Single-cell multi-omic analysis of thymocyte development reveals NFAT as a driver of CD4/CD8 lineage commitment

Zoë Steier[1,2,3,#], Laura L. McIntyre[4,#], Lydia K. Lutes[4], Tse-Shun Huang[5],
Ellen A. Robey[4,*], Nir Yosef[3,6,7,8*], Aaron Streets[1,2,3,6,*]

[1] University of California, Berkeley, Department of Bioengineering, Berkeley, CA

[2] UC Berkeley - UCSF Graduate Program in Bioengineering, Berkeley and San Francisco, CA

[3] University of California, Berkeley, Center for Computational Biology, Berkeley, CA

[4] University of California, Berkeley, Division of Immunology and Pathogenesis, Department of Molecular and Cell Biology, Berkeley, CA

[5] BioLegend, Inc., San Diego, CA, USA

[6] Chan Zuckerberg Biohub, San Francisco, CA

[7] University of California, Berkeley, Department of Electrical Engineering and Computer Sciences, Berkeley, CA

[8] Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA

[#] Co-first authors

[*] Correspondence: astreets@berkeley.edu, niryosef@berkeley.edu, erobey@berkeley.edu

Abstract

CD4 and CD8 T cells play critical roles in the mammalian immune system. While their development within the thymus from the CD4+CD8+ stage has been widely studied as a model of lineage commitment, the underlying mechanism remains unclear. To deconstruct this process, we apply CITE-seq, measuring the transcriptome and over 100 surface proteins in thymocytes from wild-type and lineage-restricted mice. We jointly analyze the paired measurements to build a comprehensive timeline of RNA and protein expression in each lineage, supporting a sequential model of lineage determination in which both lineages go through an initial phase of CD4 lineage audition, which is followed by divergence and specification of CD8 lineage cells. We identify early differences implicating T cell receptor signaling via calcineurin-NFAT in driving CD4 lineage commitment. Pharmacological inhibition validates the requirement of calcineurin-NFAT for CD4, but not CD8, lineage development, providing insight into the CD4/CD8 commitment mechanism.

Keywords

Single cell; multi-omics, thymus; thymocyte; T cell development; CD4 T cell; CD8 T cell; fate decision; lineage commitment; NFAT; calcium signaling; T cell receptor signaling

Introduction

The maturation and selection of CD4 (helper) and CD8 (cytotoxic) T cells within the thymus is critical for mammalian adaptive immunity. These two primary types of T cells, despite having different functions, arise from a common subset of thymocytes known as double positive (DP) that express both the CD4 and CD8 coreceptors on their surface. While the development of thymocytes from the DP stage into CD4+ or CD8+ single positive (SP) T cells has been widely studied as a classic model of lineage determination, the mechanism underlying commitment to the two lineages remains partially understood. CD4 T cells receive signals through T cell receptor (TCR) recognition of antigenic peptides presented on Major Histocompatibility Complex class II (MHCII), whereas CD8 T cells interact with peptide/MHCI complexes, and CD4 and CD8 surface proteins serve as coreceptors that cooperate with the TCR by binding directly to MHCII and MHCI, respectively. As T cells develop in the thymus, recognition of self-peptide/MHC can provide survival and differentiation signals in a process known as positive selection. During positive selection, the strength and dynamics of signaling through the TCR have been implicated in influencing the choice of DP precursors between the CD4 versus CD8 fates (Germain, 2002; Singer et al., 2008; Xiong and Bosselut, 2012; Moran et al., 2011). However, it remains unclear how signaling through the same receptor results in different lineage outcomes including the loss of bi-potency (referred to as lineage commitment) and the acquisition of lineage-specific characteristics (referred to as lineage specification). It is also known that the transcription factors (TFs) THPOK (encoded by *Zbtb7b*) and RUNX3 act as mutually-antagonizing master regulators of the CD4 and CD8 lineages, respectively, and help enforce subsequent differences in phenotype and effector functions. However, how differences in TCR signaling drive later differences between the two lineages and the events that link TCR signals with differential expression of the master regulators remain largely open questions (Taniuchi, 2016).

A major complicating factor in addressing these and similar questions is the heterogeneity of maturing T cells in the thymus – spanning different phases (positive selection, followed by negative or agonist selection), spatial locations (from cortex to medulla and sub-areas thereof), interactions with different types of support cells, TCR specificities, and ultimate fates (Kurd et al, 2016). The complexity of cell states in this differentiation process has long been characterized using surface protein markers including, and in addition to, CD4 and CD8 (Germain, 2002). However, this approach has been traditionally reliant on flow cytometry or fluorescence activated cell sorting (FACS) and is therefore limited in its capacity for defining intermediate stages of development or new subpopulations, as it hinges on small sets of pre-selected markers and lacks the ability to make quantitative comparisons in gene or protein expression. In contrast, advances in single-cell RNA sequencing (scRNA-seq) technologies have more recently enabled the unbiased observation of transcriptional heterogeneity in the thymus. scRNA-seq was recently used to construct a census of cell states in the mammalian thymus and helped identify new thymic subsets, as well as shed light on the kinetics of TCR rearrangement prior to positive selection (Park et al., 2020). Additional studies have used high-throughput single-cell analysis techniques to identify the early precursor populations that seed the thymus (Lavaert et al., 2020) and characterize the process of progenitor commitment (Zhou et al., 2019). A recent study added an analysis of chromatin accessibility to explore the subsequent development of the two lineages and computationally identify additional putative regulators (Chopp et al., 2020). While this study and another employing scRNA-seq (Karimi et al., 2021) led to important insights, the analyses did not provide sufficient temporal resolution to comprehensively pinpoint the early developmental stages (prior to the induction of THPOK and RUNX3). As a case in point, there is evidence for a late CD8-fated CD4+CD8+ thymocyte population (Saini et al., 2010), but this

3

population was not distinguished from earlier TCR signaled CD4+CD8+ populations that contain both CD4-fated and CD8-fated cells (Chopp et al., 2020; Karimi et al., 2021). Furthermore, putative regulators of the respective early divergence events have not been validated. Finally, these studies relied on RNA and chromatin profiles, limiting the ability to relate the results to prior protein-based knowledge on the diversity of cell states in this system. Thus, generating a high-resolution delineation of the differentiation process that provides connections to FACS-based studies, and establishing experimentally validated mechanisms of early divergence between the two lineages remain largely open endeavors.

To address these challenges, we leveraged the CITE-seq protocol (Stoeckius et al., 2017) to simultaneously measure the transcriptome and over 100 surface proteins in thousands of single thymocytes. Using the totalVI algorithm (Gayoso et al., 2021), we jointly analyzed the paired measurements in order to build a comprehensive timeline of RNA and protein expression spanning positive selection and lineage commitment. The availability of protein information allowed us to relate our findings to the foundational protein-based literature and further add to it by clarifying intermediate developmental stages and better define these stages by both transcript and surface protein composition. Furthermore, our study design included thymi from both wild-type (WT) and lineage-restricted mice, which provided a strategy to compare between lineages at the early stages when it is not possible to distinguish between lineages in WT samples. Through this analysis, we detected early differences in TCR signaling and identified signaling through calcineurin-NFAT as a putative driver of lineage differences. To validate calcineurin-NFAT as a differential driver of lineage commitment, we applied drug perturbations to an ex vivo culture system of neonatal thymic slices. While calcineurin-NFAT was necessary for CD4 lineage commitment, it was not necessary for either commitment to or maturation of the CD8 lineage. Beyond providing a high-resolution map of development, the findings presented here help fill the knowledge gap between early differences in TCR signaling at the cell surface and the subsequent differential activation of master regulator TFs in the nucleus, thus establishing a more complete model for how T cell fate commitment is controlled.

Results

A joint transcriptomic and surface protein atlas of thymocyte development in wild-type and lineage-restricted mice

To study T cell development and CD4/CD8 lineage commitment, we profiled thymocytes from both WT and lineage-restricted mice. Thymocyte populations in WT mice closely resemble those in humans (Park et al., 2020), and serve as a model of T cell development in a healthy mammalian system. However, in a WT system it is not possible to predict the ultimate fate of immature thymocytes, making it challenging to investigate the process of commitment to the CD4 or CD8 lineage. To probe the mechanism of lineage commitment, we profiled thymocytes from both WT (C57BL/6; referred to as B6) mice and mice with CD4-lineage- or CD8-lineage-restricted T cells. To track commitment to the CD4 lineage (MHCII-specific) we used mice that lack MHCI expression (β2M-/-; referred to as MHCI-/-), which have polyclonal TCR repertoires, and two TCR transgenic (TCRtg) mice that express TCRs that are specific for MHCII (AND and OT-II). To track commitment to the CD8 lineage (MHCI-specific) we used mice that lack MHCII expression (I-Aβ-/-; referred to as MHCII-/-) and two MHCI-specific TCRtg mice (F5 and OT-I). In all of these lineage-restricted mice, thymocytes are expected to pass through the same stages of development as WT thymocytes (Figure 1A). However, unlike in WT thymocytes, the fate of lineage-restricted thymocytes is known before cells present the CD4+ SP or CD8+ SP phenotype, allowing for independent characterizations of CD4 and CD8 T cell development and lineage commitment.

We characterized thymocyte development at the single-cell level by measuring transcriptomes and surface protein composition using CITE-seq (Stoeckius et al., 2017) with a panel of 111 antibodies (Table S1). We jointly analyzed these features using totalVI, which accounts for nuisance factors such as variability between samples, limited sensitivity in the RNA data and background in the protein data (Gayoso et al., 2021; Methods). We collected thymi from two biological replicates per lineage-restricted genotype and five WT biological replicates (Table S2). To enrich for thymocytes undergoing positive selection in samples from non-transgenic mice, MHC-deficient samples and three WT replicates were sorted by FACS for CD5+TCRβ+ (Figure S1A). We integrated CITE-seq data from all samples (72,042 cells) using totalVI, which allowed us to stratify cell types and states based on both RNA and protein information regardless of mouse genotype (Figure 1B). We identified the expected stages of thymocyte development including early CD4-CD8- (double negative; DN) and proliferating CD4+CD8+ (double positive proliferating; DP (P)) stages. We detected early and late stages of DP cells undergoing TCR recombination, as well as DP cells post-recombination that are downregulating *Rag* and receiving positive selection signals (signaled DP). In addition to immature and mature stages of CD4 and CD8 T cells, we observed two distinct waves of cells that appeared to be undergoing negative selection based on expression of *Bcl2l11* (BIM) and *Nr4a1* (NUR77) (Daley et al., 2013). The first appeared to emerge from the signaled DP population (lying adjacent to a cluster of dying cells), and the second from immature CD4 T cells. *Foxp3*+ regulatory T cells appeared to cluster near mature conventional CD4 T cells and the second subset of negatively selected cells. Other populations included unconventional T cells (gamma-delta T cells, NKT cells), small clusters of non-T cells (B cells, myeloid cells, and erythrocytes), a thymocyte population with high expression of interferon response genes (Xing et al., 2016), and a population of mature T cells that had returned to cycling following the cell cycle pause during thymocyte development. As expected, WT, MHCII-specific, and MHCI-specific samples were well-mixed in earlier developmental stages but segregated into CD4 and CD8 lineages in later-stage populations (Figure 1C).

Using totalVI, we defined cell populations with traditional cell type markers (Figure S1B-C) and with unbiased differential expression tests of all measured genes and proteins (Figure 1D-E, Table S3). Top differentially expressed features included classical cell surface markers of lineage (e.g., CD4, CD8), key transcription factors (e.g., *Foxp3*, *Zbtb7b*), and markers of maturation stage (e.g., *Rag1*, *Ccr4*, *S1pr1*). In addition to supporting the relevance of surface proteins in characterizing cell identities, these multi-omic definitions revealed gradual expression changes, particularly between the DP and CD4 and CD8 SP stages, that are best understood not as discrete populations, but as part of a continuous developmental process. Observation of these groups allowed us to select the populations of thymocytes receiving positive selection signals for further continuous analysis.

We focused our analysis on developing thymocytes from the signaled DP stage through mature CD4 and CD8 T cells (Methods; Figure S1D). The totalVI latent space derived from these populations captured the continuous transitions that stratified thymocytes by developmental stage and CD4/CD8 lineage (Figure 1F). This was evident through visualization of totalVI denoised protein and RNA expression of known markers including, for example, CD4 and CD8 protein expression, which revealed a gradual transition into a CD4 or CD8 SP phenotype in each of the two branches. In the CD8 branch, a CD4+CD8+ population could be seen even after a separation of the lineages in the UMAP representation of the totalVI latent space (Figure 1G; Becht et al., 2019), indicating that combining transcriptome-wide information with surface protein measurements might reveal earlier signs of lineage commitment than could have been observed from FACS-sorted populations. We also observed protein markers indicative of positive-selection-induced

TCR signaling (CD5, CD69) and maturation stage (CD24, CD62L), as well as RNA markers of TCR recombination (*Rag1)*, cell location within the thymus (chemokine receptors *Cxcr4*, *Ccr7*), and lineage regulation (TFs *Gata3*, *Zbtb7b*, *Runx3*) (Figure 1H). This single-cell analysis of positively-selecting thymocytes from WT and lineage-restricted mice thus provides a high-resolution snapshot of the continuous developmental processes, capturing cells at a variety of states that span the spectrum between the precursor (DP) and late (CD4 or CD8 SP) stages.

Pseudotime inference captures continuous maturation trajectory and clarifies intermediate stages of development

To further characterize the observed continuum of cell states, we performed pseudotime inference with Slingshot (Street et al., 2018; Saelens et al., 2019) to delineate the changes in RNA and protein expression over the course of development from the DP to SP stages (Figures 2A and S2A; Methods). Since cell-cell similarities in the reduced dimension space were based on both RNA and protein information, the placement of a cell in pseudotime reflected gradual changes in both gene and protein expression. The pseudotime inferred by Slingshot consisted of a branching trajectory with each branch corresponding to a different lineage (CD4, CD8), as can be observed by the positioning of cells from lineage-restricted mice (Methods; Table S4 and Figure 1F).

We confirmed that the pseudotime ordering determined by Slingshot correctly captured the presence and timing of many known expression changes during thymocyte development (Hogquist et al., 2015) at the RNA and protein levels (Figure 2B-C). This included events such as early downregulation of TCR recombination markers *Rag1* and *Rag2*, gradual downregulation of early markers such as *Ccr9* and *Cd24a*/CD24, transient expression of the activation and positive selection marker *Cd69*/CD69, and late upregulation of maturation markers such as *Klf2*, *S1pr1*, and *Sell*/CD62L. To explore beyond known markers, we tested for differential expression over pseudotime with totalVI (Methods) and created a comprehensive timeline of changes in RNA and protein expression separately for each lineage (Figure 2D; Tables S5 and S6). There were some expected differences visible between the two lineages in the expression of key molecules (e.g., coreceptors and master regulators; Figure 2B). However, known markers of maturation followed similar patterns and timing in both lineages (Figure 2C). Furthermore, some of the most significant differential expression events over time were common between the two lineages including the early downregulation of *Arpp21* (a negative regulator of TCR signaling (Mingueneau et al., 2013)), transient expression of the transcriptional regulator *Id2* (Cannarile et al., 2006), upregulation of Tesc, (an inhibitor of calcineurin (Perera et al., 2010)), and later upregulation of *Ms4a4b* (shown to inhibit proliferation in response to TCR stimulation (Yan et al., 2012)) (Figure 2D). This pseudotime analysis therefore provides a comprehensive view of the continuum of expression changes during the developmental process. Moreover, the consistency in timing of expression changes enables an investigation of the two lineages at comparable developmental stages.

We next sought to use pseudotime information to clarify the intermediate stages of development in the two lineages. Thymocyte populations have been commonly defined by surface protein expression using flow cytometry, and various marker combinations and gating strategies have been employed to subset thymocyte populations based on maturity and lineage (Germain, 2002; Xiong and Bosselut, 2012; Saini et al., 2010; Hu et al., 2012). However, due to the continuous nature of developmental intermediates, as well as technical variations in marker detection, no uniform consensus has emerged on how to define positive selection intermediates by flow cytometry. To address this, we performed in silico flow cytometry

analysis on totalVI denoised expression of key surface protein markers and explored their ability to distinguish between different pseudotime phases along the two lineages (Methods).

Starting with the canonical CD4 and CD8 markers, we found that MHCII-specific cells appeared to progress continuously in pseudotime from DP to CD4+CD8low to CD4+CD8- (Figures 2E-F and S2B). In contrast, MHCI-specific cells appeared to progress from DP to the CD4+CD8low gate before reversing course to reach the eventual CD4-CD8+ gate later in pseudotime, consistent with the previous literature (Lundberg et al., 1995; Lucas and Germain, 1996; Chan et al., 1993). Using pseudotime information, we detected the existence of a developmental phase (between pseudotime 6-8) at which nearly all MHCI-specific cells fall in the CD4+CD8low gate (Figure 2F). The high numbers of MHCI-specific cells in the CD4+CD8low gate underscore the fact that WT cells within the CD4+CD8low gate cannot be assumed to be committed to the CD4 lineage. Our data also provided an opportunity to explore the progression of MHCI-specific after the CD4+CD8low gate, which has been less well defined in the literature. To this end, our data indicated that at subsequent stages (pseudotime 8-12), MHCI-specific thymocytes pass again through a DP phase on their way from the CD4+CD8low gate to the CD4-CD8+ gates, while the MHCII-specific lineage does not contain late-time DP cells. Although a population of later-time MHCI-specific DP cells has been previously described ("DP3"; Saini et al., 2010), it is not commonly accounted for (Park et al., 2020; Chopp et al., 2020, Karimi et al., 2021), resulting in a missing stage of CD8 T cell development and potential contamination of the DP gate with later-time CD8 lineage cells.

We next sought to computationally identify a minimal set of surface markers for these and other stages of differentiation in our data. Our goal was twofold: first, we aimed to leverage markers that had previously been used for isolating intermediate subpopulations to establish consistency between our data and other studies; second, we aimed to better characterize the late MHCI-specific DP stage using a refined, data-driven, gating strategy. We found that four stages in time (independent of lineage) could be largely separated by in silico gating on CD69 and CD127(IL-7Ra), in which thymocytes begin with low expression of both markers, first upregulate CD69, later upregulate CD127, and finally downregulate CD69 (Figure S2C-D). The addition of CD4 and CD8 as markers allowed for the separation of lineages at later times. Finally, we established that the later-time DP population that is prominent in the CD8 lineage could be distinguished from the earlier DP cells by high expression of TCRβ (Saini et al., 2010; Marodon et al., 1994) in addition to expression of both CD69 and CD127 (Figure S2D-E). Here, we refer to this later-time DP population as DP3 to distinguish it from the earlier DP1 (CD69-, CD127-) and DP2 (CD69+, CD127-) populations.

In combination, a gating scheme based on these five surface proteins (CD4, CD8, TCRβ, CD69, and CD127) identified eight populations (DP1, DP2, CD4+CD8low, DP3, semimature CD4, semimature CD8, mature CD4, and mature CD8). This scheme, which included a marker previously unused in this setting (CD127), allows FACS to approximate the binning of thymocytes along pseudotime and lineage (Figure 2G-J). Fluorescence-based flow cytometry replicated these CITE-seq-derived gates (Methods), enabling the isolation of the eight described populations and supporting the presence of the proposed intermediate stages (Figure S2F-G). Collectively, these findings allowed us to specify an updated model of positive selection intermediates in both the CD4 and CD8 lineages (Figure S2F).

Paired measurements of RNA and protein reveal the timing of major events in CD4/CD8 lineage commitment

While defining thymocyte populations by cell surface markers provides an approximate ordering of discrete developmental stages, a quantitative and high-resolution timeline of the differences between the lineages provided by CITE-seq data could address key outstanding questions. In particular, while the downregulation of CD8 in both MHCII- and MHCI-specific thymocytes suggests that all positively selecting thymocytes initially audition for the CD4 lineage, it is not clear whether MHCII- and MHCI-specific thymocytes exhibit parallel temporal changes in CD4-defining transcription factors. In addition, the key events that lead to CD8 lineage specification have not been defined, in large part due to the lack of temporal resolution for the events that occur in MHCI-specific thymocytes after divergence from the CD4 lineage. To gain further and more nuanced insight into CD4 versus CD8 lineage commitment, we compared the continuum of expression changes of key coreceptors, transcription factors, and TCR signaling molecules involved in this process.

The expression of the CD4 and CD8 coreceptors plays an important role not only in defining the lineage of mature T cells, but also in transmitting the TCR signals that are necessary for thymocyte development (Germain, 2002). We observed that the expression of these coreceptors followed an expected pattern by which RNA expression preceded the corresponding change in protein expression, likely due to the time needed for protein translation and transport (Figures 3A and S3A). Beginning from the DP stage with high expression of both coreceptors in both lineages, we observed small dips in expression of both coreceptors ("double dull" stage (Lucas and Germain, 1996)) followed by a rise in CD4 and a continued decrease in CD8. Eventually, in the CD8 lineage, CD8 expression rose in parallel to a decrease in CD4 expression, resulting in a late transient DP stage (DP3) as the cells moved towards the CD8 SP phenotype. Differential expression between the CD4 and CD8 lineages indicated that a significant difference in *Cd8a* RNA expression began building from pseudotime point 6 (approximately in the early CD4+CD8low gate), followed by later accumulation of difference in CD8 protein expression (Figure 3B). It was not until pseudotime point 9 (the point at which the lineages can first be distinguished by flow cytometry in the semimature CD4 versus DP3 gates), that a significant difference in CD4 expression emerged. These results suggest that while the two lineages cannot be distinguished by flow cytometry at the CD4+CD8low stage, they have already begun to diverge based on *Cd8a* RNA levels.

While previous studies have characterized the expression of key lineage-defining transcription factors relative to changes in CD4 and CD8 surface expression using genetically-encoded fluorescent reporters and flow cytometry (Muroi et al., 2008, Egawa and Littman, 2008), our pseudotime analysis could more precisely characterize transcription factors' expression as they diverge between the two lineages and relate them directly to differential expression of *Cd4* and *Cd8a* RNA. To this end, we explored the timing of expression changes of the key lineage-specific transcription factor genes *Runx3* (CD8 lineage), *Zbtb7b* (encoding THPOK; CD4 lineage), and *Gata3* (a known up-stream activator of *Zbtb7b* that is more highly expressed in the CD4 lineage) (Wang et al., 2008, Taniuchi, 2016) (Figure 3C-D). Focusing on when transcription factor and coreceptor expression first diverged between the two lineages (Figure 3B,D), we observed that *Gata3* became differentially expressed in CD4 lineage cells at pseudotime 4-5, prior to differential coreceptor expression. This was followed by differential upregulation of *Zbtb7b* and downregulation of *Cd8a* in the CD4 lineage at pseudotime 6-8, consistent with the role of THPOK in repressing CD8 expression (Taniuchi, 2016). Finally, differential upregulation of *Runx3* and downregulation of *Cd4* in the CD8 lineage occurred at pseudotime 8-10, consistent with the role of RUNX3

in repressing the *Cd4* gene (Taniuchi, 2016). Examining the continuous expression changes over pseudotime (Figure 3C), we observed a parallel pattern in both CD4 and CD8 lineage cells in which *Gata3* induction was followed by a rise in *Zbtb7b*, although the expression of both CD4-associated transcription factors was lower and more transient in the CD8 lineage compared to the CD4 lineage. Interestingly, the large rise in *Runx3* expression, which occurred only in the CD8 lineage, coincided with the decrease in *Zbtb7b* in that lineage (pseudotime 7-11, corresponding to the CD4+CD8low and DP3 stages). This implied that both transcription factors may be transiently co-expressed at this stage, in spite of their ability to repress each other's expression and their reported mutually exclusive expression at later stages (Egawa and Littman, 2008; Vacchio and Bosselut, 2016; Taniuchi, 2016). In a more detailed view, in silico flow cytometry of *Zbtb7b* and *Runx3* expression for WT or lineage-restricted thymocytes (Figure 3E and S3B) showed a continuous transition in which all thymocytes initially upregulated *Zbtb7b*, whereas CD8 lineage (MHCI-specific) thymocytes subsequently and gradually downregulated *Zbtb7b*, simultaneous with *Runx3* upregulation. Intracellular flow cytometry staining supported the observed timing in differential expression of TFs (Figure S3C; Methods). Altogether, our data provide strong evidence for an initial CD4 auditioning phase for all positively selected thymocytes, and confirm that the master regulators THPOK and RUNX3 account for early lineage-specific repression of *Cd8* and *Cd4* genes respectively, with THPOK induction and CD8 repression in CD4-fated cells occurring prior to RUNX3 induction and CD4 repression in CD8-fated cells.

Previous studies have implicated TCR signaling as a driver of early differences in DP thymocytes that ultimately lead to differential expression of the lineage regulators, whereas the role of TCR signals after the CD4 lineage branch point remains controversial (Germain, 2002; Xiong and Bosselut, 2012, Singer et al., 2008). There are indications that MHCII-specific thymocytes may have a higher intensity, duration, and frequency of TCR signaling, and disruption of TCR signaling at the DP stage can promote the CD8 fate (Yasutomo et al., 2000; Liu and Bosselut, 2004; Matechak et al., 1996). Consistently, we observed a higher and more prolonged TCR response (indicated by expression of TCR target genes *Cd69 and Egr1*) in the CD4 lineage (Figure 3F), which became significantly differentially expressed simultaneous to *Gata3* and prior to *Zbtb7b* (Figure 3G). We also observed a second TCR response in the CD8 lineage after *Zbtb7b* induction (coinciding with the DP3 and semimature CD8 stages) that was not present in the CD4 lineage. This may reflect the tuning of the TCR response by MHCI-specific thymocytes to increase their sensitivity to TCR signals at later stages of positive selection (Saini et al., 2010; Au-Yeung et al., 2014, Lutes et al., 2021), and may be due in part to increased surface expression of the CD8 coreceptor and decreased expression of the negative regulator CD5 (Figures 3A and S3A). Interestingly, the second TCR signaling wave overlapped with the rise in *Runx3* and decline in *Zbtb7b* expression in MHCI-specific thymocytes after the CD4 lineage branch point, suggesting that it may be involved in driving CD8 lineage specification.

Taken together, the timing of these events encompassing coreceptors, master regulators, and TCR signaling responses support a sequential model of CD4 and CD8 T cell development, summarized in Figure S4A. In this model, all positively-selected DP thymocytes begin the process of lineage commitment by auditioning for the CD4 lineage, as GATA3 upregulation followed by THPOK induction accompanied by a drop in CD8 expression occur in parallel in both MHCII- and MHCI-specific thymocytes. Sustained TCR signals in MHCII-specific thymocytes during an initial wave of TCR signaling locks in the CD4 fate, likely due to higher GATA3 expression and the activation of the THPOK positive autoregulation loop (Muroi et al., 2008; Wang et al., 2008), and accompanied by full repression of the *Cd8* gene by THPOK. In contrast, a transient initial wave of TCR signaling in MHCI-specific thymocytes leads to an eventual drop in GATA3

and THPOK expression as the window for CD8 lineage specification begins, perhaps due to general maturation-inducing signals including the decline in E protein (E2A and HEB) transcription factor activity (Jones-Mason et al., 2012). Indeed, we observed a steady downregulation of the E protein genes *Tcf3* (E2A) and *Tcf12* (HEB) as well as transient induction of the E protein inhibitors *Id2* and *Id3* during development (Figure S3A). During this window, RUNX3 expression rises as a second wave of TCR signaling provides survival and lineage reinforcement signals to drive CD8 development. Our pseudotime analyses and this model provide a useful framework for further mechanistic dissection of the drivers of the CD4 versus CD8 lineage decision.

Emergence of differences between CD4 and CD8 lineages implicates putative drivers of lineage commitment

To better understand the process of lineage commitment, we systematically investigated how differences emerge between the CD4 and CD8 lineages by performing a totalVI differential expression test between lineage-restricted thymocytes within the same unit of pseudotime (Methods). There were no substantial differences in either RNA or protein expression between thymocytes from lineage-restricted mice at the early DP stages, and differentially expressed features gradually accumulated throughout maturation (Figure 4A; Table S7). This analysis resulted in a set of 302 genes that had significantly higher expression in MHCII-specific thymocytes ("CD4-DE") and 397 genes that had significantly higher expression in MHCI-specific thymocytes ("CD8-DE") in at least one pseudotime unit. In this test, 92 genes were included in both lists due to, for example, early up-regulation in one lineage followed by late up-regulation in the other lineage. The genes in each set were clustered by their expression across all cells in the corresponding lineage (Figure 4B-E; Tables S8 and S9). Inspection of gene expression within each cluster over pseudotime revealed characteristic temporal patterns, reflecting variation in gene expression over the course of maturation within each lineage (Figure 4B-E). For example, CD4-DE cluster 5 and CD8-DE cluster 1 showed a late divergence of expression of master regulator TFs and genes related to effector functions in their respective lineages (e.g., *Zbtb7b* and *Cd40lg* in MHCII-specific cells, and *Runx3* and *Nkg7* in MHCI-specific cells).

Considering all gene clusters, we investigated which genes shared similar variation within each lineage. We identified three clusters that were enriched for genes in TCR signaling pathways: CD4-DE clusters 4 and 7, and CD8-DE cluster 3 (hypergeometric test, Benjamini-Hochberg (BH)-adjusted P < 0.05; Methods). Both CD4-DE cluster 7 and CD8-DE cluster 3 contained many of the same TCR-target genes (e.g., *Cd69* and *Egr1*), which were initially more highly expressed in the CD4 lineage, and later more highly expressed in the CD8 lineage (Figure 3F-G). These gene clusters showed an early and sustained expression peak in CD4 lineage cells, and two temporally separated, transient peaks in the CD8 lineage, suggestive of distinct lineage-specific temporal patterns of TCR signaling. CD8-DE clusters 0 and 4 exhibited increased expression in the CD8 lineage just before the second rise in TCR signaling and contained genes implicated in modulating TCR sensitivity, providing a possible explanation for the late increase in TCR signaling in MHCI-specific thymocytes. For example, cluster 0 contained *Cd8a*, which is required for MHCI recognition, and *Themis*, which modulates TCR signal strength during positive selection (Choi et al., 2017). CD8-DE cluster 4 contained ion channel component genes *Kcna2* and *Tmie*, which have previously been proposed to play a role in enhancing TCR sensitivity in thymocytes with low self-reactivity (Lutes et al., 2021). CD4-DE cluster 4, which included *Gata3* and *Cd5*, showed sustained expression in the CD4 lineage relative to the CD8 lineage, but did not exhibit the pronounced second rise in expression in the CD8 lineage. We ordered genes by the timing of greatest mean differential expression of their respective clusters to

10

visualize the dynamics of emerging differences between lineages genome-wide (Figure 4D-E). This analysis captured the growing magnitude of phenotypic differences between MHCII- and MCHI-specific thymocytes as they develop.

To identify factors that may influence lineage commitment, we narrowed our focus to the pseudotime period just after differential gene expression was first detected and immediately upstream of master regulator expression. We performed transcription factor enrichment analysis with ChEA3 (Keenan et al., 2019), which identifies the TFs most likely to explain the expression of a set of target genes according to an integrated scoring across multiple information sources for potential regulatory activity, including ENCODE and ReMap ChIP-seq experiments (Dunham et al., 2012; Cheneby et al., 2020). We used genes differentially expressed between lineages in each unit of pseudotime as the target gene sets (Figure 4F-G; Tables S10 and S11; Methods) and ranked candidate TFs based on enrichment in each of the three pseudotime units prior to master regulator differential expression in each lineage (pseudotimes 4-7 for the CD4 lineage and pseudotimes 5-8 for the CD8 lineage). In addition to their ranking, we annotated three additional characteristics for each transcription factor that provided guidance for selecting likely regulators of lineage commitment. These characteristics included a known association with TCR signaling (Kandasamy et al., 2010; Methods), evidence of regulating *Gata3*, *Zbtb7b*, or *Runx3* according to ChEA3 databases, and differential expression of this TF itself at the relevant pseudotime stage.

In MHCI-specific cells, multiple highly ranked transcription factors such as *Ets1* and *Tcf7* have been implicated in thymocyte development (Zamisch et al., 2009), but have been previously shown to have relevance to the maturation of thymocytes in both lineages (Wang et al., 2010; Steinke et al., 2014). In MHCII-specific cells, we observed that multiple highly-ranked TFs were members of pathways associated with TCR signaling (e.g., *Egr2*, *Nfatc2*, *Egr1*, *Nfatc1*, and *Rel*), consistent with our observation that multiple TCR response genes were upregulated in the CD4 lineage relative to the CD8 lineage in the time period prior to lineage branching (Figures 3G and 4H). In particular, the top two ranked TFs in the CD4 lineage were *Egr2* and *Nfatc2*, which are known to lie downstream of two of the three main branches of the TCR signal transduction pathway: the ERK-MAPK branch and calcineurin-NFAT branch, respectively (Figure 4I; Navarro and Cantrell, 2014; Hogquist and Jameson, 2014; Malissen et al., 2014; Chakraborty and Weiss, 2014). While the ERK-MAPK branch plays a crucial role downstream of TCR signaling during positive selection in both lineages (Sharp et al., 1997; Wilkinson and Kaye, 2001; Daniels et al., 2006; McNeil et al., 2005), the roles of the other two branches (calcineurin-NFAT and PKC-NF-kB) are less clear (Gallo et al., 2007, Hettman and Leiden, 2000; Jimi et al., 2008). Investigation of the genes that were responsible for the high ranking of *Nfatc2* in the CD4 lineage revealed TCR target genes including *Egr1*, *Dusp5*, *Gata3*, *Cd5*, *Cd69*, and *Nr4a1* (Figure 4H). While *Nfatc2* was also ranked highly among early CD8-DE genes (Figure 4G), its putative targets in the CD8 lineage did not include TCR targets. Furthermore, we observed higher enrichment of the NFAT TFs in the CD4 lineage relative to the CD8 lineage (both in terms of their absolute enrichment score and their ranking compared to other TFs), suggesting that NFAT may play a larger role in the CD4 lineage. Although the NFAT TFs were not differentially expressed, NFAT activation is regulated by calcium signaling (not observed at the transcriptional level) and could therefore be differentially activated without being differentially expressed at the RNA level. These observations led us to hypothesize that TCR signaling through NFAT may be involved in differential commitment towards the CD4 rather than the CD8 lineage.

To more closely explore how branches of the TCR signaling pathway are associated with divergent transcriptional regulation between the two lineages, we focused on the three gene clusters that showed

11

the greatest enrichment for TCR target genes: CD4-DE clusters 4 and 7 and CD8-DE cluster 3 (Figure 4B-C, H). The single early peak during the CD4 audition stage in CD4-DE cluster 4, compared to the biphasic pattern of peaks during both the CD4 audition stage and the CD8 specification phase in the other two clusters (Figure 4H), suggested that the CD4-DE cluster 4 genes might be regulated by a branch of the TCR signaling pathway that is selectively active during the early CD4 audition phase. Interestingly, this gene cluster contained *Gata3*, which plays a key role in CD4 fate by activating the CD4 master regulator *Zbtb7b* (Wang et al., 2008), and has been previously implicated as a target of the TCR-associated TF NFAT (Gimferrer et al., 2011; Kandasamy et al., 2010; Scheinman and Avni, 2009). ChEA3 analysis of CD4-DE cluster 4 showed enrichment for NFAT family member *Nfatc2* (Figure S4B, Table S12), with *Gata3, Cd5, Id3, Cd28*, and *Lef1* all contributing to the enrichment score. In contrast, CD4-DE cluster 7 and CD8-DE cluster 3 showed enrichment for AP-1 transcription factors *Fosb* and *Junb*, NF-kB family members *Rel* and *Nfkb1/2*, and MEK-ERK target *Egr1* (Figure S4B, Table S13). This suggested that all three branches of the TCR signaling pathway participate during the CD4 audition phase, whereas the MEK-ERK and PKC-NF-kB branches, but not the NFAT branch, are active in the later specification of the CD8 lineage. Based on these data, together with the ranking of NFAT in driving early transcriptional differences between lineages (Figure 4F), as well as the dearth of information about the role of NFAT downstream of TCR signaling during positive selection, we chose to focus on the calcineurin-NFAT pathway for functional testing.

NFAT promotes commitment to the CD4 lineage via GATA3

To investigate the influence of calcineurin-NFAT signaling in driving CD4 versus CD8 lineage commitment, we developed an ex vivo neonatal thymic slice culture system (Methods). Since mature SP thymocytes first appear shortly after birth in mice, neonatal thymic slice cultures allowed us to manipulate TCR signaling during a new wave of CD4 and CD8 SP development. We prepared thymic slices from postnatal day 1 mice and cultured slices for up to 96 hours on tissue culture inserts (Figure S5A). We quantified populations of developing thymocytes based upon cell surface marker expression using flow cytometry (Figures 5A and S5B; Methods). As expected, at time point 0 we observed mostly DP thymocytes, whereas the frequencies of more mature populations including CD4+CD8low, CD4+ semimature (CD4+ SM), CD4+ mature (CD4+ Mat), CD4lowCD8+, and CD8+ mature (CD8+ Mat) increased over time in culture (Figure S5C-I). Consistent with previously published results (Saini et al., 2010; Lucas et al., 1993) and our pseudotime analysis, CD8 T cell development was slightly delayed compared to that of CD4 T cells (Figure S5F-I). Together, these observations validate that the neonatal slice system supports the development of both CD4 and CD8 lineage cells, thus providing an experimental time window to manipulate TCR signaling during CD4 and CD8 development with pharmacological inhibitors.

To directly test the involvement of TCR signaling through the calcineurin-NFAT branch in lineage commitment, we inhibited calcineurin activity by adding Cyclosporin A (CsA) (Liu, 1993) to the neonatal slice cultures (Figure 5B). High doses (greater than 200 ng/mL) of CsA treatment resulted in reduced cell viability and an increase in the frequency of DN thymocytes (Figure S5J-K). We therefore used low doses (50, 100, or 200 ng/mL) of CsA, which did not affect cell viability, DN, or CD69+ Sig DP (CD4+CD8+CD69+) cell populations (Figure 5C). Inhibition of calcineurin-NFAT with CsA for 96 hours in neonatal slices from WT mice led to a significant, dose dependent reduction in CD4+CD8low and CD4+ SM thymocytes (Figure 5D-E). Similarly, we observed a reduction in CD4+CD8low thymocytes in neonatal slices from MHCI-/- mice treated with CsA (Figure 5F). In neonatal slice cultures from MHCII-/- mice, the CD4+CD8low population was not impacted by CsA (Figure 5G). We did not observe a decrease in mature CD4+ T cells (Figure 5H), suggesting that many thymocytes in culture had already received sufficient positive selection signals to

complete CD4 SP development prior to CsA addition. Consistently, there was also no significant change in the frequency of CD4lowCD8+ cells and a slight increase in mature CD8+ cells (Figures 5I-J and S5L-M) upon CsA addition. To gain insight into the timing of calcineurin-NFAT signaling in CD4 lineage commitment, we tracked development over time in WT slices in the presence of CsA. We observed a reduction in CD4+CD8low and CD4+ SM cells after 48 hours of culture with CsA. These differences became significant at 72 hours and more pronounced over time (Figure 5K-L), implying that new CD4+ development was blocked in the presence of CsA. Together these data indicate that inhibition of calcineurin-NFAT prevents the emergence of new CD4 T cells, but does not impact the commitment or development of CD8 T cells.

The transcription factor *Gata3* is significantly upregulated in MHCII- versus MHCI-specific thymocytes immediately prior to CD4 lineage commitment (Figure 3C-D), and was implicated by our ChEA3 analyses and prior studies as a target of NFAT (Figure 4F,I) (Gimferrer et al., 2011; Kandasamy et al., 2010; Scheinman and Avni, 2009). Moreover, GATA3 induces the expression of the CD4 master regulator THPOK (Wang et al., 2008). Thus, we hypothesized that CsA treatment prevented CD4 development by interfering with THPOK induction via GATA3. To test this hypothesis, we performed intracellular staining to quantify GATA3 expression in neonatal slice cultures treated with CsA for 48 hours (the time point at which we first observed a significant decrease in CD4+CD8low and CD4+ SM populations with CsA addition). Similar to adult thymocytes (Figure S3C), in thymocytes from neonatal thymic slices cultured for 48 hours, GATA3 protein expression could be detected at the CD69+ Sig DP stage and increased in the CD4+CD8low stage, whereas THPOK was not detectable at the CD69+ Sig DP stage and was not fully upregulated until the CD4+ SM stage (Figure S5N). In cultures with CsA, we observed a significant reduction in GATA3 expression in CD69+ Sig DP, CD4+CD8low, and CD4+ SM cells (Figure 5M-O). Together these data validate predictions from our CITE-seq data, and implicate the calcineurin-NFAT axis as a link between TCR signals downstream of MHCII recognition and commitment to the CD4 rather than the CD8 lineage.

Discussion

In this study, we applied single-cell multi-omic analysis to investigate the development of thymocytes into CD4 and CD8 T cells. By jointly analyzing the transcriptome and surface protein expression of thymocytes from both WT and lineage-restricted mice, we comprehensively defined the continuous changes over maturation in both lineages and linked them directly to intermediate populations traditionally defined by flow cytometry. We identified key lineage-specifying differences in gene and protein expression and determined the relative ordering of single cells along their differentiation trajectory by defining a pseudotime axis based on both the RNA and protein measurements. These data support a sequential model for CD4 versus CD8 T cell lineage commitment in which both MHCII- and MHCI-specific thymocytes initially pass through a CD4 "auditioning" phase that culminates in CD4 lineage commitment for MHCII-specific cells, followed by a CD8 lineage specification phase for MHCI-specific cells (Figure S4A).

The availability of a high-resolution temporal map of gene and protein expression changes throughout positive selection and T cell lineage commitment provides a key resource for understanding mechanism, as illustrated in the current study. We used this resource to pinpoint CD4-fated and CD8-fated thymocytes just prior to differential expression of the lineage-defining master TFs *Zbtb7b* (encoding THPOK) and *Runx3*, and to compare gene expression at this critical stage. In doing so, we identified the TCR/calcium/calcineurin-regulated TF NFAT as a candidate to link differential TCR signaling upon MHCII recognition to upregulation of GATA3, THPOK, and CD4 lineage commitment. We also used this resource

13

to define two distinct temporal waves of TCR signaling during positive selection: an early wave that coincides with the CD4 auditioning phase, is initiated before THPOK induction, and is more sustained in MHCII-specific/CD4 lineage thymocytes compared to MHCI-specific/CD8 lineage thymocytes, and a later wave that overlaps with the CD8 specification stage and is specific for the CD8 lineage. While the inferred activity of some TCR-induced TFs, such as AP-1 and NF-kB, occurred in both early and late TCR signaling waves, NFAT appeared to be active primarily in the early wave, providing a potential link between TCR signals during positive selection and the CD4 T cell fate. NFAT activity has been shown to be required prior to positive selection to render DP thymocytes responsive to TCR triggering (Gallo et al., 2007), but the role of NFAT downstream of TCR signaling during positive selection has remained unknown. Here we addressed this question by a short term, low dose CsA treatment in neonatal thymic slice cultures, using conditions that minimized the potential impact on thymocytes prior to positive selection. We found that pharmacological inhibition of calcineurin-NFAT signaling with CsA prevented new development of mature CD4, but not CD8, thymocytes, and also led to decreased expression of GATA3 at the CD4+CD8low and CD4+ SM stages. Thus, computational analyses of CITE-seq data as well as experimental manipulation point to a role for NFAT downstream of TCR recognition of MHCII in driving CD4 lineage commitment.

These results raise the question, what might lead to differential activation of NFAT in MHCII-specific thymocytes? NFAT activity is regulated by TCR-induced calcium flux, and thymocytes undergoing positive selection experience serial, transient calcium signals (Bhakta et al., 2005; Melichar et al., 2013; Kurd et al,. 2016). While previous studies did not directly compare positive selection via MHCII or MHCI, the estimated duration of calcium signals was approximately 20 minutes for MHCII-specific thymocytes, but less than 5 minutes for MHCI-specific thymocytes, suggesting that positive selection signals generated upon MHCII recognition might lead to greater calcium flux compared to those generated by MHCI recognition. There is also evidence that the greater ability of the CD4 intracellular domain to recruit the TCR-associated tyrosine kinase LCK can promote the CD4 fate (Seong et al., 1992; Itano et al., 1994), and that LCK activity promotes calcium flux downstream of the TCR (Donnadieu et al., 2001). It is tempting to speculate that differences in TCR signaling due to increased LCK recruitment by CD4 could contribute to both increased calcium signals and greater NFAT activation, thus promoting the CD4 lineage choice. In addition to intrinsic differences in TCR signals due to CD4 versus CD8 coreceptor function, loss of the CD8 coreceptor expression during the CD4 auditioning phase also likely contributes to more transient TCR signaling in MHCI- compared to MHCII-specific thymocytes, as predicted by the kinetic signaling model (Singer et al., 2008). Indeed, our data show that MHCII-specific thymocytes experience a more sustained initial wave of TCR signaling compared to that observed in MHCI-thymocytes, and this may also contribute to increased NFAT activation leading to CD4 commitment.

Our data are consistent with the notion that greater NFAT activity downstream of MHCII recognition induces higher expression of GATA3, which would in turn promote THPOK expression and CD4 commitment (Wang et al., 2008). GATA3 was implicated by our ChEA3 analyses as an NFAT target, both in gene sets that were differentially expressed prior to THPOK induction, and in gene clusters associated with the early wave of TCR signaling. We also observed that GATA3 protein expression was reduced upon CsA addition to neonatal thymic slice cultures. In addition, previous in vitro studies implicated NFAT as a positive regulator of GATA3 in thymocytes (Gimferrer et al., 2011), and NFAT can bind to GATA3 cis regulatory sequences and regulates its expression in CD4 T helper type 2 cells (Scheinman et al., 2009). While previous studies have shown that genetic disruption of NFAT function or in vivo treatment with CsA indirectly impairs the ability of thymocytes to activate ERK in response to TCR triggering (Gallo et al.,

2007), this mechanism is unlikely to contribute to the effects of CsA reported here. Impairment of ERK activation in thymocytes was observed after long-term loss of NFAT function, whereas we observed reduced GATA3 expression in signaled DP and CD4+CD8low thymocytes after only 48 hours of drug exposure. In addition, under our experimental conditions, we did not observe substantial impact of CsA on expression of CD69, which is known to be regulated by ERK signaling (d'Ambrosio et al., 1994).

Our identification of calcineurin-NFAT signaling as a driver of CD4 lineage commitment leaves remaining questions about whether an alternative pathway actively drives commitment to the CD8 lineage, or rather that failure to enforce commitment to the CD4 lineage allows CD8 commitment to progress by default. This question has been challenging to address, in part because the precise timing of events for MHCI-specific thymocytes after the CD4 lineage branchpoint remained unclear. Here, we used pseudotime and in silico flow cytometry analyses to clarify the progress of MHCI-specific thymocytes from CD4+CD8low to a late-stage DP population (DP3), and eventually to the CD8 SP stage, revealing a second wave of TCR signaling unique to the CD8 lineage branch. These data are consistent with earlier indications of a prolonged requirement for TCR signaling for CD8 T cell development (Kisielow and Miazek, 1995; Liu and Bosselut, 2004; Au-Yeung et al., 2014; Sinclair and Seddon, 2014), but are at odds with the kinetic signaling model, which invokes a complete loss of TCR signals and an exclusive role for cytokine signals during CD8 lineage specification (Singer et al., 2008). Moreover, our pseudotime data reveal that the late wave of TCR signals after the branchpoint from the CD4 lineage coincides with a gradual increase in RUNX3 expression. This is consistent with earlier evidence that the amount of TCR signaling impacts RUNX3 expression at the DP3 stage of thymic development (Sinclair and Seddon, 2014) and implies that TCR signals may actively promote the CD8 fate via RUNX3 upregulation. The second wave of TCR signaling may also promote survival ((Sinclair and Seddon, 2014) and serve to ensure the elimination of any MHCII-specific thymocytes that failed the CD4 audition phase. In addition to TCR signaling it is likely that other factors, including cytokine signals (Singer et al., 2008) and the downregulation of E protein TFs (HEB and E2A) (Jones-Mason et al., 2012) also contribute to CD8 lineage specification. Importantly, factors that promote CD8 lineage specification may not need to be differentially expressed between the two lineages. This is because MHCII-specific thymocytes at the equivalent developmental stage would already express THPOK, leading to dominant repression of RUNX3 and the CD8 fate in spite of the presence of other CD8-promoting TFs (Taniuchi, 2016). We expect our data to provide a key resource for future studies of additional factors that contribute specifically to CD8 lineage specification.

While in this study we focused our analysis on CD4/CD8 lineage commitment, we anticipate that our approach using paired single-cell RNA and protein expression data could be applied to the analysis of other developmental systems such as the selection of Tregs within the thymus, the commitment of naive T cells to specialized functions within the periphery, and hematopoiesis. The simultaneous measurement of RNA and protein not only allowed us to track the differences in relative timing of RNA and protein expression events, but also enabled the direct connection between multi-omic cell profiles and tangible populations that could be isolated by FACS for further analysis, such as intracellular TF staining or chromatin profiling. We also demonstrated that in silico gating of CITE-seq protein data can inspire gating strategies for fluorescence-based flow cytometry. While differences between sequencing-based and fluorescence-based protein measurements such as noise (e.g., spectral overlap) and sensitivity (e.g., barcode amplification) might limit the direct translation of gate position, large CITE-seq panels could provide a useful platform for screening potential combinations of fluorescence-based markers. The CITE-seq method is currently limited to the measurement of surface proteins (Stoeckius et al., 2017), but

development of emerging methods such as inCITE-seq (Chung et al., 2021) to facilitate the simultaneous measurement of RNA, surface proteins, and large panels of intracellular proteins could greatly enhance the ability to generate hypotheses about molecular pathway activity, gene regulatory networks, and transcription and translation dynamics. Furthermore, since thymocytes actively traverse the thymic cortex and medulla over the course of their development, imaging could provide a valuable dimension to our current understanding of the thymocyte developmental timeline (Germain et al., 2012). Future work that integrates spatial genomic measurements with the transcriptomic and surface protein profiles generated in this study could inform how a cell's micro-environment and physical motility might influence and reflect key aspects of the thymocyte developmental trajectory.

Acknowledgments

Author Contributions

Z.S. led the study with input from E.A.R., A.S., L.L.M., L.K.L., and N.Y. Z.S. and L.K.L. performed CITE-seq experiments. T.H. contributed towards sequencing and data processing of the cDNA and ADT CITE-seq libraries. L.L.M. designed, performed, and analyzed thymic slice and flow cytometry experiments with input from all authors. Z.S. designed and implemented analysis methods with input from all authors. L.L.M. and E.A.R. analyzed flow cytometry data with input from Z.S. Z.S., L.L.M., A.S., N.Y., and E.A.R wrote the manuscript. E.A.R., A.S., and N.Y. supervised the work.

Declaration of Interests

T.H. is an employee of BioLegend Inc. The other authors declare no competing interests.

Methods

CITE-seq on mouse thymocytes

Mice: All animal care and procedures were carried out in accordance with guidelines approved by the Institutional Animal Care and Use Committees at the University of California, Berkeley and at BioLegend, Inc. WT (B6) (C57BL/6, Stock No.: 000664), β2M-/- (B6.129P2-B2m^{tm1Unc}/DcrJ, Stock No.: 002087; referred to as MHCI-/-), OT-I (C57BL/6-Tg(TcraTcrb)1100Mjb/J, Stock No.: 003831), and OT-II (B6.Cg-Tg(TcraTcrb)425Cbn/J, Stock No.: 004194) were obtained from The Jackson Laboratory. MHCII-/- (I-Aβ-/-) mice have been previously described (Grusby et al., 1991). RAG1-/-AND TCRtg mice and RAG1-/-F5 TCRtg were generated by crossing AND TCRtg (B10.Cg-Tg(TcrAND)53Hed/J, Jax Stock No.: 002761; (Kaye et al., 1989)) and F5 TCRtg (C57BL/6-Tg(CD2-TcraF5,CD2-TcrbF5)1Kio; (Mamalaki et al., 1992)) mice with RAG1-/- mice (Rag1-/-B6.129S7-Rag1tm1Mom) as previously described by (Au-Yeung et al., 2014)). All mice used in CITE-seq experiments were females between four and eight weeks of age. Samples are further described in Table S2. Mice were group housed with enrichment and segregated by sex in standard cages on ventilated racks at an ambient temperature of 26 °C and 40% humidity. Mice were kept in a dark/light cycle of 12 h on and 12 h off and given access to food and water ad libitum.

Cell preparation: Mice were sacrificed, and thymi were harvested, placed in RPMI + 10% FBS medium on ice, mechanically dissociated with a syringe plunger, and passed through a 70 μm strainer to generate a single-cell suspension.

Antibody panel preparation: We prepared a panel containing 111 antibodies (TotalSeq-A mouse antibody panel 1, BioLegend, 900003217), which are enumerated in Table S1. Immediately prior to cell staining, we centrifuged the antibody panel for 10 minutes at 14,000 g to remove antibody aggregates. We then performed a buffer exchange on the supernatant using a 50 kDa Amicon spin column (Millipore, UFC505096) following the manufacturer's protocol to transfer antibodies into RPMI + 10% FBS.

Cell sorting: To enrich for positively-selecting thymocytes in MHC-deficient and some WT samples (Table S2), live, single, TCRβ+CD5+ thymocytes were sorted by FACS. We took advantage of the fact that cells were already stained with TotalSeq (oligonucleotide-conjugated) antibodies and therefore designed oligonucleotide-fluorophore conjugates complementary to the TotalSeq barcodes (5'-CACTGAGCTGTGGAA-AlexaFluor488-3' for CD5; 5'-TCCCATAGGATGGAA-AlexaFluor647-3' for TCRb). Prior to cell staining, the TotalSeq antibody panel was mixed with oligonucleotide-fluorophore conjugates in a 1:1.5 molar ratio. This mixture was incubated for 15 minutes at room temperature to allow for oligonucleotide hybridization, and then transferred to ice. Cells were then stained with the antibody/oligonucleotide-fluorophore mixture according to the TotalSeq protocol. Cells were stained, washed, and resuspended in RPMI + 10% FBS to maintain viability. Cells were sorted using a BD FACSAria Fusion (BD Biosciences).

CITE-seq protocol and library preparation: The CITE-seq experiment was performed following the TotalSeq protocol. Cells were stained, washed, and resuspended in RPMI + 10% FBS to maintain viability. We followed the 10X Genomics Chromium Single Cell 3' v3 protocol to prepare RNA and antibody-derived-tag (ADT) libraries (Zheng et al., 2017).

Sequencing and data processing: RNA and ADT libraries were sequenced with either an Illumina NovaSeq S1 or an Illumina NovaSeq S4. Reads were processed with Cell Ranger v.3.1.0 with feature barcoding,

where RNA reads were mapped to the mouse mm10–2.1.0 reference (10X Genomics, STAR aligner (Dobin et al., 2013)) and antibody reads were mapped to known barcodes (Table S1). No read depth normalization was applied when aggregating samples.

CITE-seq data preprocessing

Prior to analysis with totalVI, we performed preliminary quality control and feature selection on the CITE-seq data. Cells with a high percentage of UMIs from mitochondrial genes (> 15% of a cell's total UMI count) were removed. We also removed cells expressing < 200 genes, and retained cells with protein library size between 1,000 and 10,000 UMI counts. We removed cells in which fewer than 70 proteins were detected of the 111 measured in the panel. An initial gene filter removed genes expressed in fewer than four cells. The top 5,000 highly variable genes (HVGs) were selected by the Seurat v3 method (Stuart et al., 2019) as implemented by scVI (Lopez et al., 2018). In addition to HVGs, we also selected genes encoding proteins in the measured antibody panel and a manually selected set of genes of interest. After all filtering, the CITE-seq dataset contained a total of 72,042 cells, 5,125 genes, and 111 proteins.

CITE-seq data analysis with totalVI

totalVI modeling of all CITE-seq data: We ran totalVI on CITE-seq data after filtering (described above), using a 20-dimensional latent space, a learning rate of 0.004, and early stopping with default parameters. Each 10X lane was treated as a batch. When generating denoised gene and protein values, we applied the *transform_batch* parameter (Gayoso et al., 2021) to view all denoised values in the context of WT samples.

Cell annotation: We stratified cells of the thymus into cell types and states based on the totalVI latent space, taking advantage of both RNA and protein information. We first clustered cells in the totalVI latent space with the Scanpy (Wolf et al., 2018) implementation of the Leiden algorithm (Traag et al., 2019) at resolution 0.6, resulting in 18 clusters. We repeated this approach to subcluster cells. We used Vision (DeTomaso et al., 2019) with default parameters for data exploration. Subclusters were manually annotated based on curated lists of cell type markers (Gayoso et al., 2021; Hogquist et al., 2015), resulting in 20 annotated clusters (excluding one cluster annotated as doublets). We visualized the totalVI latent space in two dimensions using the Scanpy (Wolf et al., 2018) implementation of the UMAP algorithm (Becht et al., 2019).

Differential expression testing of annotated cell types: We conducted a one-vs-all differential expression test between all annotated cell types, excluding clusters annotated as doublets or dying cells. We identified cell type markers by filtering for significance (log(Bayes factor) > 2.0 for genes, log(Bayes factor) > 1.0 for proteins), effect size (median log fold change (LFC) > 0.2 for both genes and proteins), and the proportion of expressing cells (detected expression in > 10% of the relevant population for genes), and sorting by the median LFC. For marker visualization, we selected the top four (if existing) differentially expressed genes and proteins per cell type, arranged by the cell type in which the LFC was highest.

totalVI modeling of positively-selecting thymocytes: To further analyze thymocyte populations with a focus on positively-selected cells, we selected the following annotated clusters: Signaled DP, Immature CD4, Immature CD8, Mature CD4, Mature CD8, Interferon signature cells, Negative selection (wave 2), and Treg. With an interest in the variation within thymocyte populations (rather than all cells in the thymus), we selected the top 5,000 HVGs in this subset, as well as genes encoding proteins in the measured antibody panel and a manually selected set of genes of interest. This resulted in a CITE-seq

dataset containing 35,943 cells, 5,108 genes, and 111 proteins. We ran totalVI on this subset dataset and generated denoised values as described above. We performed Leiden clustering and visualized the totalVI latent space in two dimensions using UMAP as described above.

Cell filtering of positively-selecting thymocytes on the CD4/CD8 developmental trajectory: After visualizing the totalVI latent space of the thymocyte subset, we applied additional filters to restrict to cells on the CD4/CD8 developmental trajectory. We used two resolutions of Leiden clustering (0.6 and 1.4) and subclustering as described above to identify and remove clusters of negatively selected cells, Tregs, gamma-delta-like cells, mature cycling cells, and outlier clusters of doublets, interferon signature cells, and CD8-transgenic-specific outlier cells. After filtering, this dataset contained 29,408 cells that were used for downstream analysis. Differential expression testing of positively-selecting thymocytes using pseudotime information is described below.

Pseudotime inference

Pseudotime inference with Slingshot: Slingshot (Street et al., 2018) was selected for pseudotime inference based on its superior performance in a comprehensive benchmarking study (Saelens et al., 2019). Slingshot pseudotime was derived from the UMAP projection of the totalVI latent space. The starting point was assigned to DP cells, and two endpoints were assigned to mature CD4 and CD8 T cells. Slingshot pseudotime derived from the full 20-dimensional totalVI latent space was highly correlated with that from the 2-dimensional space (Figure S2A), supporting our use of the 2D-derived pseudotime values for ease of visualization and analysis.

Lineage assignment: Initial lineage assignment of cells was made on the basis of their genotype (CD4 lineage for MHCI-/-, AND, and OT-II mice, CD8 lineage for MHCII-/-, F5, and OT-I mice, unassigned for WT mice). However, small numbers of cells in MHC-deficient and TCR transgenic mice develop along the alternative lineage (particularly in TCR transgenics that are *Rag* sufficient, which might express an endogenous TCR in addition to the transgenic TCR). We therefore added an additional filter of Slingshot lineage assignment weight > 0.5. Cells with a Slingshot lineage assignment weight of < 0.5 along the expected lineage based on genotype were excluded from the remaining pseudotime-based analysis.

In silico flow cytometry and gating

To perform in silico flow cytometry, totalVI denoised protein counts were log-transformed and visualized in biaxial-style scatter plots. Gates in biaxial plots were determined based on contours of cell density. An approximate alignment of gated populations to pseudotime was generated by identifying thresholds classifying adjacent populations in pseudotime by maximizing the Youden criteria.

Adult thymocyte population analysis with fluorescence-based flow cytometry

Mice: All experiments were approved by the University of California, Berkeley Animal Use and Care Committee. All mice were bred and maintained under pathogen-free conditions in an American Association of Laboratory Animal Care-approved facility at the University of California, Berkeley. WT (C57BL/6, Stock No.: 000664) and β2M-/- (B6.129P2-B2m^{tm1Unc}/DcrJ, Stock No.: 002087) were obtained from The Jackson Laboratory. MHCII-/- (I-Aβ-/-) mice have been previously described (Grusby et al., 1991). For thymocyte population analysis in adult mice, six to eight week- old animals were used. Thymi were analyzed from eight mice per genotype (four male and four female).

Flow cytometry: Thymi were mechanically dissociated into a single-cell suspension, depleted of red blood cells using ACK Lysis Buffer (0.15M NH_4Cl, 1mM KHC_3, 0.1mM Na_2EDTA). Cells were filtered, washed, and counted before being stained with a live/dead stain; Zombie NIR Fixable Viability Kit (Biolegend). Samples were blocked with anti-CD16/32 (2.4G2) and stained with surface antibodies against CD4, CD8, TCRβ, CD5, CD69, and CD127 (IL-7Ra) in FACS buffer (1% BSA in PBS) containing Brilliant Stain Buffer Plus (BD Biosciences). Intracellular staining for GATA3, THPOK, and RUNX3 was performed using the eBioscience FOXP3/Transcription Factor Staining Buffer Set (Thermo Fisher). All antibodies were purchased from BD Biosciences, Biolegend, or eBiosciences. Single-stain samples and fluorescence minus one (FMO) controls were used to establish PMT voltages, gating and compensation parameters. Cells were processed using a BD LSRFortessa or BD LSRFortessa X20 flow cytometer and analyzed using FlowJo software (Tree Star). Gates defining all populations were based on in silico-derived gates for all described proteins with the exception of CD127 in the CD4+ SM, CD8+ SM, CD4+ Mat, and CD8+ Mat populations. In these cases, the CD127 fluorescent antibody did not have comparable sensitivity to the CD127 CITE-seq measurement and was therefore excluded.

Differential expression analysis of positively-selecting thymocytes with totalVI

Testing for temporal features: Temporal features (i.e., features that are differentially expressed over time) were determined by a totalVI one-vs-all DE test within each lineage between binned units of pseudotime. DE criteria (as above) included filters for significance (log(Bayes factor) > 2.0 for genes, log(Bayes factor) > 1.0 for proteins), effect size (median log fold change > 0.2 for both genes and proteins), and the proportion of expressing cells (detected expression in > 5% of the relevant population for genes). Top temporal genes were selected as the unique set among the top three differentially expressed genes per time that were differentially expressed in both lineages.

Testing for differences between lineages: Differences between lineages were determined by a totalVI within-cluster DE test, where clusters were binned units in pseudotime and the condition was lineage assignment (i.e., cells within a given unit of pseudotime were compared between lineages). Criteria for DE were the same as above.

Clustering of differentially expressed genes: To cluster differentially expressed genes into patterns, totalVI denoised gene expression values were standard scaled, reduced dimensions across cells using PCA, and clustered genes using the Leiden algorithm (Traag et al., 2019) as implemented by Scanpy (Wolf et al., 2018). For features differentially expressed between lineages, the genes upregulated within a lineage were clustered according to expression within the lineage in which they were upregulated.

Enrichment of TCR signaling in gene clusters: A hypergeometric test (phyper) was performed to test for enrichment of TCR signaling in differentially-expressed gene clusters. TCR signaling genes were compiled from Netpath (Kandasamy et al., 2010) and a set of genes activated upon stimulation in DP thymocytes (Mingueneau et al., 2012). The background set included all genes considered in DE analysis. P-values were adjusted by the Benjamini-Hochberg procedure.

Transcription factor enrichment analysis

ChEA3 analysis: To perform transcription factor enrichment analysis with ChEA3 (Keenan et al., 2019), we first selected target gene sets as genes differentially upregulated in one lineage relative to the other in each unit of pseudotime, filtered for significance (log(Bayes factor) > 2.0), effect size (median log fold

change > 0.2), and detected expression in > 5% of the population of interest. For each target gene set, TFs were scored for enrichment by the integrated mean ranking across all ChEA3 gene set libraries (MeanRank) based on the top performance of this ranking method (Keenan et al., 2019). ChEA3 analysis on gene clusters was performed as above, but using gene clusters as the target gene set.

Ranking of candidate TFs: To generate an overall ranking of TFs for their likely involvement in CD4/CD8 lineage commitment, we focused on enrichment in the three units of pseudotime prior to master regulator differential expression in each lineage (i.e., in the CD4 lineage, the relevant pseudotime units are 4, 5, and 6, prior to the differential expression of *Zbtb7b* differential expression at pseudotime 7; in the CD8 lineage, the relevant pseudotime units are 5, 6, and 7, prior to the differential expression of *Runx3* at pseudotime 8). We excluded the pseudotime unit containing master regulator differential expression from the ranking, as genes differentially expressed at this time could be the result of the master regulator itself enforcing lineage-specific changes rather than the factors driving initial commitment to a lineage. The pseudotime unit containing master regulator differential expression is included in Figure 4F-G for visualization, but did not contribute to the ranked order of TFs. We also excluded earlier units of pseudotime since these times included very few (< 15) significantly different genes between the lineages. Finally, pseudotime bins in which a TF was not expressed in at least 5% of the population of interest, did not contribute towards that TF's ranking. The overall ranking of candidate driver TFs was then generated by taking the mean of ranks across the relevant pseudotime units.

TCR signaling pathway involvement: TFs were annotated by whether they had a known association with TCR signaling. A list of molecules involved in TCR signaling were curated from the NetPath database of molecules involved in the TCR signaling pathway and the NetPath database of genes transcriptionally upregulated by the TCR signaling pathway (Kandasamy et al., 2010). Additional genes related to TCR signaling were curated from literature sources (Shao et al., 1997; Wong et al., 2014; Lopez-Rodriguez et al., 2015; Hedrick et al., 2013; Wang et al., 2010). TFs were also annotated by whether they were known to target either *Gata3*, *Zbtb7b*, or *Runx3* according to ChEA3 databases (i.e., *Gata3*, *Zbtb7b*, or *Runx3* appeared in the Overlapping Gene list for the TF of interest in any ChEA3 query).

Neonatal thymic slice experiments

Mice: All experiments were approved by the University of California, Berkeley Animal Use and Care Committee. All mice were bred and maintained under pathogen-free conditions in an American Association of Laboratory Animal Care-approved facility at the University of California, Berkeley. WT (C57BL/6, Stock No.: 000664) and β2M-/- (B6.129P2-B2m^{tm1Unc}/DcrJ, Stock No.: 002087) were obtained from The Jackson Laboratory. MHCII-/- (I-Aβ-/-) mice have been previously described (Grusby et al., 1991). For neonatal thymic slice experiments, postnatal day 1 (P1) mice were used.

Thymic slices: We prepared thymic slices from postnatal day 1 mice, a time point that allows us to track a synchronous wave of developing CD4 and CD8 thymocytes since T cells in mice do not develop until birth. Thymic slices were prepared as previously described (Dzhagalov et al., 2012; Ross et al., 2015), with minor modifications to adjust for the smaller size of neonatal thymi compared to those of adults. Thymic lobes were dissected, removed of connective tissue, embedded in 4% low melting point agarose (GTG-NuSieve Agarose, Lonza) and sectioned into 500 μM slices using a vibratome (VT1000S, Leica). Slices were overlaid onto 0.4 μM transwell inserts (Corning, Cat. No.: 353090) and placed in a 6-well tissue culture plate with 1 mL of complete RPMI medium (RPMI-1640 (Corning), 10% FBS (Thermo), 100U/mL penicillin/streptomycin (Gibco), 1X L-glutamine (Gibco), 55μM 2-mercaptoethanol (Gibco). Slices were

cultured for indicated periods of time at 37 °C, 5% CO_2, before being prepared and analyzed by flow cytometry. For neonatal slice cultures containing Cyclosporin A (CsA; Millipore-Sigma, Cat. No.:239835), CsA was serially diluted to indicated concentrations (50-800 ng/mL) and added directly to the culture medium.

Flow cytometry: Thymic slices were mechanically dissociated into a single-cell suspension, then filtered, washed and counted before being stained with a live dead/stain; Propidium Iodine (Biolegend), Ghost Violet 510 (Tonbo), Zombie NIR, or Zombie UV Fixable Viability Kit (Biolegend). Samples were blocked with anti-CD16/32 (2.4G2) and stained with surface antibodies against CD4, CD8, TCRβ, and CD69 in FACS buffer (1% BSA in PBS) containing Brilliant Stain Buffer Plus (BD Biosciences). Intracellular staining for GATA3, RUNX3, and THPOK was performed using the eBioscience FoxP3/Transcription Factor Staining Buffer Set (Thermo Fisher). All antibodies were purchased from BD Biosciences, Biolegend, or eBiosciences. Single-stain samples and fluorescence minus one (FMO) controls were used to establish PMT voltages, gating and compensation parameters. Cells were processed using a BD LSRFortessa or BD LSRFortessa X20 flow cytometer and analyzed using FlowJo software (Tree Star). Note that the DP3 population is difficult to detect in neonatal compared to adult thymocytes samples, therefore we did not include it in our gating strategy.

Statistical analysis: Data were analyzed using Prism software (GraphPad). Comparisons were performed using an unpaired T test, one- or two-way analysis of variance, where indicated in the figure legends. For all statistical models and tests described above, the significance is displayed as follows; ns is not significant, *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Figures: Figures were created using Adobe Illustrator software. Illustrations in Figures 4I, 5B, S5A were created using Biorender.com.

Supplemental Tables

Table S1: Antibodies used in this study.

Table S2: CITE-seq sample information.

Table S3: DE test results for totalVI one-versus-all DE test between annotated thymus populations.

Table S4: Lineage information by genotype.

Table S5: DE test results for totalVI DE test across pseudotime within the CD4 lineage.

Table S6: DE test results for totalVI DE test across pseudotime within the CD8 lineage.

Table S7: DE test results for totalVI DE test within pseudotime and between CD4 and CD8 lineages.

Table S8: Cluster assignments for genes upregulated in the CD4 lineage from the totalVI DE test within pseudotime and between CD4 and CD8 lineages.

Table S9: Cluster assignments for genes upregulated in the CD8 lineage from the totalVI DE test within pseudotime and between CD4 and CD8 lineages.

Table S10: ChEA3 results for the CD4 lineage by pseudotime.

Table S11: ChEA3 test results for the CD8 lineage by pseudotime.

Table S12: ChEA3 test results for the CD4 lineage by gene cluster.

Table S13: ChEA3 test results for the CD8 lineage by gene cluster.

Data and Code Availability

CITE-seq data are being uploaded to GEO. An accession number will be provided once available. Code will be made available prior to publication.

24

References

Au-Yeung, B. B., Melichar, H. J., Ross, J. O., Cheng, D. A., Zikherman, J., Shokat, K. M., … Weiss, A. (2014). Quantitative and temporal requirements revealed for Zap70 catalytic activity during T cell development. Nature Immunology, 15(7), 687–694. https://doi.org/10.1038/ni.2918

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., … Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nature Biotechnology, 37(1), 38–44. https://doi.org/10.1038/nbt.4314

Bhakta, N. R., Oh, D. Y., & Lewis, R. S. (2005). Calcium oscillations regulate thymocyte motility during positive selection in the three-dimensional thymic environment. Nature Immunology, 6(2), 143–151. https://doi.org/10.1038/ni1161

Bosselut, R. (2004). CD4/CD8-lineage differentiation in the thymus: From nuclear effectors to membrane signals. Nature Reviews Immunology. Nature Publishing Group. https://doi.org/10.1038/nri1392

Cannarile, M. A., Lind, N. A., Rivera, R., Sheridan, A. D., Camfield, K. A., Wu, B. B., … Goldrath, A. W. (2006). Transcriptional regulator Id2 mediates CD8+ T cell immunity. Nature Immunology, 7(12), 1317–1325. https://doi.org/10.1038/ni1403

Chakraborty, A. K., & Weiss, A. (2014). Insights into the initiation of TCR signaling. Nature Immunology. Nature Publishing Group. https://doi.org/10.1038/ni.2940

Chan, S. H., Cosgrove, D., Waltzinger, C., Benoist, C., & Mathis, D. (1993). Another view of the selective model of thymocyte selection. Cell, 73(2), 225–236. https://doi.org/10.1016/0092-8674(93)90225-F

Chan, S., Waltzinger, C., Tarakhovsky, A., Benoist, C., & Mathis, D. (1999). An influence of CD5 on the selection of CD4-lineage T cells. European Journal of Immunology, 29(9), 2916–2922. https://doi.org/10.1002/(SICI)1521-4141(199909)29:09<2916::AID-IMMU2916>3.0.CO;2-I

Chèneby, J., Ménétrier, Z., Mestdagh, M., Rosnet, T., Douida, A., Rhalloussi, W., … Ballester, B. (2020). ReMap 2020: A database of regulatory regions from an integrative analysis of Human and Arabidopsis DNA-binding sequencing experiments. Nucleic Acids Research, 48(D1), D180–D188. https://doi.org/10.1093/nar/gkz945

Choi, S., Cornall, R., Lesourne, R., & Love, P. E. (2017, September 1). THEMIS: Two Models, Different Thresholds. Trends in Immunology. Elsevier Ltd. https://doi.org/10.1016/j.it.2017.06.006

Chopp, L. B., Gopalan, V., Ciucci, T., Ruchinskas, A., Rae, Z., Lagarde, M., … Bosselut, R. (2020). An Integrated Epigenomic and Transcriptomic Map of Mouse and Human ab T Cell Development. Immunity, 53(6). https://doi.org/10.1016/j.immuni.2020.10.024

Chung, H., Parkhurst, C. N., Magee, E. M., Phillips, D., Habibi, E., Chen, F., … Regev, A. (2021). Simultaneous single cell measurements of intranuclear proteins and gene expression. BioRxiv, 2021.01.18.427139. https://doi.org/10.1101/2021.01.18.427139

Daley, S. R., Hu, D. Y., & Goodnow, C. C. (2013). Helios marks strongly autoreactive CD4+ T cells in two major waves of thymic deletion distinguished by induction of PD-1 or NF-κB. Journal of Experimental Medicine, 210(2), 269–285. https://doi.org/10.1084/jem.20121458

d'Ambrosio, D., Cantrell, D. A., Frati, L., Santoni, A., & Testi, R. (1994). Involvement of p21ras activation in T cell CD69 expression. European Journal of Immunology, 24(3), 616–620. https://doi.org/10.1002/eji.1830240319

Daniels, M. A., Teixeiro, E., Gill, J., Hausmann, B., Roubaty, D., Holmberg, K., … Palmer, E. (2006). Thymic selection threshold defined by compartmentalization of Ras/MAPK signalling. Nature, 444(7120), 724–729. https://doi.org/10.1038/nature05269

DeTomaso, D., Jones, M. G., Subramaniam, M., Ashuach, T., Ye, C. J., & Yosef, N. (2019). Functional interpretation of single cell similarity maps. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-12235-0

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Donnadieu, E., Lang, V., Bismuth, G., Ellmeier, W., Acuto, O., Michel, F., & Trautmann, A. (2001). Differential Roles of Lck and Itk in T Cell Response to Antigen Recognition Revealed by Calcium Imaging and Electron Microscopy. The Journal of Immunology, 166(9), 5540–5549. https://doi.org/10.4049/jimmunol.166.9.5540

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., … Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57–74. https://doi.org/10.1038/nature11247

Dzhagalov, I. L., Melichar, H. J., Ross, J. O., Herzmark, P., & Robey, E. A. (2012). Two-Photon Imaging of the Immune System. In Current Protocols in Cytometry (Vol. 60, pp. 12.26.1-12.26.20). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471142956.cy1226s60

Egawa, T., & Littman, D. R. (2008). ThPOK acts late in specification of the helper T cell lineage and suppresses Runx-mediated commitment to the cytotoxic T cell lineage. Nature Immunology, 9(10), 1131–1139. https://doi.org/10.1038/ni.1652

Gallo, E. M., Winslow, M. M., Canté-Barrett, K., Radermacher, A. N., Ho, L., McGinnis, L., … Crabtree, G. R. (2007). Calcineurin sets the bandwidth for discrimination of signals during thymocyte development. Nature, 450(7170), 731–735. https://doi.org/10.1038/nature06305

Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K. L., Streets, A., & Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. Nature Methods, 18(3). https://doi.org/10.1038/s41592-020-01050-x

Germain, R. N. (2002). t-cell development and the CD4-CD8 lineage decision. Nature Reviews Immunology. European Association for Cardio-Thoracic Surgery. https://doi.org/10.1038/nri798

Germain, R. N., Robey, E. A., & Cahalan, M. D. (2012, June 29). A Decade of imaging cellular motility and interaction dynamics in the immune system. Science. American Association for the Advancement of Science. https://doi.org/10.1126/science.1221063

Gimferrer, I., Hu, T., Simmons, A., Wang, C., Souabni, A., Busslinger, M., … Alberola-Ila, J. (2011). Regulation of GATA-3 Expression during CD4 Lineage Differentiation. The Journal of Immunology, 186, 3892–3898. https://doi.org/10.4049/jimmunol.1003505

Grusby, M. J., Johnson, R. S., Papaioannou, V. E., & Glimcher, L. H. (1991). Depletion of CD4+ T cells in major histocompatibility complex class II-deficient mice. Science, 253(5026), 1417–1420. https://doi.org/10.1126/science.1910207

Hedrick, S. M., Michelini, R. H., Doedens, A. L., Goldrath, A. W., & Stone, E. L. (2012, September). FOXO transcription factors throughout T cell biology. Nature Reviews Immunology. NIH Public Access. https://doi.org/10.1038/nri3278

Hettmann, T., & Leiden, J. M. (2000). NF-κB Is Required for the Positive Selection of CD8 + Thymocytes . The Journal of Immunology, 165(9), 5004–5010. https://doi.org/10.4049/jimmunol.165.9.5004

Hogquist, K. A., & Jameson, S. C. (2014). The self-obsession of T cells: How TCR signaling thresholds affect fate "decisions" and effector function. Nature Immunology, 15(9), 815–823. https://doi.org/10.1038/ni.2938

Hogquist, K., Xing, Y., Hsu, F.-C., & Shapiro, V. S. (2015). T Cell Adolescence: Maturation Events Beyond Positive Selection. Journal of Immunology, 195(4), 1351–1357. https://doi.org/10.4049/jimmunol.1501050

Hu, Q., Nicol, S. A., Suen, A. Y. W., & Baldwin, T. A. (2012). Examination of thymic positive and negative selection by flow cytometry. Journal of Visualized Experiments, (68). https://doi.org/10.3791/4269

Itano, A., Salmon, P., Kioussis, D., Tolaini, M., Corbella, P., & Robey, E. (1996). The cytoplasmic domain of CD4 promotes the development of CD4 lineage T cells. Journal of Experimental Medicine, 183(3), 731–741. https://doi.org/10.1084/jem.183.3.731

Jimi, E., Strickland, I., Voll, R. E., Long, M., & Ghosh, S. (2008). Differential Role of the Transcription Factor NF-κB in Selection and Survival of CD4+ and CD8+ Thymocytes. Immunity, 29(4), 523–537. https://doi.org/10.1016/j.immuni.2008.08.010

Jones-Mason, M. E., Zhao, X., Kappes, D., Lasorella, A., Iavarone, A., & Zhuang, Y. (2012). E Protein Transcription Factors Are Required for the Development of CD4 + Lineage T Cells. Immunity, 36(3), 348–361. https://doi.org/10.1016/j.immuni.2012.02.010

Kandasamy, K., Sujatha Mohan, S., Raju, R., Keerthikumar, S., Sameer Kumar, G. S., Venugopal, A. K., … Pandey, A. (2010). NetPath: A public resource of curated signal transduction pathways. Genome Biology, 11(1), R3. https://doi.org/10.1186/gb-2010-11-1-r3

Karimi, M. M., Guo, Y., Cui, X., Pallikonda, H. A., Horková, V., Wang, Y. F., … Merkenschlager, M. (2021). The order and logic of CD4 versus CD8 lineage choice and differentiation in mouse thymus. Nature Communications, 12(1), 1–14. https://doi.org/10.1038/s41467-020-20306-w

Kaye, J., Hsu, M. L., Sauron, M. E., Jameson, S. C., Gascoigne, N. R. J., & Hedrick, S. M. (1989). Selective development of CD4+ T cells in transgenic mice expressing a class II MHC-restricted antigen receptor. Nature, 341(6244), 746–749. https://doi.org/10.1038/341746a0

Keenan, A. B., Torre, D., Lachmann, A., Leong, A. K., Wojciechowicz, M. L., Utti, V., … Ma'ayan, A. (2019). ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Research, 47(W1), W212–W224. https://doi.org/10.1093/nar/gkz446

Kernfeld, E. M., Genga, R. M. J., Neherin, K., Magaletta, M. E., Xu, P., & Maehr, R. (2018). A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation. Immunity, 48(6), 1258-1270.e6. https://doi.org/10.1016/j.immuni.2018.04.015

Kisielow, P., & Miazek, A. (1995). Positive selection of T cells: Rescue from programmed cell death and differentiation require continual engagement of the T cell receptor. Journal of Experimental Medicine, 181(6), 1975–1984. https://doi.org/10.1084/jem.181.6.1975

Kurd, N., & Robey, E. A. (2016). T-cell selection in the thymus: A spatial and temporal perspective. Immunological Reviews, 271(1), 114–126. https://doi.org/10.1111/imr.12398

Lavaert, M., Liang, K. L., Vandamme, N., Park, J.-E., Roels, J., Kowalczyk, M. S., … Taghon, T. (2020). Integrated scRNA-Seq Identifies Human Postnatal Thymus Seeding Progenitors and Regulatory Dynamics of Differentiating Immature Thymocytes. Immunity, 52(6). https://doi.org/10.1016/j.immuni.2020.03.019

Liu, J. (1993). FK506 and cyclosporin, molecular probes for studying intracellular signal transduction. Immunology Today, 14(6), 290–295. https://doi.org/10.1016/0167-5699(93)90048-P

Liu, X., & Bosselut, R. (2004). Duration of TCR signaling controls CD4-CD8 lineage differentiation in vivo. Nature Immunology, 5(3), 280–288. https://doi.org/10.1038/ni1040

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nature Methods, 15(12), 1053–1058. https://doi.org/10.1038/s41592-018-0229-2

López-Rodríguez, C., Aramburu, J., & Berga-Bolaños, R. (2015). Transcription factors and target genes of pre-TCR signaling. Cellular and Molecular Life Sciences, 72(12), 2305–2321. https://doi.org/10.1007/s00018-015-1864-8

Lucas, B., & Germain, R. N. (1996). Unexpectedly complex regulation of CD4/CD8 coreceptor expression supports a revised model for CD4+CD8+ thymocyte differentiation. Immunity, 5(5), 461–477. https://doi.org/10.1016/S1074-7613(00)80502-6

Lucas, B., Vasseur, F., & Penit, C. (1993). Normal Sequence of Phenotypic Transitions in One Cohort of 5=Bromo=2'-Deoxyuridine-Pulse-Labeled Thymocytes Correlation with T Cell Receptor Expression' (Vol. 151). Retrieved from http://www.jimmunol.org/

Lundberg, K., Heath, W., Köntgen, F., Carbone, F. R., & Shortman, K. (1995). Intermediate steps in positive selection: Differentiation of CD4+8int TCRint thymocytes into CD4−8+TCRhi Thymocytes. Journal of Experimental Medicine, 181(5), 1643–1651. https://doi.org/10.1084/jem.181.5.1643

Lutes, L. K., Steier, Z., McIntyre, L. L., Pandey, S., Kaminski, J., Hoover, A. R., … Robey, E. A. (2021). T cell self-reactivity during thymic development dictates the timing of positive selection. ELife, 10. https://doi.org/10.7554/eLife.65435

Malissen, B., Grégoire, C., Malissen, M., & Roncagalli, R. (2014). Integrative biology of T cell activation. Nature Immunology. Nature Publishing Group. https://doi.org/10.1038/ni.2959

Mamalaki, C., Norton, T., Tanaka, Y., Townsend, A. R., Chandler, P., Simpson, E., & Kioussis, D. (1992). Thymic depletion and peripheral activation of class I major histocompatibility complex-restricted T cells by soluble peptide in T-cell receptor transgenic mice. Proceedings of the National Academy of Sciences of the United States of America, 89(23), 11342–11346. https://doi.org/10.1073/pnas.89.23.11342

Marodon, G., & Rocha, B. (1994). Generation of mature T cell populations in the thymus: CD4 or CD8 down-regulation occurs at different stages of thymocyte differentiation. European Journal of Immunology, 24(1), 196–204. https://doi.org/10.1002/eji.1830240131

Matechak, E. O., Killeen, N., Hedrick, S. M., & Fowlkes, B. J. (1996). MHC class II-specific T cells can develop in the CD8 lineage when CD4 is absent. Immunity, 4(4), 337–347. https://doi.org/10.1016/S1074-7613(00)80247-2

McNeil, L. K., Starr, T. K., & Hogquist, K. A. (2005). A requirement for sustained ERK signaling during thymocyte positive selection in vivo. Proceedings of the National Academy of Sciences of the United States of America, 102(38), 13574–13579. https://doi.org/10.1073/pnas.0505110102

Melichar, H. J., Ross, J. O., Herzmark, P., Hogquist, K. A., & Robey, E. A. (2013). Distinct Temporal Patterns of T Cell Receptor Signaling During Positive Versus Negative Selection in Situ. Science Signaling, 6(297), ra92–ra92. https://doi.org/10.1126/scisignal.2004400

Mingueneau, M., Jiang, W., Feuerer, M., Mathis, D., & Benoist, C. (2012). Thymic negative selection is functional in NOD mice. Journal of Experimental Medicine, 209(3), 623–637. https://doi.org/10.1084/jem.20112593

Mingueneau, M., Kreslavsky, T., Gray, D., Heng, T., Cruse, R., Ericson, J., … Turley, S. (2013). The transcriptional landscape of αβ T cell differentiation. Nature Immunology, 14(6), 619–632. https://doi.org/10.1038/ni.2590

Moran, A. E., Holzapfel, K. L., Xing, Y., Cunningham, N. R., Maltzman, J. S., Punt, J., & Hogquist, K. A. (2011). T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. Journal of Experimental Medicine, 208(6), 1279–1289. https://doi.org/10.1084/jem.20110308

Muroi, S., Naoe, Y., Miyamoto, C., Akiyama, K., Ikawa, T., Masuda, K., … Taniuchi, I. (2008). Cascading suppression of transcriptional silencers by ThPOK seals helper T cell fate. Nature Immunology, 9(10), 1113–1121. https://doi.org/10.1038/ni.1650

Navarro, M. N., & Cantrell, D. A. (2014, August 19). Serine-threonine kinases in TCR signaling. Nature Immunology. Nature Publishing Group. https://doi.org/10.1038/ni.2941

Park, J. E., Botting, R. A., Conde, C. D., Popescu, D. M., Lavaert, M., Kunz, D. J., … Teichmann, S. A. (2020). A cell atlas of human thymic development defines T cell repertoire formation. Science, 367(6480). https://doi.org/10.1126/science.aay3224

Perera, E. M., Bao, Y., Kos, L., & Berkovitz, G. (2010). Structural and functional characterization of the mouse tescalcin promoter. Gene, 464(1–2), 50–62. https://doi.org/10.1016/J.GENE.2010.06.002

Ross, J. O., Melichar, H. J., Au-Yeung, B. B., Herzmark, P., Weiss, A., & Robey, E. A. (2014). Distinct phases in the positive selection of CD8+ T cells distinguished by intrathymic migration and T-cell receptor signaling patterns. Proceedings of the National Academy of Sciences of the United States of America, 111(25). https://doi.org/10.1073/pnas.1408482111

Ross, J. O., Melichar, H. J., Halkias, J., & Robey, E. A. (2015). Studying T cell development in thymic slices. In T-Cell Development: Methods and Protocols (Vol. 1323, pp. 131–140). Springer New York. https://doi.org/10.1007/978-1-4939-2809-5_11

Saelens, W., Cannoodt, R., Todorov, H., & Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nature Biotechnology, 37(5), 547–554. https://doi.org/10.1038/s41587-019-0071-9

Saini, M., Sinclair, C., Marshall, D., Tolaini, M., Sakaguchi, S., & Seddon, B. (2010). Regulation of Zap70 expression during thymocyte development enables temporal separation of CD4 and CD8 repertoire selection at different signaling thresholds. Science Signaling, 3(114), ra23–ra23. https://doi.org/10.1126/scisignal.2000702

Scheinman, E. J., & Avni, O. (2009). Transcriptional regulation of Gata3 in T helper cells by the integrated activities of transcription factors downstream of the interleukin-4 receptor and T cell receptor. Journal of Biological Chemistry, 284(5), 3037–3048. https://doi.org/10.1074/jbc.M807302200

Seong, R. H., Chamberlain, J. W., & Parnes, J. R. (1992). Signal for T-cell differentiation to a CD4 cell lineage is delivered by CD4 transmembrane region and/or cytoplasmic tail. Nature, 356(6371), 718–720. https://doi.org/10.1038/356718a0

Shao, H., Kono, D. H., Chen, L. Y., Rubin, E. M., & Kaye, J. (1997). Induction of the early growth response (Egr) family of transcription factors during thymic selection. Journal of Experimental Medicine, 185(4), 731–744. https://doi.org/10.1084/jem.185.4.731

Sharp, L. L., Schwarz, D. A., Bott, C. M., Marshall, C. J., & Hedrick, S. M. (1997). The influence of the MAPK pathway on T cell lineage commitment. Immunity, 7(5), 609–618. https://doi.org/10.1016/S1074-7613(00)80382-9

Sinclair, C., & Seddon, B. (2014). Overlapping and Asymmetric Functions of TCR Signaling during Thymic Selection of CD4 and CD8 Lineages. The Journal of Immunology, 192(11), 5151–5159. https://doi.org/10.4049/jimmunol.1303085

Singer, A., Adoro, S., & Park, J. H. (2008, October). Lineage fate and intense debate: Myths, models and mechanisms of CD4- versus CD8-lineage choice. Nature Reviews Immunology. Nat Rev Immunol. https://doi.org/10.1038/nri2416

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., … Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nature Methods, 14(9), 865–868. https://doi.org/10.1038/nmeth.4380

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., … Dudoit, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics, 19(1), 1–16. https://doi.org/10.1186/s12864-018-4772-0

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., … Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell, 177(7), 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Taniuchi, I. (2016). Views on helper/cytotoxic lineage choice from a bottom-up approach. Immunological Reviews, 271(1), 98–113. https://doi.org/10.1111/imr.12401

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports, 9(1), 1–12. https://doi.org/10.1038/s41598-019-41695-z

Vacchio, M. S., & Bosselut, R. (2016). Function Transcriptional Circuitry To Control Their Lineage Commitment + CD8 – + in the Thymus: How T Cells Recycle the CD4 What Happens in the Thymus Does Not Stay. J Immunol References, 196, 4848–4856. https://doi.org/10.4049/jimmunol.1600415

Wang, L., Wildt, K. F., Zhu, J., Zhang, X., Feigenbaum, L., Tessarollo, L., … Bosselut, R. (2008). Distinct functions for the transcription factors GATA-3 and ThPOK during intrathymic differentiation of CD4+ T cells. Nature Immunology, 9(10), 1122–1130. https://doi.org/10.1038/ni.1647

Wang, L., Xiong, Y., & Bosselut, R. (2010, October 1). Tenuous paths in unexplored territory: From T cell receptor signaling to effector gene expression during thymocyte selection. Seminars in Immunology. Academic Press. https://doi.org/10.1016/j.smim.2010.04.013

Weist, B. M., Kurd, N., Boussier, J., Chan, S. W., & Robey, E. A. (2015). Thymic regulatory T cell niche size is dictated by limiting IL-2 from antigen-bearing dendritic cells and feedback competition. Nature Immunology, 16(6), 635–641. https://doi.org/10.1038/ni.3171

Wilkinson, B., & Kaye, J. (2001). Requirement for sustained MAPK signaling in both CD4 and CD8 lineage commitment: A threshold model. Cellular Immunology, 211(2), 86–95. https://doi.org/10.1006/cimm.2001.1827

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. Genome Biology. https://doi.org/10.1186/s13059-017-1382-0

Wong, W. F., Looi, C. Y., Kon, S., Movahed, E., Funaki, T., Chang, L. Y., … Kohu, K. (2014). T-cell receptor signaling induces proximal Runx1 transactivation via a calcineurin-NFAT pathway. European Journal of Immunology, 44(3), 894–904. https://doi.org/10.1002/eji.201343496

Xing, Y., Wang, X., Jameson, S. C., & Hogquist, K. A. (2016). Late stages of T cell maturation in the thymus involve NF-κB and tonic type i interferon signaling. Nature Immunology, 17(5), 565–573. https://doi.org/10.1038/ni.3419

Xiong, Y., & Bosselut, R. (2012, April 1). CD4-CD8 differentiation in the thymus: Connecting circuits and building memories. Current Opinion in Immunology. Elsevier Current Trends. https://doi.org/10.1016/j.coi.2012.02.002

Yan, Y., Zhang, G. X., Williams, M. S., Carey, G. B., Li, H., Yang, J., … Xu, H. (2012). TCR stimulation upregulates MS4a4B expression through induction of AP-1 transcription factor during T cell activation. Molecular Immunology, 52(2), 71–78. https://doi.org/10.1016/j.molimm.2012.04.011

Yasutomo, K., Doyle, C., Miele, L., & Germain, R. N. (2000). The duration of antigen receptor signalling determines CD4+ versus CD8+ T-cell lineage fate. Nature, 404(6777), 506–510. https://doi.org/10.1038/35006664

Zamisch, M., Tian, L., Grenningloh, R., Xiong, Y., Wildt, K. F., Ehlers, M., … Bosselut, R. (2009). The transcription factor Ets1 is important for CD4 repression and Runx3 up-regulation during CD8 T cell differentiation in the thymus. Journal of Experimental Medicine, 206(12), 2685–2699. https://doi.org/10.1084/jem.20092024

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications, 8, 14049. https://doi.org/10.1038/ncomms14049

Zhou, W., Yui, M. A., Williams, B. A., Yun, J., Wold, B. J., Cai, L., & Rothenberg, E. V. (2019). Single-Cell Analysis Reveals Regulatory Gene Expression Dynamics Leading to Lineage Commitment in Early T Cell Development. Cell Systems, 9(4), 321-337.e9. https://doi.org/10.1016/j.cels.2019.09.008
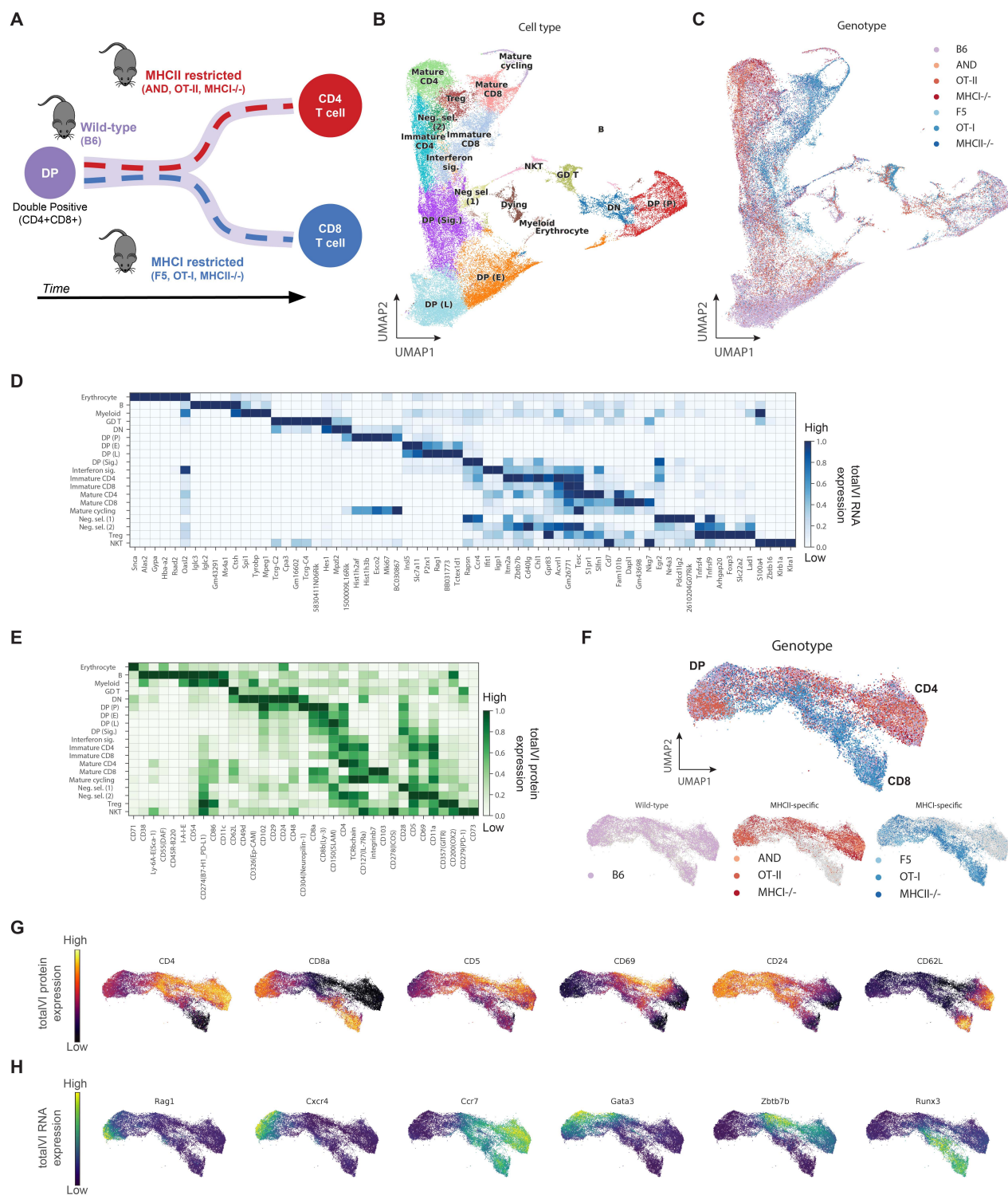
Figures, Titles, and Legends

Figure 1

Figure 1: A joint transcriptomic and surface protein atlas of thymocyte development in wild-type and lineage-restricted mice. (A) Schematic representation of thymocyte developmental trajectories in WT, MHCII-restricted (MHCI-/-, OT-II TCRtg and AND TCRtg), and MHCI-restricted (MHCII-/-, OT-I TCRtg, and F5 TCRtg) mice used in CITE-seq experiments. **(B, C)** UMAP plots of the totalVI latent space from all thymocyte CITE-seq data labeled by **(B)** cell type annotation and **(C)** mouse genotype. **(D, E)** Heatmaps of markers derived from totalVI one-vs-all differential expression test between cell types for **(D)** RNA and **(E)** proteins. Values are totalVI denoised expression. **(F)** UMAP plots of the totalVI latent space from positively-selected thymocytes with cells labeled by mouse genotype. **(G, H)** UMAP plots of the totalVI latent space from positively-selected thymocytes. **(G)** Cells colored by totalVI denoised expression of protein markers of lineage (CD4, CD8a), TCR signaling (CD5, CD69), and maturation (CD24, CD62L). **(H)** Cells colored by totalVI denoised expression of RNA markers of TCR recombination (*Rag1*), thymic location (chemokine receptors *Cxcr4*, *Ccr7*), and lineage regulation (transcription factors *Gata3*, *Zbtb7b*, *Runx3*). GD T, gamma-delta T; DN, double negative; DP (P), double positive proliferating; DP (E), DP early; DP (L), DP late; DP (Sig.), DP signaled; Interferon sig., interferon signature; Neg. sel. (1), negative selection wave 1; Neg. sel. (2), negative selection wave 2; Treg, regulatory CD4 T cell; NKT, natural killer T.
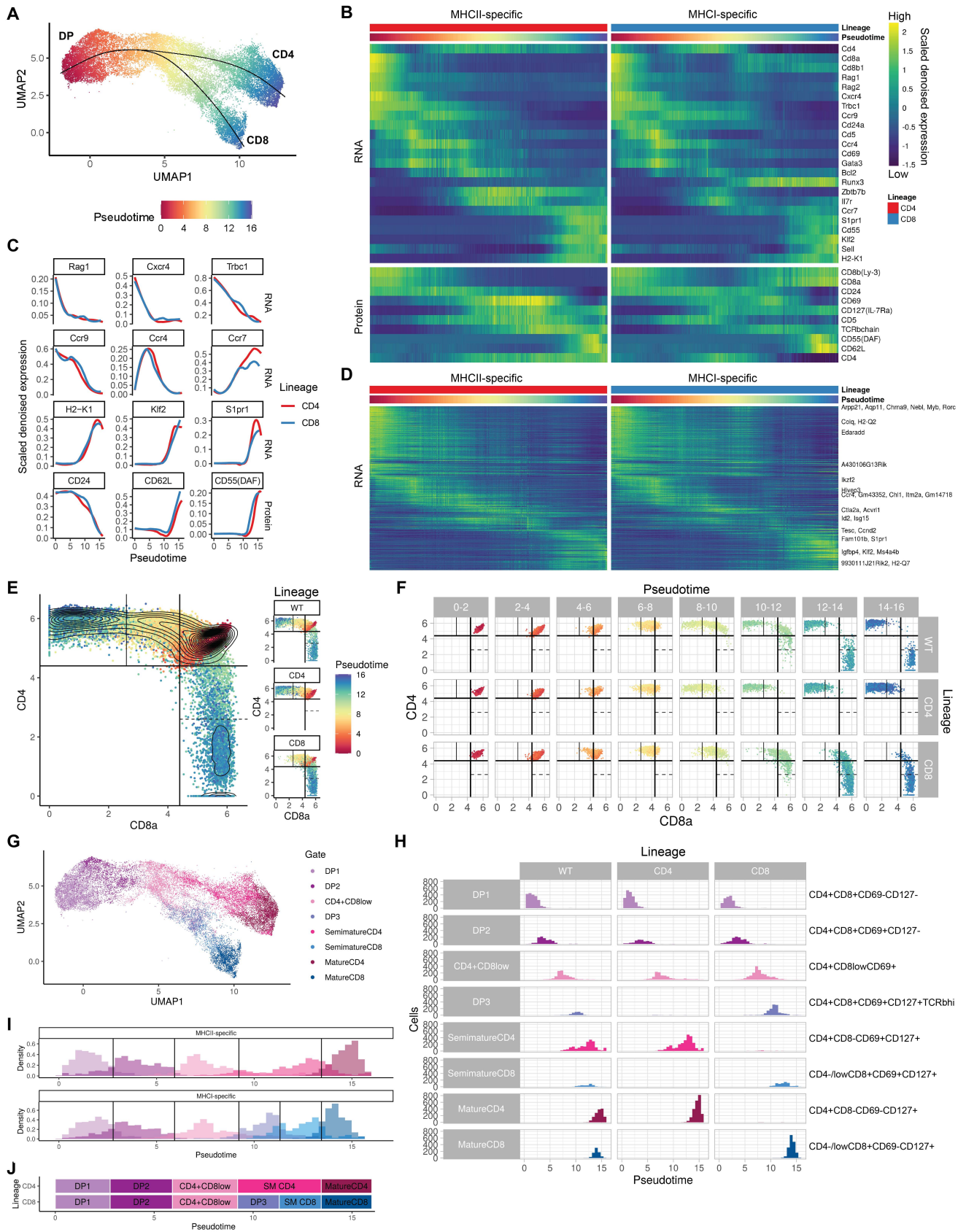
Figure 2

Figure 2: Pseudotime inference captures continuous maturation trajectory and clarifies intermediate thymocyte stages. (A) UMAP plot of the totalVI latent space from positively-selected thymocytes with cells colored by Slingshot pseudotime and smoothed curves representing the CD4 and CD8 lineages. **(B)** Heatmap of RNA (top) and protein (bottom) markers of thymocyte development over pseudotime in the CD4 and CD8 lineages. Features are colored by totalVI denoised expression, scaled per row, and sorted by peak expression in the CD4 lineage. Pseudotime axis is the same as in (A). **(C)** Expression of features in the CD4 and CD8 lineages that vary over pseudotime. Features are totalVI denoised expression values scaled per feature and smoothed by loess curves. **(D)** Heatmap of all RNA differentially expressed over pseudotime in any lineage. Features are scaled and ordered as in (B). Labeled genes are highly differentially expressed over time (Methods). **(E)** In silico flow cytometry plots of log(totalVI denoised expression) of CD8a and CD4 from positively-selected thymocytes (left) and the same cells separated by lineage (right). Cells are colored by pseudotime. Gates were determined based on contours of cell density. **(F)** In silico flow cytometry plot of data as in (E) separated by lineage and pseudotime. **(G)** UMAP plot of the totalVI latent space from positively-selected thymocytes with cells colored by gate. Cells were computationally grouped into eight gates using CD4, CD8a, CD69, CD127(IL-7Ra), and TCRβ. **(H)** Histograms of cells separated by lineage and gate with cells colored by gate as in (G). **(I)** Stacked histograms of gated populations in MHCII-specific (top) and MHCI-specific (bottom) thymocytes, with thresholds classifying gated populations over pseudotime (Methods). **(J)** Schematic timeline aligns pseudotime with gated populations, with population timing determined as in (I).
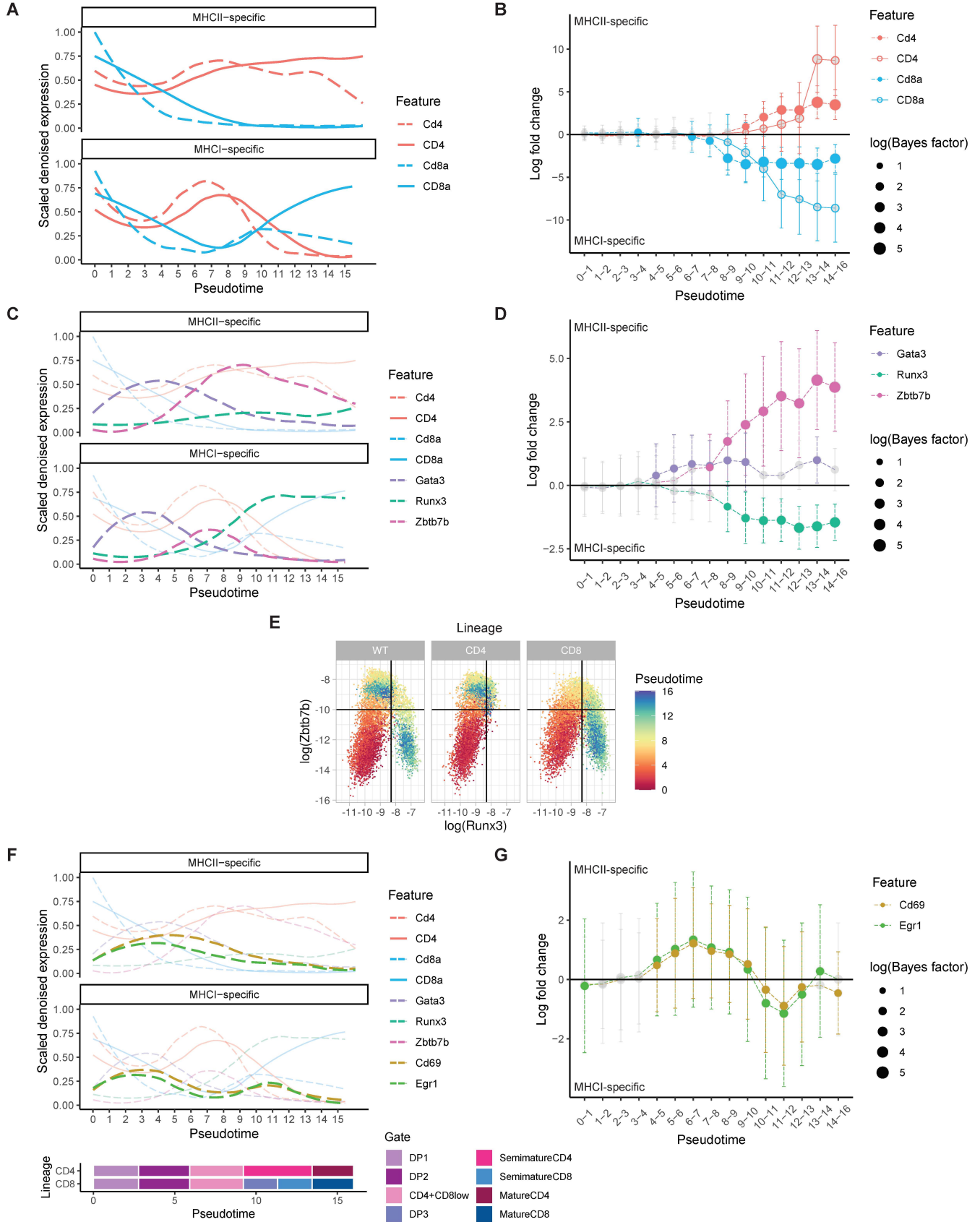
Figure 3

Figure 3: Paired measurements of RNA and protein reveal the timing of major events in CD4/CD8 lineage commitment. (A) Expression of coreceptor RNA (dashed) and protein (solid) over pseudotime in the CD4 (MHCII-specific) and CD8 (MHCI-specific) lineages. Features are totalVI denoised expression values scaled per feature and smoothed by loess curves. **(B)** Differential expression over pseudotime between CD4 and CD8 lineages for features in (A). Non-significant differences are gray, significant RNA results are filled circles, and significant protein results are open circles. Size of the circle indicates log(Bayes factor). Error bars indicate the totalVI-computed standard deviation of the median log fold change. **(C)** Expression over pseudotime as in (A), overlaying RNA expression of key transcription factors. **(D)** Differential expression over pseudotime as in (B) for features in (C). **(E)** In silico flow cytometry plots of log(totalVI denoised expression) of *Runx3* and *Zbtb7b* from positively-selected thymocytes separated by lineage and colored by pseudotime. **(F)** Expression over pseudotime as in (C), overlaying RNA expression of TCR signaling response molecules (*Cd69* and *Egr1*). **(G)** Differential expression over pseudotime as in (B) for features in (F). Schematic timeline aligns pseudotime with gated populations (see Figure 2J).
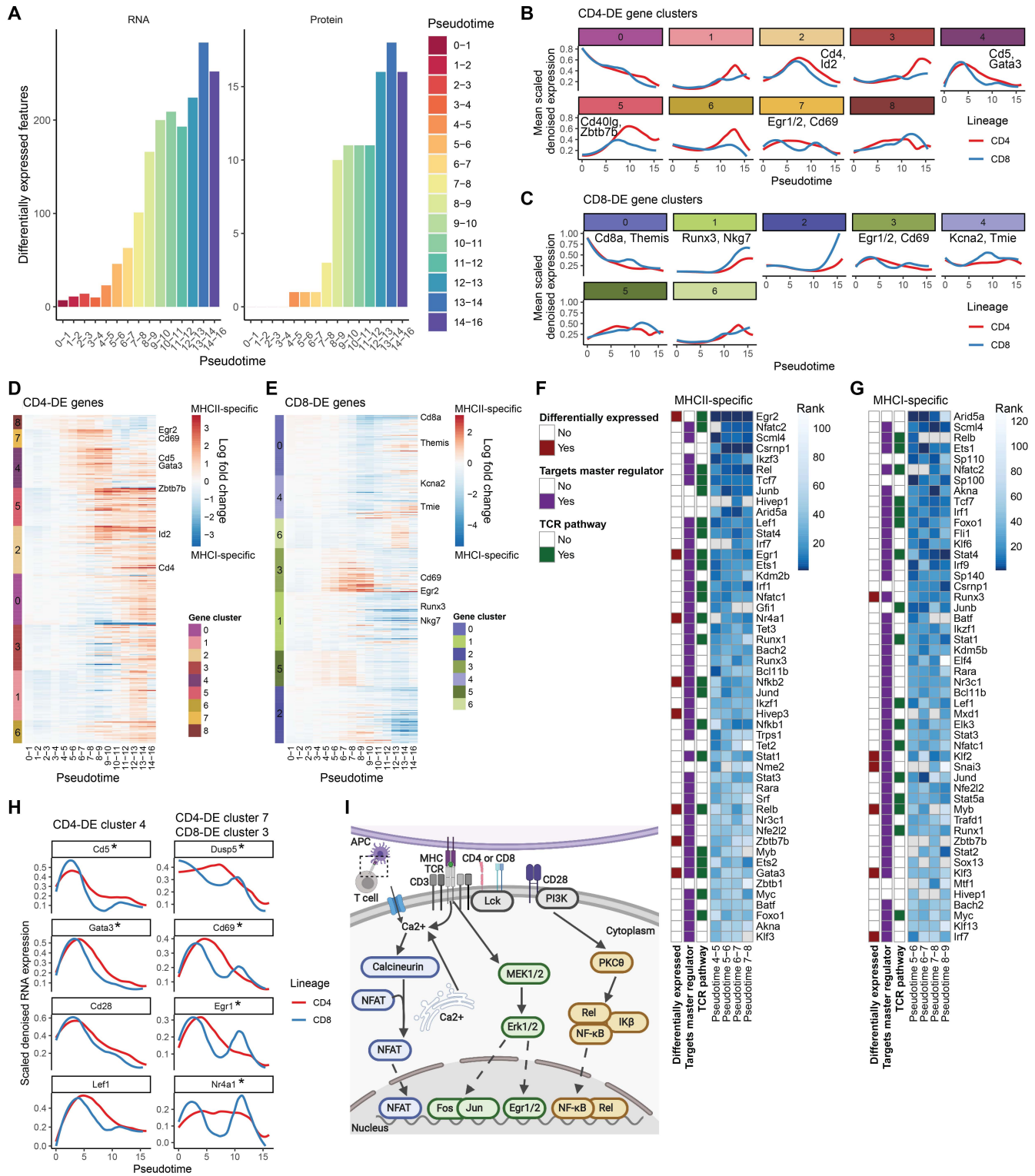
Figure 4

Figure 4: Emergence of differences between CD4 and CD8 lineages implicate putative drivers of lineage commitment. (A) Number of differentially expressed features between the CD4 and CD8 lineages across pseudotime (Methods). **(B)** Genes (RNA) upregulated in the CD4 lineage relative to the CD8 lineage scaled per gene and clustered by the Leiden algorithm according to expression in the CD4 lineage. Expression over pseudotime per cluster is displayed as the mean of scaled totalVI denoised expression per gene for genes in a cluster, smoothed by loess curves. **(C)** Same as (B), but for genes upregulated in the CD8 lineage relative to the CD4 lineage, clustered according to CD8 lineage expression. **(D)** totalVI median log fold change over pseudotime of genes upregulated in the CD4 lineage relative to the CD8 lineage. Genes are grouped by cluster in (B). Clusters are ordered by their average highest magnitude fold change. **(E)** totalVI median log fold change over pseudotime of genes downregulated in the CD4 lineage relative to the CD8 lineage (i.e., upregulated in the CD8 lineage). Genes are grouped by cluster in (C). Clusters are ordered by their average highest magnitude fold change. **(F)** Transcription factor enrichment analysis by ChEA3 for CD4-lineage-specific differentially expressed genes. TFs are ranked by mean enrichment in the three pseudotime bins prior to *Zbtb7b* differential expression (between pseudotime 4-7; pseudotime 7-8 is for visualization and does not contribute to ranking). Gray indicates a gene detected in less than 5% of cells in the relevant population. "Differentially expressed" indicates significant upregulation in at least one of the relevant time bins. "Targets master regulator" indicates a TF that targets either *Gata3*, *Runx3*, or *Zbtb7b* in ChEA3 databases. "TCR pathway" indicates membership in NetPath TCR Signaling Pathway, genes transcriptionally upregulated by TCR signaling, or genes with literature support for TCR pathway membership (Methods). **(G)** Same as in (F), but for the CD8 lineage, with ranking by mean enrichment in the three pseudotime bins prior to *Runx3* differential expression (between pseudotime 5-8; pseudotime 8-9 is for visualization and does not contribute to ranking). **(H)** Expression over pseudotime of selected TCR target genes. All genes are differentially expressed between the two lineages (with cluster membership indicated above) and are putative targets of *Nfatc2*, according to one or more of the ChEA3 analyses. Genes indicated with (*) are differentially expressed during the time windows used for ranking in (F-G), and are therefore included in the target gene sets used for the respective ChEA3 analysis. totalVI denoised expression values are scaled per gene and smoothed by loess curves. **(I)** Schematic of the three major branches of the TCR signaling pathway: calcineurin-NFAT (blue), ERK-MAPK (green), and PKC-NF-kB (orange).
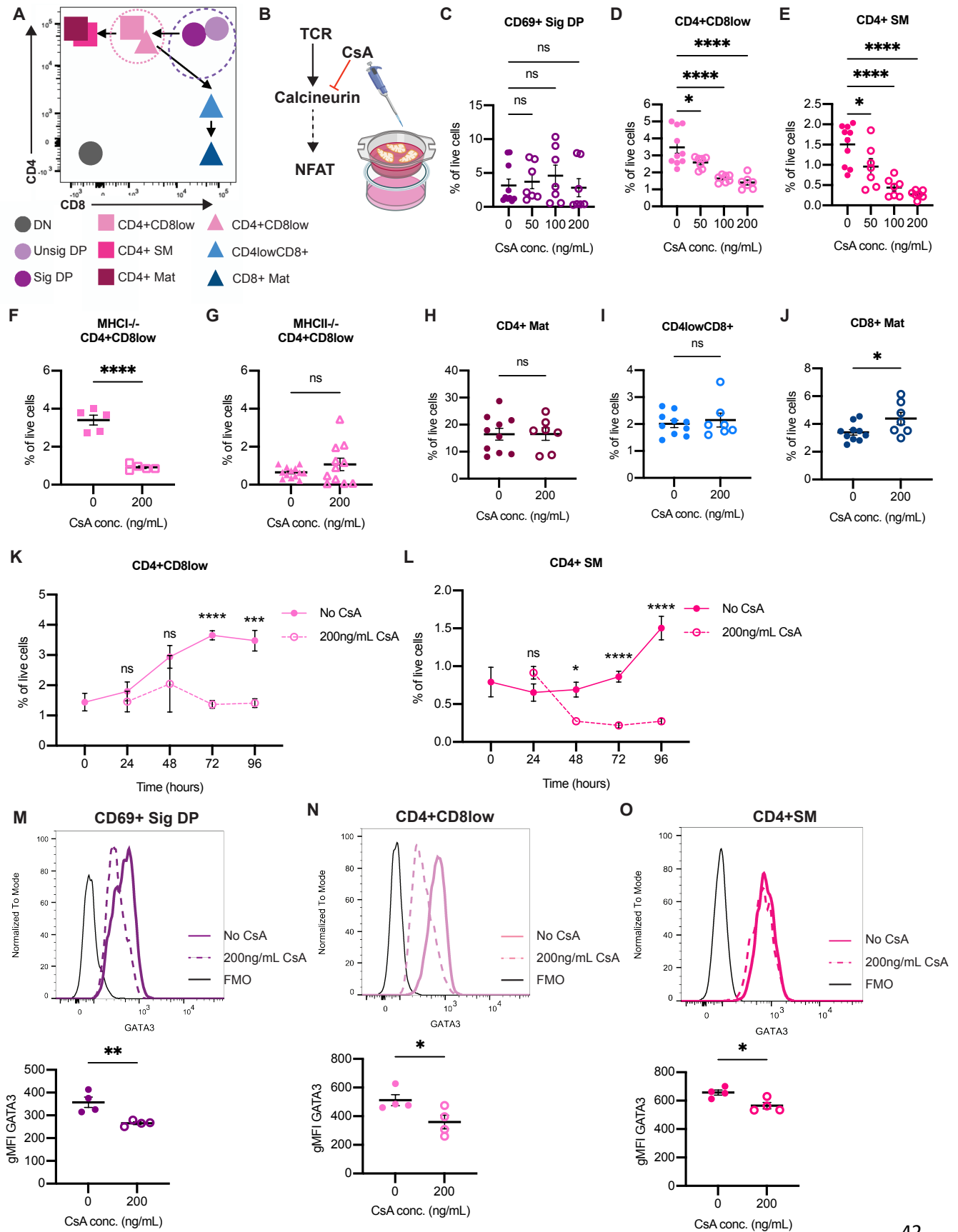
Figure 5

Figure 5: Inhibition of calcineurin-NFAT in neonatal thymic slice cultures impairs CD4, but not CD8 development. (A) Schematic showing populations of thymocytes quantified in neonatal thymic slice cultures; thymocytes were categorized into eight populations: double negative (DN; CD4-CD8-), unsignaled double positive (Unsig DP; CD4+CD8+CD69-), signaled double positive (Sig DP; CD4+CD8+CD69+), CD4+CD8low (CD4+CD8low; CD4+CD8lowTCRβ+), CD4+ semimature (CD4+ SM; CD4+CD8-TCRβhiCD69+), CD4+ mature (CD4+ Mat; CD4+CD8-TCRβhiCD69-), CD4lowCD8+ (CD4lowCD8+; CD4lowCD8+TCRβ+), CD8+ mature (CD8+ Mat; CD4-CD8+TCRβhiCD69-). For flow cytometry gating strategy, see Figure S5B. Note that the DP3 population is difficult to detect in neonatal compared to adult thymocytes samples, therefore we did not include it in our gating strategy. **(B)** Experimental overview of neonatal thymic slices cultured with a calcineurin inhibitor, Cyclosporin A (CsA). Postnatal day 1 (P1) thymic slices were harvested from mice and cultured with or without CsA at the indicated concentrations. Thymic slices were collected at indicated time points and analyzed via flow cytometry to quantify cell populations. Illustrations in (B) were created using Biorender.com. **(C-E)** Frequency (% of live cells) of **(C)** CD69+ Sig DP, **(D)** CD4+CD8low, and **(E)** CD4+ SM cells in neonatal slices from WT mice following culture in medium alone (No CsA; filled symbols) or with indicated concentrations of CsA (open symbols) for 96 hours. **(F-G)** Frequency (% of live cells) of CD4+CD8low cells in slices from **(F)** MHCI-/- (MHCII-specific; squares) or **(G)** MHCII-/- (MHCI-specific; triangles) mice following culture in medium alone (No CsA) or with 200ng/mL CsA for 96 hours. **(H-J)** Frequency (% of live cells) of **(H)** CD4+ Mat, **(I)** CD4lowCD8+, and **(J)** CD8+ Mat cells from WT slices cultured in medium alone or with 200ng/mL CsA. **(K-L)** Frequency of **(K)** CD4+CD8low and **(L)** CD4+ SM cells after 0, 24, 48, 72 and 96-hours of culture in medium alone or with 200ng/mL CsA. **(M-O)** Histograms displaying GATA3 expression (top) and quantification of geometric mean fluorescent intensity (gMFI) of GATA3 (bottom) detected by intracellular flow cytometry staining in **(M)** CD69+ Sig DP, **(N)** CD4+CD8low, and **(O)** CD4+ SM cells after 48 hours of culture in medium alone (No CsA; solid colored line/symbol) or with 200ng/mL CsA (dashed line/open symbol). Each symbol on the graphs represents a thymic slice. For (C-E), data is compiled from 3 independent experiments with WT slices. Data was analyzed using an ordinary one-way ANOVA. For (F-J), data is compiled from 2 independent experiments with MHCI-/- slices, 5 independent experiments with MHCII-/- slices, and 3 independent experiments with WT slices. Data was analyzed using an unpaired t test. In graphs (K and L), data is compiled from 9 independent experiments with WT slices. Data are displayed as the mean ± standard error of the mean (SEM). For slices cultured with no CsA for 0 hours n=6, 24 hours n=9, 48 hours n=10, 72 hours n=22, 96 hours n=10. For slices cultured with 200ng/mL CsA for 24 hours n=6, 48 hours n=7, 72 hours n=14, 96 hours n=7. Data was analyzed using an ordinary two-way ANOVA with multiple comparisons. For (M-O), graphs display representative data from 2 independent experiments with WT slices. Positive GATA3 staining was determined using the FMO control. Data was analyzed using an unpaired t test. NS is not significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Supplemental Figures, Titles, and Legends

Figure S1

Figure S1: Sorting and characterization of positively-selecting thymocytes. (A) Representative FACS plots displaying gating strategy to sort thymocytes for CITE-seq. Cell populations were gated and sorted to include lymphocytes, exclude forward scatter doublets, include Ghost Dye Violet 510 Live/Dead stain negative (live cells), then on TCRβ+CD5+ to enrich for cells that were positively-selecting. **(B-C)** Heatmaps of manually selected cell type markers for **(B)** RNA and **(C)** proteins. Values are totalVI denoised expression. **(D)** UMAP plot of totalVI latent space from positively-selected thymocytes before filtering indicating annotated populations that were retained (positively-selecting thymocytes) or removed (all other populations) from downstream analysis.
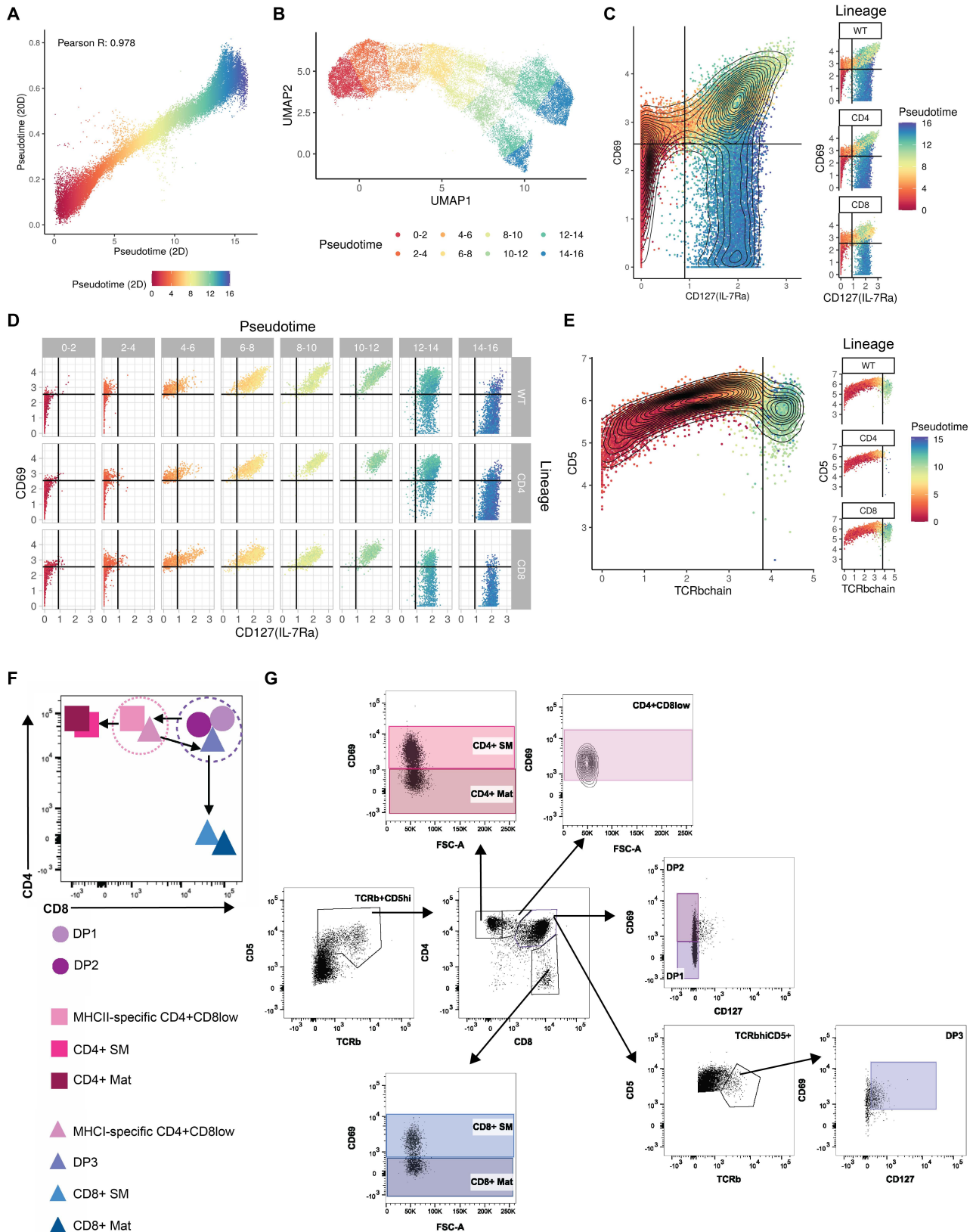
Figure S2

Figure S2: Pseudotime inference identifies intermediate thymocyte stages that can be isolated by FACS.
(A) Correlation between Slingshot pseudotime inferred from the full 20-dimensional totalVI latent space and a 2-dimensional UMAP projection of the 20-dimensional latent space. **(B)** UMAP plot of the totalVI latent space from positively-selected thymocytes. Cells are colored according to placement in one of eight bins uniformly spaced over 2D pseudotime for visualization. **(C)** In silico flow cytometry plots of log(totalVI denoised expression) of CD127(IL-7Ra) and CD69 from positively-selected thymocytes (left) and the same cells separated by lineage (right). Cells are colored by pseudotime. **(D)** In silico flow cytometry plot of data as in (C) separated by lineage and pseudotime. **(E)** In silico flow cytometry plots of log(totalVI denoised expression) of TCRβ and CD5 from DP thymocytes (left) and the same cells separated by lineage (right). Cells are colored by pseudotime. Among DP thymocytes, the DP3 population is TCRβ high, CD127+, and CD69+. **(F)** Schematic of a CD4 versus CD8 biaxial plot to identify gated populations in adult thymocytes. Cells were gated into eight subsets: DP1, DP2, CD4+CD8low, semimature CD4 (CD4+ SM), mature CD4 (CD4+ Mat), DP3, semimature CD8 (CD8+ SM), and mature CD8 (CD8+ Mat). Circles represent lineage uncommitted cells, squares represent CD4 lineage cells, and triangles represent CD8 lineage cells. **(G)** Representative flow cytometry gating strategy for thymocyte populations in adult mice. Thymocytes were harvested from 6-8-week-old WT (C57BL/6 strain), MHCI-/- or MHCII-/- mice. Cell populations were gated to include lymphocytes, exclude forward scatter and side scatter doublets, include live cells, include TCRβ+CD5int/hi, then on CD4 versus CD8. Cell populations were gated into the following subsets based upon cell surface marker expression: DP1 (CD4+CD8+CD127-CD69-), DP2 (CD4+CD8+CD127-CD69+), CD4+CD8low (CD4+CD8low; CD4+CD8lowCD69+), DP3 (CD4+CD8+TCRβhiCD5+CD127+CD69+), semimature CD4 (CD4+ SM; CD4+CD8-CD69+), mature CD4 (CD4+ Mat; CD4+CD8-CD69-), semimature CD8 (CD8+ SM; CD8+CD69+), and mature CD8 (CD8+ Mat; CD8+CD69-).
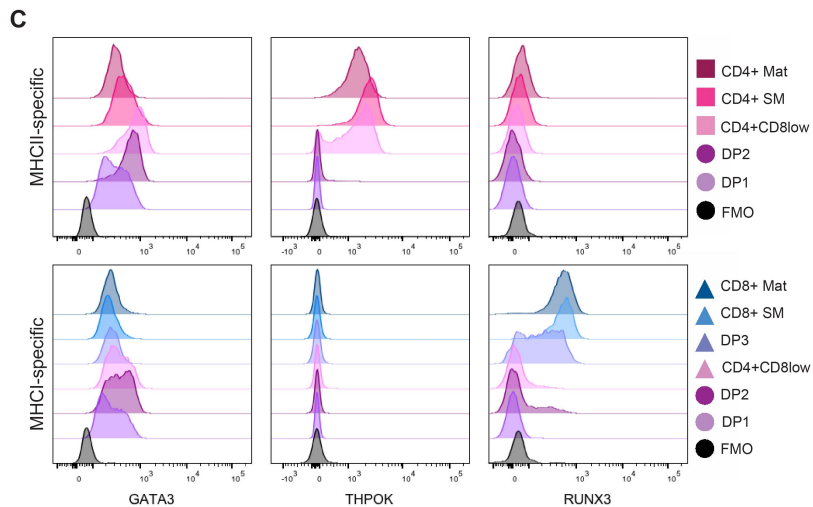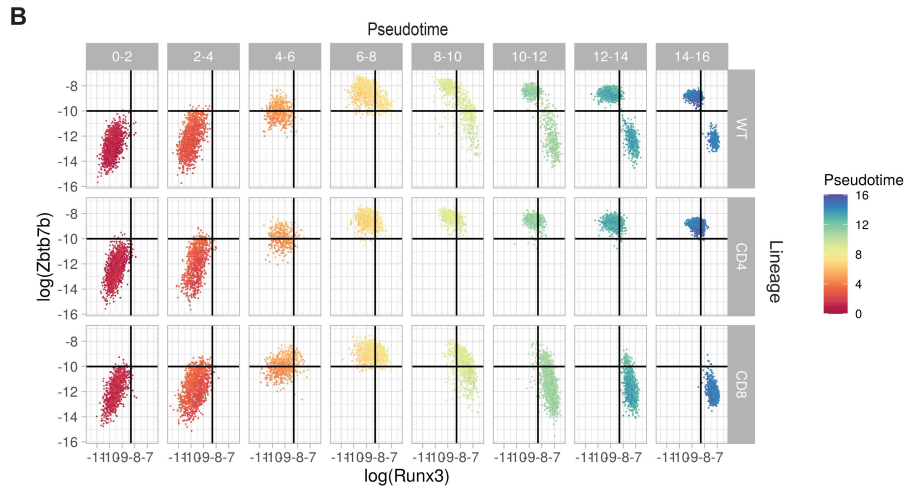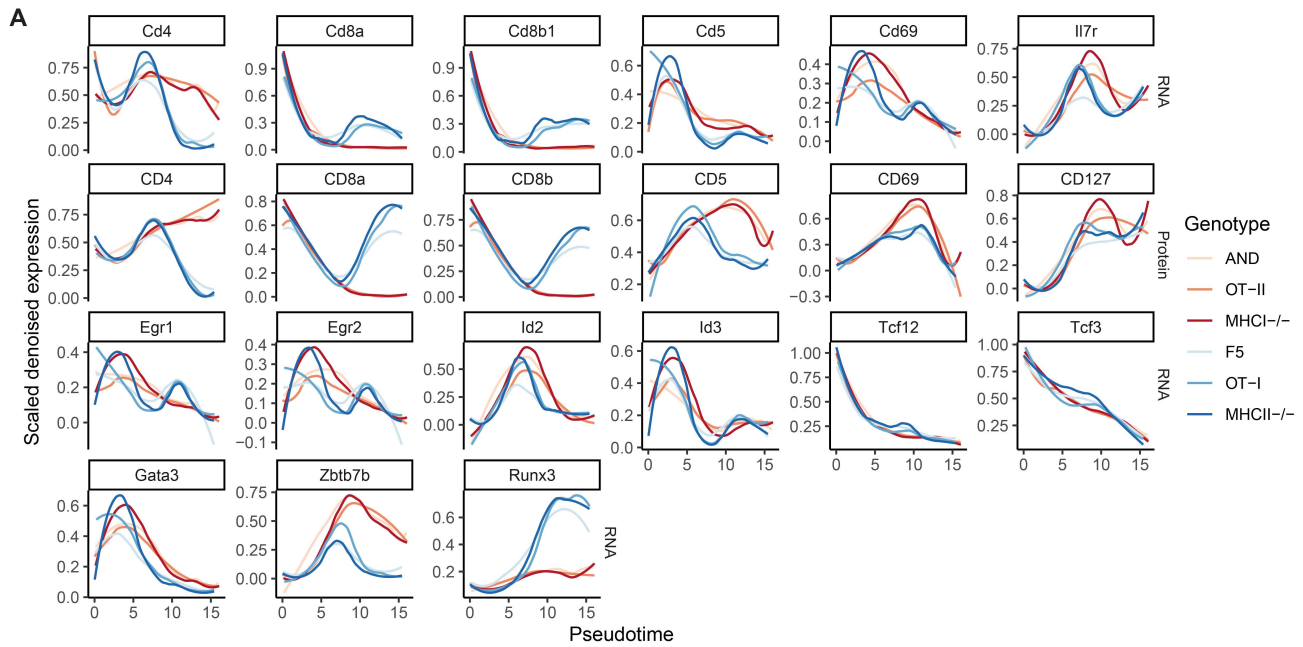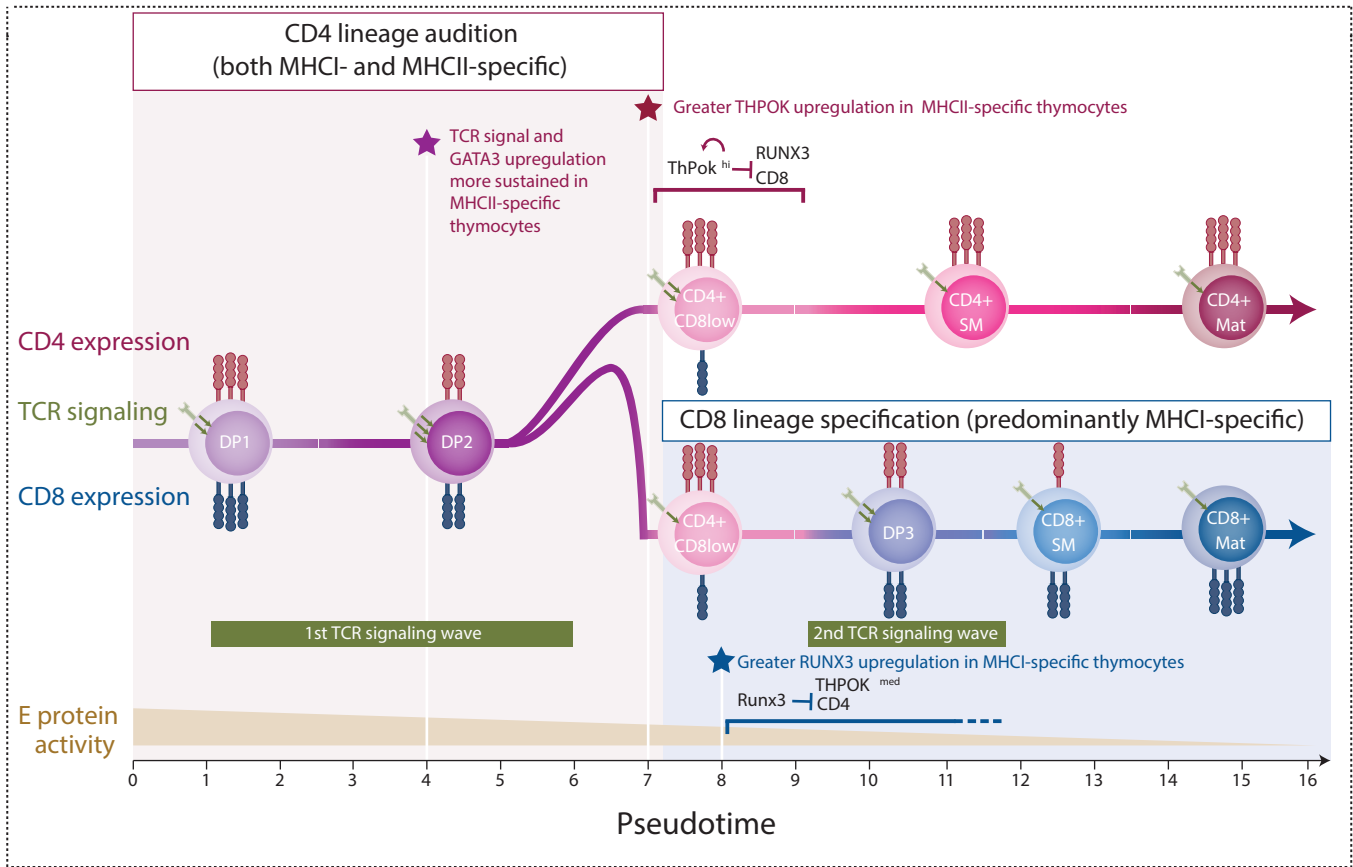
Figure S3

Figure S3: CITE-seq and fluorescence-based flow cytometry reveal the timing of expression for transcription factors and other features of CD4/CD8 development. (A) Expression of RNA and protein features over pseudotime by genotype. Features are totalVI denoised expression values scaled per feature and smoothed by loess curves. **(B)** In silico flow cytometry plots of log(totalVI denoised expression) of *Runx3* and *Zbtb7b* from positively-selected thymocytes separated by pseudotime. **(C)** Transcription factor protein expression in adult thymocyte populations. Representative histograms displaying GATA3, THPOK, and RUNX3 transcription factor expression detected by intracellular flow cytometry staining in MHCII-specific (MHCI-/-) and MHCI-specific (MHCII-/-) thymocyte populations. Thymocyte populations were gated on lymphocytes, excluding forward scatter and side scatter doublets, live cells, TCRβ+CD5int/hi then on CD4 versus CD8. Cell populations were gated into the following subsets based upon cell surface marker expression: DP1 (CD4+CD8+CD127-CD69-), DP2 (CD4+CD8+CD127-CD69+), DP3 (CD4+CD8+TCRβhiCD5+CD127+CD69+), CD4+CD8low (CD4+CD8lowCD69+), semimature CD4 (CD4+ SM; CD4+CD8-CD69+), mature CD4 (CD4+ Mat; CD4+CD8-CD69-), semimature CD8 (CD8+ SM; CD8+CD69+), and mature CD8 (CD8+ Mat; CD8+CD69-). Data is concatenated from n = 4 mice per genotype. Positive staining was determined using a fluorescence minus one (FMO) control.
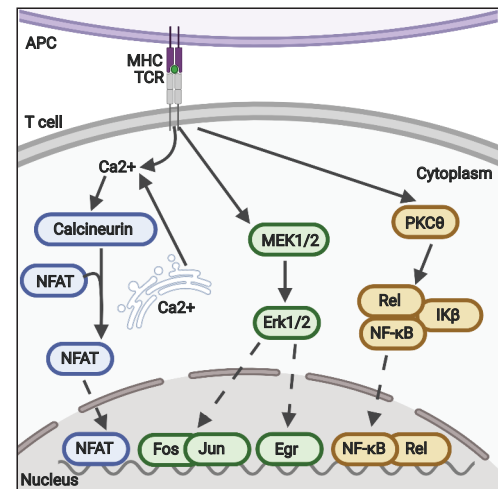
Figure S4

A

CD4 lineage audition (both MHCI- and MHCII-specific)

TCR signal and GATA3 upregulation more sustained in MHCII-specific thymocytes

Greater THPOK upregulation in MHCII-specific thymocytes

ThPok^{hi} ⊣ RUNX3 / CD8

CD4 expression

TCR signaling

CD8 expression

CD4+ CD8low · CD4+ SM · CD4+ Mat

CD8 lineage specification (predominantly MHCI-specific)

DP1 · DP2 · CD4+ CD8low · DP3 · CD8+ SM · CD8+ Mat

1st TCR signaling wave

2nd TCR signaling wave

Greater RUNX3 upregulation in MHCI-specific thymocytes

Runx3 ⊣ THPOK^{med} / CD4

E protein activity

Pseudotime

B

TCR signaling pathway

CD4-DE Cluster 4		CD4-DE Cluster 7		CD8-DE Cluster 3	
TF	Score	TF	Score	TF	Score
Nfatc2	14	Rel	10.67	Csrnp1	7.5
Gfi1	14.67	Arid5a	11	Batf3	9.333
Scml4	17	Fosb	11.67	Arid5a	14
Foxp3	29.5	Egr2	16.33	Junb	15.5
Stat4	31.75	Csrnp1	17	Fosb	15.67
Tcf7	33.8	Junb	23.25	Rel	17.33
Bach2	34.25	Batf3	23.67	Nr4a3	20.33
Lef1	39.75	Relb	33	Stat4	30
Foxo1	50	Hivep1	34	Relb	34.25
Ikzf3	57	Egr1	42.33	Plscr1	36
Csrnp1	71.5	Nr4a1	44	Egr2	37
Tbx21	83.5	Nr4a3	48.67	Fosl1	47.2
Nfatc1	97.6	Nfkb1	51.25	Pou2f2	50.2
Hey1	103	Irf7	51.67	Fos	51.6
Ets1	106.5	Nfkb2	53.25	Batf	57.2
Irf4	113.6	Pbx4	54	Tet2	59
Kdm2b	114	Egr3	57.67	Irf8	59.75
Runx3	128.6	Ddit3	61.67	Egr1	59.83
Cebpa	130.8	Irf4	64.4	Hivep1	60
Bcl11b	132.3	Fosl1	64.6	Nfkb1	62
Crebl2	146.3	Egr4	65	Gfi1	67
Pbx3	149.8	Fos	66.8	Ltf	69.5
Ascl2	152	Jun	66.83	Foxp3	71.25
Nacc2	160	Bhlhe40	67	Jun	72.33
Tfap2c	160	Pou2f2	73.6	Fosl2	74
Tet2	171.5	Runx3	76	Runx1	74.2
Tet3	175.3	Spib	77.5	Ikzf3	75
Maf	176.5	Nfatc1	77.8	Irf7	76.33
Ybx3	179	Foxp3	87.5	Irf1	82.33
Ikzf1	182.8	Stat5a	91.67	Cebpb	84.33

Calcineurin-NFAT branch

MAPK-Erk branch

PKC-NFkB branch

TCR signaling association

Figure S4: A sequential model for CD4 versus CD8 T cell lineage commitment. (A) A sequential model for CD4 versus CD8 T cell lineage commitment. Key events during positive selection inferred from CITE-seq data are displayed from left to right in their order of occurrence based on pseudotime. Colored circles indicate the order of appearance of key thymocyte stages as defined by cell surface markers. Shaded red area indicates the time window during which both MHCI- and MHCII-specific thymocytes audition for the CD4 fate, corresponding to upregulation of GATA3 followed by THPOK. Shaded blue area indicates the later time window during which those thymocytes that failed the CD4 audition (mostly MHCI-specific) receive CD8 lineage reinforcement and survival signals. Green horizontal bars indicate two distinct temporal waves of TCR signaling: a first wave that is stronger and more sustained in MHCII- compared to MHCI-specific thymocytes, and a second later wave that occurs only in MHCI-specific thymocytes during the CD8 lineage specification phase. Stars indicate the key time points of lineage divergence, including the earliest detection of greater TCR signals and GATA3 upregulation in MHCII-specific thymocytes (purple star), followed by preferential THPOK induction and CD8 repression in MHCII-specific thymocytes (red star), and finally preferential RUNX3 induction and CD4 repression in MHCI-specific thymocytes (blue star). Red bracket indicates the time window during which MHCII-specific thymocytes commit to the CD4 lineage by fully upregulating THPOK, leading to activation of a THPOK autoregulation loop (Muroi et al., 2008) and full repression of CD8. Blue bracket indicates the time window during which MHCI-specific thymocytes turn on RUNX3, leading to repression of THPOK and CD4. Brown triangle indicates the gradual downregulation of E protein transcription factor activity throughout positive selection, which eventually allows for CD8 lineage specification in thymocytes that do not express high levels of THPOK (Jones-Mason et al., 2012). **(B)** Transcription factor (TF) enrichment analysis for TCR target-enriched gene clusters. The top 30 TFs enriched in the gene sets defined by CD4-DE clusters 4 and 7 and CD8-DE cluster 3 in Figure 4B and C. The full ChEA3 enrichment analysis is in Table S11 and S12. Colored boxes correspond to TFs activated by the respective branch of TCR signaling annotated in the inset diagram. Gray boxes indicate additional TFs associated with TCR signaling based on Netpath (Kandasamy et al., 2010), and as labeled in Figure 4F-G.
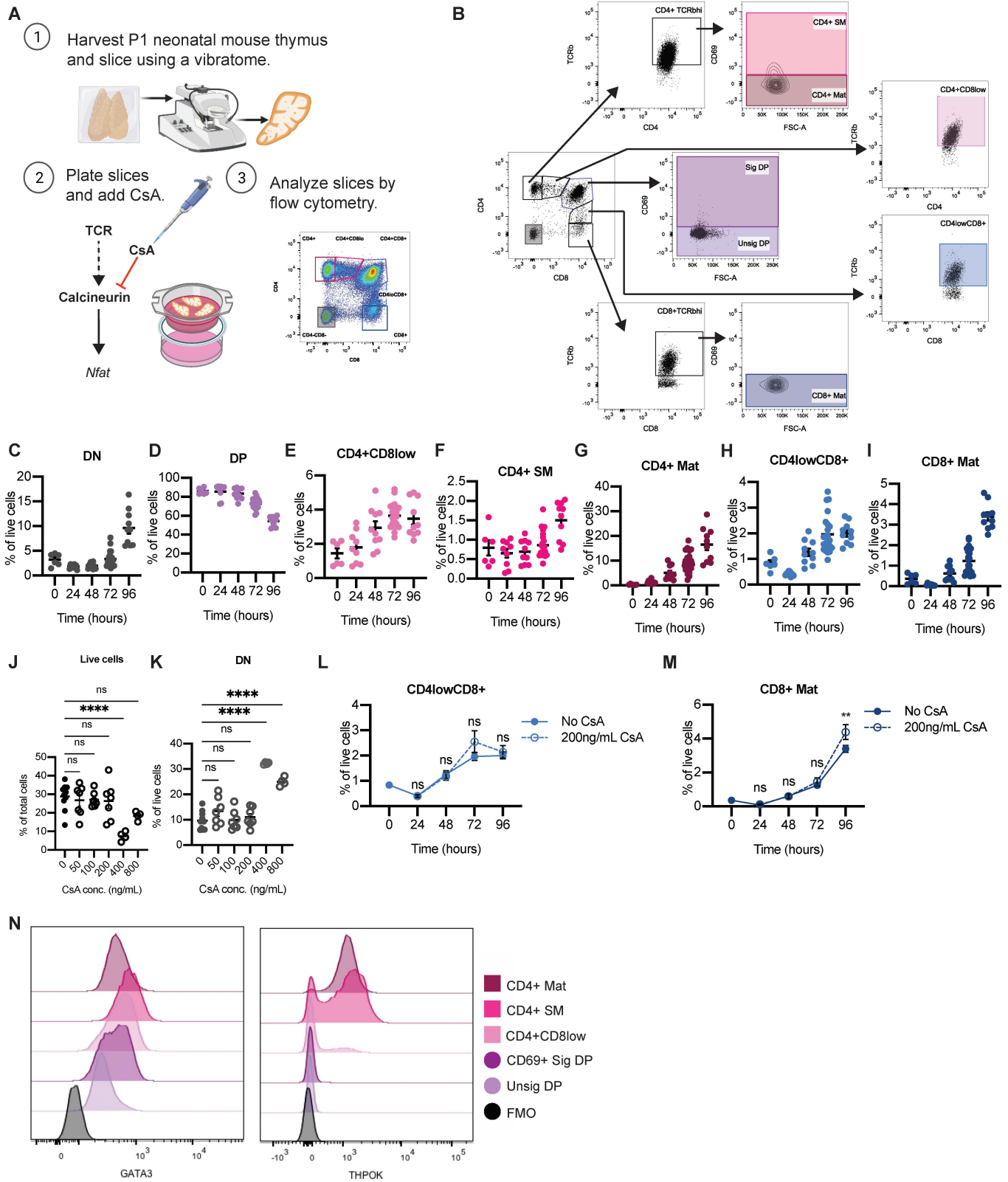
Figure S5

Figure S5: Cell populations in neonatal thymic slice cultures treated with Cyclosporin A. Thymic slices were isolated from postnatal day 1 (P1) mice and cultured in the presence or absence of Cyclosporin A (CsA) for up to 96 hours. Thymic slices were collected and analyzed at indicated time points via flow cytometry to quantify cell populations. **(A)** Experimental overview of neonatal thymic slice cultures. Illustrations in (A) were created using Biorender.com. **(B)** Representative flow cytometry gating strategy for neonatal thymic slice samples. We used a modified gating scheme to adjust for the differences in cell populations in adult versus neonatal thymocytes; e.g., the absence of a detectable DP3 population and presence of a CD4lowCD8+ population in neonatal thymic samples. Cell populations were gated on lymphocytes, excluding forward scatter and side scatter doublets, live cells, then on CD4 versus CD8. Thymocytes were further gated into eight populations: double negative (DN; CD4-CD8-), unsignaled double positive (Unsig DP; CD4+CD8+CD69-), CD69+ signaled double positive (CD69+ Sig DP; CD4+CD8+CD69+), CD4+CD8low (CD4+CD8lowTCRβ+), CD4+ semimature (CD4+ SM; CD4+CD8-TCRβhiCD69+), CD4+ mature (CD4+ Mat; CD4+CD8-TCRβhiCD69-), CD4lowCD8+ (CD4lowCD8+TCRβ+), and CD8+ mature (CD8+ Mat; CD4-CD8+TCRβhiCD69-) cell populations. **(C-I)** WT neonatal slice time course experiments. Graphs display the frequency (% of live cells) of **(C)** DN, **(D)** DP, **(E)** CD4+CD8low, **(F)** CD4+ SM, **(G)** CD4+ Mat, **(H)** CD4lowCD8+, and **(I)** CD8+ Mat cells in thymic slices after 0, 24, 48, 72 and 96 hours of culture. **(J-K)** Impact of CsA concentration **(J)** frequency of live cells (% of total cells) **(K)** frequency of DN cells (% of live cells). **(L-M)** Frequency of **(L)** CD4lowCD8+ and **(M)** CD8+ Mat cells after 0, 24, 48, 72 and 96-hours of culture in medium alone or with 200ng/mL CsA. **(N)** Transcription factor expression in neonatal thymic slice cultures after 48 hours. Representative histograms displaying concatenated (n=4) data showing GATA3 (left) and THPOK (right) transcription factor expression detected by intracellular flow cytometry staining. Each symbol on the graphs represents a thymic slice. For (C-I), graphs contain data compiled from 9 independent experiments with WT slices. Data in (J-K) is representative of 2 independent experiments. Data were analyzed using an ordinary one-way ANOVA. In graphs (L and M), data is compiled from 9 independent experiments with WT slices. Data are displayed as the mean ± standard error of the mean (SEM). For slices cultured with no CsA for 0 hours n=6, 24 hours n=9, 48 hours n=10, 72 hours n=22, 96 hours n=10. For slices cultured with 200ng/mL CsA for 24 hours n=6, 48 hours n=7, 72 hours n=14, 96 hours n=7. Data was analyzed using an ordinary two-way ANOVA with multiple comparisons. NS is not significant, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.