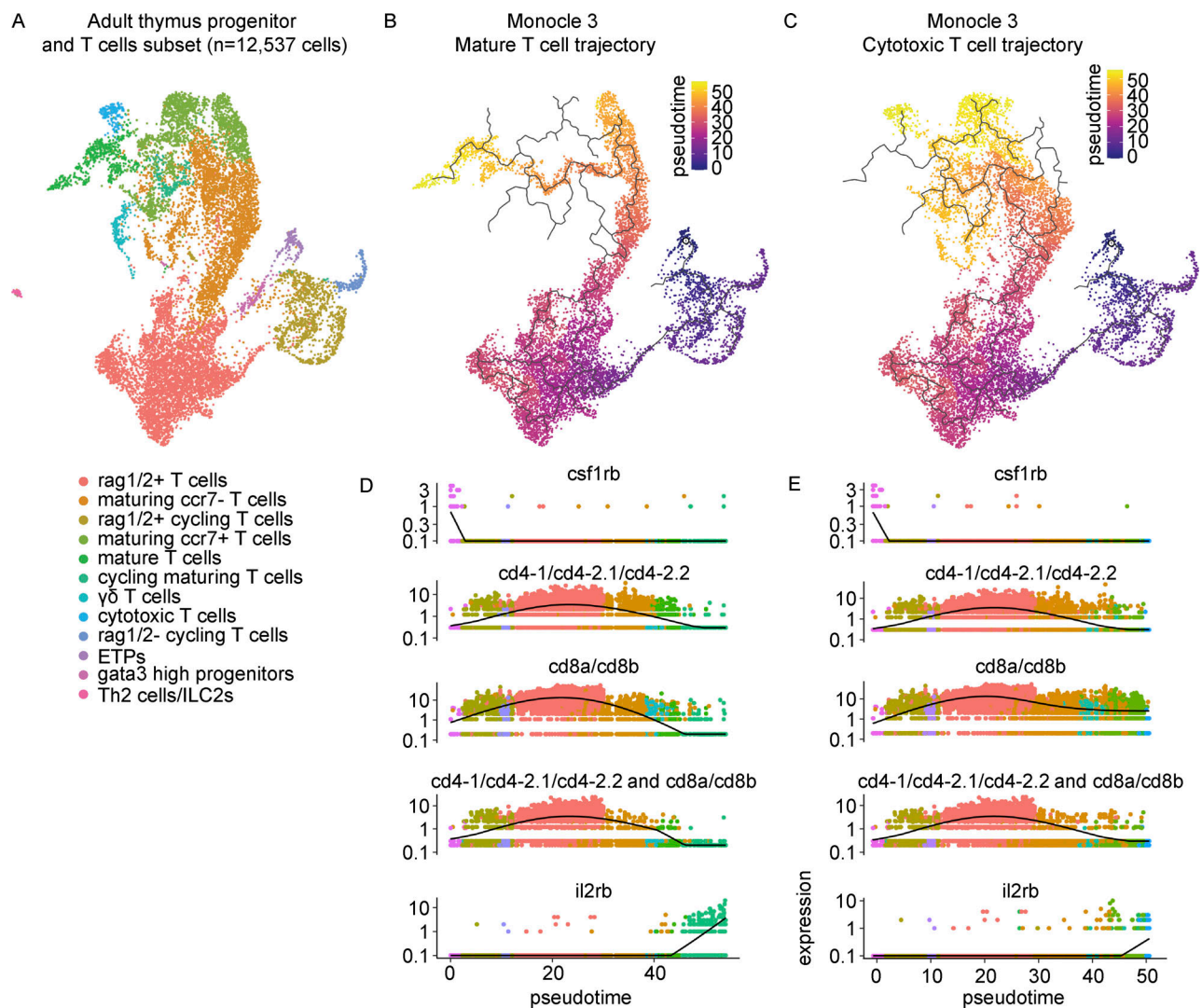


Figure 3. **Trajectory analysis of the adult thymus using Monocle 3.** **(A)** Monocle 3 UMAP visualization of the adult thymus progenitor and T cell subset. This subset of cells was derived from four adult zebrafish (3–4 mpf) dissected and processed within the same experiment. **(B and C)** UMAP visualization of the **(B)** mature T cell trajectory and **(C)** cytotoxic T cell trajectory colored by pseudotime as computed in Monocle 3. **(D and E)** Pseudotime plots of **(top)** CD4 orthologs (sum of *cd4-1*, *cd4-2.1*, *cd4-2.2* expression), **(middle)** CD8 orthologs (sum of *cd8a* and *cd8b* expression), and **(bottom)** dual CD4 and CD8 ortholog expression (minimum value of top and middle plots) in the **(D)** mature T cell trajectory and **(E)** cytotoxic T cell trajectory.

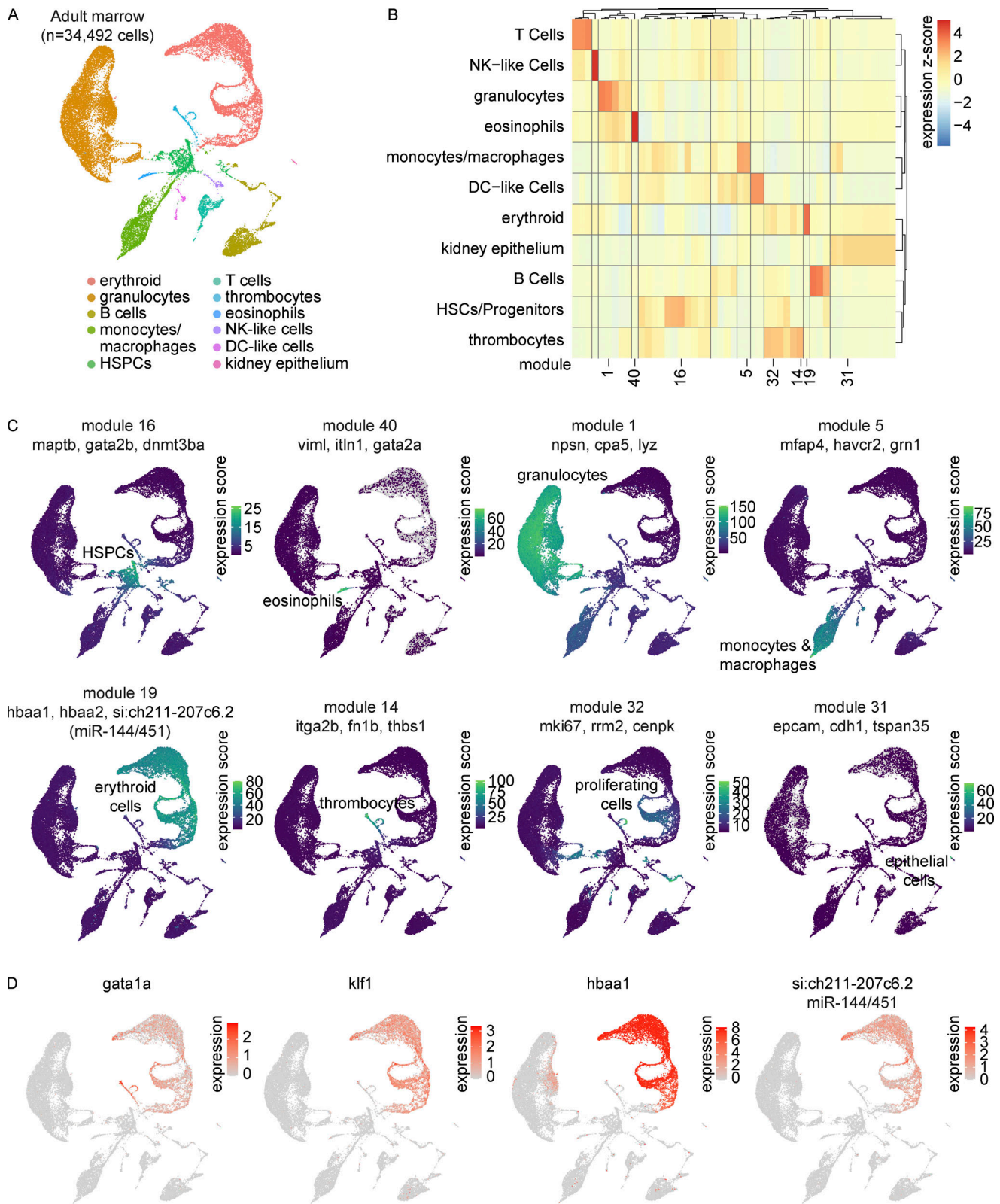


Figure 4. **Transcriptional characterization of the adult zebrafish marrow.** (A) UMAP visualization of 34,492 cells obtained from five zebrafish marrows. Marrows were dissected and processed in three independent experiments. Distinct cell lineages are denoted in different colors. Lineages were determined by differential gene expression (Wilcoxon rank sum test in Seurat) and gene module analysis in Monocle 3 (Table S8). (B) Gene module clustered heatmap of zebrafish marrow grouped by cell lineage. Color scale shows relative module enrichment (red) or depletion (blue). (C) UMAP visualizations of highlighted gene modules in panel B. (D) UMAP visualizations of canonical erythrocyte lineage genes *gata1a*, *klf1*, and *hbaa1* in addition to the newly appreciated *si:ch211-207c6.2*. This gene is one of the most specific erythrocyte genes identified at the transcriptional level in this dataset and was later determined to be a transcript containing the sequences of both *miR-144* and *miR-451*.

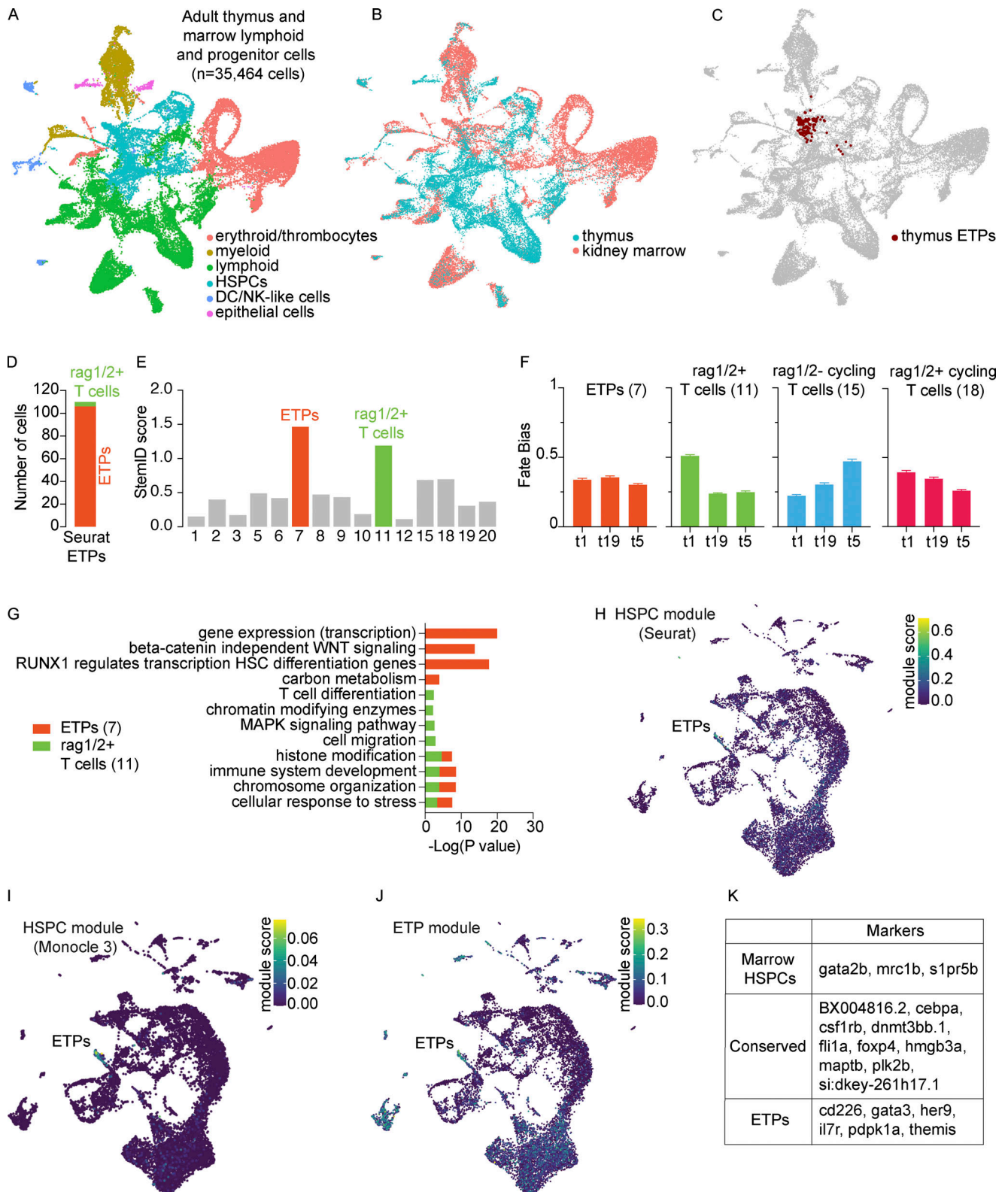


Figure 5. **Identification and characterization of ETPs.** (A and B) UMAP visualizations of the adult thymus and lymphoid and progenitor marrow fraction integrated together in Seurat ( $n = 35,464$  cells) annotated by (A) cell lineage and (B) tissue of origin. These cells were derived from seven adult zebrafish dissected and processed in three independent experiments; paired marrow and thymi were obtained from two zebrafish. (C) UMAP visualization of Seurat ETPs (cluster 38, resolution = 3; Table S1) on the integrated adult thymus and lymphoid and progenitor marrow fraction UMAP. Highlighted cells fall predominantly in the integrated HSPC cluster. (D) Assessment of the Seurat ETPs (cluster 38, resolution = 3) in RaceID3 analysis; the majority of Seurat ETPs were in RaceID3 cluster 7. (E) StemID score, the product of transcription entropy and number of inter-cluster links, displayed across all clusters; clusters 7 (ETPs) and 11 (rag1/2<sup>+</sup> T cells) are highlighted in their respective cluster colors for having the highest StemID scores. (F) Fate bias probabilities within early clusters



(highest StemID scores) across the three trajectories. Error bars depict SEM. **(G)** Enriched pathways determined using Metascape for the two earliest clusters (7 and 11) as predicted by FateID and StemID2. Negative log base 10 P values are plotted for a relevant subset of enriched pathways. **(H–J)** UMAP visualizations of gene modules in the adult thymus: (H) Seurat HSPC module (top 20 HSPC-enriched genes by specificity from Wilcoxon rank sum test), (I) Monocle 3 HSPC module, and (J) literature-based ETP gene module. See Materials and methods and Table S9 for module determination and the specific genes comprising each module, respectively. **(K)** Conserved and differentially expressed gene markers between thymic ETPs and marrow HSPCs as determined by Wilcoxon rank sum test (Tables S10 and S11). For the identification of conserved markers, P values were combined using Tippett's method (minimum P value).

and the genomic region *BX004816.2*. Differential expression analysis revealed that kidney marrow HSPCs were enriched most prominently in *gata2b*, *slpr5b*, and *mrc1b*, whereas thymic ETPs were enriched in T-cell commitment genes like *her9*, *themis*, *gata3*, *cd226*, *il7r*, and *pdpk1a* (Fig. 5 K, Fig. S4, and Table S11). Overall, these results transcriptionally define zebrafish ETPs and demonstrate that downregulation of *gata2b*, *slpr5b*, and *mrc1b* with concomitant upregulation of early T cell commitment genes are features of this population.

### Large-scale kidney marrow analysis reveals distinct subpopulations, including multiple DC-like cell and NK-like cell subsets

Having characterized zebrafish T-cell development in the thymus, we were next interested in returning to our large-scale characterization of the zebrafish marrow. Intrigued by the identification of two populations of DC-like cells and what appeared to be a heterogeneous NK-like cell population, we performed gene module analysis in Monocle 3 on the marrow subset of progenitors, DC-like cells, and NK-like cells (Fig. 6, A–C and Table S12). This analysis revealed great heterogeneity within HSPC populations (*flila*<sup>+</sup>, *gata2b*<sup>+</sup>) and identified multiple subsets, including, HSPCs (1; *nanos1*<sup>+</sup>, *spi2*<sup>+</sup>), HSPCs (2)/eosinophil progenitors (*cfid*<sup>+</sup>, *MFAP4* (1 of many).8<sup>+</sup>), cycling HSPCs (*mki67*<sup>+</sup> or *mcm5*<sup>+</sup>), and HSPCs (4)/lymphoid progenitors (*ackr3a*<sup>+</sup>, *runx3*<sup>+</sup>; Fig. 6 C). The myeloid progenitor and HSPCs (4)/lymphoid progenitor expression patterns were particularly interesting for their shared expression of genes with the DC-like cells (1) and DC-like cells (2) populations, respectively. We identified both pan-DC cell modules and subset specific modules suggestive of significant gene expression differences between these populations (Fig. 6 D). Pan-DC markers included *spock3*, *xcr1a.1*, and *trib1*, whereas subset analysis suggested the existence of both conventional-like DCs (cDCs), specifically cDC1s expressing *znf366*, *batf3*, *snx22*, and *hepacam2*, and plasmacytoid-like DCs enriched for endosomal TLRs (e.g., *tlr9*), *irf8*, *ctsb*, *flt3*, and *tcfl2* (Fig. 6 D). Gene module analysis for the NK-like cells revealed distinct sets of genes in multiple NK-like cell states, e.g., NK-like cells (1), enriched in *ccl33.2*, *nudt6*, *il21*, and *cyth4a* (Fig. 7 A). A very small subcluster of this population appeared to be *eomesa*-negative and was enriched in the macrophage/DC-like markers *grn1*, *grn2*, *ptgs2b*, and *mchr1b* (module 40). Additionally, although not separate clusters at this resolution, distinct gene sets were found within the NK-like progenitors and cells (2), separating into progenitors (2; module 31) and NK-like cells (2; module 10). These findings provide valuable insight into the great transcriptional heterogeneity within understudied zebrafish immune cell populations and prime future investigations.

### Zebrafish T and NK-like cells are distinguished transcriptionally by multiple genes including the *SPI1/PU.1* ortholog, *spi1b*

To explore NK-like cell biology in greater detail and to specifically relate these innate lymphoid cells to their T-cell counterparts, we subsetted all marrow NK-like cells and T cells (Fig. 7 B and Table S13). This analysis revealed rich diversity of T and NK-like states and demonstrated great consistency with our previous NK-like cell analysis, thereby allowing similar nomenclature to be used (Fig. 7, C and D). In addition to the diversity appreciated at the cluster level, we also noted a small population of *rorc*<sup>+</sup> lymphocytes that appeared to be a distinct population (Fig. 7 C). To better support the assignments of these populations as T- and NK-like cells, we first assessed the expression of TCR  $\alpha/\beta/\gamma/\delta$  chains in these cells in the full marrow and a close-up analysis of this subset (Fig. 8 A). In the subset, we noted a mix of  $\alpha\beta$ - and  $\gamma\delta$ -expressing T cells in the main T-cell cluster, distinguished from a smaller side cluster of cells predominantly expressing  $\gamma$  and/or  $\delta$  chains. In total, within the clusters deemed to be T cells, 365 out of 1,244 cells (29.3%) expressed at least one TCR gene, on par with the detection efficiency of other T-cell associated genes like *lck* (detection efficiency = 27.4%). This contrasted with the rare recovery of these TCR transcripts in the clusters denoted as NK-like cells, with only 3 out of 289 cells (1%) expressing any of these transcripts. Further investigation into the transcriptional differences between these populations revealed specific enrichments in each population, including *ckbb*, *spi1b*, *cfbl*, and *fcer1g1* in NK-like cells and *dusp2*, *il2rb*, *sla2*, and *FP236356.1* in T cells (Fig. 8 B and Table S14). To contextualize these results, we note that innate lymphoid cells (e.g., NK-like cells) here were *lck*<sup>-</sup>, *spi1b*<sup>+</sup>, and *spil1a*<sup>-</sup>, whereas T cells were *lck*<sup>+</sup>, *spi1b*<sup>-</sup>, and *spil1a*<sup>-</sup> (Fig. 8 C). These data identify key distinguishing markers between marrow NK-like cells and T cells in zebrafish.

### B lymphocyte development demonstrates transcriptional conservation in key developmental genes and includes a pre-B cell state

To investigate zebrafish B cell development, we subsetted B and progenitor cells from the marrow for further analysis (Fig. 9 A). HSPC populations, multiple stages of B cell development, and two mature B cell populations (IgT and IgD) were distinguished by differential gene expression (Fig. 9 B and Table S15). Relative to the thymus, marrow IgT B cells only comprised 15% of mature B cells (vs. 63%). Of note, while *ighd* and *ighz* were detected in all B cell populations, including *rag1/2*<sup>+</sup> B cells (1) and *rag1/2*<sup>-</sup> B cells, light chain genes *igllc3* and *iglv5* were not detected in these two populations (Fig. 9 B). Gene module analysis identified additional genes enriched at different stages of B cell development (Fig. 9 C

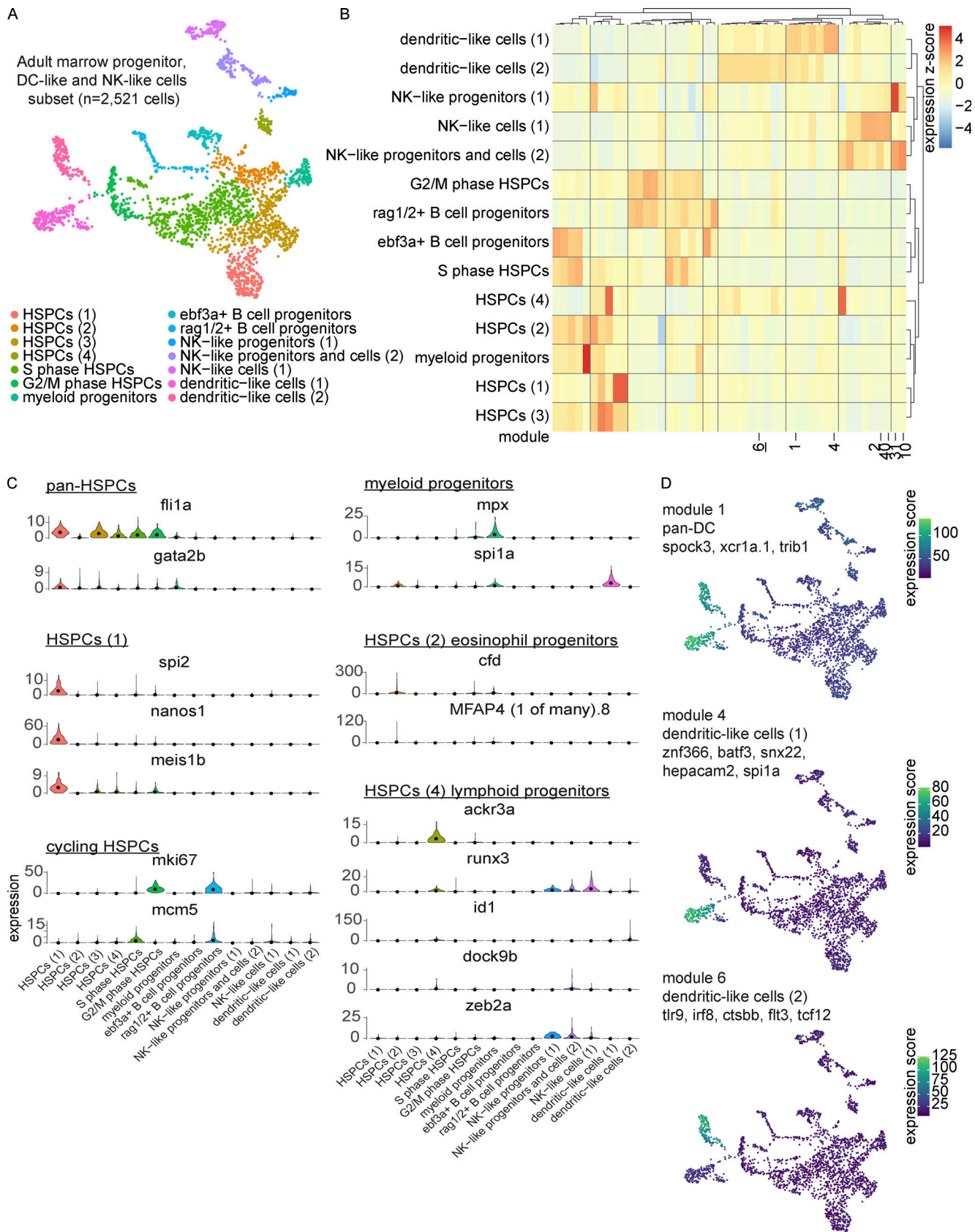


Figure 6. **Transcriptional characterization of marrow subset of HSPCs, DC-like cells, and NK-like cells. (A)** UMAP visualization of 2,521 DC-like cells, NK-like cells, and progenitors subsetted from the larger marrow UMAP. Cell type annotations were based on differential expression, as determined by Wilcoxon rank sum test in Seurat and gene module analysis in Monocle 3 (Table S12). **(B)** Gene module clustered heatmap of DC-like cells, NK-like cells, and progenitors

grouped by cell type. Color scale shows relative module enrichment (red) or depletion (blue). **(C)** Violin plots of differentially expressed genes as determined by Wilcoxon rank sum test among HSPC clusters. **(D)** UMAP visualizations of gene modules enriched across all DC-like cell populations (module 1) and specific to either the cDC population (DC-like cells [1]; module 4) or the plasmacytoid-like DC population (DC-like cells [2]; module 6).

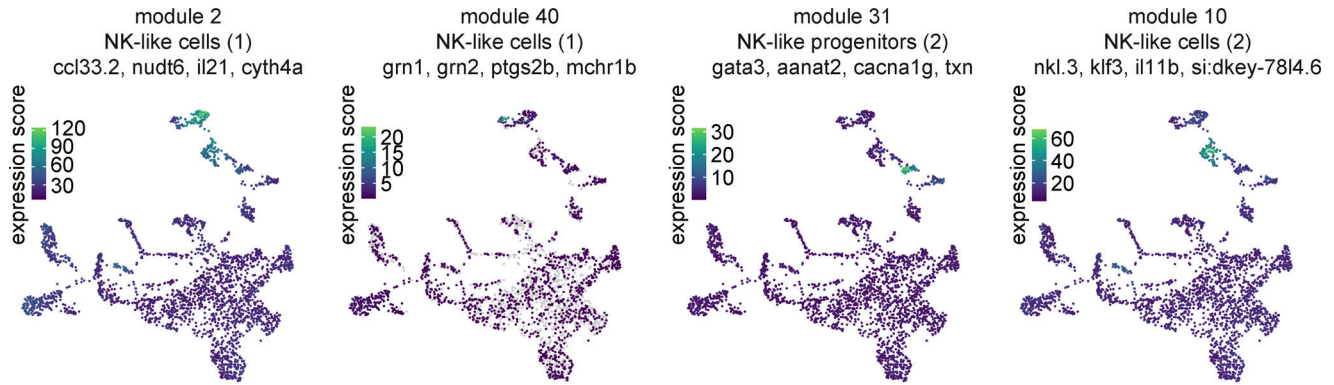
and Table S16). Module 9, expressed in *rag1/2<sup>+</sup>* B cells (1), included the genes *dntt*, *sid1*, and *si:ch211-1a19.2*, the former encoding the enzyme terminal deoxynucleotidyl transferase, important in generating junctional diversity in V(D)J recombination, and the latter two containing immunoglobulin domains (Alt and Baltimore, 1982; Komori et al., 1993; Yoder et al., 2002; Fig. 9 D). Later developing B cells were enriched in migratory genes including *capga*, *gadd45ga*, and *pcdh7b* (module 37). Additional genes found to be differentially expressed between mature IgD and IgT B cells included *adipor2* and *swap70b* enriched in module 32 and *mrap2a* and *nfil3-4* enriched in module 7 (Fig. 9 D). These data illuminate great transcriptional heterogeneity throughout zebrafish B cell development and show differences in thymic and marrow mature B cell isotype abundances.

Interested in the differential detection of heavy and light chain genes, we investigated the transcriptional dynamics of B cell development in greater detail through Monocle 3 pseudotime analysis (Fig. 9, E and F). We selected 11 genes that would allow us to explore lineage commitment (*csf1rb*, *ebf3a*, *pax5*, *cd79a*), B-cell receptor rearrangement (*rag1*, *rag2*, *sid1*, *dntt*), light chain expression (*igl3v5*, *igl1c3*), and cell state (*mki67*). Specifically focused on the timing of B cell heavy and light chain gene expression where IgD and IgT B cells would not be expected to differ greatly, we chose to explore gene expression in pseudotime for B cell development as a whole and not along any particular branched structure. We observed a concomitant decrease in *csf1rb* expression with increase in *ebf3a* expression at the transition between multipotent progenitor and commitment to the B cell lineage. Multiple stages of proliferation were evidenced by *mki67* expression, the first at the juncture of B cell commitment and the second following a phase of *sid1*, *dntt*, and *rag1/2* gene expression. The expression of the B cell genes *pax5* and *cd79a* followed that of the expression of *ebf3a* and they remained expressed onward in development. The expression of *rag1/2* was noted in two distinct populations of B cells, with *rag1/2<sup>+</sup>* B cells (1), the earlier population in pseudotime, enriched in *sid1* and *dntt* expression and notably negative for immunoglobulin light chain expression. While largely overlapping, the detection of *rag1* and *rag2* expression was not identical, with *rag2* expression remaining elevated in *rag1/2* low cycling B cells. An analogous analysis using FateID revealed similar pseudotime assignment and gene expression dynamics throughout B cell development (Fig. 10, A–D). The most noticeable differences in these analyses were the relative ordering of HSPC populations, which were not relevant to our consideration of B cell development post-commitment. Taken together, these results show the dynamic gene expression changes that occur in zebrafish B cell development and strongly support the existence of a heavy chain<sup>+</sup> and light chain<sup>-</sup> pre-B cell state.

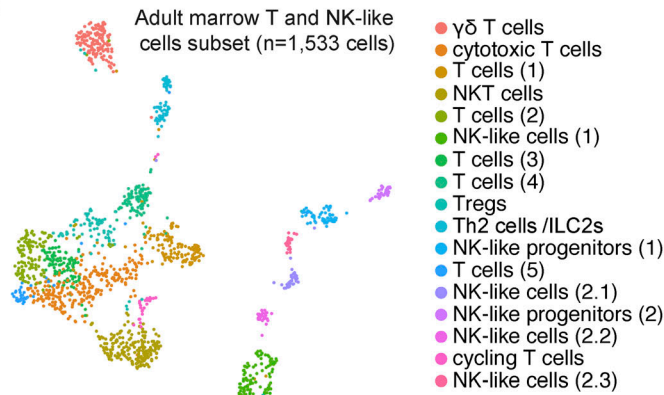
## Discussion

scRNA-seq has proven to be a powerful technique not only for revealing cell type and tissue heterogeneity but also inferring cellular developmental trajectories. By profiling ~70,000 total cells from the zebrafish primary lymphoid organs, the kidney marrow and thymus, we painted a high-resolution picture of zebrafish lymphocyte development. Beginning with the zebrafish thymus, an understudied organ, we explored two different timepoints of T cell development and noted substantial cell type similarity but finer differences in thymic composition and gene expression. We selected the early timepoint of 4 wpf as the earliest feasible dissection of a morphologically mature thymus (Lam et al., 2002; Willett et al., 1997b), balancing desired cell number, dissection quality, and cellular preparation efficiency. While we anticipated many of the T-cell developmental changes that we observed, namely the relative decrease in *rag1/2<sup>+</sup>* developing T cell subsets over time consistent with involution, we were surprised that the juvenile thymi contained a similarly rich array of non-T cell populations from DC-like cell subsets to NK-like cells. The presence, albeit low, of thymic B cells at 4 wpf was particularly interesting as mature B cells are only first detected ~3 wpf (Page et al., 2013). The observation that erythrocytes detected in juvenile and adult thymic samples not only differed in embryonic vs. adult hemoglobin expression but also in ribosomal subunit expression suggests that there may be temporal ribosomal heterogeneity (Kondrashov et al., 2011). Zebrafish TECs exhibited rich diversity, transcriptionally reminiscent of multiple mammalian subsets including mTECs, cTECs, and tuft-like mTECs. Although not entirely consistent transcriptionally due to the lack of *psmb11a* and *psmb11b* detection, our high T cell TECs could represent the zebrafish equivalent of mammalian thymic nurse cells (Nakagawa et al., 2012). Thymic fibroblasts appeared to be activated, as suggested by the expression of genes commonly expressed in cancer-associated fibroblasts, *twist1a* and its downstream transcription factor *prrx1a/b* (Yeo et al., 2018). We note that these findings, like all others of this study, are interpreted in relation to the assumed roles of these genes in zebrafish based on known mammalian biology and that there may be species-specific differences that account for such described transcriptional “inconsistencies.” We speculate that these results only touch the surface of zebrafish TEC heterogeneity due to our lack of enzymatic digestion of the thymus tissue and believe this could be an exciting area of future research. ETPs were identified by their expression of early T cell differentiation genes and retention of some HSPC markers. Notably, *gata2b*, one of the best markers of marrow HSPCs, was detected in <2% of thymic ETPs, potentially reflecting an early and required downregulation to commit to the T cell lineage. The subset-specific markers identified here provide a springboard for future studies to visualize early T-cell development including T cell-TEC/T cell-fibroblast interactions through the generation of ETP-, TEC-, and fibroblast-specific transgenic lines and

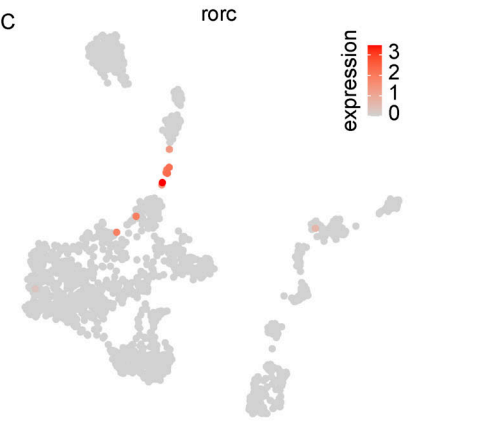
A Adult marrow progenitor, DC-like and NK-like cells subset (n=2,521 cells)



B Adult marrow T and NK-like cells subset (n=1,533 cells)



C *rorc*



D

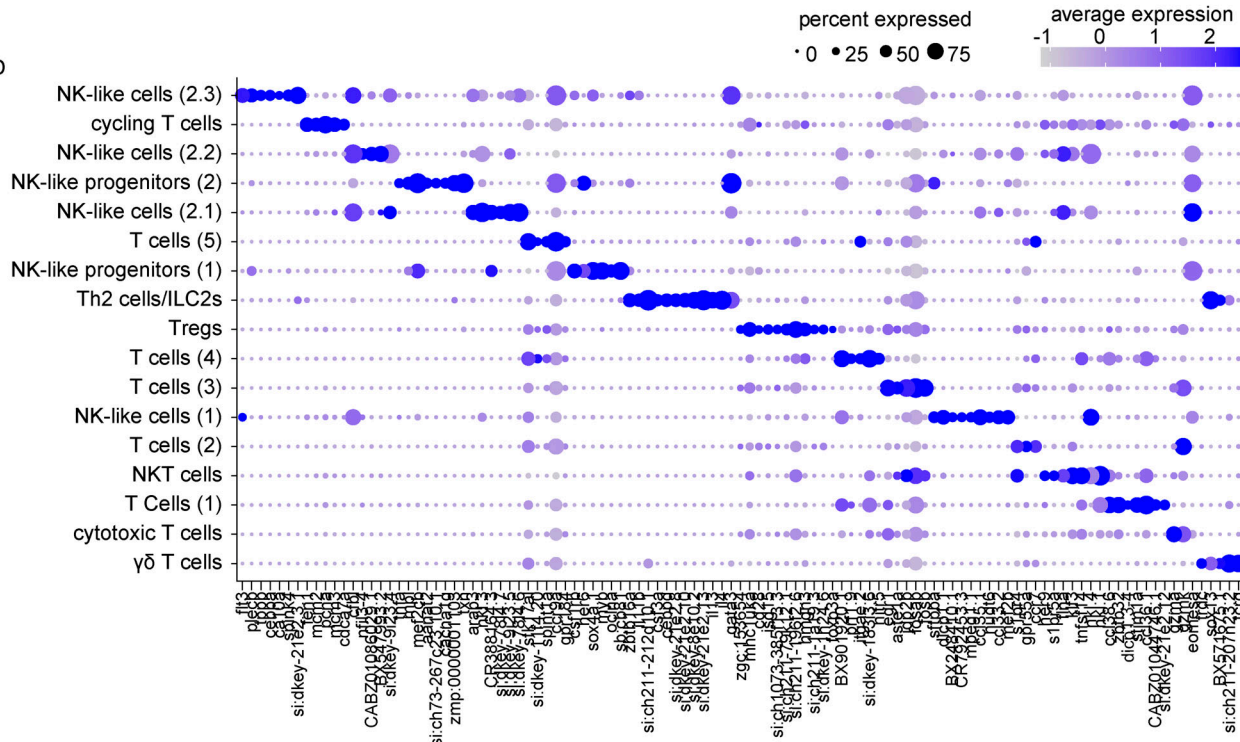


Figure 7. **Transcriptional characterization of marrow NK-like and T cell subpopulations.** (A) UMAP visualizations of gene modules enriched across NK-like populations in the progenitor, DC-like cell, and NK-like cell marrow subset. (B) UMAP visualization of 1,533 T and NK-like cells as subsetted, re-integrated, and clustered from the larger marrow UMAP in Fig. 4 A. Cell type annotations were based on differential expression analysis as determined by Wilcoxon rank sum test in Seurat (Table S13). (C) UMAP visualization highlighting *rorc*<sup>+</sup> T cells, too small a population to be resolved as an independent cluster, but nevertheless demonstrating clear heterogeneity. (D) Dot plot of selected markers differentially expressed across T and NK-like cell populations. Dot size reflects the percentage of cells within a population expressing a given marker and dot color shows the average expression within that population.

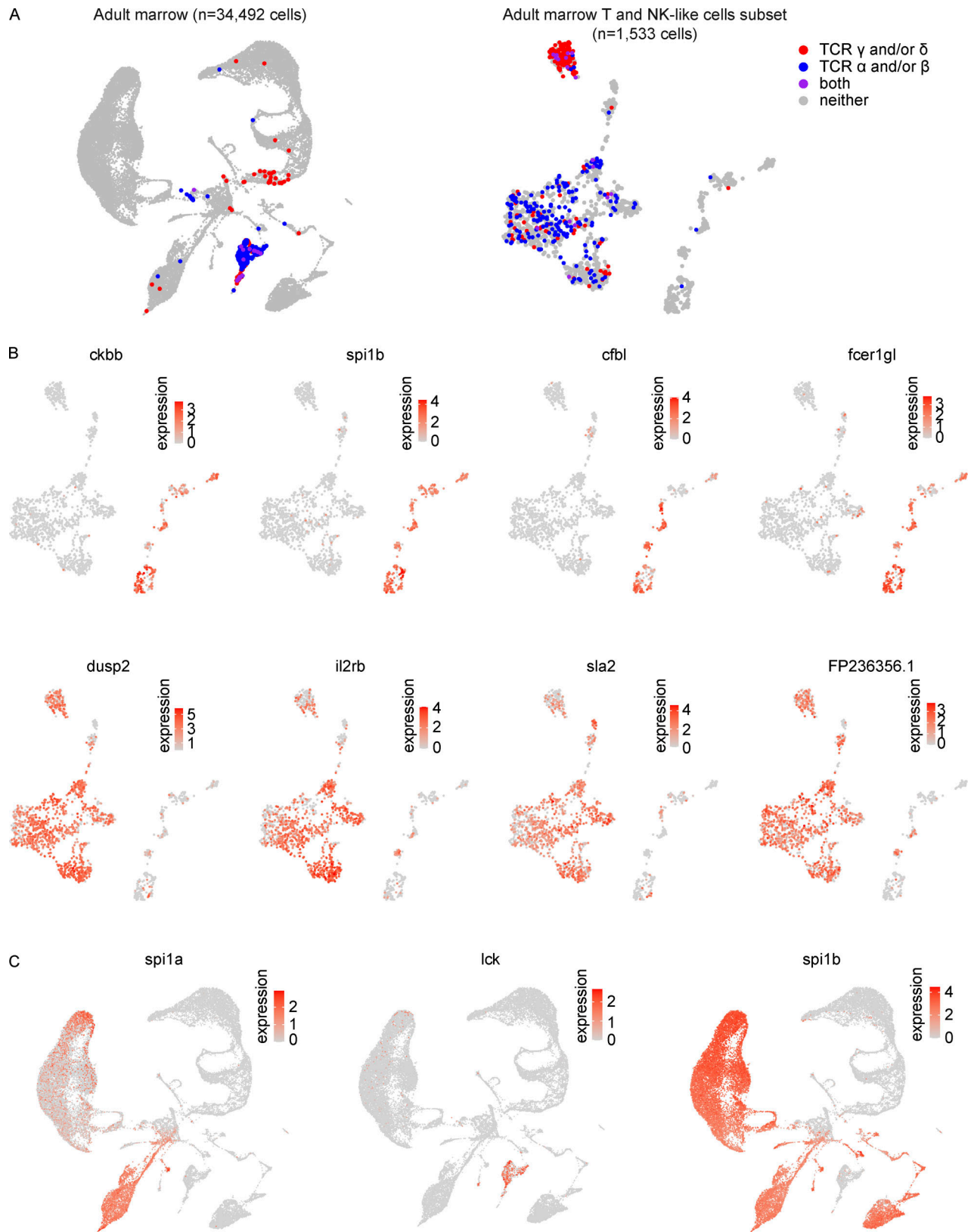


Figure 8. **Distinguishing markers of NK-like cells and T cells.** (A) UMAP visualization of detection of TCR  $\gamma$  and/or  $\delta$  (red),  $\alpha$  and/or  $\beta$  (blue), or a combination of both sets of receptors (purple) in the (left) larger marrow object and (right) T and NK-like cell populations. (B) Top: UMAP visualizations of subset of genes identified in both Monocle 3 top markers analysis (limited to top 25 most specific genes as ranked by Jensen-Shannon divergence) and Seurat differential expression analysis (Wilcoxon rank sum test) as being significantly enriched in NK-like cells vs. T cells (Table S14). Bottom: UMAP visualizations of subset of genes identified in both Monocle 3 top markers analysis and Seurat differential expression analysis as being significantly enriched in T cells vs. NK-like

cells. *FP236356.1* is a 109,805 bp region that includes TCR  $\beta$  constant 1 (*trbc1*; Table S14). **(c)** UMAP visualizations of expression of *spi1a*, *spi1b*, and *lck* in the marrow. Expression of *spi1a* is observed in a subset of *spi1b*-expressing cells and may distinguish myeloid and lymphoid cells. NK-like cells here are *spi1b*<sup>+</sup> while T cells are *spi1b* negative. DC-like cells (2) are *spi1a*<sup>-</sup>*lck*<sup>+</sup> whereas DC-like cells (1) are *spi1a*<sup>+</sup>*lck*<sup>-</sup>, further consistent with the hypothesis that DC-like cells (2) are plasmacytoid DCs and DC-like cells (1) are myeloid DCs and specifically akin to cDC1s.

provide a larger window into the evolution of T-cell immunity and T-cell development.

Building upon prior transcriptional characterizations of the zebrafish marrow, our work highlights the existence of at least two DC-like subsets and extensive heterogeneity within the marrow T, NK-like, and B cell populations. The two DC-like populations seem to reflect a lymphoid-myeloid split, with the population identified as resembling plasmacytoid DCs expressing a lymphoid signature and the type 1 cDC-like population expressing a myeloid signature. Intriguingly, an *XCRI* ortholog was identified as a pan-zebrafish marrow DC marker, while it has been reported to be an exclusive cDC1 marker in mice (Bachem et al., 2012). Zebrafish T and NK-like cell diversity within the marrow richly paralleled that of mammals with recovery of transcriptional profiles consistent with cytotoxic T cells, NKT cells, Th2s/ILC2s, regulatory T cells, multiple subsets of NK-like cells, and even a small population of *rorc*<sup>+</sup> TCR-lacking ILC3s. TCR expression clearly distinguished T and innate lymphoid cells such as NK-like cells in our data. Although *il2rb* has been reported to be a marker of NK cells in past studies, our results strongly suggest *il2rb* as a marker of mature thymic and marrow T cells (Carmona et al., 2017; Tang et al., 2017). Our B cell development investigations show that zebrafish *ebf3a* may act like mammalian *EBF1* (Hagman et al., 1991) as a B cell lineage commitment factor. Pseudotime analysis tracing the development of B cells from multipotent progenitors to mature B cells further revealed intriguing findings. The presence of a *sid1*<sup>+</sup>, *dntt*<sup>+</sup>, *rag1/2*<sup>+</sup>, *igl3v5*<sup>-</sup>, *igllc3*<sup>-</sup> population of developing B cells is highly suggestive of the existence of a pre-B cell state. Although underexplored, *sid1* has been proposed as an ortholog to *VPREB1* (Yoder et al., 2002), which would be consistent with this finding. We also speculate whether the gene *si:ch211-1a19.2* may encode an additional surrogate light chain subunit and perhaps be a  $\lambda 5$  homolog, as it has been classified previously (Garcia et al., 2018). Collectively, our work provides a rich transcriptional atlas of the zebrafish marrow and thymus and sets the stage for future investigations to better understand teleost adaptive immunity and the immunological modeling of human disease in zebrafish.

## Materials and methods

### Zebrafish husbandry and lines

Zebrafish were maintained in accordance with Boston Children's Hospital Institutional Animal Care and Use Committee protocols and in line with Animal Resources at Children's Hospital guidelines. In total, seven adult zebrafish (3–5 mpf) and 21 juvenile zebrafish (4 wpf) were analyzed in these studies. Five of the adult zebrafish were generated through the mating of a pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA guide fish with a pTol2-DRv7 GESTALT (Genome Editing of Synthetic Target

Arrays for Lineage Tracing) barcode fish (McKenna et al., 2016; Raj et al., 2018b). The parental strains were gifts from the Schier and Gagnon labs, respectively (Harvard University, Cambridge, MA; University of Utah, Salt Lake City, UT). Two-stage barcoding was induced in these fish, first by injection at the single-cell stage with sgRNAs 1–4 and again at 28 h after fertilization by heat shock as previously described (Raj et al., 2018a). The remaining two adult fish and all 21 juvenile zebrafish were Tg(*lck:eGFP*), utilized for the purpose of identifying and isolating T cells, as previously described (Langenau et al., 2004). Of the adult fish, six were female and one was male. Sex of the juvenile fish could not be determined at the time of analysis.

### Tissue collection and cell sorting

Adult and juvenile zebrafish were euthanized by rapid chilling and confirmed deceased prior to dissecting thymi and/or kidney marrows. To reduce erythrocyte contamination from the peripheral blood, adult fish were cardiac bled prior to tissue dissection using a heparin-coated (1,000 USP U/ml in 1× [PBS] p10 or p20 pipette tip). The juvenile fish were deemed too fragile for this procedure to be performed. For the thymi dissections, the zebrafish were positioned on their sides on a Styrofoam lid under a fluorescent dissecting scope and held in place/braced with the aid of one 23-gauge needle and gentle counterpressure applied with the aid of dissection scissors and/or forceps. The operculum was then removed, allowing access to and removal of one of the thymi, and placed in room temperature 0.5% BSA in HBSS without Ca<sup>2+</sup> and Mg<sup>2+</sup>. Each fish was then carefully flipped over to the other side and this procedure was repeated to obtain the second thymus. When fluorescent reporters were not employed, brightfield microscopy and anatomical landmarks were used instead. Kidney marrows were dissected under a dissecting microscope by placing the zebrafish on a paper towel and using dissection scissors to make a superficial incision along the ventral midline to expose the internal organs. Forceps were used to remove all exposed organs including the intestines, liver, and swim bladder. In female zebrafish, eggs were also removed to clear the surgical field. Once visible along the dorsum, the kidneys were removed using forceps and placed in 0.5% BSA in DPBS without Ca<sup>2+</sup> and Mg<sup>2+</sup> on ice. In the case of joint thymus and kidney marrow dissection (two adult fish), the thymi were dissected first. The tissues were then mechanically dissociated into single-cell suspensions by pipetting and filtered through a 50- $\mu$ m disposable filter (CellTrics). The filtered samples were then briefly centrifuged at 400 *g* × 5 min to remove debris and resuspended in their respective buffers for flow cytometry. Sytox blue was added to the samples immediately prior to analysis to assess cell viability. Live cells were sorted (85  $\mu$ m nozzle) on a BD FACSAriaII into 300  $\mu$ l of 0.5% BSA in DPBS and kept on ice following the sort. In the case of the marrows, live cell sorts were divided into two fractions based on forward

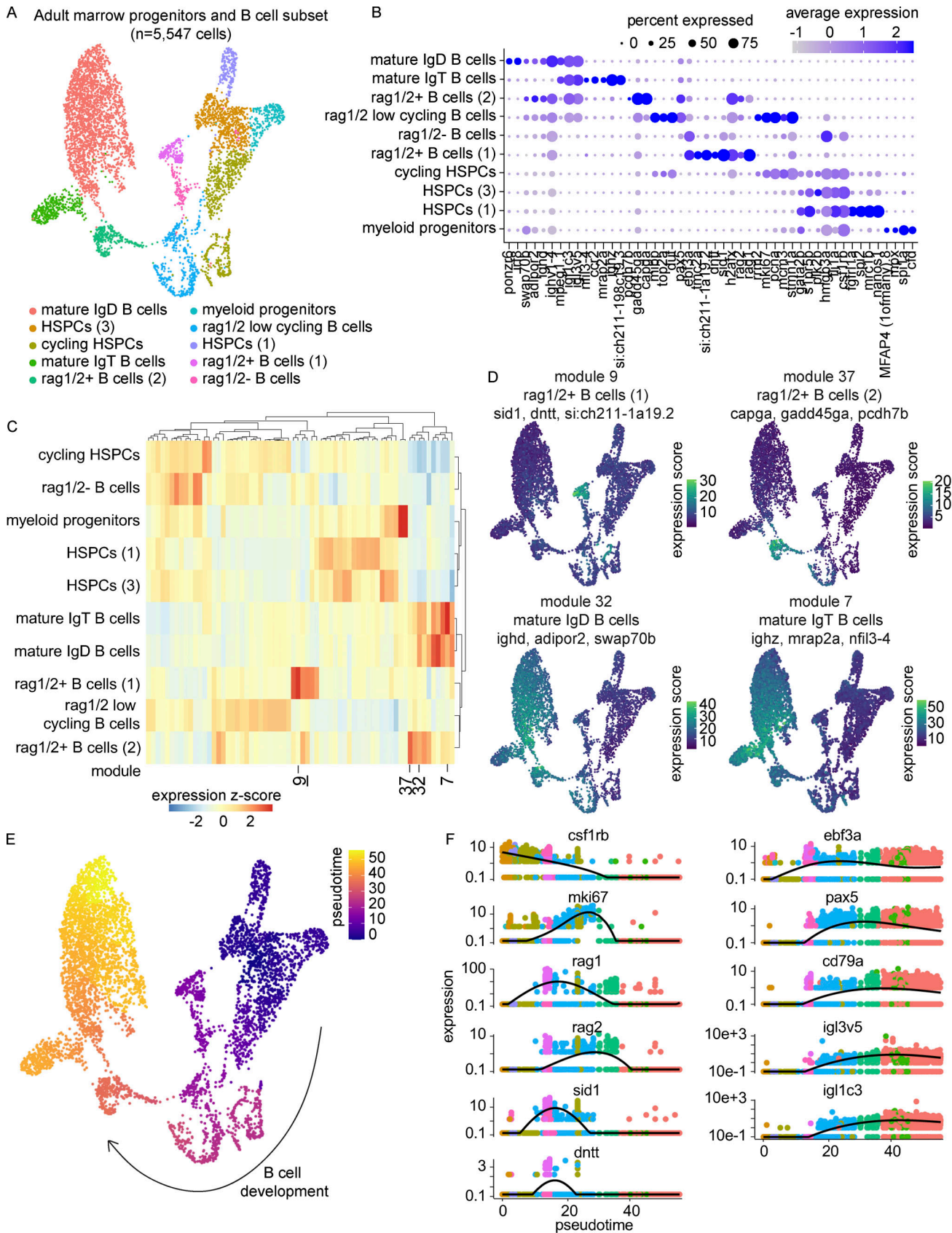


Figure 9. **Transcriptional characterization and trajectory analysis of B cell development using Monocle 3.** (A) UMAP visualization of 5,547 B cells and progenitors from the marrow. Cell type annotations were based on differential expression as determined by the Wilcoxon rank sum test in Seurat and gene

module analysis in Monocle 3 (Tables S15 and S16). Nomenclature consistent with Fig. 6 A was used for the HSPC populations when appropriate. **(B)** Dot plot of selected markers differentially expressed across B and progenitor cell populations. Dot size reflects the percentage of cells within a population expressing a given marker and dot color shows the average expression within that population. **(C)** Gene module clustered heatmap of zebrafish B cells and progenitors grouped by cell type demonstrating cell type enrichments. Color scale shows relative module enrichment (red) or depletion (blue). **(D)** UMAP visualizations of gene modules enriched in different stages of B cell development. **(E)** UMAP visualization of B cells and progenitors colored by pseudotime as computed in Monocle 3. **(F)** Gene expression dynamics over pseudotime computed in Monocle 3 colored by cluster assignment for 11 genes related to B cell development.

scatter and side scatter characteristics—a lymphoid and progenitor fraction and a granulocyte fraction (Fig. S5, A and B). This allowed for the relative enrichment of lymphoid and

progenitor cells in the overall analysis, advantageous for our goal to investigate B cell developmental trajectories. By nature of this sorting strategy, our marrow data does not reflect whole

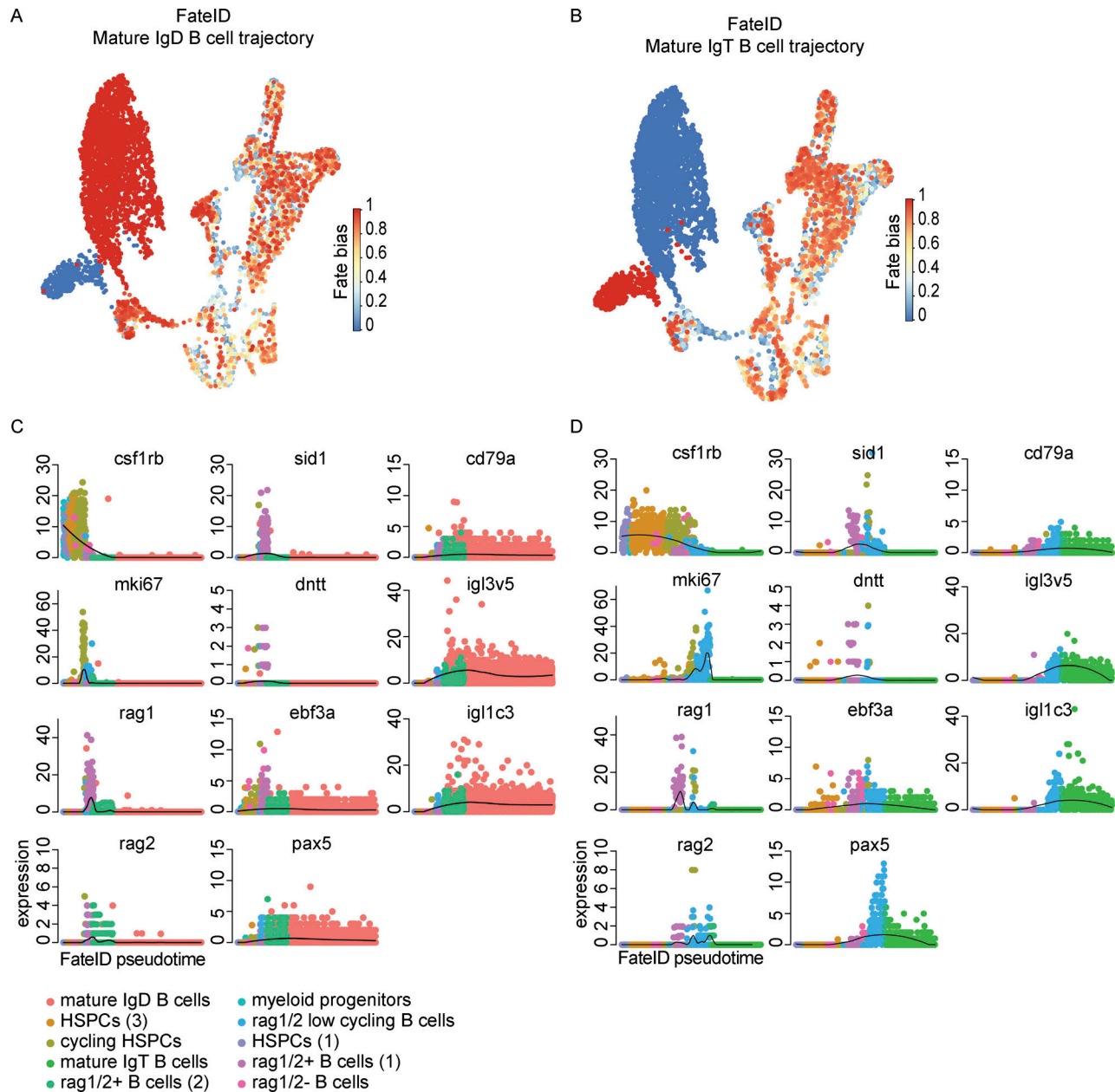


Figure 10. **Trajectory analysis of B cell development using FateID.** **(A and B)** UMAP visualizations where fate bias is depicted in color (red = high, blue = low) and represents the probability that a cell will be assigned to a given lineage. The lineage-defining clusters **(A)** mature IgD B cells and **(B)** mature IgT B cells have a fate bias of 1 in their respective trajectory maps and a fate bias of 0 in the other. **(C and D)** Pseudotime plots of the expression of *csf1rb*, *mki67*, *rag1*, *rag2*, *sid1*, *dntt*, *ebf3a*, *pax5*, *cd79a*, *igl3v5*, and *igl1c3* in the **(C)** mature IgD B cell trajectory and **(D)** mature IgT B cell trajectory.



kidney marrow composition. Between 50,000 and 100,000 cells were sorted per sample.

### scRNA-seq and genome alignment

Sorted cells were pelleted by centrifugation at  $400 g \times 5$  min and resuspended to a final concentration of 1,000 cells/ $\mu$ l following Trypan blue staining. Transcriptome profiling was performed using the Chromium Next GEM Single Cell 3' v3.1 Reagent Kits (10x Genomics). A targeted recovery of 6,000 cells was set for loading the Chromium Controller in all the assays, and steps were performed as described in the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide (Rev D). The final transcriptome libraries were pooled in sets of 4 or 5 and sequenced at the Harvard University Bauer Core (<https://bauercore.fas.harvard.edu/>) on an Illumina NovaSeq 6000 using an SP flow cell (Read 1: 28 cycles, i7 index: 8 cycles, i5 index: 0 cycles, Read 2: 91 cycles). For transcriptome analysis, demultiplexed FASTQ files were processed through the Cell Ranger 3.1.0 pipeline (Zheng et al., 2017) and aligned to a Zon lab custom genome. More specifically, this Zon lab custom genome consisted of zebrafish Ensembl GRCz11 and five additional genes (EGFP, mCherry, dsRed, TagBFP, and BRAFv600e). This was done for consistency and to retain the ability to assess EGFP expression from the *lck:eGFP* zebrafish. (Note: BRAFv600e denotes the activating missense mutation in human B-Raf proto-oncogene [BRAF] commonly seen in some cancers. This gene is included as part of our standard alignment procedure when recovery of EGFP is desired. At processing, the BRAFv600e “gene” contained both the BRAFv600e sequence in addition to part of the SV40 poly(A) signal also found in the *lck:eGFP* plasmid. For this reason, “BRAFv600e” is recovered in this dataset in a similar set of cells where EGFP is identified.)

### Quality control and data processing in SoupX and Seurat

Unfiltered feature-barcode matrices from the Cell Ranger outputs for all samples were loaded into the SoupX package (Young and Behjati, 2020) to create SoupChannel objects. Multiple methods of contamination fraction estimation were applied and assessed qualitatively for success of contamination removal and sparing of true signal using built-in functions. SoupX processing required a more recent version of Seurat (Stuart et al., 2019) to be used. Therefore, Seurat library v3.2.2 was attached prior to the analysis. Seurat v3.1.3 was used for all other analyses. The following contamination fraction settings were used for each sample: all four adult thymus datasets  $\rho = 0.02$ ; lymphoid 27, myeloid 27  $\rho = 0.05$ ; lymphoid 22, myeloid 22, lymphoid 24, myeloid 24, all four juvenile thymus datasets  $\rho = 0.07$ ; myeloid 44, myeloid 49  $\rho = 0.10$ ; lymphoid 44, lymphoid 49  $\rho = 0.18$ . The corrected count matrices were then loaded into Seurat v3.1.3. This included the application of quality control filters (see below), normalization (scTransform; Hafemeister and Satija, 2019), finding integration anchors (reciprocal PCA), and integrating the data. Metadata on tissue of origin (marrow vs. thymus) or age of fish (juvenile vs. adult) were stored when useful.

The following quality control filters were applied: (1) Adult thymus: At least 200 genes but no more than 4,000, at least 500 UMIs but no more than 25,000, and a mitochondrial gene

percentage <15%; (2) Integrated adult and juvenile thymi: At least 200 genes but no more than 4,000, at least 500 UMIs but no more than 30,000, and a mitochondrial gene percentage <15%; (3) Lymphoid/progenitor and myeloid marrow fractions: At least 200 genes but no more than 4,000, at least 1,500 UMIs but no more than 50,000, and a mitochondrial gene percentage <15%. A more stringent lower bound on UMIs was applied here than on other objects due to clusters appearing to contain cell fragments at lower cutoffs that affected cell characterization; (4) Integrated lymphoid/progenitor marrow fraction and adult thymus: At least 200 genes but no more than 4,000, at least 500 UMIs but no more than 38,000, and a mitochondrial gene percentage <15%.

Following integration, standard processing steps of PCA dimensionality reduction, clustering, and visualization by UMAP were conducted. Cell identities were assigned by differential expression analysis and comparison to reported markers and/or as described in the text. We provide a table of the number of cells assigned to each cell type for our adult and juvenile thymus object and lymphoid/progenitor and myeloid marrow fractions object in Table S17. Additional subsetting on these objects was performed for subsequent analyses and standard processing procedures were applied.

Specifically, further subsetting was performed on the integrated lymphoid/progenitor and myeloid fractions processed with 39 principal components and clustered at a resolution of 0.8 to obtain the DC/NK development object (clusters 8, 21, 24, 28, 29, 32), B cell development object (clusters 3, 8, 19, 20, 21, 26, 28, 30), and T and NK cell objects (clusters 12, 24, 35). These subsetted objects were split and then re-integrated in the same manner as the full object. Note: clusters 8, 21, and 28 were progenitor clusters and were thus included in both the DC/NK and B cell development objects. Additionally, TECs and fibroblasts were subsetted from the adult and juvenile thymi processed with 50 principal components and a resolution of 1.5 by taking clusters 29, 31, 32, 39, and 42, splitting the object and re-integrating. We also subsetted, split, and re-integrated the T cell populations from the adult thymus object to perform T cell development trajectory analysis in Monocle 3. This subset excluded the non-T cell populations, such as B cells, TECs, DC-like cells (1) and (2), mixed, erythrocytes, NK-like cells, and macrophages. Cell type assignments were made similar to the full objects, and consistency of nomenclature was checked by extracting cell ids from the subsetted objects and confirming their location on the larger UMAP. All Seurat UMAPs of gene expression were generated with the FeaturePlot() function run on the RNA assay using default parameters except for the following specifications: cols = c(“lightgrey”, “red”), min.cutoff = 0, and order = TRUE. Genes visualized by dot plot were selected from among those identified through differential expression analysis using the FindAllMarkers() function in Seurat. In the majority of cases, we limited our results to genes expressed in a minimum of 25% of cells of a given population; but, in smaller datasets like our TEC, T, and NK-like cell, and B cell analyses, we reduced this minimum percentage to 15%. For the B cell analysis only when we also performed gene module analysis in Monocle 3 (see Gene module and trajectory analysis in Monocle 3), we included all

genes highlighted from the gene module analysis individually for clarity, even if they did not meet these minimum cutoffs in Seurat (e.g., *pcdh7b*). For final inclusion, we considered the statistical significance, fold change, specificity, relation to known mammalian markers, and general interest. We include the full lists of differentially expressed genes for all dot plots generated in this fashion as supplemental datasets: Tables S1 (adult thymus), S2 (adult and juvenile thymus), S5 (TECs), S13 (T and NK-like cells), and S15 (B cells). Additionally, we provide a summary table of the lineage defining genes mentioned in our results section and how they relate to past work in zebrafish and known mammalian lineage markers in Table S18 (Agulleiro et al., 2010; Athanasiadis et al., 2017; Avagyan et al., 2021; Bachem et al., 2012; Bajoghli et al., 2009; Balla et al., 2010; Baracho et al., 2014; Baran-Gale et al., 2020; Baranov et al., 2016; Barbier et al., 2019; Baron et al., 2019; Barratt and Weitz, 2021; Bautista et al., 2021; Beetz et al., 2007; Ben-Moshe et al., 2010; Bhandari et al., 2016; Boller and Grosschedl, 2014; Bouwman et al., 2021; Brahler et al., 2018; Burda et al., 2010; Butko et al., 2015; Cadieux et al., 2005; Calvete, 1994; Carmona et al., 2017; Challen et al., 2014; Chan et al., 1997; Chen et al., 2021; Chong et al., 1995; Chopin et al., 2019; Dale and Topczewski, 2011; Danger et al., 2022; Danilova et al., 2005; De Decker et al., 2021; Dee et al., 2016; Di et al., 2017; Donovan et al., 2013; Ellmeier et al., 2013; Fallatah et al., 2022; Farahani and Xaymardan, 2015; Ferrero et al., 2020; Fu et al., 2009; Ganis et al., 2012; Garcia et al., 2018; Gioacchino et al., 2021; Gonzalez-Leal et al., 2014; Gordon et al., 2012; Gore et al., 2016; Hartmann et al., 2013; Hason et al., 2022; Haunerding et al., 2021; He et al., 2014; Hernandez et al., 2018; Hess and Boehm, 2012; Hey and O'Neill, 2016; Hildner et al., 2008; Hinton et al., 2004; Ho et al., 2009; Huang et al., 2020; Ikeda et al., 2017; Iwanami et al., 2011; Jeffrey et al., 2006; Jojic et al., 2013; Kadouri et al., 2020; Kalev-Zylinska et al., 2003; Karsunky et al., 2003; Kim et al., 2006; Kim et al., 2019; Klebanoff et al., 2013; Koenen et al., 2019; Kopranner et al., 2001; Kowalczyk et al., 2015; Kruse et al., 2009; Kuil et al., 2020; Lancaster et al., 2018; Langenau et al., 2004; Lee et al., 1994; Liao and Wang, 2021; Lieberman, 2010; Lieschke et al., 2002; Liu et al., 2015; Lloyd and Snelgrove, 2018; Lyons et al., 2001; Ma et al., 2012; Mahony et al., 2021 *Preprint*; Marazuela et al., 2004; Martinez-Pomares, 2012; Matsumoto et al., 2008; Melichar et al., 2007; Mirabello et al., 2014; Mitchell et al., 2019; Modrell et al., 2017; Monney et al., 2002; Morath and Schamel, 2020; Murphy and Knee, 1994; Naganuma et al., 2011; Nemeth et al., 2003; Nitta et al., 2009; Nutt and Kee, 2007; Oettinger et al., 1990; Ohtani et al., 2008; Ong et al., 2020; Pandey et al., 2002; Parker et al., 2022; Parrish-Novak et al., 2000; Patel, 1999; Pearce et al., 2003; Perucho et al., 2014; Rajan et al., 2020; Remnant et al., 2021; Ridge et al., 2022; Rothenfusser et al., 2002; Rougeot et al., 2019; Ryu et al., 2005; Samborska et al., 2022; Schatz et al., 1989; Seeger et al., 1996; Shiau et al., 2015; Sichien et al., 2016; Sidney et al., 2014; Silacci et al., 2004; Smith et al., 1996; Stanley and Chitu, 2014; Sutoh et al., 2012; Takizawa et al., 2007; Tang et al., 2017; Trzpis et al., 2007; Tsai et al., 1994; Uechi et al., 2008; Ueno et al., 2004; Vanlandewijck et al., 2018; Venkateswarlu, 2003; Verlaet et al., 2002; Walton et al., 2015; Walzer et al., 2007; Wan et al., 2016; Wang et al., 2008; Wang and

Peng, 2021; Wei et al., 2011; Wei et al., 2002; White et al., 2010; Willett et al., 1997a; Yeo et al., 2018; Ying et al., 2005; Yoder et al., 2002; Yoder et al., 2007; Yokota et al., 2003; Yoon et al., 2015; Zakrzewska et al., 2010; Zaslavsky et al., 2010; Zhang et al., 2013; Zhou et al., 2017; Zhu et al., 2012; Zimmerman et al., 2011). Our publicly available cell browsers also allow for the expression of any mapped gene of interest to be visualized (see Data availability).

R version 3.6.2 was used for all Seurat analyses. Monocle 3 analyses were carried out in R version 3.6.2 or 4.0.2.

### Comparing the composition of the adult and juvenile thymi

Cells of the adult and juvenile thymi were compared in two ways: (i) performing a two-group beta-binomial test of significance on a cell-type-by-cell-type basis using the `countdata` R package (Tables S3 and S6; Pham et al., 2010) and (ii) using the `FindAllMarkers` and `FindMarkers` functions in Seurat on the full object and within each cell type split by timepoint (adult vs. juvenile). Default parameters were used in all cases unless otherwise specified. Point (i) was applied to both the full adult and juvenile thymus in addition to the TEC/fibroblast subset. The full list of genes resulting from comparing adult vs. juvenile erythrocytes is provided in Table S4. To confirm that the ribosomal subunits identified as being differentially expressed in adult and juvenile thymi were not a technical artifact of inconsistent SoupX contamination removal, the uncorrected data was examined. Genes listed in the main text as being enriched in the adult thymus erythrocytes were likewise enriched in the uncorrected data.

### Clustering with RaceID3 and VarID for T cell trajectory analysis

For the analysis of T cell trajectories and the identification of ETPs, we performed an independent clustering of the adult thymi in RaceID3 (Herman et al., 2018). Following batch correction and the identification of outlier clusters, the RaceID3 SCseq object was subjected to integrated processing with VarID (Grun, 2020). More specifically, the RaceID3 SCseq object was generated from the raw data by calling the `filterdata()` function, with the parameters `mintotal = 1,000`, `minexpr = 3`, `minnumber = 1`, and the application of the internal RaceID batch correction method. This method identifies batch-associated genes through a local neighborhood approach and successive pair-wise merging (Herman et al., 2018). For this analysis, the size of the neighborhood, `knn`, was set to 10. Mitochondrial genes were regressed out for cell type inference, with `ccor` set to 0.15. Next, `varRegression()` was run to remove residual batch effects. Following these quality control steps, a distance matrix was computed and initial clusters determined by calling `compdist()` and `clustexp()` with default parameters. Outliers of the initial `k`-medoids clusters were identified with `findoutliers()`, a function that makes use of an internally computed background model of the expected gene expression variability. This model assumes a negative binomial distribution for transcript counts defined by the mean and variance of expression of each gene per cluster and has been validated for raw count data (Grun et al., 2014). Following the identification of outliers, `rforcorrect()` was run to improve the robustness of the final clustering. Next,

clusters corresponding to cells in the T cell developmental pathway were subsetted for trajectory analysis; thus, clusters 4, 13, 14, 16, and 17 (corresponding to non-T cell populations, e.g., TECs, NK-like cells, B cells, etc.) do not appear in the UMAP shown in Fig. 2 A. VarID was used for subsequent processing on this subsetted UMAP. In brief, the function `pruneKnn` was run on the filtered expression matrix with `large = TRUE`, `pcaComp = 100`, `regNB = TRUE`, `knn = 10`,  $\alpha = 1$ , and `ngenes = 2,000`, and links with probabilities lower than P value  $< 0.01$  were removed. The adjacency matrix was computed, Louvain clustering was performed, and the clusters were visualized. Gene expression was visualized by calling a color-scale modified version of the function `fractDotPlot()` with  $\log_2$  transformation. FateID analysis was performed on this updated RaceID3 SCseq object. R version 4.1.0 was used for this analysis.

### Exploring T cell developmental trajectories with FateID

To investigate T cell trajectories, we applied the FateID algorithm to the RaceID3/VarID-generated object (Grun, 2020; Herman et al., 2018). In brief, FateID takes as input the user-defined most differentiated cells and implements an iterative random forest classification to calculate fate bias, working its way back to progenitor populations. Based on gene expression patterns consistent with mature populations, clusters 1, 19, and 5 were set as the endpoints in this analysis for trajectory 1 (mature T cells), trajectory 19 (cytotoxic T cells), and trajectory 5 ( $\gamma\delta$  T cells), respectively. In addition to specifying these endpoints, `fatebias()` was called using the following non-default parameters: the VarID distance matrix, `minnr = 5`, `minnrh = 10`, `confidence = 0.85`, and `seed = 12345`, and `compdr()` was used to visualize the dimensional reduction by UMAP. To observe the differentiation trajectories, principal curves were computed using the results from `fateBias()` and `compdr()` with a threshold of 0.15 for the fraction of random forest votes ascribed to a given trajectory. To investigate gene expression changes along pseudotime, cells with fate bias toward a given trajectory were extracted from the principal curve. Lowly expressed genes were removed by filtering the RaceID3-normalized transcript expression values using `filterset()` with `minexp = 2` and `minnumber = 1`. A SOM of the pseudo-temporal order was generated using `getsom()` with the following inputs: the filtered gene set, `nb = 100`, and  $\alpha = 0.5$ . This function groups genes with similar expression profiles into modules for further analysis. The SOM matrix was further processed with `procsom()` to group the initial modules into larger nodes, for which genes having  $> 0.9$  correlation of the SOM z-scores were combined into the same node. The minimal number of genes per node was set to 5. The processed SOMs were plotted with `ploheatmap()`. Table S7 lists the genes within each node. Pseudotime plots were generated using the FateID `plotexpression()` function for each gene in each trajectory, separately. Y axis upper limits were manually set to ensure these plots could be directly compared for each gene between trajectories 1 and 19; for some genes this upper limit excluded a small number of outlier cells for which inclusion in the plot would obscure the interpretation of finer expression level changes in pseudotime.  $\alpha$  was set to 0.3 for all expression

curves. R version 4.1.0 was used for the fate bias and pseudotime analysis of T cell development.

### Exploring B cell developmental trajectories with FateID

To investigate B cell developmental trajectories, we directly applied FateID to our B cell development Seurat object. Specifically, we stored the RNA count data and cell type classification as determined in Seurat and used these data as input to the `fateBias()` function with target clusters set to the mature IgD and IgT B cells and the following non-default parameters: `minnr = 30`, `minnrh = 45`, `confidence = 0.95`, and `seed = 12345`. Next, a dimensional reduction representation was generated using the `compdr()` function with `m = c("umap")` and `umap$D2` was set using the cell embeddings from Seurat to preserve UMAP coordinates. Principal curves were computed using the results from `fateBias()` and `compdr()` and a threshold of 0.8 for the fraction of random forest votes ascribed to a given trajectory. This was purposefully set to be more stringent than in our T cell trajectory analysis due to the clearer assignment of B cells based on their immunoglobulin heavy chain gene expression. Investigations of gene expression changes in pseudotime were performed by calling the `plotexpression()` function for each gene of interest on the cells assigned to each trajectory independently. Y axis upper limits were manually set as above to allow for direct comparison of expression between trajectories and  $\alpha$  values were set on a gene-by-gene basis to highlight expression changes in pseudotime. R version 4.1.0 was used for the fate bias and pseudotime analysis of B cell development.

### Identification of ETPs in thymus

To identify zebrafish ETPs, we employed multiple methods in addition to standard differential expression analysis which initially led to our use of this terminology. First, we integrated data from the adult thymus with that of the lymphoid/progenitor fraction of the adult marrow and highlighted the cell IDs of the cells from a high-resolution clustering of the adult thymus (resolution = 3) that were the strongest candidate ETPs (cluster 38). Clustering within the HSPC cluster of this integrated object would suggest that these cells were most similar transcriptionally to HSPCs in the marrow and offer additional support for this nomenclature. Next, we ensured the consistency between our ETP cell type calling within our Seurat and RaceID3 analyses and took advantage of our RaceID3 clustering to use StemID2 to infer the most progenitor-like population within the zebrafish thymus (Grun et al., 2016; Herman et al., 2018). The StemID2 score is computed based on the transcription entropy and number of inter-cluster links of each population, with less differentiated cell types tending to have greater transcription entropy and a higher number of inter-cluster links than more differentiated cell types (Grun et al., 2016; Herman et al., 2018). Transcriptional entropy was computed using default parameters; projections of cells onto inter-cluster links were computed with `projcells()` using default parameters except for setting `nmode = FALSE`. A P value threshold (`pthr`) of 0.01 was used as the cutoff for link significance. The StemID score, the product of the two values, was calculated for all clusters with `compscore()` and `scthr = 0.8`. Mean fate bias toward each trajectory was visualized

for the clusters with the highest StemID scores (clusters 7, 11, 15, and 18). Error bars depict standard error of the mean (SEM). Additionally, for the two clusters with the highest StemID scores, we performed pathway analysis using Metascape (Zhou et al., 2019). Finally, we were able to take advantage of our kidney marrow analysis and previously reported markers of mouse and human ETPs in the literature to compute gene module scores in Seurat for three gene modules: (1) an HSPC module as determined by Seurat differential expression analysis (the 20 most specific HSPC genes (smallest pct.2 values) were assigned to this module), (2) an HSPC module as determined by Monocle 3 gene module analysis (module 16), and (3) an ETP module including orthologs of highly specific mouse and human ETP markers as determined by literature review (Table S9; Cante-Barrett et al., 2017; Cordes et al., 2022 Preprint; Porritt et al., 2004; Rothenberg et al., 2008; Zeng et al., 2019). In most cases, all zebrafish orthologs were included in this gene module with the exception of *gata2a*, previously shown to mark eosinophils (Balla et al., 2010), and *cd44c* due its lack of inclusion in the current mapping.

Once the ETPs were identified in the combined adult lymphoid/progenitor marrow fraction and thymus object, the Seurat FindMarkers() and FindConservedMarkers() functions were called to determine differentially expressed genes between the thymic ETPs and marrow HSPCs (cluster 4 in object visualized with 35 PCs and resolution = 0.6) and also the conserved genes between these populations (Tables S10 and S11). All cells within cluster 4 were used for the purpose of assessing conserved markers and the markers of highest conservation are reported here, defined by expression in <5% of non-cluster 4 cells derived from the thymus and <25% of non-cluster 4 cells derived from the marrow. A more stringent cutoff was used for the thymus due to the relatively sparse recovery of most transcripts (low pct.1 values for specific markers); however, we note that this did eliminate genes like *si:ch21l-16lc3.6*, which were less specific but more sensitive, in this case recovered in 94% of thymic ETPs but also 15% of other thymic cells. Our criteria also favored the recovery of myeloid lineage genes due to the lower stringency in the marrow. This approach was advantageous from the point of view of identifying highly specific ETP markers that were also expressed in HSPCs. For other analyses and goals, different filtering criteria may be desired. Upon noting a broader population of thymic cells included in this cluster beyond our putative ETPs, for differential expression analysis, only the most specific subset of thymic ETPs from the adult thymus object (cluster 38, resolution = 3) were compared to cluster 4 marrow cells. Significantly enriched genes in HSPCs or ETPs had to meet the following criteria: be detected in >40% of the population, have a ln fold change >0.4 over the other tissue's population, and be detected in <5% of the other tissue's population. Wishing to minimize the recovery of differentially expressed genes simply due to tissue differences, FindMarkers() with a min.pct of 0.02 was additionally called on the B cells of the marrow vs. the B cells of the thymus (cells within clusters 2 and 19). Genes that met the criteria above but were identified as being differentially expressed in thymic vs. marrow B cells and were recovered in <10% of marrow B cells were not included in the final gene lists for HSPC or ETP enrichment. None of the marrow HSPC enriched

genes were eliminated from this analysis; however *lck*, *runx3*, *FP236356.1*, *ccr9b*, and *bcl11a* were removed from the ETP enrichment list. While a true enrichment in any of these genes may be present, this elimination allows us to be more conservative in our analysis. We note that although *pdpkl1a* was also identified as a differentially expressed gene in marrow vs. thymic B cells, we have chosen to include it in our enrichment list because its expression in thymic ETPs is less likely to be due to tissue-specific transcript contamination as detection in marrow B cells was non-negligible at 14.4% (vs. 30.5% in thymus). Eosinophilic markers were excluded from the set of markers enriched in kidney marrow HSPCs because it was observed that they were present only in a specific subset of eosinophil progenitors.

### Gene module and trajectory analysis in Monocle 3

The developer branch of Monocle 3 was used in all analyses (Cao et al., 2019). Seurat objects were first converted to Monocle 3 cell\_data\_set objects using the as.cell\_data\_set() function, thus preserving the dimensionality reduction, clustering, and cell annotation from Seurat. The nearest neighbor graph was rebuilt in Monocle 3 and used to conduct trajectory analysis and compute pseudotime. Standard functions were employed for differential gene expression analysis.

To enable plotting in Monocle 3, the gene\_name and gene\_short\_name metadata slots were populated manually via calling the rownames() and rowData() functions for the desired inputs and estimate\_size\_factors() was run.

For each of the following objects, the following deviations from the default parameters were used to define clusters and learn trajectory graphs: full marrow object: resolution 0.003; DC/NK object: resolution = 0.009, ncenter = 1,300; B cell object: resolution = 0.009, ncenter = 400, minimal\_branch\_len = 15; thymus T cell subset: resolution = 0.004, ncenter = 1,200. (Note: Trajectory analysis was not performed on the T and NK cell object, a subset of the full marrow object, so the Seurat clustering was used in downstream Monocle 3 analysis.)

Following clustering and trajectory determination, differential gene expression analysis was conducted using the graph\_test function, which performs spatial autocorrelation analysis using the Moran's I test. As recommended, the principal\_graph was selected as the neighbor graph to use in this analysis and identifies genes that vary over the computed trajectory instead of between clusters. To find modules of co-regulated genes, find\_gene\_modules() was called on a subset of the original object only containing differentially expressed genes with q\_values set to <0.05 and the resolutions listed above (Tables S8, S12, and S16). Standard processing was followed for visualizing the computed gene modules and exploring module specificity. Verification of individual genes identified as being specific to populations of interest was performed in Seurat.

The most differentially expressed genes between the T and NK cell populations were identified through calling the top\_markers() function grouping the cells by lineage, as assigned from the full marrow object (Table S14).

For B cell pseudotime analysis, all cells from the B cell and progenitor object in addition to 11 genes of interest: *pax5*, *cd79a*,

*sidl*, *dntt*, *rag1*, *rag2*, *csflrb*, *ebf3a*, *mki67*, *igl3v5*, and *igllc3* were subsetted. An ordering of the cells was determined by interactively selecting the root state to be within the HSPC population from within the `order_cells()` function. The pseudotime ordering was then observed for all genes of interest with the genes ordered according to time of expression and a minimum untransformed expression level of 0.1. Note: *ebf3a*, although not ascribed as the direct ortholog of the B cell commitment factor *EBF1* in mammals, appeared to have the most consistent expression with *EBF1* and was thus selected as a B cell gene. Further, we note that although the IgT and IgD populations are two independent lineages, we did not separate them on the pseudotime plot for the purposes of exploring early B cell development when they appeared intermixed. The genes we selected for subsetting would not be anticipated to be expressed differently between these two lineages.

For T cell pseudotime analysis, all cells from the subsetted adult thymus object containing only T cell populations along with the following seven genes of interest: *cd4-1*, *cd4-2.1*, *cd4-2.2*, *cd8a*, *cd8b*, *csflrb*, and *il2rb* were subsetted. Based on the expression patterns of the *CD4* and *CD8* orthologs, cells were selected using the `choose_cells()` function for pathway analysis of the mature T cells and cytotoxic T cells, respectively. These two subsets were ordered using the same root selected from within the ETP population. Next, we modified the `plot_genes_in_pseudotime()` function to visualize single cells expressing orthologs of *CD4* and *CD8* simultaneously. Specifically, we plot the sum of all *CD4* orthologs, the sum of all *CD8* orthologs, and the minimum of those sums as the joint *CD4-CD8* expression using a minimum untransformed expression level of 0.1.

### TCR alignment and analysis

TCR  $\alpha$  and  $\beta$  constant regions were not annotated in zebrafish Ensembl GRCz11 (the closest annotation is FP236356.1, a 109,805 bp region that includes TCR  $\beta$  constant 1 [*trbc1*]). Therefore, an independent alignment was performed to recover TCR  $\alpha$  and TCR  $\beta$  transcripts based on the recovery of their constant regions, the most 3'. A custom genome was defined with three genes, two individually for the four Cb1 (TCRb.C1) and three Cb2 exons (TCRb.C2), respectively (2), and one for the TCR  $\alpha$  constant region (TCRa.C) as annotated in GenBank listing AF424545.1. The custom alignment was processed through Cell Ranger, similarly to the full transcriptome mapping, and assigned as “threeprime” chemistry. The filtered `feature_bc_matrix` files were loaded into Seurat in a way that preserved all of the cellular barcodes as identified by Cell Ranger (CreateSeuratObject settings `min.cells = -1`, `min.features = -1`). All datasets from which more than one cell was recovered were merged, and the cell ids corresponding to counts of  $TCRa.C > 0$ ,  $TCRb.C1 > 1$ , and  $TCRb.C2 > 0$  were retrieved.  $\alpha\beta$  T cells were defined as meeting any of the above criteria. Similarly, cell ids corresponding to counts of *trdc*  $> 0$  and *tcrg* (*si:dkeyp-13d12.23*)  $> 1$  were also retrieved from the original transcriptome mappings.  $\gamma\delta$  T cells were defined as meeting either or both of those requirements for *trdc* and *tcrg* recovery. These transcript cutoffs were set based on the observed sensitivity and specificity of cutoff choice in the marrow and used

consistently in the thymus. We note that this choice appears less specific in the thymus but we maintained the marrow cutoffs for sensitivity, consistency, and transparency.

### Online supplemental material

Fig. S1 shows an extended cell type characterization of the adult zebrafish thymus in addition to supporting characterizations of the adult and juvenile thymus, including visualization of cells by timepoint, an investigation into the mixed granulocyte population, and ribosomal subunit heterogeneity. Fig. S2 shows supporting visualizations and analysis for the TEC and fibroblast subset of the adult and juvenile thymus along with supplemental visualizations of the FateID analysis and identification of zebrafish ETPs. Fig. S3 shows conserved marker genes between marrow HSPCs and ETPs. Fig. S4 shows differentially expressed genes between marrow HSPCs and ETPs. Fig. S5 shows representative plots of the fluorescence-activated cell sorting gates for marrows and thymi. Table S1 shows differentially expressed genes in the adult zebrafish thymus. Table S2 shows differentially expressed genes from the adult and juvenile zebrafish thymi. Table S3 shows beta-binomial test results for the adult and juvenile zebrafish thymi. Table S4 shows differentially expressed genes in adult and juvenile erythrocytes from the adult and juvenile thymi analysis. Table S5 shows differentially expressed genes in the TEC and fibroblast subset of the adult and juvenile thymi. Table S6 shows beta-binomial test results for the TEC and fibroblast subset from the adult and juvenile thymi. Table S7 shows gene lists for SOM nodes from the FateID analysis of the adult thymus. Table S8 shows Monocle 3 gene modules of the adult zebrafish kidney marrow. Table S9 shows gene modules used to identify zebrafish ETPs. Table S10 shows conserved marker genes between marrow HSPCs and ETPs. Table S11 shows differentially expressed genes between marrow HSPCs and ETPs. Table S12 shows Monocle 3 gene modules of the DC-like cell and NK-like cell marrow subset. Table S13 shows differentially expressed genes for marrow T and NK-like cell subset. Table S14 shows top distinguishing genes between marrow T and NK-like cells. Table S15 shows differentially expressed genes for marrow B cell subset. Table S16 shows Monocle 3 gene modules of the marrow B cell subset. Table S17 shows cell counts by cell type for full adult and juvenile thymi and marrow analyses. Table S18 shows marker gene references.

### Data availability

The data reported in this article have been deposited in the Gene Expression Omnibus database under accession no. GSE190794. Additionally, cell browser objects (Speir et al., 2021) have been generated for ease of data exploration and are available at the following URL: <https://dr-marrow-thymus.cells.ucsc.edu>.

### Acknowledgments

The authors thank Serine Avagyan, Jon Henninger, and Aya Ludin Tal for their training and technical expertise at the outset of this work; Jeff Mito for his analytical/computational guidance; Emmet Flynn for his assistance with generating cell browsers; Shiv Pillai, Alejandro Gutierrez, and Vijay Sankaran for their

helpful feedback and experimental suggestions; Yi Zhou, Marlies Rossman, Avik Choudhuri, Teshika Jayewickreme, Haitong Mao, Alec Walker, Nancy Manley, Angel Edgardo Flores, and Ana Cvejic for their helpful discussions for data interpretation; Ronald Mathieu, Mahnaz Paktinat, and the entire Boston Children's Hospital Flow Cytometry core for their assistance in data collection; Chris Lawrence, Shane Hurley, Li-Kun Zhang, Jason Best, and the rest of the Aquatics Research Facilities Staff for their help with fish maintenance; and Sam Watrus for his critical reading of the manuscript and helpful guidance along the way. The authors are grateful to all of their colleagues who contributed ideas and support throughout this work. The graphical abstract was created with [BioRender.com](https://www.biorender.com).

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award numbers F30HL152628 (to S.A. Rubin), T32HL007574 (to C. Pessoa Rodrigues), and R01HL144780 (to L.I. Zon). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional funding for this work was provided by the Human Frontier Science Program (HFSP LT000494/2020-L to C.S. Baron). L.I. Zon is a Howard Hughes Medical Institute Investigator.

Author contributions: S.A. Rubin, C.S. Baron, and L.I. Zon designed the study. S.A. Rubin, C.S. Baron, and A.F. Corbin conducted the experiments. C. Pessoa Rodrigues, M. Duran, and C. Trapnell contributed new methods/analytic tools. S.A. Rubin, C. Pessoa Rodrigues, M. Duran, and S.P. Yang analyzed the data. S.A. Rubin, C.S. Baron, C. Pessoa Rodrigues, and M. Duran created and/or prepared visualizations of the data. S.A. Rubin wrote the original draft of the manuscript with guidance from C.S. Baron and L.I. Zon. All authors critically reviewed and edited the manuscript.

Disclosures: L.I. Zon reported personal fees from Fate Therapeutics, CAMP4 Therapeutics, Amagma Therapeutics, Scholar Rock, Branch Biosciences, Celularity, and Cellarity outside the submitted work. No other disclosures were reported.

Submitted: 6 January 2022

Revised: 11 June 2022

Accepted: 6 July 2022

## References

- Agulleiro, M.J., S. Roy, E. Sanchez, S. Puchol, N. Gallo-Payet, and J.M. Cerda-Reverter. 2010. Role of melanocortin receptor accessory proteins in the function of zebrafish melanocortin receptor type 2. *Mol. Cell. Endocrinol.* 320:145–152. <https://doi.org/10.1016/j.mce.2010.01.032>
- Alt, F.W., and D. Baltimore. 1982. Joining of immunoglobulin heavy chain gene segments: Implications from a chromosome with evidence of three D-JH fusions. *Proc. Natl. Acad. Sci. USA.* 79:4118–4122. <https://doi.org/10.1073/pnas.79.13.4118>
- Athanasiadis, E.I., J.G. Bothhof, H. Andres, L. Ferreira, P. Lio, and A. Cvejic. 2017. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat. Commun.* 8:2045. <https://doi.org/10.1038/s41467-017-02305-6>
- Avagyan, S., M.C. Weber, S. Ma, M. Prasad, W.P. Mannherz, S. Yang, J.D. Buenrostro, and L.I. Zon. 2021. Single-cell ATAC-seq reveals GATA2-dependent priming defect in myeloid and a maturation bottleneck in

- lymphoid lineages. *Blood Adv.* 5:2673–2686. <https://doi.org/10.1182/bloodadvances.2020002992>
- Bachem, A., E. Hartung, S. Guttler, A. Mora, X. Zhou, A. Hegemann, M. Plantinga, E. Mazzini, P. Stoitzner, S. Gurka, et al. 2012. Expression of XCR1 characterizes the batf3-dependent lineage of dendritic cells capable of antigen cross-presentation. *Front. Immunol.* 3:214. <https://doi.org/10.3389/fimmu.2012.00214>
- Bajoghli, B., N. Aghaallaei, I. Hess, I. Rode, N. Netuschil, B.H. Tay, B. Venkatesh, J.K. Yu, S.L. Kaltenbach, N.D. Holland, et al. 2009. Evolution of genetic networks underlying the emergence of thymopoiesis in vertebrates. *Cell.* 138:186–197. <https://doi.org/10.1016/j.cell.2009.04.017>
- Bajoghli, B., A.M. Dick, A. Claasen, L. Doll, and N. Aghaallaei. 2019. Zebrafish and Medaka: Two teleost models of T-cell and thymic development. *Int. J. Mol. Sci.* 20:E4179. <https://doi.org/10.3390/ijms20174179>
- Balla, K.M., G. Lugo-Villarino, J.M. Spitsbergen, D.L. Stachura, Y. Hu, K. Banuelos, O. Romo-Fewell, R.V. Aroian, and D. Traver. 2010. Eosinophils in the zebrafish: Prospective isolation, characterization, and eosinophilia induction by helminth determinants. *Blood.* 116:3944–3954. <https://doi.org/10.1182/blood-2010-03-267419>
- Baracho, G.V., M.H. Cato, Z. Zhu, O.R. Jaren, E. Hobeika, M. Reth, and R.C. Rickert. 2014. PDK1 regulates B cell differentiation and homeostasis. *Proc. Natl. Acad. Sci. USA.* 111:9573–9578. <https://doi.org/10.1073/pnas.1314562111>
- Baran-Gale, J., M.D. Morgan, S. Maio, F. Dhalla, I. Calvo-Asensio, M.E. Deadman, A.E. Handel, A. Maynard, S. Chen, F. Green, et al. 2020. Ageing compromises mouse thymus function and remodels epithelial cell differentiation. *Elife.* 9:e56221. <https://doi.org/10.7554/eLife.56221>
- Baranov, M.V., N.H. Revelo, I. Dingjan, R. Maraschini, M. Ter Beest, A. Honigsmann, and G. van den Bogaart. 2016. SWAP70 organizes the actin cytoskeleton and is essential for phagocytosis. *Cell Rep.* 17:1518–1531. <https://doi.org/10.1016/j.celrep.2016.10.021>
- Barbier, P., O. Zejnelli, M. Martinho, A. Lasorsa, V. Belle, C. Smet-Nocca, P.O. Tsvetkov, F. Devred, and I. Landrieu. 2019. Role of tau as a microtubule-associated protein: Structural and functional aspects. *Front. Aging Neurosci.* 11:204. <https://doi.org/10.3389/fnagi.2019.00204>
- Baron, C.S., A. Barve, M.J. Muraro, R. van der Linden, G. Dharmadhikari, A. Lyubimova, E.J.P. de Koning, and A. van Oudenaarden. 2019. Cell type purification by single-cell transcriptome-trained sorting. *Cell.* 179:527–542.e19. <https://doi.org/10.1016/j.cell.2019.08.006>
- Barratt, J., and I. Weitz. 2021. Complement factor D as a strategic target for regulating the alternative complement pathway. *Front. Immunol.* 12:712572. <https://doi.org/10.3389/fimmu.2021.712572>
- Barraza, F., R. Montero, V. Wong-Benito, H. Valenzuela, C. Godoy-Guzman, F. Guzman, B. Kollner, T. Wang, C.J. Secombes, K. Maisey, and M. Imarai. 2020. Revisiting the teleost thymus: Current knowledge and future perspectives. *Biology.* 10:8. <https://doi.org/10.3390/biology10010008>
- Bautista, J.L., N.T. Cramer, C.N. Miller, J. Chavez, D.I. Berrios, L.E. Byrnes, J. Germino, V. Ntranos, J.B. Sneddon, T.D. Burt, et al. 2021. Single-cell transcriptional profiling of human thymic stroma uncovers novel cellular heterogeneity in the thymic medulla. *Nat. Commun.* 12:1096. <https://doi.org/10.1038/s41467-021-21346-6>
- Beetz, S., D. Diekhoff, and L.A. Steiner. 2007. Characterization of terminal deoxynucleotidyl transferase and polymerase mu in zebrafish. *Immunogenetics.* 59:735–744. <https://doi.org/10.1007/s00251-007-0241-7>
- Ben-Moshe, Z., G. Vatine, S. Alon, A. Tovim, P. Mracek, N.S. Foulkes, and Y. Goshilf. 2010. Multiple PAR and E4BP4 bZIP transcription factors in zebrafish: Diverse spatial and temporal expression patterns. *Chronobiol. Int.* 27:1509–1531. <https://doi.org/10.3109/07420528.2010.510229>
- Bertrand, J.Y., A.D. Kim, S. Teng, and D. Traver. 2008. CD41<sup>+</sup> cmyb<sup>+</sup> precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis. *Development.* 135:1853–1862. <https://doi.org/10.1242/dev.015297>
- Bhandari, S., J.N. Lee, Y.I. Kim, I.K. Nam, S.J. Kim, S.J. Kim, S. Kwak, G.S. Oh, H.J. Kim, H.J. Yoo, et al. 2016. The fatty acid chain elongase, Elov11, is required for kidney and swim bladder development during zebrafish embryogenesis. *Organogenesis.* 12:78–93. <https://doi.org/10.1080/15476278.2016.1172164>
- Boller, S., and R. Grosschedl. 2014. The regulatory network of B-cell differentiation: A focused view of early B-cell factor 1 function. *Immunol. Rev.* 261:102–115. <https://doi.org/10.1111/imr.12206>
- Bouwman, A.C., K.R. van Daalen, S. Crnk, T. Ten Broeke, and N. Boven-schen. 2021. Intracellular and extracellular roles of granzyme K. *Front. Immunol.* 12:677707. <https://doi.org/10.3389/fimmu.2021.677707>
- Brahler, S., B.H. Zinselmeyer, S. Raju, M. Nitschke, H. Suleiman, B.T. Saunders, M.W. Johnson, A.M.C. Bohner, S.F. Viehmann, D.J. Theisen,

- et al. 2018. Opposing roles of dendritic cell subsets in experimental GN. *J. Am. Soc. Nephrol.* 29:138–154. <https://doi.org/10.1681/ASN.2017030270>
- Burda, P., P. Laslo, and T. Stopka. 2010. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia.* 24:1249–1257. <https://doi.org/10.1038/leu.2010.104>
- Butko, E., M. Distel, C. Pouget, B. Weijts, I. Kobayashi, K. Ng, C. Mosimann, F.E. Poulain, A. McPherson, C.W. Ni, et al. 2015. Gata2b is a restricted early regulator of hemogenic endothelium in the zebrafish embryo. *Development.* 142:1050–1061. <https://doi.org/10.1242/dev.119180>
- Cadieux, B., B.P. Chitramuthu, D. Baranowski, and H.P.J. Bennett. 2005. The zebrafish progranulin gene family and antisense transcripts. *BMC Genom.* 6:156. <https://doi.org/10.1186/1471-2164-6-156>
- Calvete, J.J. 1994. Clues for understanding the structure and function of a prototypic human integrin: The platelet glycoprotein IIb/IIIa complex. *Thromb. Haemostas.* 72:1–15. <https://doi.org/10.1055/s-0038-1648803>
- Cante-Barrett, K., R.D. Mendes, Y. Li, E. Vroegindewij, K. Pike-Overzet, T. Wabeke, A.W. Langerak, R. Pieters, F.J.T. Staal, and J.P.P. Meijerink. 2017. Loss of CD44(dim) expression from early progenitor cells marks T-cell lineage commitment in the human thymus. *Front. Immunol.* 8:32. <https://doi.org/10.3389/fimmu.2017.00032>
- Cao, J., M. Spielmann, X. Qiu, X. Huang, D.M. Ibrahim, A.J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F.J. Steemers, et al. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature.* 566: 496–502. <https://doi.org/10.1038/s41586-019-0969-x>
- Carmona, S.J., S.A. Teichmann, L. Ferreira, I.C. Macaulay, M.J.T. Stubbington, A. Cvejic, and D. Gfeller. 2017. Single-cell transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune cell types. *Genome Res.* 27:451–461. <https://doi.org/10.1101/gr.207704.116>
- Challen, G.A., D. Sun, A. Mayle, M. Jeong, M. Luo, B. Rodriguez, C. Mallaney, H. Celik, L. Yang, Z. Xia, et al. 2014. Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. *Cell Stem Cell.* 15:350–364. <https://doi.org/10.1016/j.stem.2014.06.018>
- Chan, F.Y., J. Robinson, A. Brownlie, R.A. Shivdasani, A. Donovan, C. Brugnara, J. Kim, B.C. Lau, H.E. Witkowska, and L.I. Zon. 1997. Characterization of adult  $\alpha$ - and  $\beta$ -globin genes in the zebrafish. *Blood.* 89: 688–700. <https://doi.org/10.1182/blood.v89.2.688>
- Chen, D., T.X. Tang, H. Deng, X.P. Yang, and Z.H. Tang. 2021. Interleukin-7 biology and its effects on immune cells: Mediator of generation, differentiation, survival, and homeostasis. *Front. Immunol.* 12:747324. <https://doi.org/10.3389/fimmu.2021.747324>
- Chong, J.P., H.M. Mahbubani, C.Y. Khoo, and J.J. Blow. 1995. Purification of an MCM-containing complex as a component of the DNA replication licensing system. *Nature.* 375:418–421. <https://doi.org/10.1038/375418a0>
- Chopin, M., A.T. Lun, Y. Zhan, J. Schreuder, H. Coughlan, A. D'Amico, L.A. Mielke, F.F. Almeida, A.J. Kueh, R.A. Dickins, et al. 2019. Transcription factor PU.1 promotes conventional dendritic cell identity and function via induction of transcriptional regulator DC-SCRIPT. *Immunity.* 50: 77–90.e5. <https://doi.org/10.1016/j.immuni.2018.11.010>
- Cordes, M., K. Cante-Barrett, E.B. van den Akker, F.A. Moretti, S.M. Kielbasa, S. Vloemans, L. Garcia-Perez, C. Teodosio, J.J.M. van Dongen, K. Pike-Overzet, et al. 2022. Single-cell immune profiling reveals novel thymus-seeding populations, T cell commitment, and multi-lineage development in the human thymus. *bioRxiv.* (Preprint posted February 19, 2022). <https://doi.org/10.1101/2022.02.18.481026>
- Dale, R.M., and J. Topczewski. 2011. Identification of an evolutionarily conserved regulatory element of the zebrafish col2a1a gene. *Dev. Biol.* 357: 518–531. <https://doi.org/10.1016/j.ydbio.2011.06.020>
- Danger, R., Y. Feseha, and S. Brouard. 2022. The pseudokinase TRIB1 in immune cells and associated disorders. *Cancers.* 14:1011. <https://doi.org/10.3390/cancers14041011>
- Danilova, N., J. Bussmann, K. Jekosch, and L.A. Steiner. 2005. The immunoglobulin heavy-chain locus in zebrafish: Identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat. Immunol.* 6: 295–302. <https://doi.org/10.1038/nii166>
- De Decker, M., M. Lavaert, J. Roels, L. Tilleman, B. Vandekerckhove, G. Lecclercq, F. Van Nieuwerburgh, P. Van Vlierberghe, and T. Taghon. 2021. HES1 and HES4 have non-redundant roles downstream of Notch during early human T-cell development. *Haematologica.* 106:130–141. <https://doi.org/10.3324/haematol.2019.226126>
- Dee, C.T., R.T. Nagaraju, E.I. Athanasiadis, C. Gray, L. Fernandez Del Ama, S.A. Johnston, C.J. Secombes, A. Cvejic, and A.F.L. Hurlstone. 2016. CD4-Transgenic zebrafish reveal tissue-resident Th2- and regulatory T cell-like populations and diverse mononuclear phagocytes. *J. Immunol.* 197: 3520–3530. <https://doi.org/10.4049/jimmunol.1600959>
- Di, Q., Q. Lin, Z. Huang, Y. Chi, X. Chen, W. Zhang, and Y. Zhang. 2017. Zebrafish nephrosin helps host defence against Escherichia coli infection. *Open Biol.* 7:170040. <https://doi.org/10.1098/rsob.170040>
- Donovan, J., X. Shiwen, J. Norman, and D. Abraham. 2013. Platelet-derived growth factor  $\alpha$  and  $\beta$  receptors have overlapping functional activities towards fibroblasts. *Fibrogenesis Tissue Repair.* 6:10. <https://doi.org/10.1186/1755-1536-6-10>
- Dore, L.C., J.D. Amigo, C.O. Dos Santos, Z. Zhang, X. Gai, J.W. Tobias, D. Yu, A.M. Klein, C. Dorman, W. Wu, et al. 2008. A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc. Natl. Acad. Sci. USA.* 105: 3333–3338. <https://doi.org/10.1073/pnas.0712312105>
- Dornburg, A., T. Ota, M.F. Criscitiello, I. Salinas, J.O. Sunyer, S. Magadan, P. Boudinot, Z. Xu, M.F. Flajnik, A. Singer, et al. 2021. From IgZ to IgT: A call for a common nomenclature for immunoglobulin heavy chain genes of ray-finned fish. *Zebrafish.* 18:343–345. <https://doi.org/10.1089/zeb.2021.0071>
- Ellmeier, W., L. Haust, and R. Tschisnarov. 2013. Transcriptional control of CD4 and CD8 coreceptor expression during T cell development. *Cell Mol. Life Sci.* 70:4537–4553. <https://doi.org/10.1007/s00018-013-1393-2>
- Fallatah, W., R. De, D. Burks, R.K. Azad, and P. Jagadeeswaran. 2022. Analysis of transcribed sequences from young and mature zebrafish thrombocytes. *PLoS One.* 17:e0264776. <https://doi.org/10.1371/journal.pone.0264776>
- Farahani, R.M., and M. Xaymardan. 2015. Platelet-derived growth factor receptor  $\alpha$  as a marker of mesenchymal stem cells in development and stem cell biology. *Stem Cells Int.* 2015:362753. <https://doi.org/10.1155/2015/362753>
- Ferrero, G., E. Gomez, S. Lyer, M. Rovira, M. Miserocchi, D.M. Langenau, J.Y. Bertrand, and V. Wittamer. 2020. The macrophage-expressed gene (mpeg) 1 identifies a subpopulation of B cells in the adult zebrafish. *J. Leukoc. Biol.* 107:431–443. <https://doi.org/10.1002/jlba.11119-223R>
- Fu, G., S. Vallee, V. Rybakina, M.V. McGuire, J. Ampudia, C. Brockmeyer, M. Salek, P.R. Fallen, J.A.H. Hoerter, A. Munshi, et al. 2009. Thms controls thymocyte selection through regulation of T cell antigen receptor-mediated signaling. *Nat. Immunol.* 10:848–856. <https://doi.org/10.1038/ni.1766>
- Ganis, J.J., N. Hsia, E. Trompouki, J.L.O. de Jong, A. DiBiase, J.S. Lambert, Z. Jia, P.J. Sabo, M. Weaver, R. Sandstrom, et al. 2012. Zebrafish globin switching occurs in two developmental stages and is controlled by the LCR. *Dev. Biol.* 366:185–194. <https://doi.org/10.1016/j.ydbio.2012.03.021>
- Garcia, E.G., S. Iyer, S.P. Garcia, S. Loontjens, R.I. Sadreyev, F. Speleman, and D.M. Langenau. 2018. Cell of origin dictates aggression and stem cell number in acute lymphoblastic leukemia. *Leukemia.* 32:1860–1865. <https://doi.org/10.1038/s41375-018-0130-0>
- Gioacchino, E., C. Koyunlar, J. Zink, H. de Looper, M. de Jong, T. Dobrzycki, C.B. Mahony, R. Hoogenboezem, D. Bosch, P.M.H. van Strien, et al. 2021. Essential role for Gata2 in modulating lineage output from hematopoietic stem cells in zebrafish. *Blood Adv.* 5:2687–2700. <https://doi.org/10.1182/bloodadvances.2020002993>
- Gonzalez-Leal, I.J., B. Roger, A. Schwarz, T. Schirmeister, T. Reinheckel, M.B. Lutz, and H. Moll. 2014. Cathepsin B in antigen-presenting cells controls mediators of the Th1 immune response during Leishmania major infection. *PLoS Negl. Trop. Dis.* 8:e3194. <https://doi.org/10.1371/journal.pntd.0003194>
- Gordon, S.M., J. Chaix, L.J. Rupp, J. Wu, S. Madera, J.C. Sun, T. Lindsten, and S.L. Reiner. 2012. The transcription factors T-bet and Eomes control key checkpoints of natural killer cell maturation. *Immunity.* 36:55–67. <https://doi.org/10.1016/j.immuni.2011.11.016>
- Gore, A.V., B. Athans, J.R. Iben, K. Johnson, V. Russanova, D. Castranova, V.N. Pham, M.G. Butler, L. Williams-Simons, J.T. Nichols, et al. 2016. Epigenetic regulation of hematopoiesis by DNA methylation. *Elife.* 5:e11813. <https://doi.org/10.7554/eLife.11813>
- Grun, D. 2020. Revealing dynamics of gene expression variability in cell state space. *Nat. Methods.* 17:45–49. <https://doi.org/10.1038/s41592-019-0632-3>
- Grun, D., L. Kester, and A. van Oudenaarden. 2014. Validation of noise models for single-cell transcriptomics. *Nat. Methods.* 11:637–640. <https://doi.org/10.1038/nmeth.2930>
- Grun, D., M.J. Muraro, J.C. Boisset, K. Wiebrands, A. Lyubimova, G. Dharmadhikari, M. van den Born, J. van Es, E. Jansen, H. Clevers, et al. 2016. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell.* 19:266–277. <https://doi.org/10.1016/j.stem.2016.05.010>
- Hafemeister, C., and R. Satija. 2019. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial

- regression. *Genome Biol.* 20:296. <https://doi.org/10.1186/s13059-019-1874-1>
- Hagman, J., A. Travis, and R. Grosschedl. 1991. A novel lineage-specific nuclear factor regulates mb-1 gene transcription at the early stages of B cell differentiation. *EMBO J.* 10:3409-3417. <https://doi.org/10.1002/j.1460-2075.1991.tb04905.x>
- Haire, R.N., J.P. Rast, R.T. Litman, and G.W. Litman. 2000. Characterization of three isotypes of immunoglobulin light chains and T-cell antigen receptor  $\alpha$  in zebrafish. *Immunogenetics.* 51:915-923. <https://doi.org/10.1007/s002510000229>
- Hartmann, U., H. Hulsmann, J. Seul, S. Roll, H. Midani, I. Breloy, D. Hechler, R. Muller, and M. Paulsson. 2013. Testican-3: A brain-specific proteoglycan member of the BM-40/SPARC/osteonectin family. *J. Neurochem.* 125:399-409. <https://doi.org/10.1111/jnc.12212>
- Hason, M., T. Mikulasova, O. Machonova, A. Pombinho, T.J. van Ham, U. Irion, C. Nusslein-Volhard, P. Bartunek, and O. Svoboda. 2022. M-CSFR/CSF1R signaling regulates myeloid fates in zebrafish via distinct action of its receptors and ligands. *Blood Adv.* 6:1474-1488. <https://doi.org/10.1182/bloodadvances.2021005459>
- Haueneringer, V., M.D. Moccia, L. Opitz, S. Vavassori, H. Dave, and M.M. Hauri-Hohl. 2021. Novel. Combination of surface markers for the reliable and comprehensive identification of human thymic epithelial cells by flow cytometry: Quantitation and transcriptional characterization of thymic stroma in a pediatric cohort. *Front. Immunol.* 12:740047. <https://doi.org/10.3389/fimmu.2021.740047>
- He, B.L., X. Shi, C.H. Man, A.C.H. Ma, S.C. Ekker, H.C.H. Chow, C.W.E. So, W.W.L. Choi, W. Zhang, Y. Zhang, and A.Y.H. Leung. 2014. Functions of flt3 in zebrafish hematopoiesis and its relevance to human acute myeloid leukemia. *Blood.* 123:2518-2529. <https://doi.org/10.1182/blood-2013-02-486688>
- Herman, J.S., Sagar, and D. Grun. 2018. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat. Methods.* 15:379-386. <https://doi.org/10.1038/nmeth.4662>
- Hernandez, P.P., P.M. Strzelecka, E.I. Athanasiadis, D. Hall, A.F. Robalo, C.M. Collins, P. Boudinot, J.P. Levraud, and A. Cvejic. 2018. Single-cell transcriptional analysis reveals ILC-like cells in zebrafish. *Sci. Immunol.* 3:eau5265. <https://doi.org/10.1126/sciimmunol.aau5265>
- Hess, I., and T. Boehm. 2012. Intravital imaging of thymopoiesis reveals dynamic lympho-epithelial interactions. *Immunity.* 36:298-309. <https://doi.org/10.1016/j.immuni.2011.12.016>
- Hey, Y.Y., and H.C. O'Neill. 2016. Antigen presenting properties of a myeloid dendritic-like cell in murine spleen. *PLoS One.* 11:e0162358. <https://doi.org/10.1371/journal.pone.0162358>
- Hildner, K., B.T. Edelson, W.E. Purtha, M. Diamond, H. Matsushita, M. Kohyama, B. Calderon, B.U. Schraml, E.R. Unanue, M.S. Diamond, et al. 2008. Batf3 deficiency reveals a critical role for CD8 $\alpha$  dendritic cells in cytotoxic T cell immunity. *Science.* 322:1097-1100. <https://doi.org/10.1126/science.1164206>
- Hinton, H.J., D.R. Alessi, and D.A. Cantrell. 2004. The serine kinase phosphoinositide-dependent kinase 1 (PDK1) regulates T cell development. *Nat. Immunol.* 5:539-545. <https://doi.org/10.1038/ni1062>
- Ho, I.C., T.S. Tai, and S.Y. Pai. 2009. GATA3 and the T-cell lineage: Essential functions before and after T-helper-2-cell differentiation. *Nat. Rev. Immunol.* 9:125-135. <https://doi.org/10.1038/nri2476>
- Huang, Z., G. Qi, J.S. Miller, and S.G. Zheng. 2020. CD226: An emerging role in immunologic diseases. *Front. Cell Dev. Biol.* 8:564. <https://doi.org/10.3389/fcell.2020.00564>
- Ikeda, F., K. Yoshida, T. Toki, T. Uechi, S. Ishida, Y. Nakajima, Y. Sasahara, Y. Okuno, R. Kanezaki, K. Terui, et al. 2017. Exome sequencing identified RPS15A as a novel causative gene for Diamond-Blackfan anemia. *Haematologica.* 102:e93-e96. <https://doi.org/10.3324/haematol.2016.153932>
- Iwanami, N., F. Mateos, I. Hess, N. Riffel, C. Soza-Ried, M. Schorpp, and T. Boehm. 2011. Genetic evidence for an evolutionarily conserved role of IL-7 signaling in T cell development of zebrafish. *J. Immunol.* 186:7060-7066. <https://doi.org/10.4049/jimmunol.1003907>
- Jeffrey, K.L., T. Brummer, M.S. Rolph, S.M. Liu, N.A. Callejas, R.J. Grumont, C. Gillieron, F. Mackay, S. Grey, M. Camps, et al. 2006. Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nat. Immunol.* 7:274-283. <https://doi.org/10.1038/ni1310>
- Jojic, V., T. Shay, K. Sylvia, O. Zuk, X. Sun, J. Kang, A. Regev, D. Koller, Immunological Genome Project Consortium, A.J. Best, et al. 2013. Identification of transcriptional regulators in the mouse immune system. *Nat. Immunol.* 14:633-643. <https://doi.org/10.1038/ni.2587>
- Kadouri, N., S. Nevo, Y. Goldfarb, and J. Abramson. 2020. Thymic epithelial cell heterogeneity: TEC by TEC. *Nat. Rev. Immunol.* 20:239-253. <https://doi.org/10.1038/s41577-019-0238-0>
- Kalev-Zylinska, M.L., J.A. Horsfield, M.V.C. Flores, J.H. Postlethwait, J.Y.M. Chau, P.M. Cattin, M.R. Vitas, P.S. Crosier, and K.E. Crosier. 2003. Runx3 is required for hematopoietic development in zebrafish. *Dev. Dyn.* 228:323-336. <https://doi.org/10.1002/dvdy.10388>
- Karsunky, H., M. Merad, A. Cozzio, I.L. Weissman, and M.G. Manz. 2003. Flt3 ligand regulates dendritic cell development from Flt3<sup>+</sup> lymphoid and myeloid-committed progenitors to Flt3<sup>+</sup> dendritic cells in vivo. *J. Exp. Med.* 198:305-313. <https://doi.org/10.1084/jem.20030323>
- Kim, H.P., J. Imbert, and W.J. Leonard. 2006. Both integrated and differential regulation of components of the IL-2/IL-2 receptor system. *Cytokine Growth Factor Rev.* 17:349-366. <https://doi.org/10.1016/j.cytogfr.2006.07.003>
- Kim, J.H., J. Hwang, J.H. Jung, H.J. Lee, D.Y. Lee, and S.H. Kim. 2019. Molecular networks of FOXP family: Dual biologic functions, interplay with other molecules and clinical implications in cancer progression. *Mol. Cancer.* 18:180. <https://doi.org/10.1186/s12943-019-1110-3>
- Kissa, K., E. Murayama, A. Zapata, A. Cortes, E. Perret, C. Machu, and P. Herbomel. 2008. Live imaging of emerging hematopoietic stem cells and early thymus colonization. *Blood.* 111:1147-1156. <https://doi.org/10.1182/blood-2007-07-099499>
- Klebanoff, S.J., A.J. Kettle, H. Rosen, C.C. Winterbourn, and W.M. Nauseef. 2013. Myeloperoxidase: A front-line defender against phagocytosed microorganisms. *J. Leukoc. Biol.* 93:185-198. <https://doi.org/10.1189/jlb.0712349>
- Koehn, J., F. Bachelier, K. Balabanian, G. Schlecht-Louf, and C. Gallego. 2019. Atypical chemokine receptor 3 (ACKR3): A comprehensive overview of its expression and potential roles in the immune system. *Mol. Pharmacol.* 96:809-818. <https://doi.org/10.1124/mol.118.115329>
- Komori, T., A. Okada, V. Stewart, and F.W. Alt. 1993. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science.* 261:1171-1175. <https://doi.org/10.1126/science.8356451>
- Kondrashov, N., A. Pusic, C.R. Stumpf, K. Shimizu, A.C. Hsieh, J. Ishijima, T. Shiroishi, and M. Barna. 2011. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell.* 145:383-397. <https://doi.org/10.1016/j.cell.2011.03.028>
- Koprunner, M., C. Thisse, B. Thisse, and E. Raz. 2001. A zebrafish nanos-related gene is essential for the development of primordial germ cells. *Genes Dev.* 15:2877-2885. <https://doi.org/10.1101/gad.212401>
- Kowalczyk, M.S., I. Tirosh, D. Heckl, T.N. Rao, A. Dixit, B.J. Haas, R.K. Schneider, A.J. Wagers, B.L. Ebert, and A. Regev. 2015. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* 25:1860-1872. <https://doi.org/10.1101/gr.192237.115>
- Kruse, E.A., S.J. Loughran, T.M. Baldwin, E.C. Josefsson, S. Ellis, D.K. Watson, P. Nurden, D. Metcalf, D.J. Hilton, W.S. Alexander, and B.T. Kile. 2009. Dual requirement for the ETS transcription factors Flt-1 and Erg in hematopoietic stem cells and the megakaryocyte lineage. *Proc. Natl. Acad. Sci. USA.* 106:13814-13819. <https://doi.org/10.1073/pnas.0906556106>
- Kuil, L.E., N. Oosterhof, G. Ferrero, T. Mikulasova, M. Hason, J. Dekker, M. Rovira, H.C. van der Linde, P.M. van Strien, E. de Pater, et al. 2020. Zebrafish macrophage developmental arrest underlies depletion of microglia and reveals Csf1r-independent metaphocytes. *Elife.* 9:e53403. <https://doi.org/10.7554/eLife.53403>
- Lam, S.H., H.L. Chua, Z. Gong, Z. Wen, T.J. Lam, and Y.M. Sin. 2002. Morphologic transformation of the thymus in developing zebrafish. *Dev. Dyn.* 225:87-94. <https://doi.org/10.1002/dvdy.10127>
- Lancaster, J.N., Y. Li, and L.I.R. Ehrlich. 2018. Chemokine-Mediated choreography of thymocyte development and selection. *Trends Immunol.* 39:86-98. <https://doi.org/10.1016/j.it.2017.10.007>
- Langenau, D.M., A.A. Ferrando, D. Traver, J.L. Kutok, J.P.D. Hezel, J.P. Kanki, L.I. Zon, A.T. Look, and N.S. Trede. 2004. In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc. Natl. Acad. Sci. USA.* 101:7369-7374. <https://doi.org/10.1073/pnas.0402248101>
- Lee, M.G., S.O. Sharrow, A.G. Farr, A. Singer, and M.C. Udey. 1994. Expression of the homotypic adhesion molecule E-cadherin by immature murine thymocytes and thymic epithelial cells. *J. Immunol.* 152:5653-5659
- Liao, M., and J. Wang. 2021. Tcf12 balances the reconstitution and differentiation capacity of hematopoietic stem cell. *Blood Sci.* 3:14-19. <https://doi.org/10.1097/BS9.0000000000000059>



- Lieberman, J. 2010. Granzyme A activates another way to die. *Immunol. Rev.* 235:93–104. <https://doi.org/10.1111/j.0105-2896.2010.00902.x>
- Lieschke, G.J., A.C. Oates, B.H. Paw, M.A. Thompson, N.E. Hall, A.C. Ward, R.K. Ho, L.I. Zon, and J.E. Layton. 2002. Zebrafish SPI-1 (PU.1) marks a site of myeloid development independent of primitive erythropoiesis: Implications for axial patterning. *Dev. Biol.* 246:274–295. <https://doi.org/10.1006/dbio.2002.0657>
- Liu, B., L. Qu, and S. Yan. 2015. Cyclooxygenase-2 promotes tumor growth and suppresses tumor immunity. *Cancer Cell Int.* 15:106. <https://doi.org/10.1186/s12935-015-0260-7>
- Liu, X., Y.S. Li, S.A. Shinton, J. Rhodes, L. Tang, H. Feng, C.A. Jette, A.T. Look, K. Hayakawa, and R.R. Hardy. 2017. Zebrafish B cell development without a pre-B cell stage, revealed by CD79 fluorescence reporter transgenes. *J. Immunol.* 199:1706–1715. <https://doi.org/10.4049/jimmunol.1700552>
- Lloyd, C.M., and R.J. Snelgrove. 2018. Type 2 immunity: Expanding our view. *Sci. Immunol.* 3:eaat1604. <https://doi.org/10.1126/sciimmunol.aat1604>
- Lugo-Villarino, G., K.M. Balla, D.L. Stachura, K. Banuelos, M.B.F. Werneck, and D. Traver. 2010. Identification of dendritic antigen-presenting cells in the zebrafish. *Proc. Natl. Acad. Sci. USA.* 107:15850–15855. <https://doi.org/10.1073/pnas.1000494107>
- Lyons, S.E., B.C. Shue, L. Lei, A.C. Oates, L.I. Zon, and P.P. Liu. 2001. Molecular cloning, genetic mapping, and expression analysis of four zebrafish *c/ebp* genes. *Gene.* 281:43–51. [https://doi.org/10.1016/s0378-1119\(01\)00774-0](https://doi.org/10.1016/s0378-1119(01)00774-0)
- Ma, D., L. Wang, S. Wang, Y. Gao, Y. Wei, and F. Liu. 2012. Foxn1 maintains thymic epithelial cells to support T-cell development via *mcm2* in zebrafish. *Proc. Natl. Acad. Sci. USA.* 109:21040–21045. <https://doi.org/10.1073/pnas.1217021110>
- Macaulay, I.C., V. Svensson, C. Labalette, L. Ferreira, F. Hamey, T. Voet, S.A. Teichmann, and A. Cvejic. 2016. Single-cell RNA-sequencing reveals a continuous spectrum of differentiation in hematopoietic cells. *Cell Rep.* 14:966–977. <https://doi.org/10.1016/j.celrep.2015.12.082>
- Mahony, C.B., B. Noyvert, P. Vrljicak, S. Ott, M. Higgs, and R. Monteiro. 2021. Gata2a maintains *cebpa* and *npmla* in haematopoietic stem cells to sustain lineage differentiation and genome stability. (Preprint posted July 19, 2021). [bioRxiv. https://doi.org/10.1101/2021.07.19.452890](https://doi.org/10.1101/2021.07.19.452890)
- Marazuela, M., A. Acevedo, M.A. Garcia-Lopez, M. Adrados, M.C. de Marco, and M.A. Alonso. 2004. Expression of MAL2, an integral protein component of the machinery for basolateral-to-apical transcytosis, in human epithelia. *Journal Histochem. Cytochem.* 52:243–252. <https://doi.org/10.1177/002215540405200212>
- Martinez-Pomares, L. 2012. The mannose receptor. *J. Leukoc. Biol.* 92:1177–1186. <https://doi.org/10.1189/jlb.0512231>
- Matsumoto, F., S.I. Saitoh, R. Fukui, T. Kobayashi, N. Tanimura, K. Konno, Y. Kusumoto, S. Akashi-Takamura, and K. Miyake. 2008. Cathepsins are required for Toll-like receptor 9 responses. *Biochem. Biophys. Res. Commun.* 367:693–699. <https://doi.org/10.1016/j.bbrc.2007.12.130>
- McKenna, A., G.M. Findlay, J.A. Gagnon, M.S. Horwitz, A.F. Schier, and J. Shendure. 2016. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science.* 353:aaf7907. <https://doi.org/10.1126/science.aaf7907>
- Meeker, N.D., A.C.H. Smith, J.K. Frazer, D.F. Bradley, L.A. Rudner, C. Love, and N.S. Trede. 2010. Characterization of the zebrafish T cell receptor  $\beta$  locus. *Immunogenetics.* 62:23–29. <https://doi.org/10.1007/s00251-009-0407-6>
- Melichar, H.J., K. Narayan, S.D. Der, Y. Hiraoka, N. Gardiol, G. Jeannot, W. Held, C.A. Chambers, and J. Kang. 2007. Regulation of  $\gamma\delta$  versus  $\alpha\beta$  T lymphocyte differentiation by the transcription factor SOX13. *Science.* 315:230–233. <https://doi.org/10.1126/science.1135344>
- Mirabello, L., E.R. Macari, L. Jessop, S.R. Ellis, T. Myers, N. Giri, A.M. Taylor, K.E. McGrath, J.M. Humphries, B.J. Ballew, et al. 2014. Whole-exome sequencing and functional studies identify RPS29 as a novel gene mutated in multicase Diamond-Blackfan anemia families. *Blood.* 124:24–32. <https://doi.org/10.1182/blood-2013-11-540278>
- Mitchell, D.M., C. Sun, S.S. Hunter, D.D. New, and D.L. Stenkamp. 2019. Regeneration associated transcriptional signature of retinal microglia and macrophages. *Sci. Rep.* 9:4768. <https://doi.org/10.1038/s41598-019-41298-8>
- Mitra, S., A. Alnabulsi, C.J. Secombes, and S. Bird. 2010. Identification and characterization of the transcription factors involved in T-cell development, *t-bet*, *stat6* and *foxp3*, within the zebrafish, *Danio rerio*. *FEBS J.* 277:128–147. <https://doi.org/10.1111/j.1742-4658.2009.07460.x>
- Modrell, M.S., M. Lyne, A.R. Carr, H.H. Zakon, D. Buckley, A.S. Campbell, M.C. Davis, G. Micklem, and C.V. Baker. 2017. Insights into electrosensory organ development, physiology and evolution from a lateral line-enriched transcriptome. *Elife.* 6:e24197. <https://doi.org/10.7554/eLife.24197>
- Monney, L., C.A. Sabatos, J.L. Gaglia, A. Ryu, H. Waldner, T. Chernova, S. Manning, E.A. Greenfield, A.J. Coyle, R.A. Sobel, et al. 2002. Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature.* 415:536–541. <https://doi.org/10.1038/415536a>
- Moore, F.E., E.G. Garcia, R. Lobbardi, E. Jain, Q. Tang, J.C. Moore, M. Cortes, A. Molodtsov, M. Kasheta, C.C. Luo, et al. 2016. Single-cell transcriptional analysis of normal, aberrant, and malignant hematopoiesis in zebrafish. *J. Exp. Med.* 213:979–992. <https://doi.org/10.1084/jem.20152013>
- Morath, A., and W.W. Schamel. 2020.  $\alpha\beta$  and  $\gamma\delta$  T cell receptors: Similar but different. *J. Leukoc. Biol.* 107:1045–1055. <https://doi.org/10.1002/JLB.2MR1219-233R>
- Murayama, E., K. Kissa, A. Zapata, E. Mordelet, V. Briolat, H.F. Lin, R.I. Handin, and P. Herbomel. 2006. Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development. *Immunity.* 25:963–975. <https://doi.org/10.1016/j.immuni.2006.10.015>
- Murphy, P.R., and R.S. Knee. 1994. Identification and characterization of an antisense RNA transcript (gfg) from the human basic fibroblast growth factor gene. *Mol. Endocrinol.* 8:852–859. <https://doi.org/10.1210/mend.8.7.7984147>
- Naganuma, T., Y. Sato, T. Sassa, Y. Ohno, and A. Kihara. 2011. Biochemical characterization of the very long-chain fatty acid elongase ELOVL7. *FEBS Lett.* 585:3337–3341. <https://doi.org/10.1016/j.febslet.2011.09.024>
- Nakagawa, Y., I. Ohigashi, T. Nitta, M. Sakata, K. Tanaka, S. Murata, O. Kanagawa, and Y. Takahama. 2012. Thymic nurse cells provide microenvironment for secondary T cell receptor  $\alpha$  rearrangement in cortical thymocytes. *Proc. Natl. Acad. Sci. USA.* 109:20572–20577. <https://doi.org/10.1073/pnas.1213069109>
- Nemeth, M.J., D.J. Curtis, M.R. Kirby, L.J. Garrett-Beal, N.E. Seidel, A.P. Cline, and D.M. Bodine. 2003. Hmgb3: an HMGB-box family member expressed in primitive hematopoietic cells that inhibits myeloid and B-cell differentiation. *Blood.* 102:1298–1306. <https://doi.org/10.1182/blood-2002-11-3541>
- Nitta, T., S. Nitta, Y. Lei, M. Lipp, and Y. Takahama. 2009. CCR7-mediated migration of developing thymocytes to the medulla is essential for negative selection to tissue-restricted antigens. *Proc. Natl. Acad. Sci. USA.* 106:17129–17133. <https://doi.org/10.1073/pnas.0906956106>
- Nutt, S.L., and B.L. Kee. 2007. The transcriptional regulation of B cell lineage commitment. *Immunity.* 26:715–725. <https://doi.org/10.1016/j.immuni.2007.05.010>
- Oettinger, M.A., D.G. Schatz, C. Gorka, and D. Baltimore. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science.* 248:1517–1523. <https://doi.org/10.1126/science.2360047>
- Ohtani, M., N. Hayashi, K. Hashimoto, T. Nakanishi, and J.M. Dijkstra. 2008. Comprehensive clarification of two paralogous interleukin 4/13 loci in teleost fish. *Immunogenetics.* 60:383–397. <https://doi.org/10.1007/s00251-008-0299-x>
- Ong, S.L.M., I.J.H.M. de Vos, M. Meroshini, Y. Poobalan, and N.R. Dunn. 2020. Microfibril-associated glycoprotein 4 (Mfap4) regulates haematopoiesis in zebrafish. *Sci. Rep.* 10:11801. <https://doi.org/10.1038/s41598-020-68792-8>
- Page, D.M., V. Wittamer, J.Y. Bertrand, K.L. Lewis, D.N. Pratt, N. Delgado, S.E. Schale, C. McGue, B.H. Jacobsen, A. Doty, et al. 2013. An evolutionarily conserved program of B-cell development and activation in zebrafish. *Blood.* 122:e1–e11. <https://doi.org/10.1182/blood-2012-12-471029>
- Pandey, A., N. Ibarrola, I. Kratchmarova, M.M. Fernandez, S.N. Constantinescu, O. Ohara, S. Sawasdikosol, H.F. Lodish, and M. Mann. 2002. A novel Src homology 2 domain-containing molecule, Src-like adapter protein-2 (SLAP-2), which negatively regulates T cell receptor signaling. *J. Biol. Chem.* 277:19131–19138. <https://doi.org/10.1074/jbc.M110318200>
- Parker, J., N.C. Guslund, S. Jentoft, and O. Roth. 2022. Characterization of pipefish immune cell populations through single-cell transcriptomics. *Front. Immunol.* 13:820152. <https://doi.org/10.3389/fimmu.2022.820152>
- Parrish-Novak, J., S.R. Dillon, A. Nelson, A. Hammond, C. Sprecher, J.A. Gross, J. Johnston, K. Madden, W. Xu, J. West, et al. 2000. Interleukin 21 and its receptor are involved in NK cell expansion and regulation of lymphocyte function. *Nature.* 408:57–63. <https://doi.org/10.1038/35040504>

- Patel, Y.C. 1999. Somatostatin and its receptor family. *Front. Neuroendocrinol.* 20:157–198. <https://doi.org/10.1006/frne.1999.0183>
- Pearce, E.L., A.C. Mullen, G.A. Martins, C.M. Krawczyk, A.S. Hutchins, V.P. Zediak, M. Banica, C.B. DiCioccio, D.A. Gross, C.A. Mao, et al. 2003. Control of effector CD8<sup>+</sup> T cell function by the transcription factor Eomesodermin. *Science*. 302:1041–1043. <https://doi.org/10.1126/science.1090148>
- Perucho, L., A. Artero-Castro, S. Guerrero, S. Ramon y Cajal, M.E. LLeonart, and Z.Q. Wang. 2014. RPLP1, a crucial ribosomal protein for embryonic development of the nervous system. *PLoS One*. 9:e99956. <https://doi.org/10.1371/journal.pone.0099956>
- Pham, T.V., S.R. Piersma, M. Warmoes, and C.R. Jimenez. 2010. On the  $\beta$ -binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics*. 26:363–369. <https://doi.org/10.1093/bioinformatics/btp677>
- Porritt, H.E., L.L. Rumpf, S. Tabrizifard, T.M. Schmitt, J.C. Zuniga-Pflucker, and H.T. Petrie. 2004. Heterogeneity among DNI prothymocytes reveals multiple progenitors with different capacities to generate T cell and non-T cell lineages. *Immunity*. 20:735–745. <https://doi.org/10.1016/j.immuni.2004.05.004>
- Qiu, X., Q. Mao, Y. Tang, L. Wang, R. Chawla, H.A. Pliner, and C. Trapnell. 2017. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*. 14:979–982. <https://doi.org/10.1038/nmeth.4402>
- Raj, B., J.A. Gagnon, and A.F. Schier. 2018a. Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR-Cas9 barcodes by scGESTALT. *Nat. Protoc.* 13:2685–2713. <https://doi.org/10.1038/s41596-018-0058-x>
- Raj, B., D.E. Wagner, A. McKenna, S. Pandey, A.M. Klein, J. Shendure, J.A. Gagnon, and A.F. Schier. 2018b. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* 36:442–450. <https://doi.org/10.1038/nbt.4103>
- Rajan, A.M., R.C. Ma, K.M. Kocha, D.J. Zhang, and P. Huang. 2020. Dual function of perivascular fibroblasts in vascular stabilization in zebrafish. *PLoS Gen.* 16:e1008800. <https://doi.org/10.1371/journal.pgen.1008800>
- Remnant, L., N.Y. Kochanova, C. Reid, F. Cisneros-Soberanis, and W.C. Earnshaw. 2021. The intrinsically disorderly story of Ki-67. *Open Biol.* 11:210120. <https://doi.org/10.1098/rsob.210120>
- Ridge, K.M., J.E. Eriksson, M. Pekny, and R.D. Goldman. 2022. Roles of vimentin in health and disease. *Genes Dev.* 36:391–407. <https://doi.org/10.1101/gad.349358.122>
- Rothenberg, E.V., J.E. Moore, and M.A. Yui. 2008. Launching the T-cell-lineage developmental programme. *Nat. Rev. Immunol.* 8:9–21. <https://doi.org/10.1038/nri2232>
- Rothenfusser, S., E. Tuma, S. Endres, and G. Hartmann. 2002. Plasmacytoid dendritic cells: The key to CpG. *Hum. Immunol.* 63:1111–1119. [https://doi.org/10.1016/s0198-8859\(02\)00749-8](https://doi.org/10.1016/s0198-8859(02)00749-8)
- Rougeot, J., V. Torraca, A. Zakrzewska, Z. Kanwal, H.J. Jansen, F. Sommer, H.P. Spalink, and A.H. Meijer. 2019. RNAseq profiling of leukocyte populations in zebrafish larvae reveals a cxcl11 chemokine gene as a marker of macrophage polarization during mycobacterial infection. *Front. Immunol.* 10:2720. <https://doi.org/10.3389/fimmu.2019.02720>
- Ryu, S., J. Holzschuh, S. Erhardt, A.K. Ettl, and W. Driever. 2005. Depletion of minichromosome maintenance protein 5 in the zebrafish retina causes cell-cycle defect and apoptosis. *Proc. Natl. Acad. Sci. USA*. 102:18467–18472. <https://doi.org/10.1073/pnas.0506187102>
- Samborska, B., D.G. Roy, J.F. Rahbani, M.F. Hussain, E.H. Ma, R.G. Jones, and L. Kazak. 2022. Creatine transport and creatine kinase activity is required for CD8(+) T cell immunity. *Cell Rep.* 38:110446. <https://doi.org/10.1016/j.celrep.2022.110446>
- Schatz, D.G., M.A. Oettinger, and D. Baltimore. 1989. The V(D)J recombination activating gene, RAG-1. *Cell*. 59:1035–1048. [https://doi.org/10.1016/0092-8674\(89\)90760-5](https://doi.org/10.1016/0092-8674(89)90760-5)
- Schorpp, M., M. Bialecki, D. Diekhoff, B. Walderich, J. Odenthal, H.M. Maischein, A.G. Zapata, and T. Boehm. 2006. Conserved functions of ikaros in vertebrate lymphocyte development: Genetic evidence for distinct larval and adult phases of T cell development and two lineages of B cells in zebrafish. *J. Immunol.* 177:2463–2476. <https://doi.org/10.4049/jimmunol.177.4.2463>
- Seeger, A., W.E. Mayer, and J. Klein. 1996. A complement factor B-like cDNA clone from the zebrafish (*Brachydanio rerio*). *Mol. Immunol.* 33:511–520. [https://doi.org/10.1016/0161-5890\(96\)00002-8](https://doi.org/10.1016/0161-5890(96)00002-8)
- Seelye, S.L., P.L. Chen, T.C. Deiss, and M.F. Criscitiello. 2016. Genomic organization of the zebrafish (*Danio rerio*) T cell receptor  $\alpha/\delta$  locus and analysis of expressed products. *Immunogenetics*. 68:365–379. <https://doi.org/10.1007/s00251-016-0904-3>
- Shiau, C.E., Z. Kaufman, A.M. Meireles, and W.S. Talbot. 2015. Differential requirement for irf8 in formation of embryonic and adult macrophages in zebrafish. *PLoS One*. 10:e0117513. <https://doi.org/10.1371/journal.pone.0117513>
- Sichien, D., C.L. Scott, L. Martens, M. Vanderkerken, S. Van Gassen, M. Plantinga, T. Joeris, S. De Prijck, L. Vanhoutte, M. Vanheerswyngheles, et al. 2016. IRF8 transcription factor controls survival and function of terminally differentiated conventional and plasmacytoid dendritic cells, respectively. *Immunity*. 45:626–640. <https://doi.org/10.1016/j.immuni.2016.08.013>
- Sidney, L.E., M.J. Branch, S.E. Dunphy, H.S. Dua, and A. Hopkinson. 2014. Concise review: Evidence for CD34 as a common marker for diverse progenitors. *Stem Cells*. 32:1380–1389. <https://doi.org/10.1002/stem.1661>
- Silacci, P., L. Mazzolai, C. Gauci, N. Stergiopoulos, H.L. Yin, and D. Hayoz. 2004. Gelsolin superfamily proteins: Key regulators of cellular functions. *Cell Mol. Life Sci.* 61:2614–2623. <https://doi.org/10.1007/s00018-004-4225-6>
- Smith, L.T., S. Hohaus, D.A. Gonzalez, S.E. Dziennis, and D.G. Tenen. 1996. PU.1 (Spi-1) and C/EBP  $\alpha$  regulate the granulocyte colony-stimulating factor receptor promoter in myeloid cells. *Blood*. 88:1234–1247. <https://doi.org/10.1182/blood.v88.4.1234.bloodjournal8841234>
- Speir, M.L., A. Bhaduri, N.S. Markov, P. Moreno, T.J. Nowakowski, I. Papa-theodorou, A.A. Pollen, B.J. Raney, L. Senjega, W.J. Kent, and M. Haeussler. 2021. UCSC cell browser: Visualize your single-cell data. *Bioinformatics*. 37:4578–4580. <https://doi.org/10.1093/bioinformatics/btab503>
- Stanley, E.R., and V. Chitu. 2014. CSF-1 receptor signaling in myeloid cells. *Cold Spring Harb. Perspect. Biol.* 6:a021857. <https://doi.org/10.1101/cshperspect.a021857>
- Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W.M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. 2019. Comprehensive integration of single-cell data. *Cell*. 177:1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>
- Sutoh, Y., M. Kondo, Y. Ohta, T. Ota, U. Tomaru, M.F. Flajnik, and M. Kasahara. 2012. Comparative genomic analysis of the proteasome  $\beta$ 5t subunit gene: Implications for the origin and evolution of thymoproteasomes. *Immunogenetics*. 64:49–58. <https://doi.org/10.1007/s00251-011-0558-0>
- Svensson, V., R. Vento-Tormo, and S.A. Teichmann. 2018. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13:599–604. <https://doi.org/10.1038/nprot.2017.149>
- Takizawa, F., K. Araki, K. Ito, T. Moritomo, and T. Nakanishi. 2007. Expression analysis of two Eomesodermin homologues in zebrafish lymphoid tissues and cells. *Mol. Immunol.* 44:2324–2331. <https://doi.org/10.1016/j.molimm.2006.11.018>
- Tang, Q., S. Iyer, R. Lobbardi, J.C. Moore, H. Chen, C. Lareau, C. Hebert, M.L. Shaw, C. Neftel, M.L. Suva, et al. 2017. Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J. Exp. Med.* 214:2875–2887. <https://doi.org/10.1084/jem.20170976>
- Trapnell, C., D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N.J. Lennon, K.J. Livak, T.S. Mikkelsen, and J.L. Rinn. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32:381–386. <https://doi.org/10.1038/nbt.2859>
- Trede, N.S., D.M. Langenau, D. Traver, A.T. Look, and L.I. Zon. 2004. The use of zebrafish to understand immunity. *Immunity*. 20:367–379. [https://doi.org/10.1016/s1074-7613\(04\)00084-6](https://doi.org/10.1016/s1074-7613(04)00084-6)
- Trzpis, M., P.M.J. McLaughlin, L.M.F.H. de Leij, and M.C. Harmsen. 2007. Epithelial cell adhesion molecule: More than a carcinoma marker and adhesion molecule. *Am. J. Pathol.* 171:386–395. <https://doi.org/10.2353/ajpath.2007.070152>
- Tsai, F.Y., G. Keller, F.C. Kuo, M. Weiss, J. Chen, M. Rosenblatt, F.W. Alt, and S.H. Orkin. 1994. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*. 371:221–226. <https://doi.org/10.1038/371221a0>
- Uechi, T., Y. Nakajima, A. Chakraborty, H. Torihara, S. Higa, and N. Kenmochi. 2008. Deficiency of ribosomal protein S19 during early embryogenesis leads to reduction of erythrocytes in a zebrafish model of Diamond-Blackfan anemia. *Hum. Mol. Gen.* 17:3204–3211. <https://doi.org/10.1093/hmg/ddn216>
- Ueno, T., F. Saito, D.H.D. Gray, S. Kuse, K. Hieshima, H. Nakano, T. Kakiuchi, M. Lipp, R.L. Boyd, and Y. Takahama. 2004. CCR7 signals are essential

- for cortex-medulla migration of developing thymocytes. *J. Exp. Med.* 200:493–505. <https://doi.org/10.1084/jem.20040643>
- Vanlandewijck, M., L. He, M.A. Mae, J. Andrae, K. Ando, F. Del Gaudio, K. Nahar, T. Lebouvier, B. Lavina, L. Gouveia, et al. 2018. A molecular atlas of cell types and zonation in the brain vasculature. *Nature*. 554:475–480. <https://doi.org/10.1038/nature25739>
- Venkateswarlu, K. 2003. Interaction protein for cytohesin exchange factors 1 (IPCEF1) binds cytohesin 2 and modifies its activity. *J. Biol. Chem.* 278: 43460–43469. <https://doi.org/10.1074/jbc.M304078200>
- Verlaet, M., A. Adamantidis, B. Coumans, G. Chanas, W. Zorzi, E. Heinen, T. Grisar, and B. Lakaye. 2002. Human immune cells express ppMCH mRNA and functional MCHRI receptor. *FEBS Lett.* 527:205–210. [https://doi.org/10.1016/s0014-5793\(02\)03232-5](https://doi.org/10.1016/s0014-5793(02)03232-5)
- Walton, E.M., M.R. Cronan, R.W. Beerman, and D.M. Tobin. 2015. The macrophage-specific promoter mfp4 allows live, long-term analysis of macrophage behavior during mycobacterial infection in zebrafish. *PLoS One*. 10:e0138949. <https://doi.org/10.1371/journal.pone.0138949>
- Walzer, T., L. Chiossone, J. Chaix, A. Calver, C. Carozzo, L. Garrigue-Antar, Y. Jacques, M. Baratin, E. Tomasello, and E. Vivier. 2007. Natural killer cell trafficking in vivo requires a dedicated sphingosine 1-phosphate receptor. *Nat. Immunol.* 8:1337–1344. <https://doi.org/10.1038/ni1523>
- Wan, F., C.B. Hu, J.X. Ma, K. Gao, L.X. Xiang, and J.Z. Shao. 2016. Characterization of  $\gamma\delta$  T cells from zebrafish provides insights into their important role in adaptive humoral immunity. *Front. Immunol.* 7:675. <https://doi.org/10.3389/fimmu.2016.00675>
- Wang, H., C.H. Lee, C. Qi, P. Taylor, J. Feng, S. Abbasi, T. Atsumi, and H.C. Morse 3rd. 2008. IRF8 regulates B-cell lineage specification, commitment, and differentiation. *Blood*. 112:4028–4038. <https://doi.org/10.1182/blood-2008-01-129049>
- Wang, Z., and M. Peng. 2021. A novel prognostic biomarker LCP2 correlates with metastatic melanoma-infiltrating CD8(+) T cells. *Sci. Rep.* 11:9164. <https://doi.org/10.1038/s41598-021-88676-9>
- Wei, G., B.J. Abraham, R. Yagi, R. Jothi, K. Cui, S. Sharma, L. Narlikar, D.L. Northrup, Q. Tang, W.E. Paul, et al. 2011. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity*. 35:299–311. <https://doi.org/10.1016/j.immuni.2011.08.007>
- Wei, S., S. Segura, J. Vendrell, F.X. Aviles, E. Lanoue, R. Day, Y. Feng, and L.D. Fricker. 2002. Identification and characterization of three members of the human metalloproteinase gene family. *J. Biol. Chem.* 277: 14954–14964. <https://doi.org/10.1074/jbc.M112254200>
- White, D.T., K.M. McShea, M.A. Attar, and L.C. Santy. 2010. GRASP and IPCEF promote ARF-to-Rac signaling and cell migration by coordinating the association of ARNO/cytohesin 2 with Dock180. *Mol. Biol. Cell*. 21: 562–571. <https://doi.org/10.1091/mbc.e09-03-0217>
- Willett, C.E., J.J. Cherry, and L.A. Steiner. 1997a. Characterization and expression of the recombination activating genes (rag1 and rag2) of zebrafish. *Immunogenetics*. 45:394–404. <https://doi.org/10.1007/s002510050221>
- Willett, C.E., A.G. Zapata, N. Hopkins, and L.A. Steiner. 1997b. Expression of zebrafish rag genes during early development identifies the thymus. *Dev. Biol.* 182:331–341. <https://doi.org/10.1006/dbio.1996.8446>
- Wolf, F.A., F.K. Hamey, M. Plass, J. Solana, J.S. Dahlin, B. Gottgens, N. Rajewsky, L. Simon, and F.J. Theis. 2019. PAGA: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* 20:59. <https://doi.org/10.1186/s13059-019-1663-x>
- Yeo, S.Y., K.W. Lee, D. Shin, S. An, K.H. Cho, and S.H. Kim. 2018. A positive feedback loop bi-stably activates fibroblasts. *Nat. Commun.* 9:3016. <https://doi.org/10.1038/s41467-018-05274-6>
- Ying, J., G. Srivastava, W.S. Hsieh, Z. Gao, P. Murray, S.K. Liao, R. Ambinder, and Q. Tao. 2005. The stress-responsive gene GADD45G is a functional tumor suppressor, with its response to environmental stresses frequently disrupted epigenetically in multiple tumors. *Clin. Cancer Res.* 11:6442–6449. <https://doi.org/10.1158/1078-0432.CCR-05-0267>
- Yoder, J.A., M.E. Nielsen, C.T. Amemiya, and G.W. Litman. 2002. Zebrafish as an immunological model system. *Microb. Infect.* 4:1469–1478. [https://doi.org/10.1016/s1286-4579\(02\)00029-1](https://doi.org/10.1016/s1286-4579(02)00029-1)
- Yoder, J.A., T.M. Orcutt, D. Traver, and G.W. Litman. 2007. Structural characteristics of zebrafish orthologs of adaptor molecules that associate with transmembrane immune receptors. *Gene*. 401:154–164. <https://doi.org/10.1016/j.gene.2007.07.014>
- Yokota, T., C.S.R. Meka, T. Kouro, K.L. Medina, H. Igarashi, M. Takahashi, K. Oritani, T. Funahashi, Y. Tomiyama, Y. Matsuzawa, and P.W. Kincade. 2003. Adiponectin, a fat cell product, influences the earliest lymphocyte precursors in bone marrow cultures by activation of the cyclooxygenase-prostaglandin pathway in stromal cells. *J. Immunol.* 171:5091–5099. <https://doi.org/10.4049/jimmunol.171.10.5091>
- Yoon, S., S. Mitra, C. Wyse, A. Alnabulsi, J. Zou, E.M. Weerdenburg, A.M. van der Sar, D. Wang, C.J. Secombes, and S. Bird. 2015. First demonstration of antigen induced cytokine expression by CD4-1+ lymphocytes in a poikilotherm: Studies in zebrafish (*Danio rerio*). *PLoS One*. 10:e0126378. <https://doi.org/10.1371/journal.pone.0126378>
- Young, M.D., and S. Behjati. 2020. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience*. 9:gaa151. <https://doi.org/10.1093/gigascience/giaa151>
- Zakrzewska, A., C. Cui, O.W. Stockhammer, E.L. Benard, H.P. Spaink, and A.H. Meijer. 2010. Macrophage-specific gene functions in Spil-directed innate immunity. *Blood*. 116:e1–e11. <https://doi.org/10.1182/blood-2010-01-262873>
- Zaslavsky, A., K.H. Baek, R.C. Lynch, S. Short, J. Grillo, J. Folkman, J.E. Italiano Jr., and S. Ryeom. 2010. Platelet-derived thrombospondin-1 is a critical negative regulator and potential biomarker of angiogenesis. *Blood*. 115:4605–4613. <https://doi.org/10.1182/blood-2009-09-242065>
- Zeng, Y., C. Liu, Y. Gong, Z. Bai, S. Hou, J. He, Z. Bian, Z. Li, Y. Ni, J. Yan, et al. 2019. Single-cell RNA sequencing resolves spatiotemporal development of pre-thymic lymphoid progenitors and thymus organogenesis in human embryos. *Immunity*. 51:930–948.e6. <https://doi.org/10.1016/j.immuni.2019.09.008>
- Zhang, H., M. Alberich-Jorda, G. Amabile, H. Yang, P.B. Staber, A. Di Ruscio, A. Diruscio, R.S. Welner, A. Ebralidze, J. Zhang, et al. 2013. Sox4 is a key oncogenic target in C/EBP $\alpha$  mutant acute myeloid leukemia. *Cancer Cell*. 24:575–588. <https://doi.org/10.1016/j.ccr.2013.09.018>
- Zheng, G.X.Y., J.M. Terry, P. Belgrader, P. Ryvkin, Z.W. Bent, R. Wilson, S.B. Ziraldo, T.D. Wheeler, G.P. McDermott, J. Zhu, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8: 14049. <https://doi.org/10.1038/ncomms14049>
- Zhou, X., B.L. Updegraff, Y. Guo, M. Peyton, L. Girard, J.E. Larsen, X.J. Xie, Y. Zhou, T.H. Hwang, Y. Xie, et al. 2017. PROTOCADHERIN 7 acts through SET and PP2A to potentiate MAPK signaling by EGFR and KRAS during lung tumorigenesis. *Cancer Res.* 77:187–197. <https://doi.org/10.1158/0008-5472.CAN-16-1267-T>
- Zhou, Y., B. Zhou, L. Pache, M. Chang, A.H. Khodabakhshi, O. Tanaseichuk, C. Benner, and S.K. Chanda. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10: 1523. <https://doi.org/10.1038/s41467-019-09234-6>
- Zhu, L.Y., P.P. Pan, W. Fang, J.Z. Shao, and L.X. Xiang. 2012. Essential role of IL-4 and IL-4Ra interaction in adaptive immunity of zebrafish: Insight into the origin of Th2-like regulatory mechanism in ancient vertebrates. *J. Immunol.* 188:5571–5584. <https://doi.org/10.4049/jimmunol.1102259>
- Zimmerman, A.M., K.E. Romanowski, and B.J. Maddox. 2011. Targeted annotation of immunoglobulin light chain (IgL) genes in zebrafish from BAC clones reveals kappa-like recombining/deleting elements within IgL constant regions. *Fish Shellfish Immunol.* 31:697–703. <https://doi.org/10.1016/j.fsi.2010.09.015>

## Supplemental material

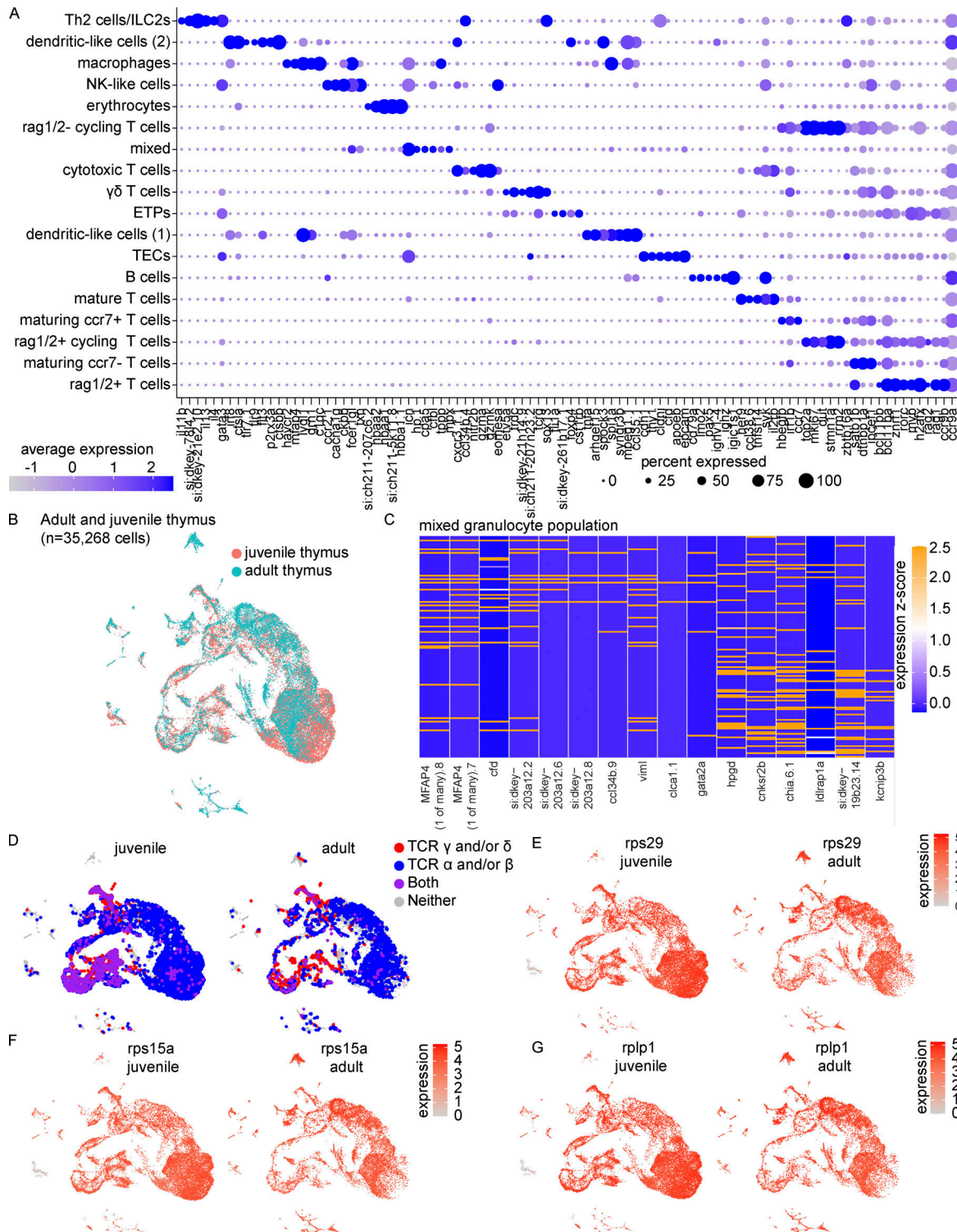


Figure S1. **Extended cell type characterizations of the zebrafish thymus.** (A) Dot plot of selected markers of the adult thymus showing expression across all cell populations, an expanded version of Fig. 1 B (see also Table S1). Dot size reflects the percentage of cells within a population expressing a given marker and dot color shows the average expression within that population. Similarly to Fig. 1, A and B, thymi were derived from four adult zebrafish (3–4 mpf) processed within the same experiment. (B) UMAP visualization by timepoint of the 35,268 adult and juvenile thymus cells integrated in Seurat. These cells were derived from the thymi of four adult zebrafish and four technical replicates of juvenile zebrafish (4 wpf; pool of cells from 21 zebrafish) processed in two independent experiments as in Fig. 1 C. (C) Heatmap visualization of selected genes within the mixed granulocyte population to demonstrate heterogeneity within the integrated adult and juvenile thymus cells. (D) UMAP visualization of detection of TCR  $\gamma$  and/or  $\delta$  (red),  $\alpha$  and/or  $\beta$  (blue), or a combination of both sets of receptors (purple) in the integrated adult and juvenile thymi. (E–G) UMAP visualizations of the expression of ribosomal genes (E) *rps29*, (F) *rps15a*, and (G) *rplp1* that demonstrated differential expression in erythrocytes derived from juvenile vs. adult thymi as determined by Wilcoxon rank sum test in Seurat (Table S4).

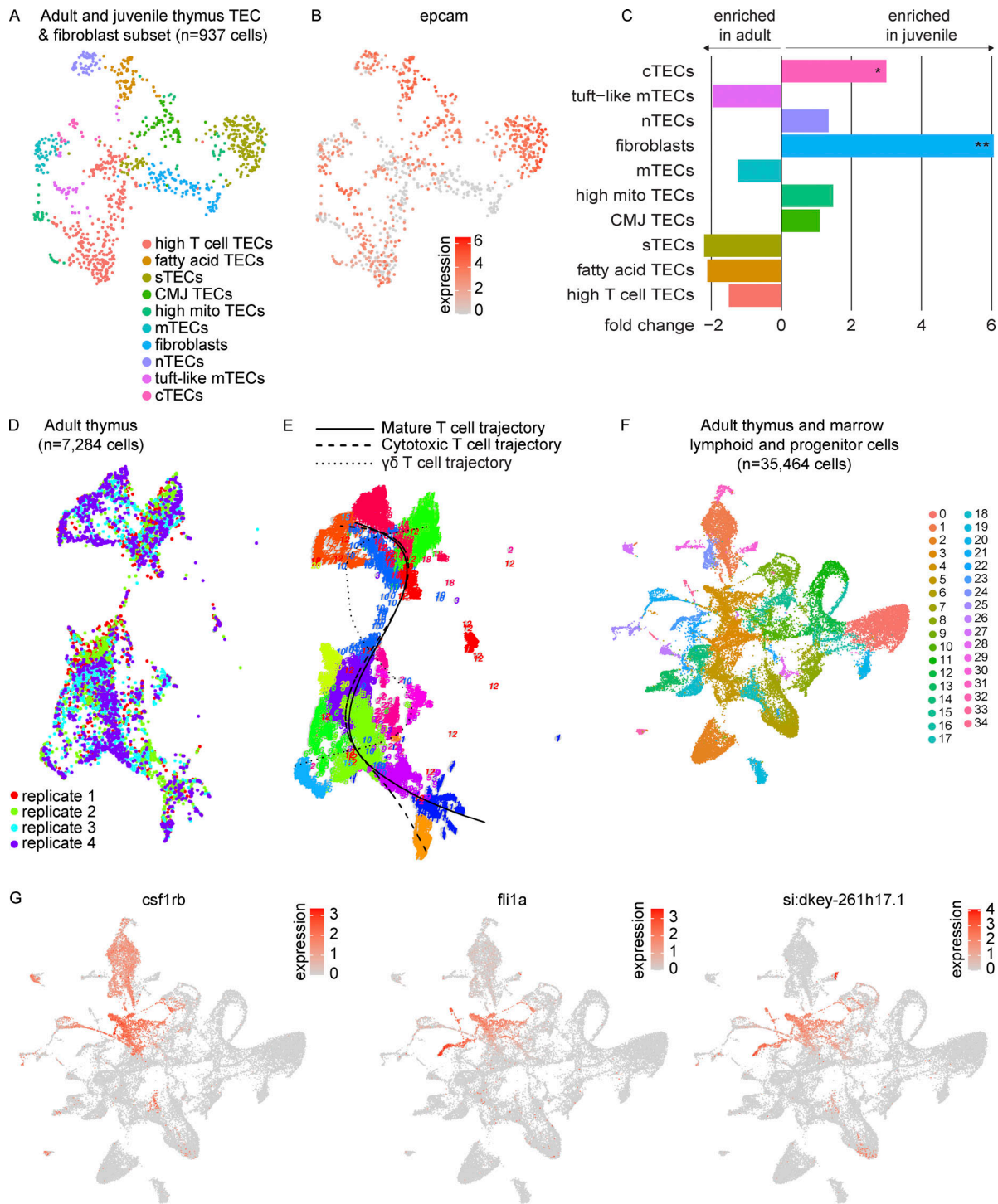


Figure S2. **Zebrafish TECs and T cell developmental trajectories.** (A) UMAP visualization of the 937 cells within the TEC and fibroblast populations of the combined adult and juvenile thymi as subsetted and re-analyzed to improve resolution of cell populations. Cell type annotations were based on differential expression analysis as determined by Wilcoxon rank sum test in Seurat (Table S5). (B) UMAP visualization of the expression of the epithelial marker, *epcam*, within the TEC and fibroblast subset. (C) Bar plot of beta-binomial test comparing composition of epithelial and fibroblast subset between juvenile and adult thymi (Table S6). Average fold change for each subpopulation is depicted by bar length and direction, with negative values enriched in the adult thymus and positive values enriched in the juvenile thymus. Statistical significance is shown for Benjamini-Hochberg adjusted P values: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ . (D) UMAP visualization of RaceID3 clustering of cells derived from four adult thymi ( $n = 7,284$  cells, independent analysis from Seurat) by replicate. (E) FateID principal curves fitted to mature T cell, cytotoxic T cell, and  $\gamma\delta$  T cell trajectories. (F) UMAP visualization of adult thymus and lymphoid and progenitor marrow fraction integrated together in Seurat ( $n = 35,464$  cells) colored by cluster. These cells were derived from seven adult zebrafish dissected and processed in three independent experiments; paired marrow and thymi were obtained from two zebrafish as in Fig. 5, A–C. Cluster 4 is where the majority of thymus ETPs were highlighted in Fig. 5 C. (G) UMAP visualizations of marker genes used to identify the HSPC cluster (cluster 4). nTEC, neural TEC; mito, mitochondrial-gene expressing; CMJ, corticomedullary junction; sTEC, structural TEC.

## Conserved markers



Adult thymus and marrow  
lymphoid and progenitor cells  
(n=35,464 cells)

Figure S3. **Conserved genes between marrow HSPCs and ETPs.** UMAP visualizations of subset of genes identified as being conserved across thymic marrow HSPCs and ETPs; 7 of 10 conserved genes listed in Fig. 5 K are shown here as the other three conserved genes were visualized in Fig. S2 G (see also Table S10). Marker conservation was determined by Wilcoxon rank sum test; P values for the ETPs and HSPCs were combined using Tippett's method (minimum P value).

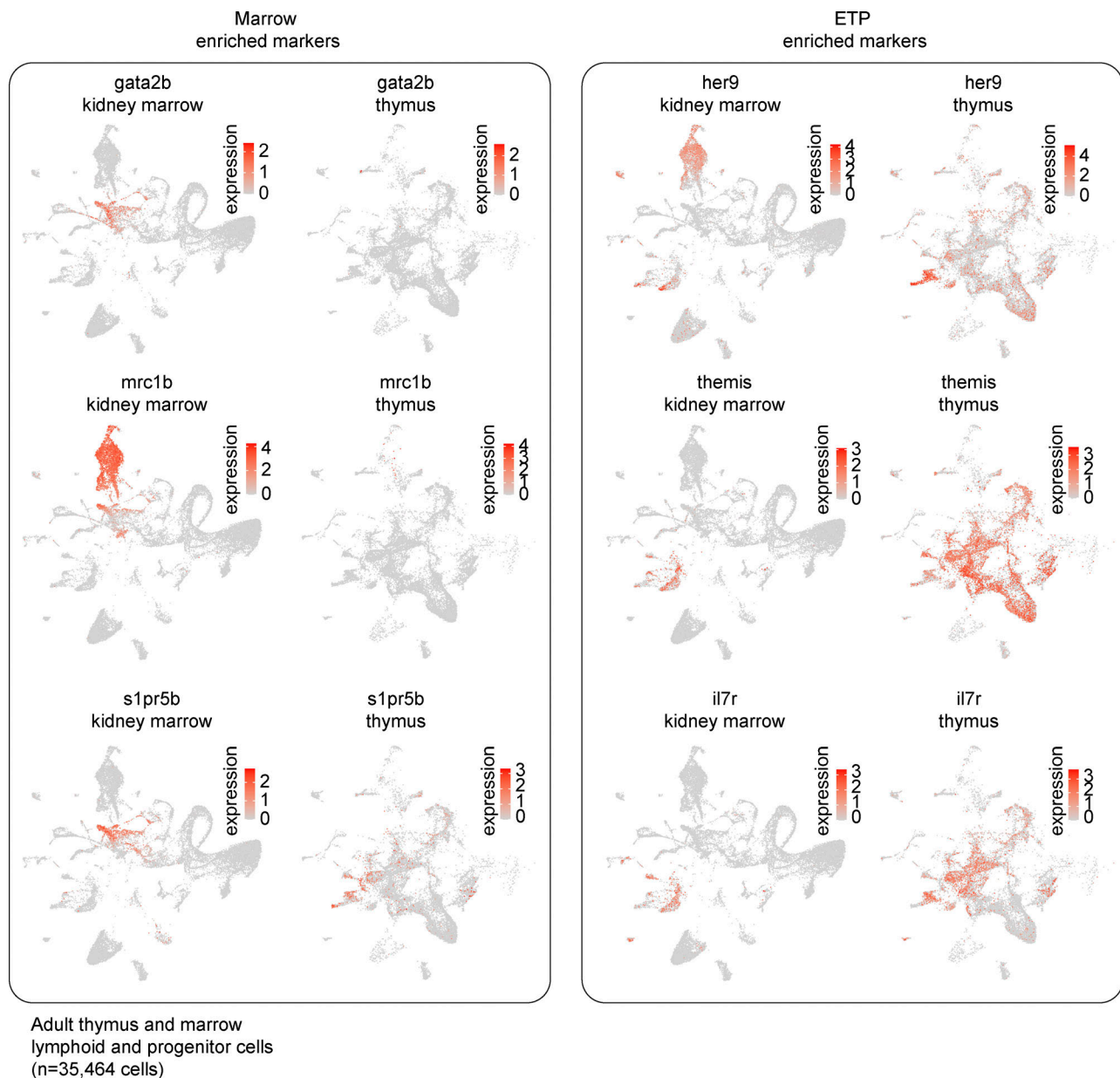


Figure S4. **Differentially expressed genes between marrow HSPCs and ETPs.** UMAP visualizations of subset of genes identified as being differentially expressed between marrow HSPCs and ETPs by Wilcoxon rank sum test (three shown for each; see also Table S11).

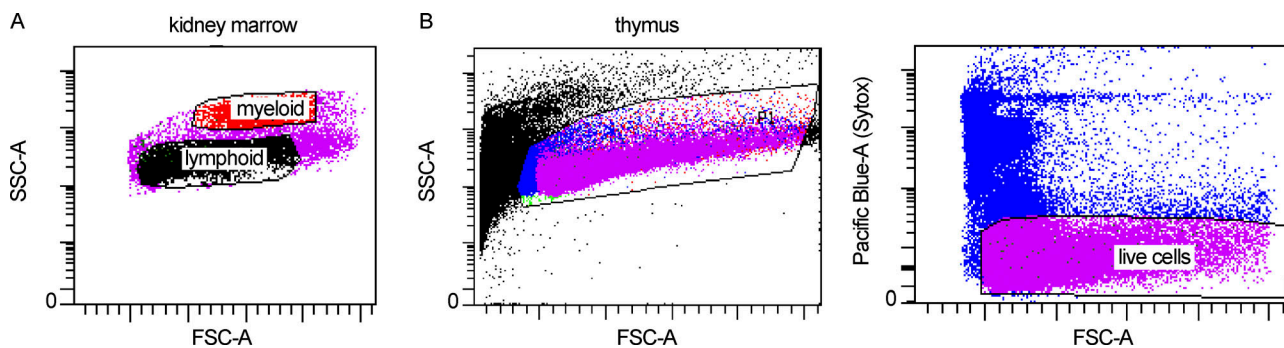


Figure S5. **Fluorescence-activated cell sorting gates for marrow and thymus.** (A and B) Representative fluorescence-activated cell sorting gating for (A) kidney marrow and (B) thymus sorts. The lymphoid gate of the marrow encompasses both lymphoid and progenitor cells whereas the myeloid gate contains mostly granulocytes. All live, single cells were sorted from the thymi. SSC-A, side scatter area; FSC-A, forward scatter area.



Provided online are Table S1, Table S2, Table S3, Table S4, Table S5, Table S6, Table S7, Table S8, Table S9, Table S10, Table S11, Table S12, Table S13, Table S14, Table S15, Table S16, Table S17, and Table S18. Table S1 shows differentially expressed genes for adult zebrafish thymus. Table S2 shows differentially expressed genes for adult and juvenile zebrafish thymi. Table S3 shows beta-binomial test results for adult and juvenile zebrafish thymi. Table S4 shows differentially expressed genes in adult and juvenile erythrocytes from thymus analysis. Table S5 shows differentially expressed genes for TEC and fibroblast subset of adult and juvenile thymi. Table S6 shows beta-binomial test results for TEC and fibroblast subset of adult and juvenile thymi. Table S7 shows gene lists for SOM nodes in FateID (adult thymus). Table S8 shows Monocle 3 gene modules for adult zebrafish kidney marrow. Table S9 shows gene modules used to identify zebrafish ETPs. Table S10 shows conserved marker genes between marrow HSPCs and ETPs. Table S11 shows differentially expressed genes between marrow HSPCs and ETPs. Table S12 shows Monocle 3 gene modules for marrow DC-like and NK-like cell subset. Table S13 shows differentially expressed genes for marrow T and NK-like cell subset. Table S14 shows top distinguishing genes between marrow T and NK-like cells. Table S15 shows differentially expressed genes for marrow B cell subset. Table S16 shows Monocle 3 gene modules for marrow B cell subset. Table S17 shows cell counts by cell type for full adult and juvenile thymi and marrow analyses. Table S18 shows marker gene references.