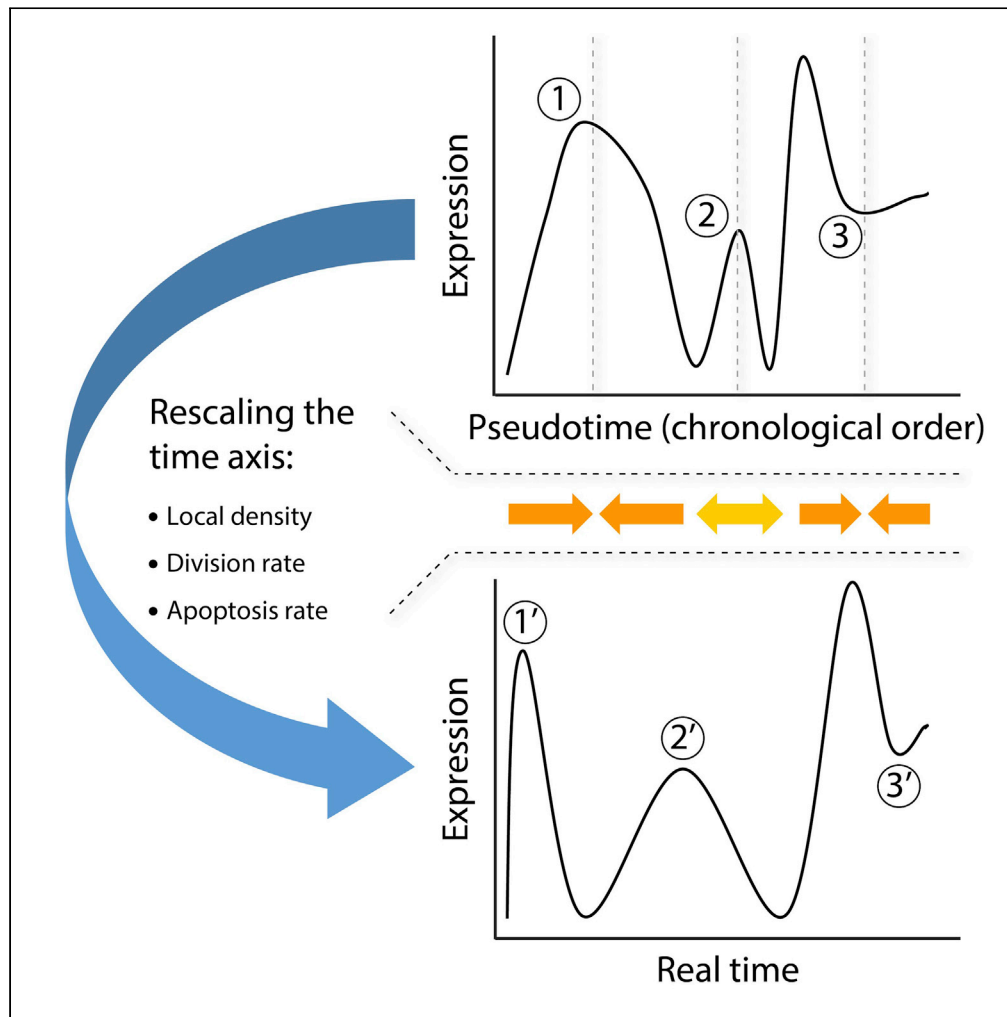**Article**

From pseudo to real-time dynamics of T cell thymic differentiation

Avishai Gavish,
Benny Chain,
Tomer M. Salame,
Yaron E. Antebi,
Shir Nevo, Shlomit
Reich-Zeliger, Nir
Friedman

avishai.gavish@weizmann.ac.il
(A.G.)
shlomit.reich-zeliger@
weizmann.ac.il (S.R.-Z.)

Highlights
Pseudotime inference
based on local expression
densities in high
dimensional space

Rescaling the pseudotime
axis by incorporating cell
division and apoptosis
rates

Experimental validation
using CyTOF T cell data
from the mouse thymus

Article

From pseudo to real-time dynamics of T cell thymic differentiation

Avishai Gavish,[1,*] Benny Chain,[2] Tomer M. Salame,[3] Yaron E. Antebi,[4] Shir Nevo,[5] Shlomit Reich-Zeliger,[5,6,7,*] and Nir Friedman[5,6]

SUMMARY

Numerous methods have recently emerged for ordering single cells along developmental trajectories. However, accurate depiction of developmental dynamics can only be achieved after rescaling the trajectory according to the relative time spent at each developmental point. We formulate a model which estimates local cell densities and fluxes, and incorporates cell division and apoptosis rates, to infer the real-time dimension of the developmental trajectory. We validate the model using mathematical simulations and apply it to experimental high dimensional cytometry data obtained from the mouse thymus to construct the true time profile of the thymocyte developmental process. Our method can easily be implemented in any of the existing tools for trajectory inference.

INTRODUCTION

Technological advancement in measuring expression profiles of multiple variables at single-cell resolution has paved the way for the creation of new analytical frameworks called trajectory inference. Based on the similarity in high dimensions, dozens of methods exist today that allow sorting cells along the temporal course of biological processes.[1–5] Techniques are constantly evolving to address systems of growing complexity, as when cells bifurcate or when development proceeds in a tree-like topology. Recent studies have compared the strengths and weaknesses of these methods to aid users to select the method that best suits their data.[6–8] However, trajectory inference, also known as "pseudotime" analysis, can only capture the topology and sequence of the developmental process. In order to capture the time dimension, incorporating the length of time spent by each cell in each developmental stage, additional information on the fluxes within the system (the number of cells entering and leaving the compartments) is required. Several studies attempted to describe true temporal dynamics using some *a-priori* knowledge about the system,[9] reconstruct the trajectory using experiments to infer some time-points along it,[10,11] or ignored cell division and cell death.[12]

A well-studied system for trajectory inference is the mouse thymus. Bone marrow hematopoietic progenitors migrate to the thymus, where they commit to the T-cell lineage and further mature into functional T lymphocytes.[6] T-cell development proceeds through a series of discrete phenotypic stages that are characterized by the expression of several important membrane molecules, most notably CD4 and CD8. Thymocytes initially express low CD4 and CD8 levels (CD4$^-$CD8$^-$) in the double-negative stage (DN). The DN stage can be subdivided into four distinct phases (DN1 to DN4), each characterized by a different marker expression profile.[13–15] The cells gradually start to express higher levels of both molecules (CD4$^+$CD8$^+$) and transition into the double-positive stage (DP). During the DP stage, cells massively proliferate and undergo two selection processes: positive and negative. In the course of positive selection, only cells that recognize MHC molecules by their T-cells receptors (TCR) survive, while cells that fail to do so die by neglect.[16–22] In the course of negative selection, cells that have high affinity to the MHC self-peptide complex undergo apoptosis.[22,23] At this point the developmental trajectory bifurcates toward one out of two single-positive stages (SP), where some cells suppress their high CD8 levels while accumulating further CD4 (CD4$^+$CD8$^-$CD3$^+$), or vice versa (CD4$^-$CD8$^+$CD3$^+$) (Figure 1A).[7,8,14,15,24] During this maturation process, proliferation and apoptosis occur at varying rates, while many more surface markers are upregulated or downregulated upon the differentiating T-cells.[8,22,24–27] Cell development and selection in the thymus play key roles in shaping the adaptive immune system and maintaining self-tolerance.[28–30]

Here, we use mass cytometry (CyTOF) to measure high dimensional protein expression in thousands of unsynchronized differentiating single-cells simultaneously. Data derived for each cell includes a comprehensive panel

[1]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel

[2]Division of Infection and Immunity, University College London, London, UK

[3]Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel

[4]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

[5]Department of Immunology, Weizmann Institute of Science, Rehovot, Israel

[6]These authors contributed equally

[7]Lead contact

*Correspondence:
avishai.gavish@weizmann.ac.il (A.G.),
shlomit.reich-zeliger@weizmann.ac.il (S.R.-Z.)

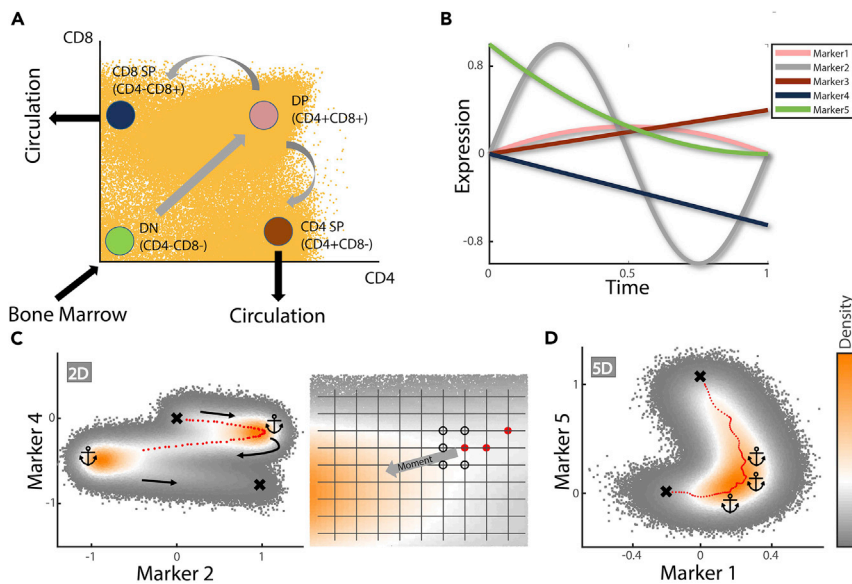https://doi.org/10.1016/j.isci.2022.105826

Figure 1. Dimensionality reduction and segmentation

(A) Cell differentiation in the thymus. Scatterplot of cells in representative mouse (yellow) depicts a snapshot of the maturation process, with naive cells entering the thymus in the double-negative state (DN, CD4⁻CD8⁻), sequentially differentiating into double-positive cells (DP, CD4⁺CD8⁺), and finally acquiring one of the single-positive fates (SP, CD4$^{+/-}$ CD8−/+). Arcsine transformation was used here and for gating purposes (see STAR Methods).

(B) Five simulated markers. Profiles are shown without adding noise. Each time step (dt = 0.01), 100 new cells were added and assigned with t = 0. Initial marker values for each cell were chosen from a normal distribution centered at the values plotted at t = 0, with a SD of 0.1. Noise was also added to the cycle time length of each cell (after which the cell was removed) which was chosen from a normal distribution centered at T = 1 with a SD of 0.1. Division and apoptosis were not implemented in these simulations, which were run until reaching the steady state.

(C) Trajectory course calculated upon density plot of two markers in B. Starting and ending points are predetermined (upper and lower crosses respectively). Anchors (corresponding symbols) are placed in positions of local maxima. The trajectory (red circles) travels upon the crest connecting path edges and anchors. Panel to the right magnifies the last position obtained and illustrates how new position coordinates are chosen; new coordinates are chosen from nearest neighbors surrounding the previous position on a grid, where progression in the direction of the previous position is forbidden (optional new coordinates are surrounded by black circles). Default choice is for the neighbor in the direction aligned toward the next anchor (moment). Deviation is permitted only if the difference in densities between the default neighbor and some other permitted neighbor exceeds a predetermined value, thereby justifying a detour. See methods for more details.

(D) Trajectory calculation in five dimensions. The principle is similar to calculation in two dimensions (see text), allowing for trajectory dynamics in any region where at least one of the markers changes. The calculated trajectory is projected here upon a two-dimensional density plot, with anchors and crosses representing landmarks as in C. Note that anchors are placed near, but not in the exact position, of the global maxima in two dimensions (see methods for superfluous anchor removal in higher dimensions).

of surface marker levels, together with transcription factors and markers for cell-cycle stage and cell activation. Using ^{127}IDU-injected mice,[31] we were able to infer division and apoptosis rates throughout the process.

Similar to previous pseudotime methods, we first formulate an intuitive model for sorting cells along a developmental axis.[5,32] Our pseudotime ordering method emphasizes the importance of local cell densities, which is a key concept in the consecutive step (however, our real-time method can be implemented directly on one of the prevailing methods). Next, by incorporating compartment-specific proliferation and death rates, we capture quantitatively the flow of cells through the developmental process. In this way, we add a true time dimension to the basic developmental sequence and topology captured by previous approaches.

RESULTS

Trajectory inference in multidimensional data

The initial step is to order cells or segments of "similar" cells, according to chronological order along the developmental course of events. To illustrate how this method works, we start by simulating an imaginary

process where cells enter the system at a constant flux at time $t = 0$, and begin expressing five markers as illustrated (Figure 1B). Random noise was added so that initial values of each marker were chosen from a distribution centered at $f^i_{(t=0)}$ with a SD of 0.1 (f^i is the expression function for each $i = [1, 5]$ marker). Noise was also added to the entire cycle time length T of each cell after which cells were removed, so that T was chosen from a random distribution centered at $T = 1$ with a SD equal to 0.1. Assuming no cell division or apoptosis during the process at first, marker values were updated every $dt = 0.01$ time interval. The simulation was run until reaching the steady state, in which the total cell number was constant up to some level of fluctuations due to noise. Importantly, at steady state, local cell density as a function of time was also constant (SI Figures S1 and S2).

For simplicity, we first calculate the trajectory in two-dimensions, choosing the markers illustrated in Figure 1C. Marker expression space is divided into a $m^i \cdot m^j$ grid, with D^{ij} denoting cell density at each grid position. Our working assumption is that the trajectory must pass through points of local maxima in cell density (anchors in Figure 1C). We also assume that the initial and final positions of the trajectory are known approximately and are manually determined (e.g. the assumed location of the starting and ending cells). The trajectory, denoted as l, then progress between points on the grid from one anchor to the next, traveling upon the density crest. The default trajectory is in the direction of the next anchor ("moment"), while a detour is only permitted when the difference in densities between the grid point aligned with the moment and that at some other neighboring point exceeds some predetermined value (see supplemental information section 2 for more details).

Computing trajectory coordinates l^{ij} effectively reduces dimensionality to one. Each cell is now associated with a single point upon l to which it is closest based on Euclidean distance. Assuming the trajectory is composed of k evenly spaced points, each point can generaly affiliate many cells; the k cell-segments are ordered according to the chronological course of the systems' evolution. Inside a given segment, there is no internal ordering of cells. Thus, the number of sections (or initial grid density) and the extent of trajectory curvature, dictate the resolution by which cells are sorted.

Trajectory inference in this method is closely dependent on dynamics in marker expression. E.g., if the two markers in Figure 1C happen to be stationary at some point, density in two-dimensional space around that point will be relatively high. Extending the method to higher dimensions by adding more markers is therefore beneficial, as local kinetics in the expression of even a single marker is able to unfold regions where seemingly no dynamics occur. Following a similar line as in two dimensions, the trajectory can be calculated in higher dimensions by dividing each dimension into a grid and passing from a given initial position through regions of local maxima. Notably, the removal of superfluous local maxima points might be necessary after expanding the number of dimensions (SI section 2). Figure 1D illustrates the trajectory that was calculated in 5 dimensions (using all 5 markers) projected upon a two-dimensional density plot.

From pseudotime to real-time: Theory and simulation

The method described above segments similar cells according to Euclidian proximity in high dimensions with respect to the trajectory. To illustrate the principle by which real-time can be calculated, it is useful to first focus on fluxes and dynamics within a single segment of cells (Figure 2A). Within every time interval, a fraction of cells mature to a point where they exit the segment and enter the neighboring one ahead. Similarly, some other fraction of cells enters the segment by exiting the preceding segment. For a segment with index i, these density fluxes of entering and exiting cells can be written respectively as $f_{i-1} \cdot v_{i-1}$ and $f_i \cdot v_i$, where f denotes cell density (the number of cells in a segment divided by the total number of cells), and v denotes the progression rate along the trajectory l. Other sources for cell turnover within a segment are cell division and apoptosis, whose rates are denoted as γ_1 and γ_2 respectively. Every time interval, $f_i \cdot \gamma_1^i$ cells are added to the segment, while $f_i \cdot \gamma_2^i$ cells are removed. Thus, the local density within a segment, which in the steady state must remain constant over time, is ultimately defined by the different fluxes of entering and exiting cells, and the local proliferation and apoptosis rates. A high local density, in the presence of a low division-to-apoptosis rate ratio, is indicative of a prolonged dwelling time of cells in the segment, and vice versa. This conservation principle is presented compactly in the equation shown Figure 2B. A closed-form expression in the case when γ_1 and γ_2 are constant in space is highlighted below the equation (rectangle), where t is the cumulative average time spent up tp point l on the trajectory, and represents the average developmental time of cells in that segment (More derivatives of this equation are given in supplemental information section 3).
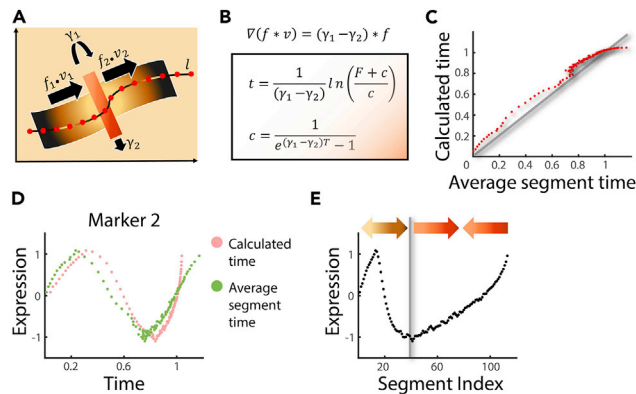
Figure 2. Time calculation

(A and B) The time equation. Cartoon outlining dynamics within a single segment is shown in A. Red circles represent trajectory points in some imaginary data, with the red stripe magnifying a given segment. The local density in each segment is determined by the flux of entering cells $f_1 \cdot v_1$ (f and v represent local cell density and velocity, respectively), flux of cells exiting the segment $f_2 \cdot v_2$, and cell division and apoptosis rates (γ_1 and γ_2, respectively). High local density can be a manifestation of lengthy dwell-time of cells in a segment, or of a high division to apoptosis rate ratio. Similarly, low local density can be a result of rapid progression through the segment, or of a high apoptosis to division ratio. (B) gives the time equation in its general form with ∇ denoting the differential operator. The formulation for the time as function of location along the trajectory, $t(l)$, is given after solving the time equation assuming constant $\gamma_1 - \gamma_2 > 0$, with F denoting the cumulative density up to location $l = l'$, and T denoting the time reached when cells arrive at the trajectory end. See methods for more derivations including numeric formulation for when division is not constant.

(C) Validation of formulation accuracy. Simulations were performed as in Figure 1B, this time incorporating constant division and apoptosis rates ($\gamma_1 = 0.09$, $\gamma_2 = 0.03$). The "Average segment time" directly averages simulation times of cells in each segment and represents real time. "Calculated time" represents time in each segment using the formulation in B. The deviation toward the end is due to an increasing number of cells exiting the cycle, with the few remaining cells adding negligibly to the cumulative density. Correlation coefficient reaches 0.99.

(D) Average values of Marker2 (Figure 1B), using real segment average time (green) and calculated times (pink).

(E) Average marker times as the function of segment index (pseudotime) distorts the expression profile. Arrows indicate rescaling applied by the time equation.

To test the theory formulated above, we simulated the dynamics in Figure 1B, this time adding proliferation and apoptosis. Figure 2C compares the "real" dwell time, which was calculated by averaging the known simulated times of all cells within each segment, to the time calculated using the equation given in Figure 2B. The simulation and prediction agree to within a threshold determined by noise (Figure 2C) and by a deviation as t approaches 1 that can be explained by the open-ended model, allowing cells to leave the system, rather than dying. The trajectory of marker expression plotted as a function of calculated or simulated times are also in good agreement (Figure 2D), but differ substantially when plotted as a function of pseudotime, which does not incorporate division and death methods (Figure 2E). Notably, by incorporating known timing landmarks and adjusting points along the pseudotime trajectory accordingly (e.g. via a designated experiment) a better estimate of the real-time profile could be achieved, but the pseudotime points in-between the landmarks would still not be scaled. Rescaling the entire pseudotime axis (as in Figure 2D) is equivalent to adjusting each of the time points to its correct relative position along the axis.

Experimental determination of proliferation and apoptosis rates

To measure cell proliferation *in-vivo*, we injected mice with 5-Iodo-2'-deoxyuridine (^{127}IDU), a thymidine analog incorporated into DNA during the S-phase. Accordingly, only proliferating cells incorporate IDU and integrate it into the DNA. Upon further division, half of the initially incorporated IDU is passed on to each daughter cell. Following injection, we harvested the mouse thymus after 1, 3, 6, and 12 h (3 mice in each group) and performed CyTOF analysis (Figure 3A). IDU levels are readily detected as elemental Iodine by CyTOF, with no additional experimental requirements.

Plotting IDU expression as a function of DNA levels allows the identification of two cell populations—dividing and non-dividing cells, with and without IDU uptake respectively (Figure 3B). Gating the dividing (IDU positive) cells selects the cells that have divided in the time interval that lapsed between IDU injection
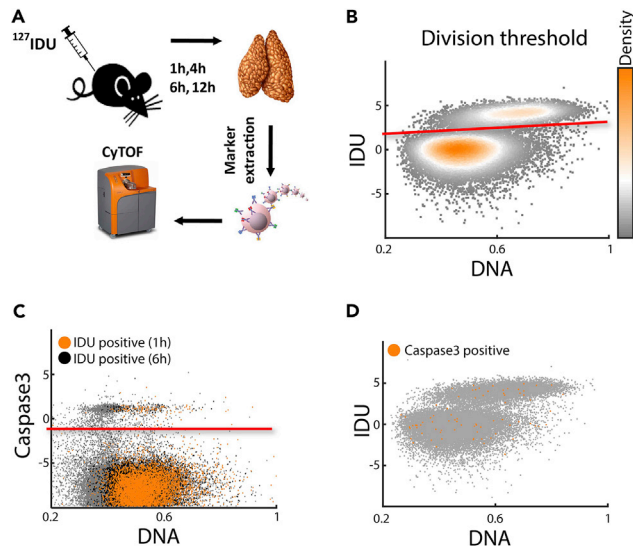
Figure 3. Determining division and apoptosis rates

(A) Scheme describing the IDU injection process.

(B and C) Threshold for dividing and apoptotic cells. The logarithm of markers used to estimate the thresholds for division and apoptosis (IDU and Caspase3, respectively) are plotted against normalized DNA. Panel B shows data from a mouse 1 h post-IDU injection, and panel C shows data from two mice 1 and 6 h post-IDU injection. Cells above the red line in B were considered as dividing cells within 1 h. Cells above the red line in C, that were also positive for IDU, were considered as cells that will undergo apoptosis within 1 or 6 h. γ_1 and γ_2 in each segment for this mouse were calculated as the percentage of IDU or IDU-Caspase3 positive cells in the segment divided by the corresponding δt in day units.

(D) Caspase3-positive cells in a mouse 1 h post-IDU injection are roughly evenly distributed between the two populations of IDU positive and negative cells (SI Section 4).

and thymus harvesting (we denote this time interval as δt). We assume that the number of cells dividing more than once within δt (max 12h) is negligible. Thus, the proliferation parameter in each segment γ_1^i can be calculated by dividing the proportion of IDU-positive cells within the segment by δt.

Our chosen apoptosis marker was Caspase3, a 32 kDa cysteine protease that is activated during the early stages of programmed cell death.[33] Caspase3-positive cells could be observed as a scatter of cells above the remaining non-apoptotic cells (Figure 3C). Since calculating the apoptosis parameter necessitates dividing by a time interval within which apoptosis will occur, we chose only cells that were positive for both Caspase3 and IDU. To check whether these cells indeed reliably represent the entire population destined for apoptosis within δt, and not some unique sub-population, we verified that cells considered Caspase3-positive are evenly distributed throughout the IDU-DNA plane, and are not skewed into some specific region (Figure 3D).

Hence, the gated cells positive to both Caspase3 and IDU were assumed to undergo apoptosis within the same δt used for assessing the division rate. Similar to γ_1^i, the apoptosis parameter in each segment was calculated by dividing the proportion of IDU-Caspase3-positive cells within the segment by δt. Notably, this approach allows the apoptosis rate to be locally higher than the division rate.

Assessing the proliferation and apoptosis parameters should be relatively robust to the time of IDU injection (which is arbitrary). Indeed, the relative population of IDU-positive cells roughly increased linearly with δt, resulting in comparable similar proliferation and apoptosis parameters irrespective of injection time (Figures 3C and supplemental information Figure S3).

It is well established that in a short time interval during the positive selection process that occurs in the DP phase, approximately 90% of the thymocytes undergo death by neglect.[8,16,34] The direct measurement of death-by-neglect using established apoptosis markers such as Caspase3 is nevertheless challenging, since dying cells are rapidly removed by scavenger macrophages, leading to significant under-estimates of the number of dying cells. We verified this extensive apoptosis incident by an *in-vivo* experiment in which we
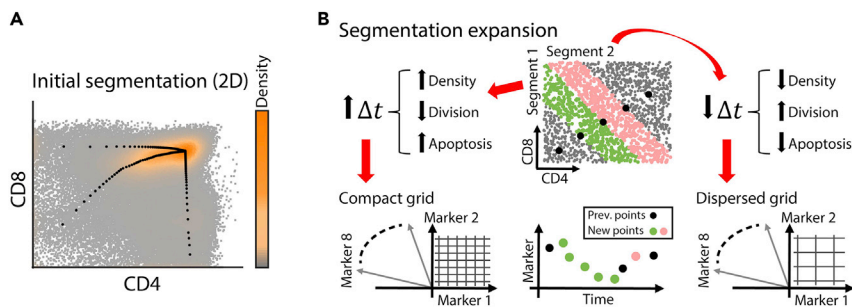
Figure 4. A two-step segmentation process

(A and B) Since one-step segmentation in higher dimensions can lead to significant dispersal of cells along the projected trajectory in CD4/CD8 space (e.g. assigning DN cells to points far along the trajectory, supplemental information Figure S4), segmentation was first performed in 2D. Panel in A shows the trajectory upon the density plot in CD4/CD8 space, leading from DN to DP, and splitting toward SP states. The time equation was solved separately for the DN + DP phase and each SP phase. Sequentially, each segment was sub-segmented using more markers (8 dimensions) as illustrated in B. A zoomed portion of the trajectory calculated in A is shown, with cells between two pairs of points colored differently (green and pink). The grid size for sub-segmentation was proportionally chosen with respect to the relative time spent in the original segment (Δt), potentially resulting in the addition of more or less time points.

adoptively transferred DN thymocytes from CD45.1 (B6SJL) mice into the thymus of CD45.2 (C57Bl) mice, and followed their development (Figure 5A). The observed reduction in CD45.1 cell-number reduction between 21 and 28 days supported the hypothesis that about 90% of cells are eliminated (Figures 5B and supplemental information Figure S3), Based on the literature, and these additional experiments, we therefore modified the parameter controlling the rate of cell death γ_2 in the DP phase so as to model the loss of 90% of the thymocytes during this stage (SI section 3).

Imputation of a real-time dimension in the thymus developmental trajectories

CD4 and CD8 are the two most pivotal markers that define the final outcome as well as key transitional phases along a thymocyte developemental trajectory. Trajectory calculation in high dimensions and cell-affiliation to trajectory points according to Euclidian distance can result in cell dispersion across CD4/CD8 space, so that for instance some cells expressing low CD4 and CD8 (and therefore considered to be in DN phase) are assigned to relatively distant points along *l* (SI Figure S4). This results from the fact that a simple Euclidian calculation gives all markers a similar weight in defining cell proximity to trajectory points. *A priori* knowledge of CD4 and CD8 importance in specifying differentiation states can therefore help refine the segmentation process.

We thus performed trajectory calculation and cell assignment in two steps. At first, we calculated the trajectory in CD4/CD8 plane and assigned cells to trajectory points using these two markers alone (Figure 4A). The bifurcation point was identified manually, along with the initial and final trajectory points. The estimated dwelling time of cells in each segment was calculated as described above. We next performed sub-segmentation by considering more markers. For each given segment, aside from CD4 and CD8, we chose another six markers with the highest local variability in the segment, which thus have the potential of contributing the most to understanding the systems' dynamics. In segments where cells dwelled for a relatively long period according to the calculation in 2-dimensions (high Δt), we wanted a higher resolution and therefore divided the segment into more points by choosing a compact grid. On the other hand, in regions where cells spent less time according to the calculation in 2-dimensions (low Δt), we allowed a sparser grid and hence less points were added within the segment (Figure 4B).

Dividing the segmentation process into two steps enforced the distribution of cells according to the CD4/CD8 expression. Sub-segmentation in the second step allowed "unfolding" of regions where additional transitions could be discovered by the addition of more markers.

Construction of the time axis allows the comparison of different marker profiles along the two SP branches. CD4 and CD8 expression profiles follow the expected trend in which both markers rise up to the bifurcation point (end of DP phase), after which the counter marker starts to decrease along the respective branch
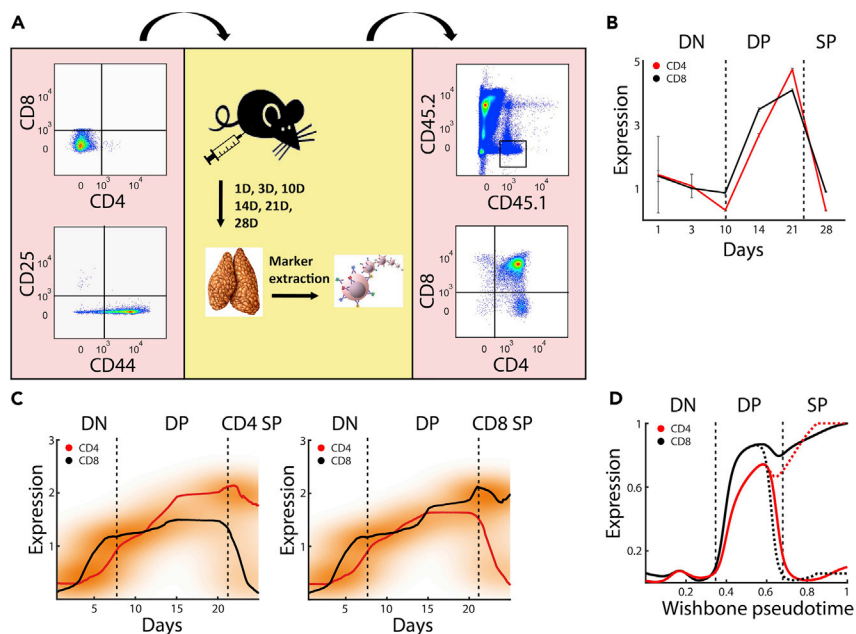
Figure 5. Real-time axis validation

(A) *In-vivo* validation was performed by intrathymic, adoptive transfer of DN1-3 (CD4⁻CD8⁻CD44+/−CD25+/−) cells from CD45.1 mice (B6SJL) into CD45.2 (C57BL) mice (left and center panel). Thymi were harvested 1, 3, 10, 14, 21, and 28 days post-transfer and stained with antibodies against CD45.1, CD45.2, CD8, CD4, and other developmental markers (center). Further analysis was performed by gating the CD45.1+/CD45.2- cells. Each time point presents Thymi from two mice, and the experiment was repeated twice.

(B) CD8 (Black) and CD4 (Red) expression levels were calculated by tracking the CD45.1+/CD45.2- gated cells. The expression levels for CD4 and CD8 were separately normalized to the median fluorescence of all the cells from all the time points. Black broken vertical lines separate here (and in later figures) between the DN, DP, and SP phases as indicated on top.

(C) CD4 and CD8 profiles as constructed by our method for real-time axis inference averaged across 12 mice.

(D) Wishbone pseud-temporal ordering captures the major stages in T cell development but scales differently than the real-time axis.

(Figure 5C). To compare our expression profiles to those obtained using pseudotime, we used the well-established algorithm Wishbone, originally used to infer developmental trajectories for maturing thymocytes[5] (Figure 5D). The two approaches were successful in detecting the major developmental states, but the relevant time spent in each state was apparently different. While the Wishbone trajectory suggested that cells spent ∼30% of the time in the DN state and ∼20% of the time in the DP state, our real-time trajectory indicates that cells spent less than 20% in the DN state and about 60% of the time in the DP state. One limitation of our segmentation approach is in detecting relatively rare cells whose population size is much lower than the number of cells in the segment to which they belong. For example, the DN1 (CD4-/CD8-/CD3-/CD25-/CD44+) population size is considered to be about 1% of all DN cells (and ∼0.03-0.05% [37]), while the initial segments in the trajectory are more than an order of magnitude larger in terms of cell numbers. Indeed, the initial segments in our model seem to capture the dynamics of the DN3 phase in which cell density starts to increase rapidly, which results in an effective shortening of the DN phase. This is seen in Figures S5, S6, and S7 where CD25 and CD3 levels are relatively high, consistent with the DN3 phase (CD4-/CD8-/CD3+/CD25+/CD44-).

To validate the predicted dwell times in the different developmental phases predicted by the model, we sought to determine the timeline of differentiation experimentally. To this end, we used the adoptive-transfer model outlined above, detecting T-cells 1,3,10,14,21 and 28 days post-transfer (Figure 5A). We observed that CD45.1 T cells remained in the DN state for approximately 7 days, after which they began to express CD4 and CD8 in parallel, thereby entering the DP state. About 25 days post-transfer we could not detect ∼90% of cells. We speculate that the reduction in cell number was because cells either underwent apoptosis (via positive or negative selection), or matured to CD4 or CD8 SP cells and left the
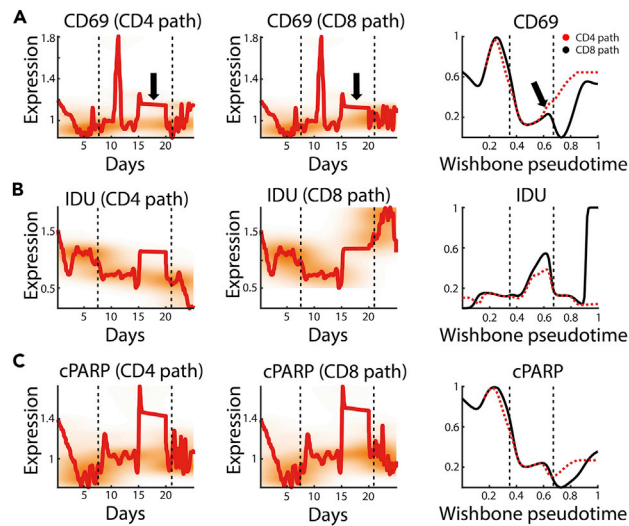
Figure 6. The real-time axis is able to unfold regions of biological significance

(A-C) CD69, IDU, and cPARP real-time trajectory profiles along the CD4 SP and CD8 SP paths (left and center panels respectively, averaging data from 12 mice.), and along Wishbone pseudotime trajectory (right, one mouse). Black broken vertical lines separate between DN, DP, and SP phases as in Figure 5. Black arrows in A indicate the second elevation in CD69 expression which is spread over several days along the real-time axis as opposed to the relative condensed portion it occupies in the Wishbone trajectory.

thymus.[8,23,34] The inferred dwell times, therefore, correspond well to the experimental observations, and fit the data better than the standard pseudotime projections.

The full profiles of all markers we used can be found in supplemental information Figures S6 and S7.

A real-time trajectory recognizes dynamics overseen using pseudotime

We examined the inferred dynamics of expression of all the markers and compared them to the Wishbone trajectories. The general trends were similar for most markers (SI Figures S6 and S7). However, some specific differences emerged. CD69, a marker of TCR signaling[17,18] exhibited complex dynamics during the DP stage, with a high expression peak around day ten, which declined sharply and then rose again between days 15-20 (Figure 6A). The second peak was less obvious in the Wishbone trajectory (Figure 6A, right panel).

A similar but delayed pattern is seen in the expression of cleaved-PARP1 (cPARP), which is produced during both apoptosis and necrosis[35] (Figure 6C).

Proliferation, identified by IDU incorporation, also exhibits a biwave profile, where cells divide more extensively around day 5 during the third phase of the DN state (DN3),[36,37] and then again at the DP state[38,39] between days 15 and 20 (Figure 6B). The cases in which the two methods exhibited different capacities in detecting peaks or falls in marker expression, rather than merely differing in the scaling of the time axis, are due to differences in pre-processing and in pseudotime axis construction. While Wishbone relies on nearest-neighbor graphs, diffusion maps and sampled waypoints along the trajectory to ultimately achieve ordering at a single-cell level and detect bifurcation points, we rely on local densities and assign cells into segments (rather order all cells) in which marker values are averaged. In addition, we averaged values from 12 mice that were harvested at different times after IDU injection, whereas Wishbone was performed on a selected mouse that was harvested 3 h post-IDU injection.

CD25 and CD44 are two markers that are known to have higher expression levels during the DN phase and decrease during the DP phase (SI Figure S5). Notably, the profiles of CD25 and CD44 were inferred primarily via dynamics in the expression of other markers who defined the trajectory route, as these two markers did not happen to participate in the sub-segmentation of any of the segments. Interestingly, while CD25 seems to decrease toward the DP phase and remain relatively low in the SP phase along both branches, CD44 decreases during the DP phase along the CD8-SP branch but rises toward the end of the CD4-SP branch.

The method described in this article not only infers the experimentally observed dynamics of the overall developmental trajectory, but identifies and highlights some more complex dynamic features of individual molecules (such as CD27, Helios, TCF1, and Notch1) which are compatible with the known complex biology of thymocyte differentiation.

DISCUSSION

Trajectory inference has gained wide popularity as a tool for analyzing single-cell omics data. Although many ordering approaches have evolved during the last few years, most methods are comprised of some sort of dimensionality reduction followed by trajectory inference based on expression profile similarities, and ultimately allocate cells along a continuous pseudo-temporal axis.[2,40]

Our study builds on this work by estimating cell fluxes along the developmental trajectory and hence rescaling the pseudotime axis to more closely reflect the real dynamics of thymic differentiation.

We initially formulate a method for pseudo-temporal ordering based upon expression densities, which is conceptually similar to existing methods but highlights the importance of local cell densities as preparation for the next step. While this method is less sensitive in detecting relatively rare cell populations that can be diluted within segments (such as the DN1 and DN2 cell populations) and better performs after initially outlining the trajectory in CD4/CD8 space, it is much less sensitive to initial condition selection (e.g. as seen in cases where an initial single cell is selected). Selection of the first cell in the approximate region of the initial segment is sufficient for the trajectory to later converge with the maximal density crest. We further incorporate apoptosis and division, and assuming steady-state, infer the real-time dimension. One limitation of our model is detecting subpopulations of cells that progress at different rates along the same trajectory. Mean marker values of the different populations in each segment would be calculated, and the inferred time would reflect a weighted average of the different cell populations. Future models better fit for detecting such cases would allow splitting these cell populations and rescaling the pseudotime axis separately. Another limitation of this study is its reliance on Caspase3 as the sole apoptosis marker in the thymus, which does not capture "death by neglect" known to occur during the DP phase. This limitation, however, is not unique to our model, since to our knowledge there is no well-established marker for tracking that mode of apoptosis. We were hence obliged to manually introduce this event in our model based on published estimates as to the rates of cell elimination during positive selection.

Developing T-cells in the mouse thymus is an ideal system for trajectory inference, since cells are known to continuously travel across developmental stages (characterized by different markers) in a well-defined period of time. Implementing our method, we were able to depict the observed course of T cell differentiation and extract temporal profiles for all markers that we measured. Importantly, regions of the developmental landscape that were compacted by traditional pseudotime inference could be expanded by incorporating additional markers revealing events of biological significance. This temporal reconstruction of the differentiation pathway can readily be applied to other complex developmental systems, and the trajectories can be further refined as higher dimensional datasets become available.

Limitations of the study

Since all cells with a similar marker expression profile are grouped into the same segment, our model does not detect subpopulations of cells that progress along the same trajectory at different rates. Instead, the inferred time would reflect a weighted average of the different populations. Another limitation of the study is detecting "death by neglect" that occurs during the DP phase, where apoptotic cells are rapidly eliminated by scavenger macrophages, with standard apoptosis markers such as Caspase3. We were hence obliged to explicitly adjust the apoptotic rate during the DP phase according to what is known in the literature. Finally, since cells in the early DN (DN1-2) phase are sparse in number, we had difficulty in detecting them, and our trajectory initiates at a later stage within the DN phase.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105826.

AUTHOR CONTRIBUTIONS

Conceptualization, N.F, S.RZ, and A.G.; methodology, A.G., S.RZ., and N.F.; software, A.G; formal analysis, A.G; investigation, S.N., T.M.S, and S.RZ.; writing – original draft, A.G. and S.RZ.; writing – review & editing, B.C and Y.E.A.; visualization, A.G.; supervision, N.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

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

2. Cannoodt, R., Saelens, W., and Saeys, Y. (2016). Computational methods for trajectory inference from single-cell transcriptomics. Eur. J. Immunol. 46, 2496–2506. https://doi.org/10.1002/eji.201646347.

3. Liu, Z., Lou, H., Xie, K., Wang, H., Chen, N., Aparicio, O.M., Zhang, M.Q., Jiang, R., and Chen, T. (2017). Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. Nat. Commun. 8, 22–28. https://doi.org/10.1038/s41467-017-00039-z.

4. Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 20, 1–9. https://doi.org/10.1186/s13059-019-1663-x.

5. Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'Er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat. Biotechnol. 34, 637–645. https://doi.org/10.1038/nbt.3569.

6. Koch, U., and Radtke, F. (2011). Mechanisms of T cell development and transformation. Annu. Rev. Cell Dev. Biol. 27, 539–562. https://doi.org/10.1146/annurev-cellbio-092910-154008.

7. Yui, M.A., and Rothenberg, E.v. (2014). Developmental gene networks: a triathlon on the course to T cell identity. Nat. Rev. Immunol. 14, 529–545. https://doi.org/10.1038/nri3702.

8. Sawicka, M., Stritesky, G.L., Reynolds, J., Abourashchi, N., Lythe, G., Molina-París, C., and Hogquist, K.A. (2014). From pre-DP, post-DP, SP4, and SP8 thymocyte cell counts to a dynamical model of cortical and medullary selection. Front. Immunol. 5, 14–19. https://doi.org/10.3389/fimmu.2014.00019.

9. Weinreb, C., Wolock, S., Tusi, B.K., Socolovsky, M., and Klein, A.M. (2018). Fundamental limits on dynamic inference from single-cell snapshots. Proc. Natl. Acad. Sci. USA 115, E2467–E2476. https://doi.org/10.1073/pnas.1714723115.

10. Fischer, D.S., Fiedler, A.K., Kernfeld, E.M., Genga, R.M.J., Bastidas-Ponce, A., Bakhti, M., Lickert, H., Hasenauer, J., Maehr, R., and Theis, F.J. (2019). Inferring population

dynamics from single-cell RNA-sequencing time series data. Nat. Biotechnol. *37*, 461–468. https://doi.org/10.1038/s41587-019-0088-0.

11. Farbehi, N., Patrick, R., Dorison, A., Xaymardan, M., Janbandhu, V., Wystub-Lis, K., Ho, J.W., Nordon, R.E., and Harvey, R.P. (2019). Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. Elife *8*, e43882. https://doi.org/10.7554/eLife.43882.

12. Kuritz, K., Stöhr, D., Maichl, D.S., Pollak, N., Rehm, M., and Allgöwer, F. (2020). Reconstructing temporal and spatial dynamics from single-cell pseudotime using prior knowledge of real scale cell densities. Sci. Rep. *10*, 3619. https://doi.org/10.1038/s41598-020-60400-z.

13. Seo, W., and Taniuchi, I. (2016). Transcriptional regulation of early T-cell development in the thymus. Eur. J. Immunol. *46*, 531–538. https://doi.org/10.1002/eji.201545821.

14. Germain, R.N. (2002). t-cell development and the CD4-CD8 lineage decision. Nat. Rev. Immunol. *2*, 309–322. https://doi.org/10.1038/nri798.

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

16. Szondy, Z., Garabuczi, É., Tóth, K., Kiss, B., and Köröskényi, K. (2012). Thymocyte death by neglect: contribution of engulfing macrophages. Eur. J. Immunol. *42*, 1662–1667. https://doi.org/10.1002/eji.201142338.

17. Yamashita, I., Nagata, T., Tada, T., and Nakayama, T. (1993). CD69 cell surface expression identifies developing thymocytes which audition for T cell antigen receptor-mediated positive selection. Int. Immunol. *5*, 1139–1150. https://doi.org/10.1093/intimm/5.9.1139.

18. Ross, J.O., Melichar, H.J., Au-Yeung, B.B., Herzmark, P., Weiss, A., and Robey, E.A. (2014). Distinct phases in the positive selection of CD8+ T cells distinguished by intrathymic migration and T-cell receptor signaling patterns. Proc. Natl. Acad. Sci. USA *111*, E2550–E2558. https://doi.org/10.1073/pnas.1408482111.

19. Kurd, N., and Robey, E.A. (2016). T-cell selection in the thymus: a spatial and temporal perspective. Immunol. Rev. *271*, 114–126. https://doi.org/10.1111/imr.12398.

20. Robert, P.A., Kunze-Schumacher, H., Greiff, V., and Krueger, A. (2021). Modeling the dynamics of t-cell development in the thymus. Entropy *23*, 437. https://doi.org/10.3390/e23040437.

21. Melichar, H.J., Ross, J.O., Herzmark, P., Hogquist, K.A., and Robey, E.A. (2013). Distinct temporal patterns of T cell receptor signaling during positive versus negative selection in situ. Sci. Signal. *6*, ra92. https://doi.org/10.1126/scisignal.2004400.

22. Hernandez, J.B., Newton, R.H., and Walsh, C.M. (2010). Life and death in the thymus-cell death signaling during T cell development. Curr. Opin. Cell Biol. *22*, 865–871. https://doi.org/10.1016/j.ceb.2010.08.003.

23. Breed, E.R., Watanabe, M., and Hogquist, K.A. (2019). Measuring thymic clonal deletion at the population level. J. Immunol. *202*, 3226–3233. https://doi.org/10.4049/jimmunol.1900191.

24. Yates, A.J. (2014). Theories and quantification of thymic selection. Front. Immunol. *5*, 13–15. https://doi.org/10.3389/fimmu.2014.00013.

25. le Campion, A., Lucas, B., Dautigny, N., Léaument, S., Vasseur, F., and Pénit, C. (2002). Quantitative and qualitative adjustment of thymic T cell production by clonal expansion of premigrant thymocytes. J. Immunol. *168*, 1664–1671. https://doi.org/10.4049/jimmunol.168.4.1664.

26. Le, J., Park, J.E., Ha, V.L., Luong, A., Branciamore, S., Rodin, A.S., Gogoshin, G., Li, F., Loh, Y.H.E., Camacho, V., et al. (2020). Single-cell RNA-seq mapping of human thymopoiesis reveals lineage specification trajectories and a commitment spectrum in T cell development. Immunity *52*, 1105–1118.e9. https://doi.org/10.1016/j.immuni.2020.05.010.

27. Chopp, L.B., Gopalan, V., Ciucci, T., Ruchinskas, A., Rae, Z., Lagarde, M., Gao, Y., Li, C., Bosticardo, M., Pala, F., et al. (2020). An integrated epigenomic and transcriptomic map of mouse and human αβ T cell development. Immunity *53*, 1182–1201.e8. https://doi.org/10.1016/j.immuni.2020.10.024.

28. Griesemer, A.D., Sorenson, E.C., and Hardy, M.A. (2010). The role of the thymus in tolerance. Transplantation *90*, 465–474. https://doi.org/10.1097/TP.0b013e3181e7e54f.

29. Hogquist, K.A., Baldwin, T.A., and Jameson, S.C. (2005). Central tolerance: learning self-control in the thymus. Nat. Rev. Immunol. *5*, 772–782. https://doi.org/10.1038/nri1707.

30. Goodnow, C.C., Sprent, J., Fazekas de St Groth, B., and Vinuesa, C.G. (2005). Cellular and genetic mechanisms of self tolerance and autoimmunity. Nature *435*, 590–597. https://doi.org/10.1038/nature03724.

31. Behbehani, G.K., Bendall, S.C., Clutter, M.R., Fantl, W.J., and Nolan, G.P. (2012). Single-cell mass cytometry adapted to measurements of the cell cycle. Cytometry A. *81*, 552–566. https://doi.org/10.1002/cyto.a.22075.

32. Kafri, R., Levy, J., Ginzberg, M.B., Oh, S., Lahav, G., and Kirschner, M.W. (2013). Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. Nature *494*, 480–483. https://doi.org/10.1038/nature11897.

33. Tyas, L., Brophy, V.A., Pope, A., Rivett, A.J., and Tavaré, J.M. (2000). Rapid caspase-3 activation during apoptosis revealed using fluorescence-resonance energy transfer. EMBO Rep. *1*, 266–270. https://doi.org/10.1093/embo-reports/kvd050.

34. Stritesky, G.L., Xing, Y., Erickson, J.R., Kalekar, L.A., Wang, X., Mueller, D.L., Jameson, S.C., and Hogquist, K.A. (2013). Murine thymic selection quantified using a unique method to capture deleted T cells. Proc. Natl. Acad. Sci. USA *110*, 4679–4684. https://doi.org/10.1073/pnas.1217532110.

35. Elmore, S. (2007). Apoptosis: a Review of programmed cell death. Toxicol. Pathol. *35*, 495–516. https://doi.org/10.1080/01926230701320337.

36. Laurent, J., Bosco, N., Marche, P.N., and Ceredig, R. (2004). New insights into the proliferation and differentiation of early mouse thymocytes. Int. Immunol. *16*, 1069–1080. https://doi.org/10.1093/intimm/dxh108.

37. Paiva, R.A., Sousa, A.G.G., Ramos, C.v., Ávila, M., Lilue, J., Paixão, T., and Martins, V.C. (2021). Self-renewal of double-negative 3 early thymocytes enables thymus autonomy but compromises the β-selection checkpoint. Cell Rep. *35*, 108967. https://doi.org/10.1016/j.celrep.2021.108967.

38. Ciofani, M., and Zúñiga-Pflücker, J.C. (2007). The thymus as an inductive site for T lymphopoiesis. Annu. Rev. Cell Dev. Biol. *23*, 463–493. https://doi.org/10.1146/annurev.cellbio.23.090506.123547.

39. Li, Y., Li, K., Zhu, L., Li, B., Zong, D., Cai, P., Jiang, C., Du, P., Lin, J., and Qu, K. (2021). Development of double-positive thymocytes at single-cell resolution. Genome Med. *13*, 49. https://doi.org/10.1186/s13073-021-00861-7.

40. van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., and Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. Nat. Commun. *11*, 1201. https://doi.org/10.1038/s41467-020-14766-3.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Mouse CD4 (RM4-5)-172 Yb	FLUIDIGM	RRID:AB_2924892.
Anti-Mouse CD8a (53-6.7)-153Eu	FLUIDIGM	RRID:AB_2924893
Anti-Mouse CD45 (30-F11)-89Y	FLUIDIGM	RRID:AB_2924894
Anti-Cleaved Caspase 3 (D3E9)-142Nd	FLUIDIGM	Cat: #3142004; RRID:AB_2847863
Anti-Human cleaved PARP (F21-852)-143Nd	FLUIDIGM	Cat: #3143011; RRID:AB_2927562.
Anti-Mouse ikaros 144Nd	Biolegend	Cat: #653302; RRID:AB_2561711
Anti-Mouse CD69 145Nd	FLUIDIGM	Cat: #3145005; RRID:AB_2895115
Anti-Mouse Notch1 147Sm	Biolegend	Cat: #130602; RRID:AB_1227721
anti-mouse CD90.2 149sm	FLUIDIGM	Cat: #105333; RRID:AB_2927564
Anti-Mouse CD24 150Nd	FLUIDIGM	Cat: #3150009; RRID:AB_2916042
Anti-Mouse CD25 151Eu	FLUIDIGM	Cat: #3151007; RRID:AB_2827880
Anti-Mouse CD3e 152Sm	FLUIDIGM	Cat: #3152004; RRID:AB_2687836
Anti-Human/Mouse Bcl11b - 154Sm	Biolegend	Cat: #6506; RRID:AB_10915967
Anti-Mouse CD366 (Tim-3) - 155Gd	Biolegend	Cat: #1197; RRID:AB_1626128
Anti-Mouse CD196-156Gd	FLUIDIGM	Cat: #3156016; RRID:AB_2895120
Anti-Human/Mouse pStat3 158Gd	FLUIDIGM	Cat: #3158005, RRID:AB_2661827
Anti-Mouse TCRgd (GL3) 159 Tb	FLUIDIGM	Cat: #3159012; RRID:AB_2922919
Anti-Mouse CD5 160Gd	FLUIDIGM	Cat: #3160002; RRID:AB_2927568
anti-mouse/human CD11b 161Dy	Biolegend	Cat: #101202,PRID:AB_312785
Anti-Mouse CD11c 161Dy	Biolegend	Cat: #117341; RRID:AB_256280
Anti-Mouse CD19 161Dy	Biolegend	Cat: #1155; RRID:AB_2629714
Anti-Human/Mouse/Rat pLck 162Dy	FLUIDIGM	Cat: #3162004; RRID:AB_2827886
Anti-Human/Mouse RUNX3 163Dy	Biolegend	Cat: #697902; RRID:AB_2687207
Anti-Mouse CD62L 164Dy	FLUIDIGM	Cat: #3164003; RRID:AB_2885021
Anti-Mouse Foxp3-165Ho	FLUIDIGM	Cat: #3165024; RRID:AB_2687843
Anti-Mouse CD117 (ckit) - 166Er	FLUIDIGM	Cat: #3166004; RRID:AB_2801435
Anti-Human/Mouse GATA3 167Er	FLUIDIGM	Cat: #3167007; RRID:AB_2927569
Anti-Human/Mouse/Rat Ki-67 168Er	FLUIDIGM	Cat: #3168007; RRID:AB_2800467
Anti-Mouse TCR b 169 Tm	FLUIDIGM	Cat: #3169002; RRID:AB_2827883
Anti-Mouse NK1.1 170Er	FLUIDIGM	Cat: #3170002; RRID:AB_2885023
Anti-Human/Mouse CD44 171 Yb	FLUIDIGM	Cat: #3171003; RRID:AB_2895121
anti Mouse TCF7/TCF1-173Yb	R&D Systems	Cat: #MAB8224; RRID:AB_1771413
Anti-Mouse CD197 (CCR7) 174 Yb	Biolegend	Cat: #1201; RRID:AB_389229
Anti-Mouse CD127 (IL7Ra) 175Lu	FLUIDIGM	Cat: #3175006; RRID:AB_2927570
Anti-Human/Mouse Helios - 176 Yb	Biolegend	Cat: #1372; RRID:AB_10900638
Chemicals, peptides, and recombinant proteins		
IDU (*5-Iodo-2′-deoxyuridine*)	Sigma-Aldrich	Cat: #I7125
Critical commercial assays		
Cell-ID™ 20-Plex Pd Barcoding Kit	FLUIDIGM	Cat# 201060
cell-ID Cisplatin	FLUIDIGM	Cat: #201064
Cell-ID Intercalator-Ir	FLUIDIGM	Cat: #201192

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MIBItag Conjugation Kit_144Nd	FLUIDIGM	Cat: #600144
MIBItag Conjugation Kit_147Sm	FLUIDIGM	Cat: #600147
MIBItag Conjugation Kit_154Sm	FLUIDIGM	Cat: #600154
MIBItag Conjugation Kit_155Gd	FLUIDIGM	Cat: #600155
MIBItag Conjugation Kit_161Dy	FLUIDIGM	Cat: #600161
MIBItag Conjugation Kit_163Dy	FLUIDIGM	Cat: #600163
MIBItag Conjugation Kit_173 Yb	FLUIDIGM	Cat: #600173
MIBItag Conjugation Kit_174 Yb	FLUIDIGM	Cat: #600174
Deposited data		
CyTOF data	FLOWRespository	http://flowrepository.org/id/FR-FCM-Z5UP
Experimental models: Organisms/strains		
C57BL/6JOlaHsd	Envigo	Cat:#: 057
B6.SJL-*Ptprc*ᵃ*Pepc*ᵇ/BoyJ	Jackson	Cat:#:002,014
Software and algorithms		
MATLAB	MathWorks	https://www.mathworks.com/products/matlab.html
FLOWJO	BD Biosciences	
WISHBONE	Dana Peer lab	https://github.com/dpeerlab/wishbone
Trajectory Inference	Avishai Gavish	https://github.com/avishai-g/trajectory_inference

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Shlomit Reich-Zeliger (Shlomit.reich-zeliger@weizmann.ac.il).

Materials availability

This study did not generate new unique reagents.

There are no restrictions to the availability of newly generated materials in this study.

Data and code availability

- All original code has been deposited on our GitHub repository https://github.com/avishai-g/trajectory_inference, and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- CyTOF data have been deposite at FLLOWRespository, DOIs are listed in the key resources tableAny additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Female C57BL/6 and B6SJL mice (six weeks old) were obtained from Envigo and Jackson Laboratories. All mice were housed at the Weizmann Institute in compliance with institutional ethics approvals (from the Institutional Animal Use Committee of Weizmann institute).

METHOD DETAILS

Thymus data and mass cytometry

Thymocytes were isolated from 6-week-old C57Bl mice thymus. Thymocytes were treated in red blood cell lysis buffer and then resuspended in staining media and counted. Cells were stained with metal-conjugated antibodies according to manufacturer's protocol (Table S1). Briefly around 200k cells were stained

with cell-ID TM Cisplatin (Fluidigm) (5 min RT). Next cells were stained with surface antibodies (30 min RT), and fixed with 1.6% PFA (10 min RT). After permeabilization with 100% ice-cold Methanol (15 min, 4C), the cells were stained with intracellular antibodies (30 min, RT). Finally the cells were labeled with Iridium DNA intercalator for DNA content and analyzed by CyTOF mass cytometry using CyTOF2. Data was normalized using bead normalized with bead standards.

We collected data from 12 independent C57Bl mice thymuses

IDU injection

IDU (5-Iodo-2'-deoxyuridine) was purchased from Sigma-Aldrich (ref. I7125). All mice were injected in their peritoneum with a solution of 0.4 mg of IDU diluted in 200 μL of sterile PBS, adjusted to a pH of 8.5–9. Thymocytes were isolated 1, 3, 6 and 12h post injection. Cells were stained with metal-conjugated antibodies according to manufacturer's protocol (Table S1). Briefly, around 200k cells were stained with cell-ID TM Cisplatin (Fluidigm) (5 min RT). Next cells were stained with surface antibodies (30 min RT), and fixed with 1.6% PFA (10 min RT). After permeabilization with 100% ice-cold Methanol (15 min, 4 °C), the cells were stained with intracellular antibodies (30 min, RT). Finally, the cells were labeled with Iridium DNA intercalator for DNA content and analyzed by CyTOF mass cytometry using CyTOF2. Data were normalized using bead normalized with bead standard.

CD45.1 Thymocyte adoptive transfer

Thymucytes were purified from B6SJL (CD45.1) mice. The cells were stained with CD4, CD8, CD44 and CD25 and sorted to the CD4⁻CD8 double-negative population. 10,000 cells in 50 mL of the DN population were adoptively transferred directly to the thymus of each C57BL/Hj (CD45.2) mice. Thymocytes were isolated 1, 3, 10, 14, 21 and 28 days post adoptive transfer. Cells were stained with various antibodies and analyzed by Flow cytometer.

Data processing and parameters

Here we first describe the pipeline for pre-processing data before inferring the trajectory path and calculating time. We later give the parameters we chose for results presented in the main text and sections above.

Data pre-processing

Pre-processing the data before trajectory inference included the following steps:

Raw data

Raw data from a single mouse thymus, including expression levels of unsynchronized differentiating single-cells measured by CyTOF, was loaded to a custom written MATLAB program. We analyzed altogether 12 mice, including mice harvested 1, 4, 6 and 12 h after IdU injection (3 mice in each group). Initial cell counts for each mouse ranged between 40,000 and 174,000 (with a mean of 97,000 and a SE of 16,500). Notably, cells were not synchronized, representing all stages of maturation in the thymus.

Expression levels of the following markers were measured for every cell

Gating

To gate out only T-cells from the raw data, we first used the inverse hyperbolic sine transformation, $Marker \rightarrow arsinh(\frac{Marker}{cofactor})$, choosing cofactor = 5. T-cells were gated by choosing cells with low levels of 'Non T-Cells' markers (see Table S1), with the transformed marker threshold set at ~5 (Figure S8).

Outlier removal

we removed any cell with expression levels above the 99.99 percentile, for any of these markers:

CD4, CD8, CD45, CD27, CD69, CD90, CD24, CD25, CD3, CD336, CD62l, CD5, CD117, CD44, CD197, CD127.

Normalization

Normalizing each marker by its mean value was performed, since marker value outputs from CyTOF are arbitrary and are only significant for comparing signaling strengths between cells for a given marker, but not for comparing different marker values.

Parameters

The trajectory was initially calculated in two-dimensions (CD4/CD8 space) as described, where the trajectory was divided into 60 evenly-spaced points from the initial starting point until the point of bifurcation. Using the same interval between points, the trajectory was extended from the bifurcation point along each of the SP states. A detour from the path leading to the next anchor was allowed if the density in a neighboring position on the grid was 1.5 times that of the mean of all neighboring cells and twice that of the density in the default point.

After the trajectory position was determined, cells were assigned to the closest trajectory point based on Euclidian distance, and the time equation was solved numerically (Figure 2B and supplemental information section 3). For sub-segmentation we picked the 8 most varying markers in each segment. We chose a different grid density for each segment based on the time spent in the segment that was calculated in two dimensions (Figure 4). The maximal allowed grid size in 8-dimensions was of size 10 for the segment where the maximal time was spent, and other segments were assigned grid sizes proportional to the ratio in times compared to the maximal time.

QUANTIFICATION AND STATISTICAL ANALYSIS

CyTOF data analysis

After the acquisition, data were performed using the CyTOF software for normalization. Barcodes were deconvoluted using the CyTOF "debarcoding" function. The residual normalization beads, debris, dead cells, and doublets were exclusively gated using CYTOBANK.