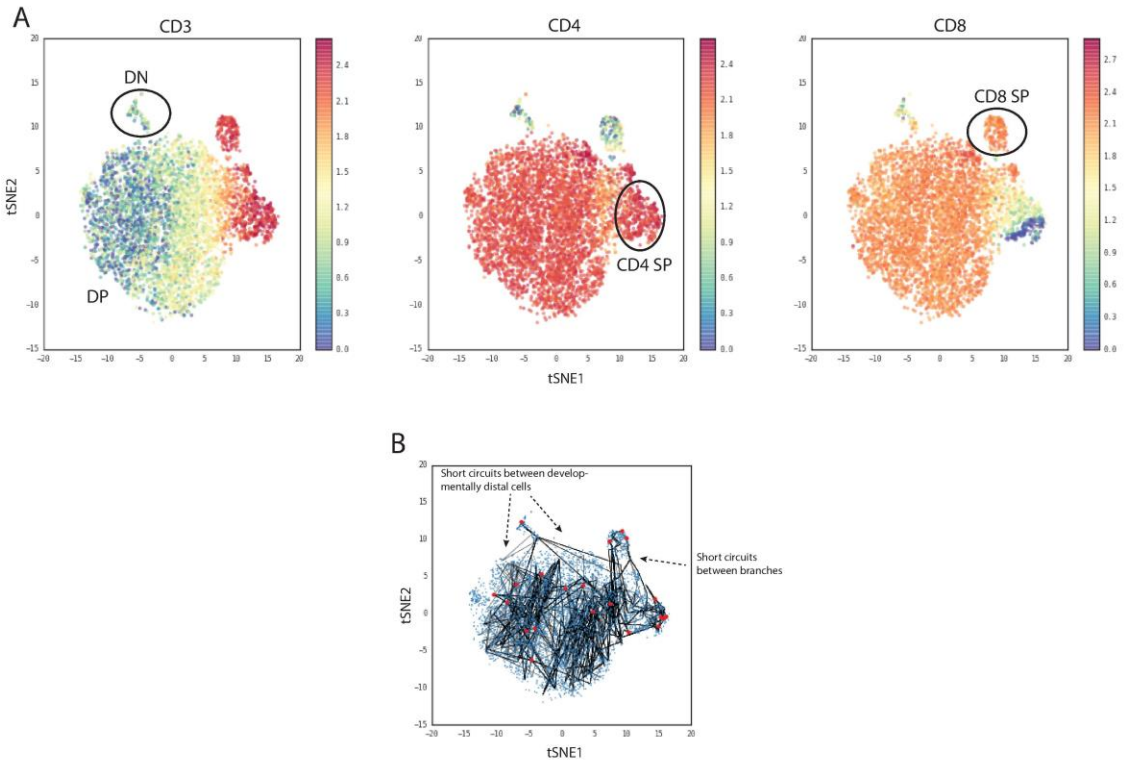
Table of contents

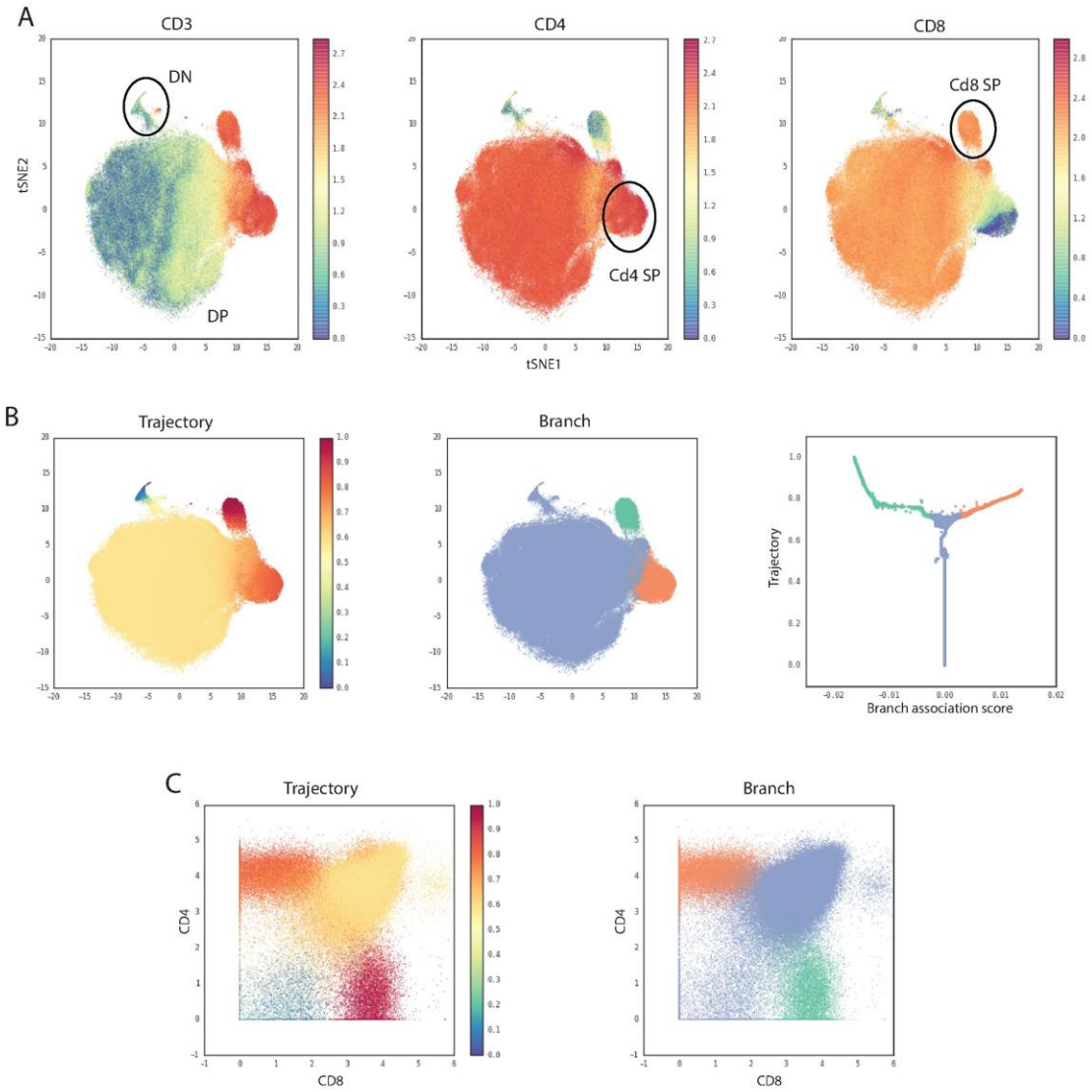
Supplementary figures



Supplementary Fig. 1: Short-circuits in nearest neighbor graphs

(A) tSNE maps showing the expression of CD3, CD4 and CD8

(B) Plot highlighting the short circuits in the data between developmentally distal cells and in between cells of the two SP branches. The waypoints are shown in red and black lines represent shortest paths between all pairs of waypoints.

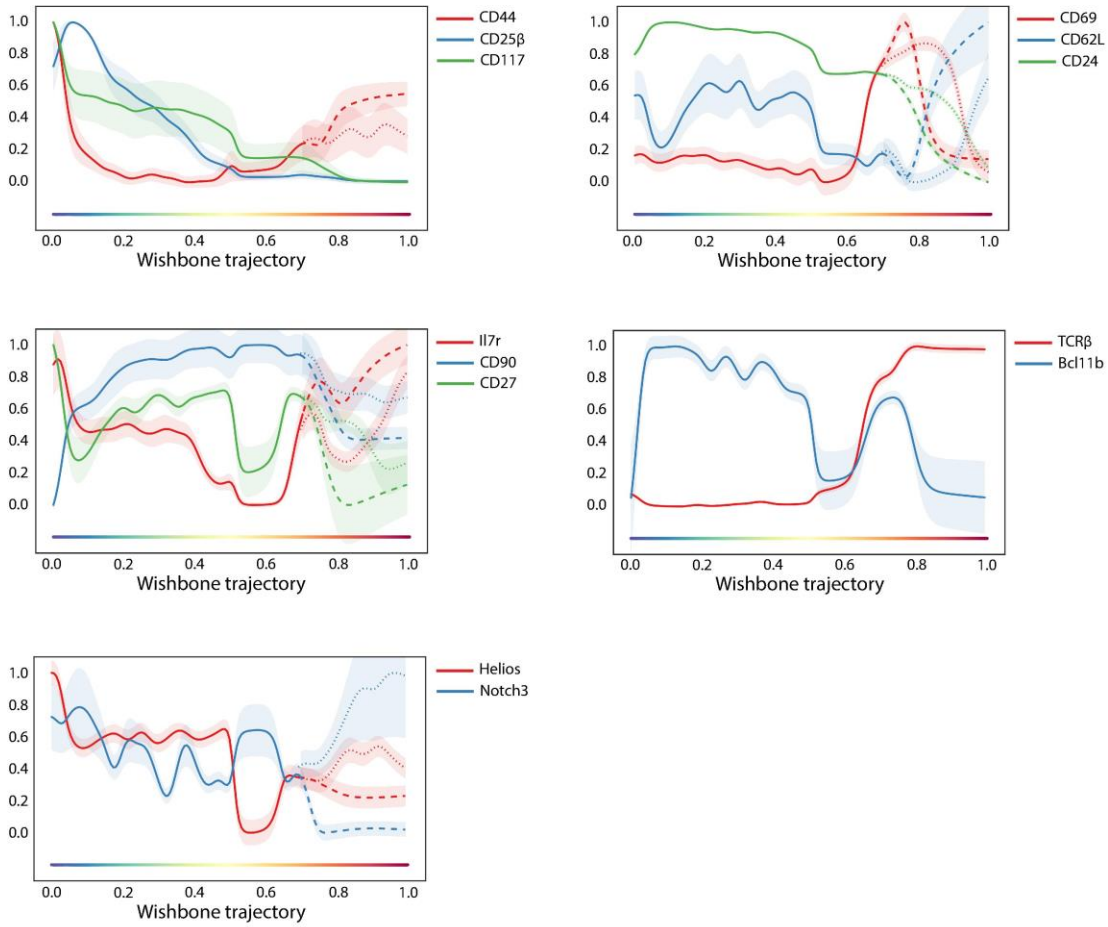


Supplementary Fig. 2: Thymus T cell populations and Wishbone results

(A) tSNE map of mouse thymus replicate 1 showing the DN, DP and two SP populations based on expression of CD3, CD4 and CD8.

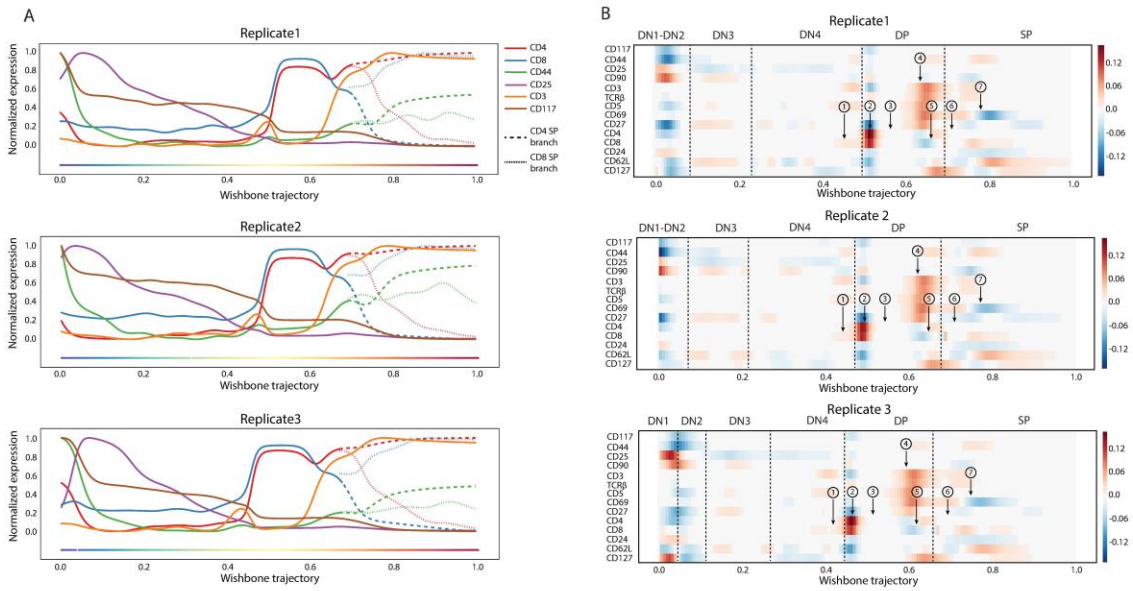
(B) tSNE map colored by the inferred Wishbone trajectory (left panel) and the branch associations (middle panel). The branch association scores as a function of Wishbone trajectory are shown in the right panel.

(C) Trajectory and branch associations shown on the CD8-CD4



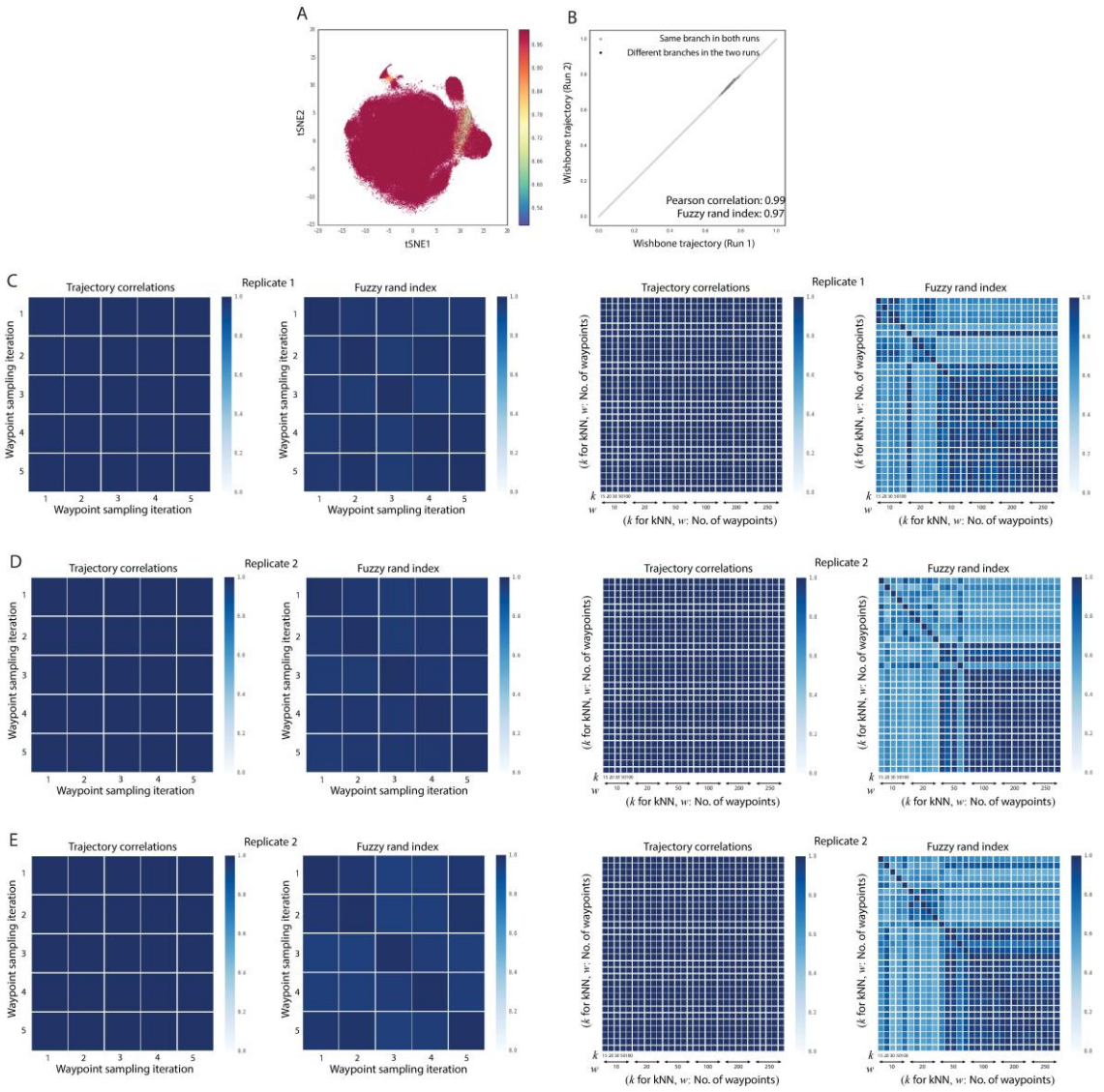
Supplementary Fig. 3: Variance of markers along trajectory

Variance of the markers along Wishbone trajectory is very tight irrespective of whether the marker is used for learning. This is a companion plot to Figure 2D and shows the variance of markers not shown in Figure 2D.



Supplementary Fig. 4: Wishbone results are reproducible across replicates

The marker trends and their derivatives are reproducible across three independent mouse thymus replicates.



Supplementary Fig. 5: Wishbone results are robust to parameter choice

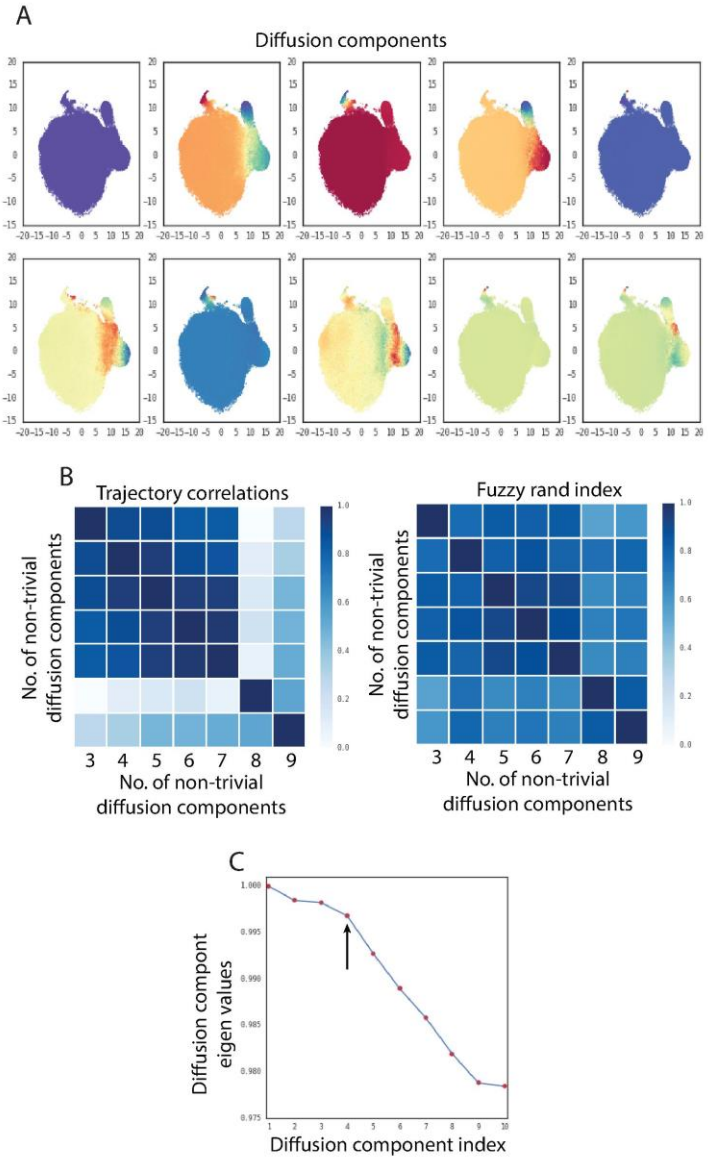
(A) tSNE map showing the probability of most confident branch association for each cell. This probability as expected is low close to the branch point and high elsewhere.

(B) Comparison of Wishbone results for two independent samples of waypoints. Cells with disagreement in branch associations in the two runs are shown in darker shade. The trajectories are highly correlated (Pearson correlation of 0.99)

and the branch associations are also consistent between the two runs (Fuzzy rand index of 0.97).

(C) Heatmaps showing the robustness of Wishbone trajectory and branches for different waypoint samples in replicate 1, with the correlation never falling below 0.99 and fuzzy rand-index never below 0.95 (Two left panels). (Right panels) Similar plots showing robustness of trajectory and branches for different combinations of parameters: k, the number of neighbors for nearest neighbor graph construction and nW, the number of waypoints. The trajectories are correlated across combinations of k and nW (Pearson correlation > 0.99). The branch associations are stable for nW > 100 (Fuzzy rand-index > 0.87, with Fuzzy rand-index > 0.95 for nW =250).

(D-E) Same as in C, for replicates 2 and 3.

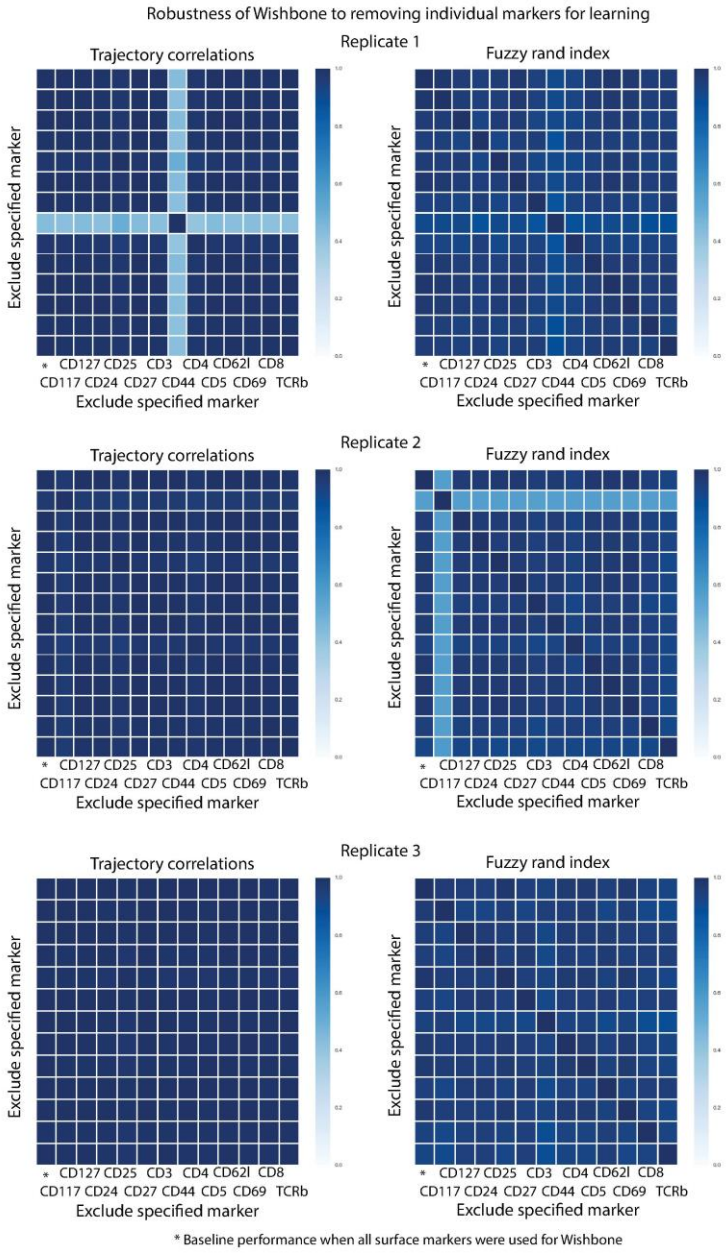


Supplementary Fig. 6: Diffusion components of the mouse thymus data and robustness of Wishbone results to number of components

(A) Top 10 diffusion components of replicate 1 of the mouse thymus. The first component is associated with eigen value 1 and is the trivial component.

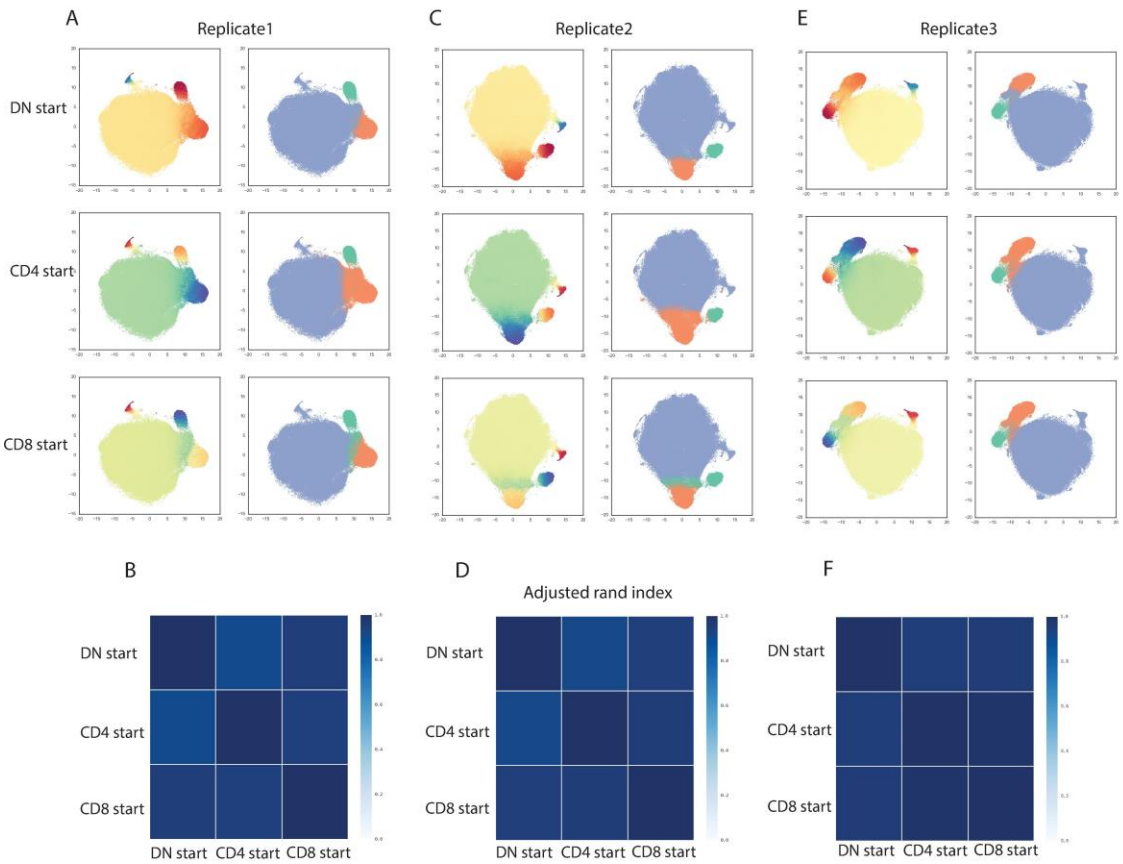
(B) Wishbone trajectory and branch associations are robust to the number of non-trivial components used for learning

(C) A large eigen gap between the 4th and 5th eigen values indicates that components 2, 3 and 4 are suitable for learning.



Supplementary Fig. 7: Wishbone results are robust to exclusion of individual markers

Wishbone trajectory and branch associations are largely robust to exclusion of individual markers across the three replicates. The results are compared to baseline performance derived when all the surface markers were used for Wishbone.

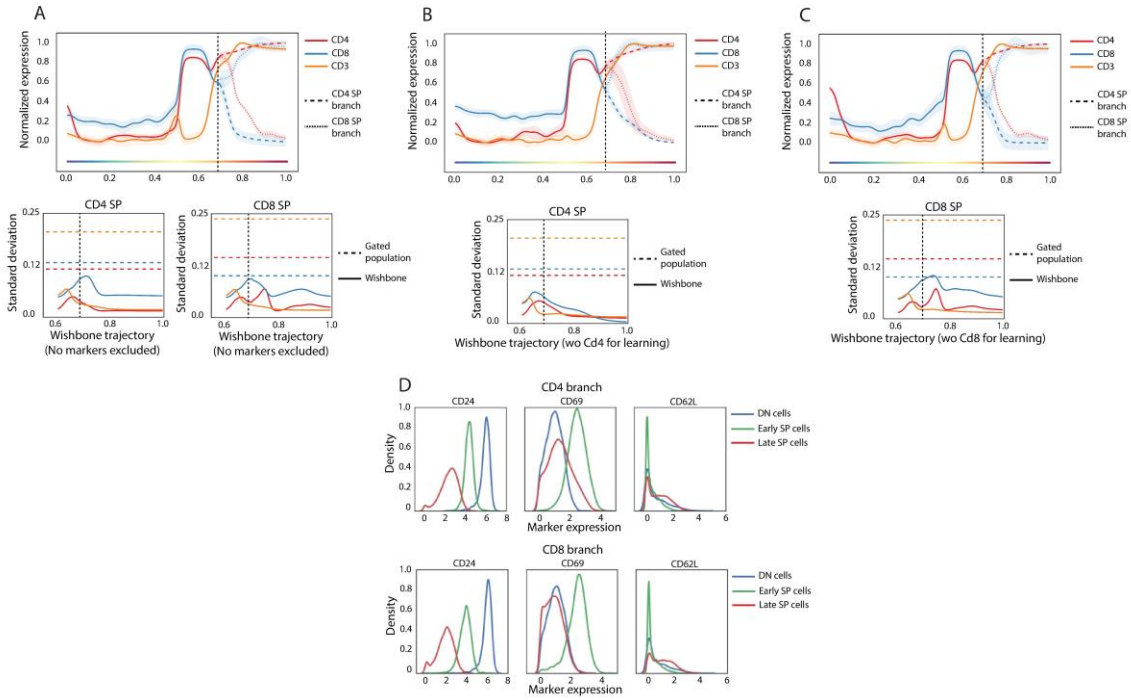


Supplementary Fig. 8: Wishbone results to choice of starting branch

(A) tSNE maps showing the trajectory and branches for replicate 1 using one of the DN cells (top panel), CD4 cells (middle panel) and CD8 cells (bottom panel) as the input early cell.

(B) The branches across different choice of start cells are consistent except for an expected degree of uncertainty at the branch point (Fuzzy rand index > 0.85).

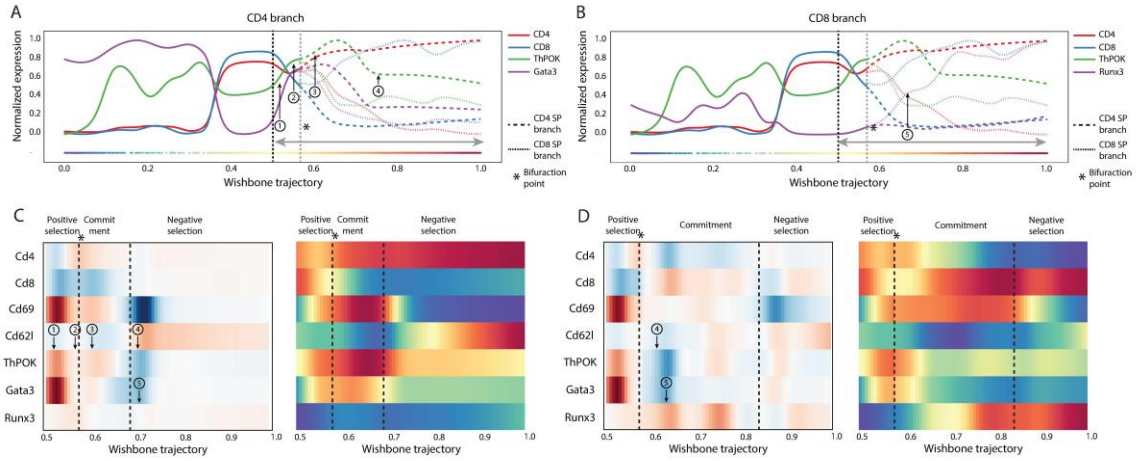
(C-F) Same as in (A-B) for replicates 2 and 3.



Supplementary Fig. 9: Variance of markers in gated populations is in part explained by comparison of developmentally distinct cells

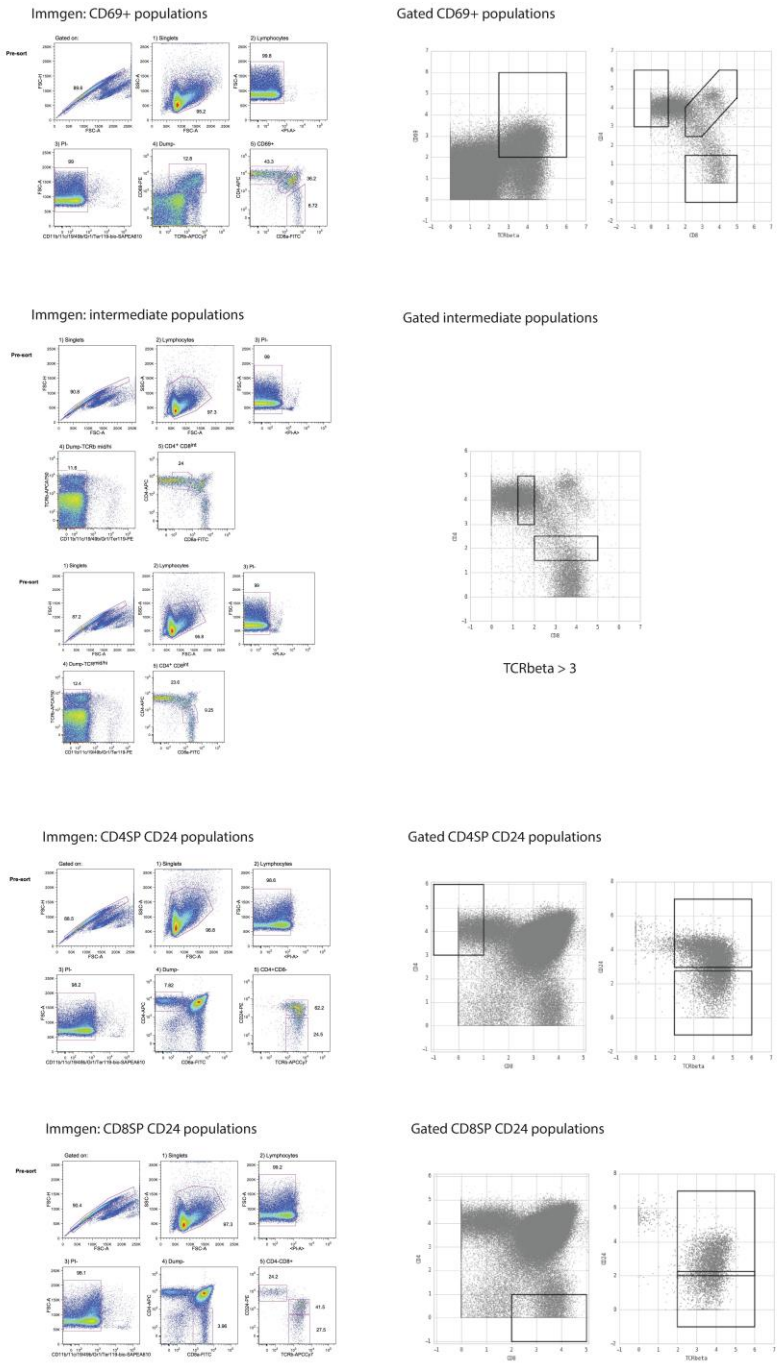
(A) Top panel shows the marker trends for CD3, CD4 and CD8 along T cell differentiation trajectory. CD4 and CD8 SP populations were identified by gating (Figure 4A) and the marker variances within the gated populations were compared variance along the trajectory. Bottom left panel shows this comparison for the CD4 branch. Variance along trajectory is shown in solid lines whereas the population variance is shown in dotted lines. Bottom right panel shows corresponding results for the CD8 branch.

(B-C) Same as in A with CD4 and CD8 respectively excluded for learning.



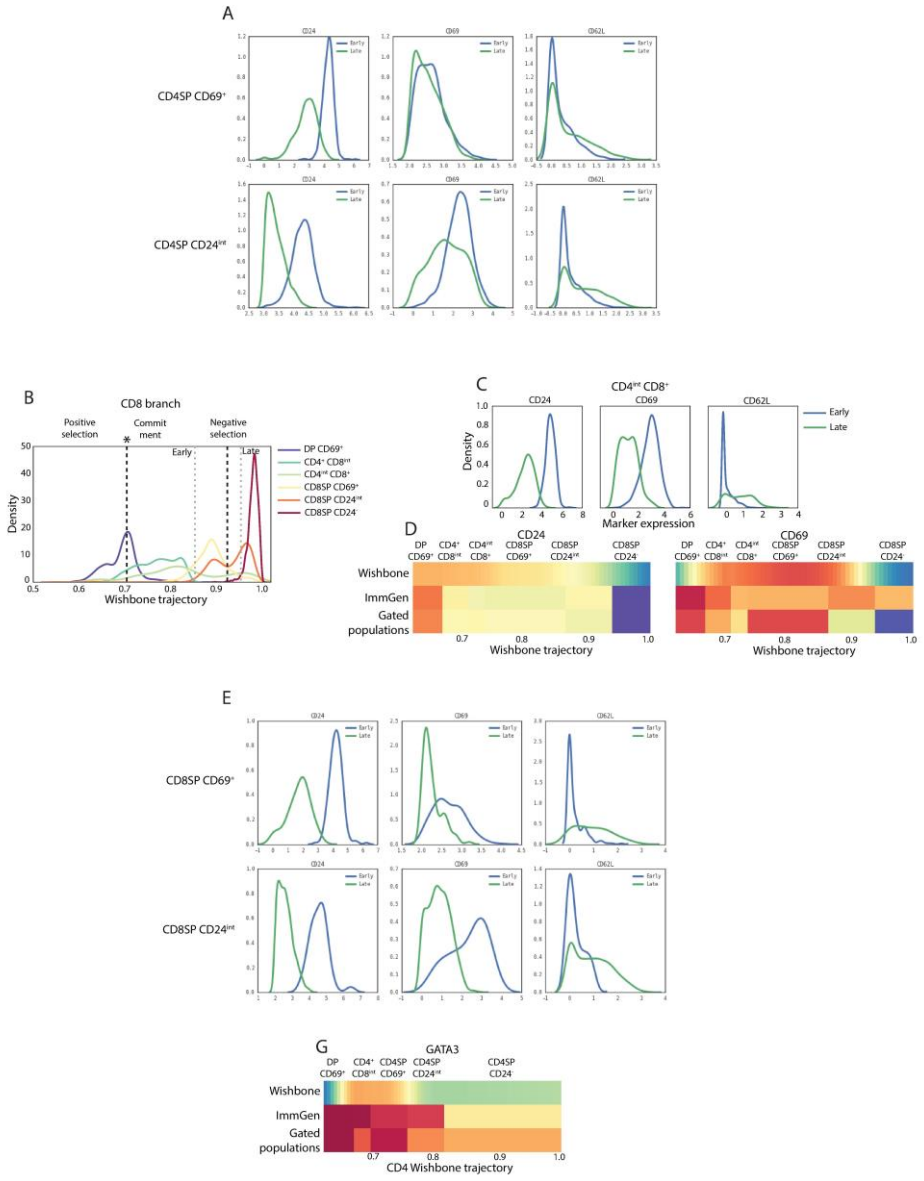
Supplementary Fig. 10: Transcription factor dynamics in SP populations

Companion to Figure 4A-D showing reproducibility of TF dynamics in an independent replicate.



Supplementary Fig. 11: Gating of cells using the ImmGen gating scheme

The left panels show the gating scheme used by ImmGen to identify different populations along SP maturation. Right panels show the corresponding gating of the mass cytometry data. Note that we used Phenograph (Materials and methods) to remove debris and identify lymphoid cells.

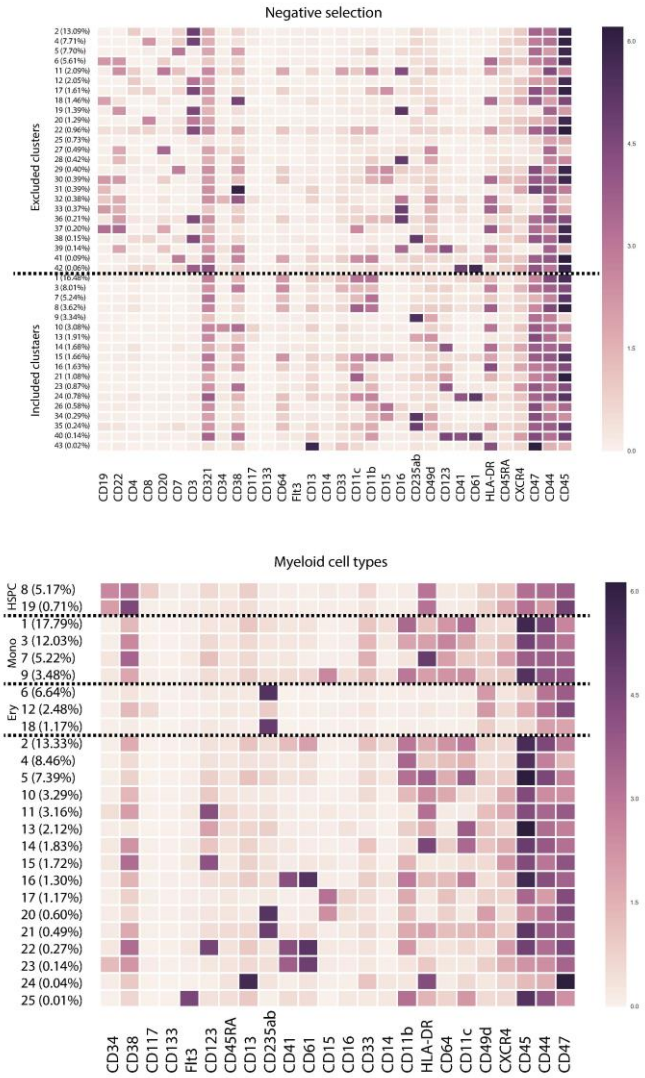


Supplementary Fig. 12: Comparison of Wishbone trajectories to gated populations

(A) Companion to Figure 4E-G showing that the Early and Late cells of CD4SP CD24^{int} and CD4SP CD24^{-} gates demonstrate landmarks of immature and mature cells respectively (p < 1e-6, Kolmogorov-Smirnov test).

(B-E) Similar to Figure 4E-G with results for the CD8 branch instead of CD4 branch.

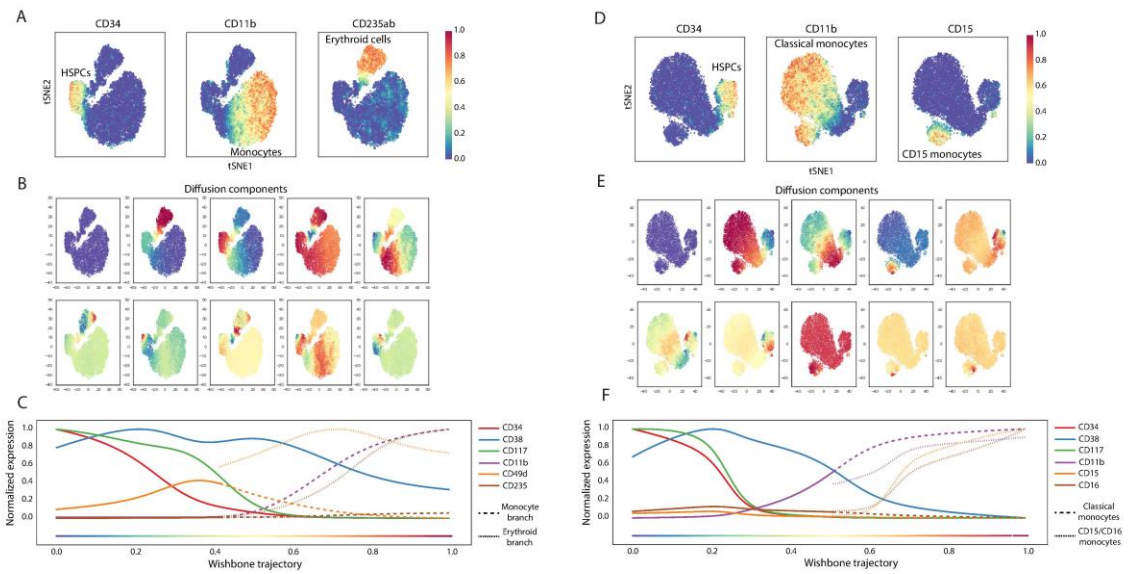
(G) Comparison of GATA3 marker dynamics in the CD4 branch.



Supplementary Fig. 13: Identification of cell types in the human myeloid dataset

Phenograph was run on all the cells and the non-myeloid clusters were removed based on expression of the following markers (Top panel): CD19, CD22, CD20 (Bcells), CD4, CD8, CD3 (T cells), CD7 (NK cells).

Phenograph was then rerun on the remaining clusters using only the myeloid markers to identify the different myeloid populations (Bottom panel) based on expression of: CD34, CD38 (HSPC – hematopoietic stem and progenitor cells), CD64, CD11b, CD11c (Monocytes and their progenitors) and CD235ab (Erythroid lineage).



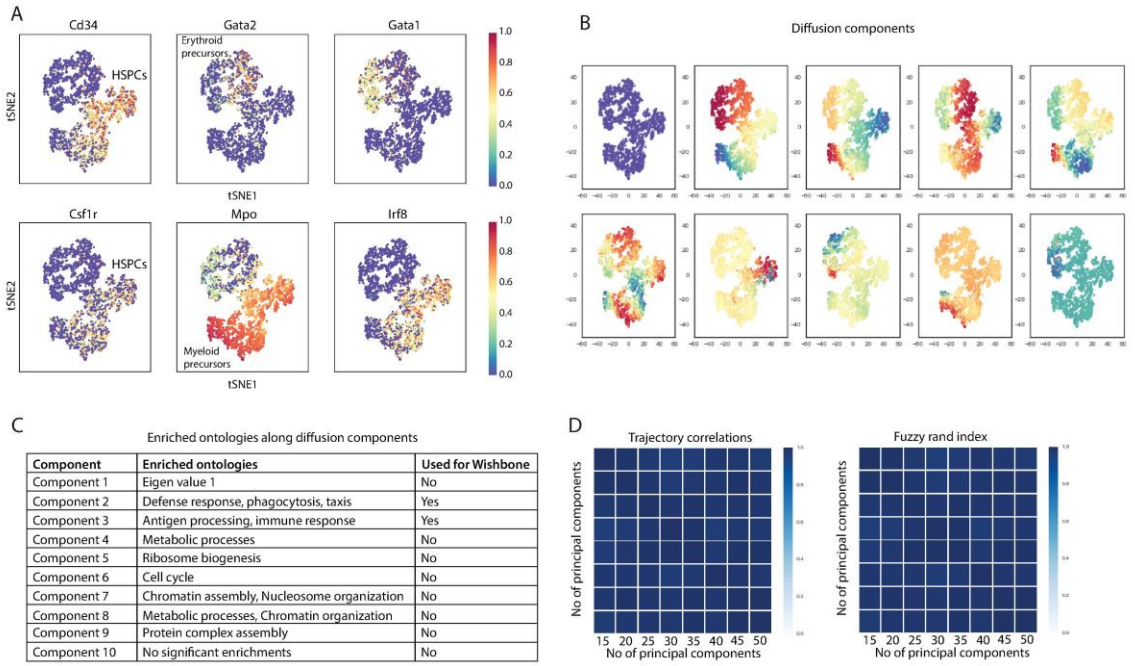
Supplementary Fig. 14: Myeloid datasets to evaluate Wishbone generalization

(A) tSNE map showing the expression of characteristic markers used to build the monocyte-erythroid cell dataset.

(B) Top 10 diffusion components of the monocyte-erythrocyte dataset.

(C) Marker trends show the trajectory starting at HSPCs (high CD34). This is followed by a downregulation of CD34 and CD117. The two mature populations: Monocytes and erythroid cells are correctly identified as the two branches with respective upregulation of CD11b and CD235ab.

(D-F) Same as in (A-C) showing marker expression and trends for the classical monocyte-CD15 monocyte dataset.



Component	Enriched ontologies	Used for Wishbone
Component 1	Eigen value 1	No
Component 2	Defense response, phagocytosis, taxis	Yes
Component 3	Antigen processing, immune response	Yes
Component 4	Metabolic processes	No
Component 5	Ribosome biogenesis	No
Component 6	Cell cycle	No
Component 7	Chromatin assembly, Nucleosome organization	No
Component 8	Metabolic processes, Chromatin organization	No
Component 9	Protein complex assembly	No
Component 10	No significant enrichments	No

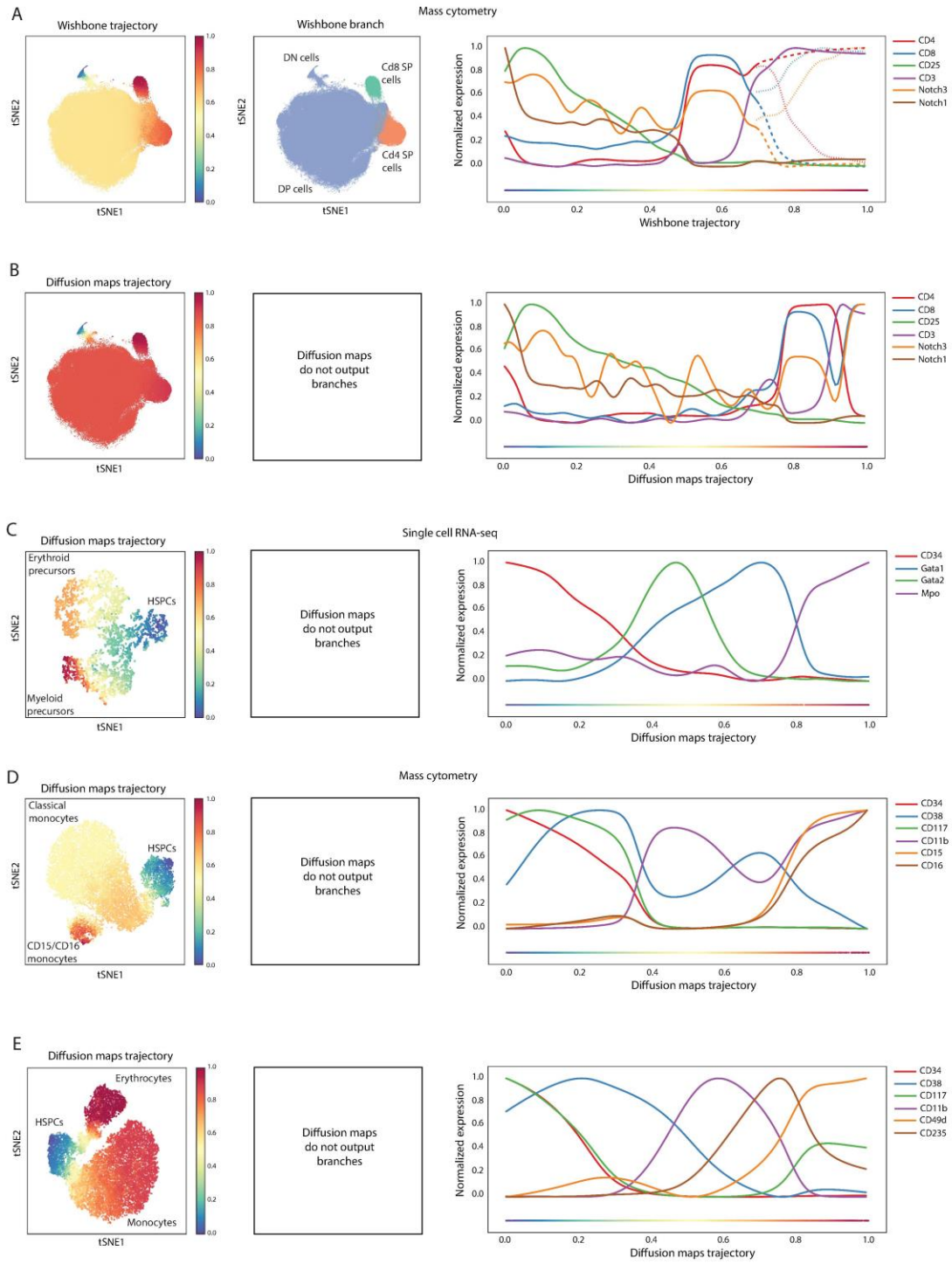
Supplementary Fig. 15: Wishbone generalizes to single-cell RNA-seq data

(A) tSNE maps showing the characteristic expression genes used to identify the cell type identities of single-cell RNA-seq profiles

(B) Top 10 diffusion components of the single-cell RNA-seq dataset shown in A.

(C) Gene ontology enrichments for the top 10 components derived using gene set enrichment of analysis.

(D) Wishbone results are robust to the number of principal components, which are used as a preprocessing step to account for the drop-outs in single cell RNA-seq data.



Supplementary Fig. 16: Comparison of Wishbone results to diffusion maps

(A) tSNE maps showing the trajectory and branch associations of Wishbone for the mouse thymus data (left and middle panels). The right panel shows the

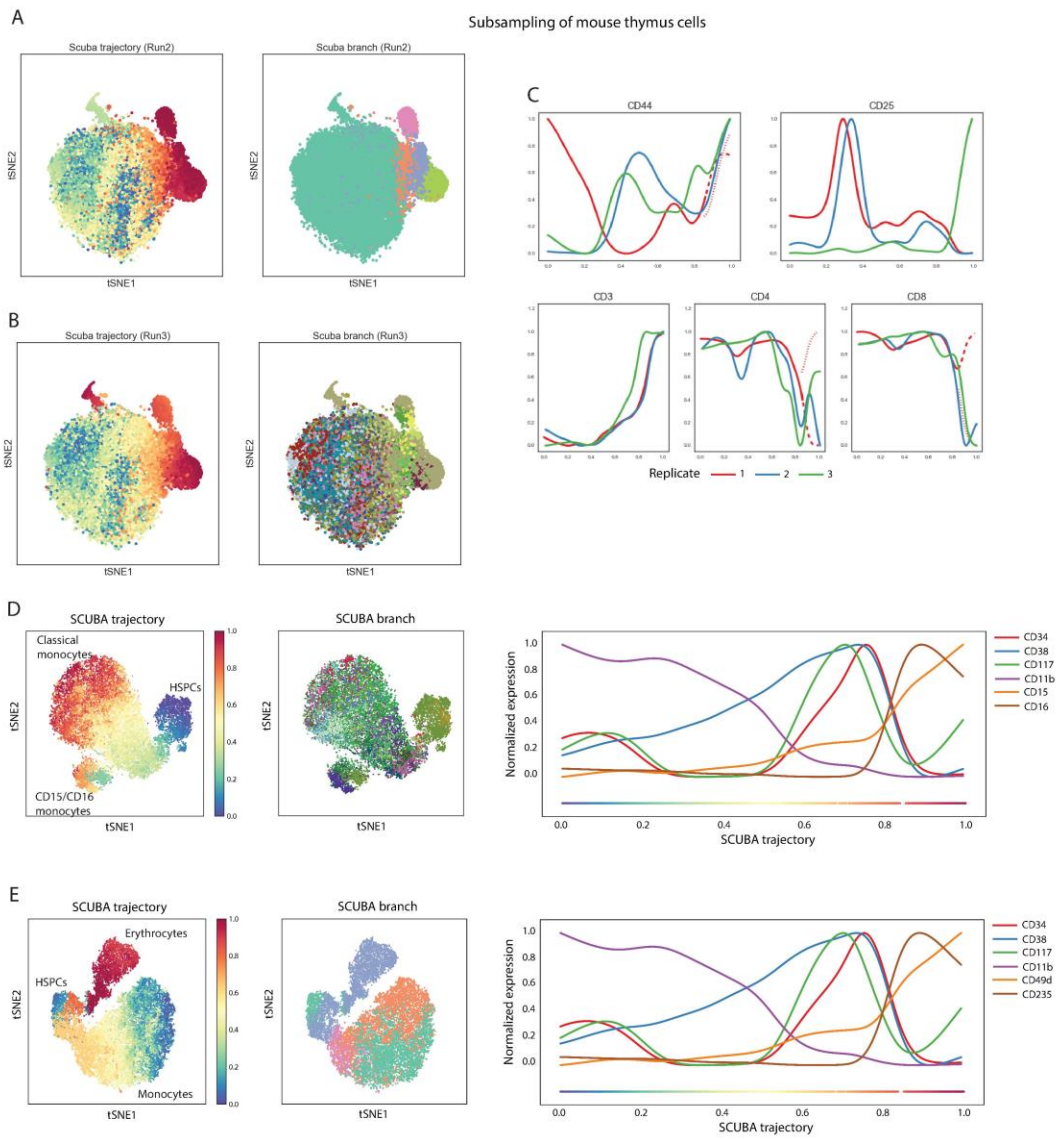
dynamics of markers used for learning such as CD4, CD8, CD25 and CD3 and markers not used for learning such as Notch1 and Notch3.

(B) Diffusion map ordering of cells determined as a Euclidean distance of all cells from the start cell in the reduced dimensionality space. Diffusion maps correctly recover the known stages in T cell development but suffers from a loss of resolution particularly in DP and SP stages. The loss of resolution is also highlighted by the noisy fluctuations of Notch1 and Notch3 in the DN stage.

(C) Diffusion maps accurately recover the ordering of cells in the single-cell RNA-seq dataset.

(D) Diffusion map trajectory of the monocyte dataset in human myeloid system places the mature classical monocytes earlier to their progenitors (Compare to Figure 5D).

(E) Diffusion map trajectory derived as Euclidean distance in the reduced dimensionality space accurately recovers the ordering in the monocyte-erythroid cell dataset.

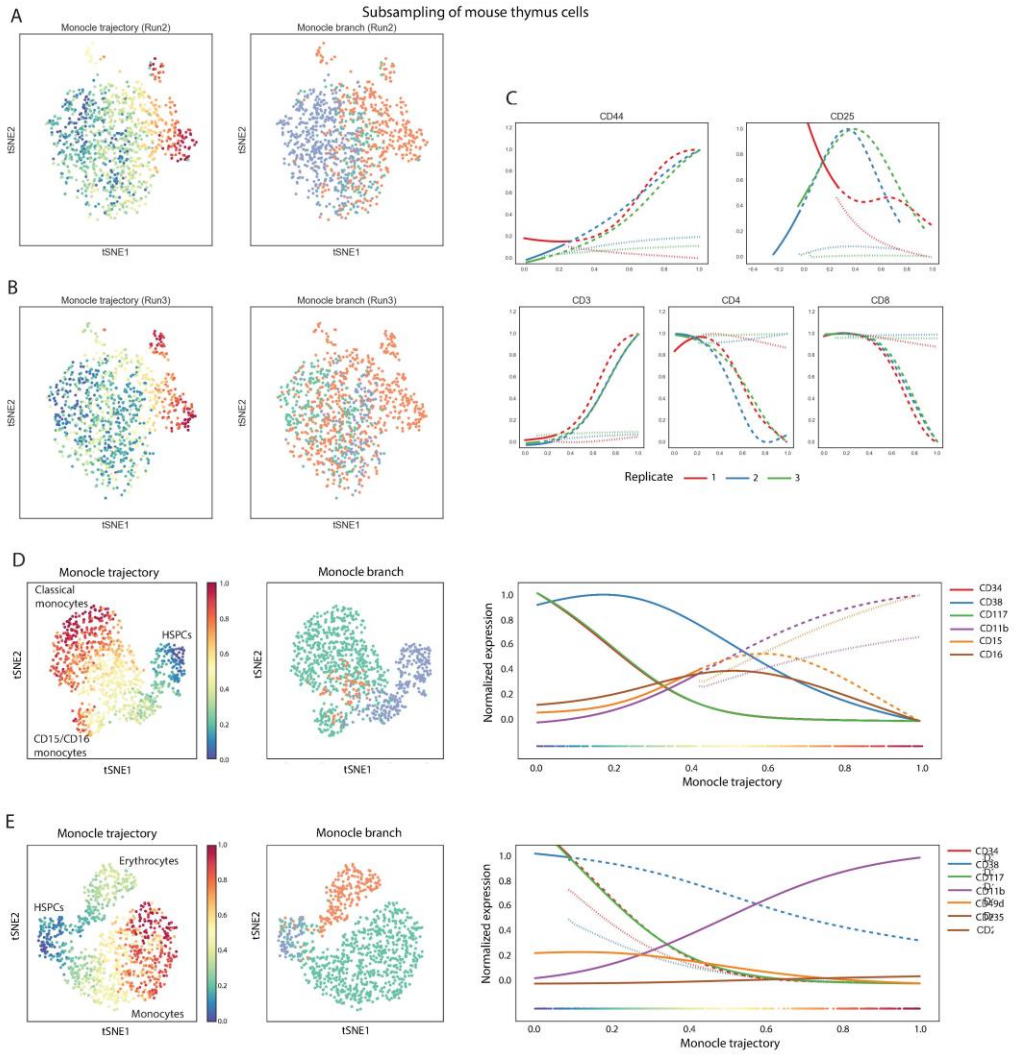


Supplementary Fig. 17: Comparison of Wishbone results to SCUBA

(A-C) Repeated subsampling of cells from replicate 1 of mouse thymus results in inconsistent ordering and branching of cells by SCUBA. Some runs do however recover coherent branches (Figure 6C).

(D) SCUBA accurately recovers the ordering of human myeloid cells and the marker dynamics are largely consistent with known biology. SCUBA however results in a large number of incoherent branches.

(E) SCUBA fails to correctly order the cells and SCUBA branches do not correspond to the underlying populations.

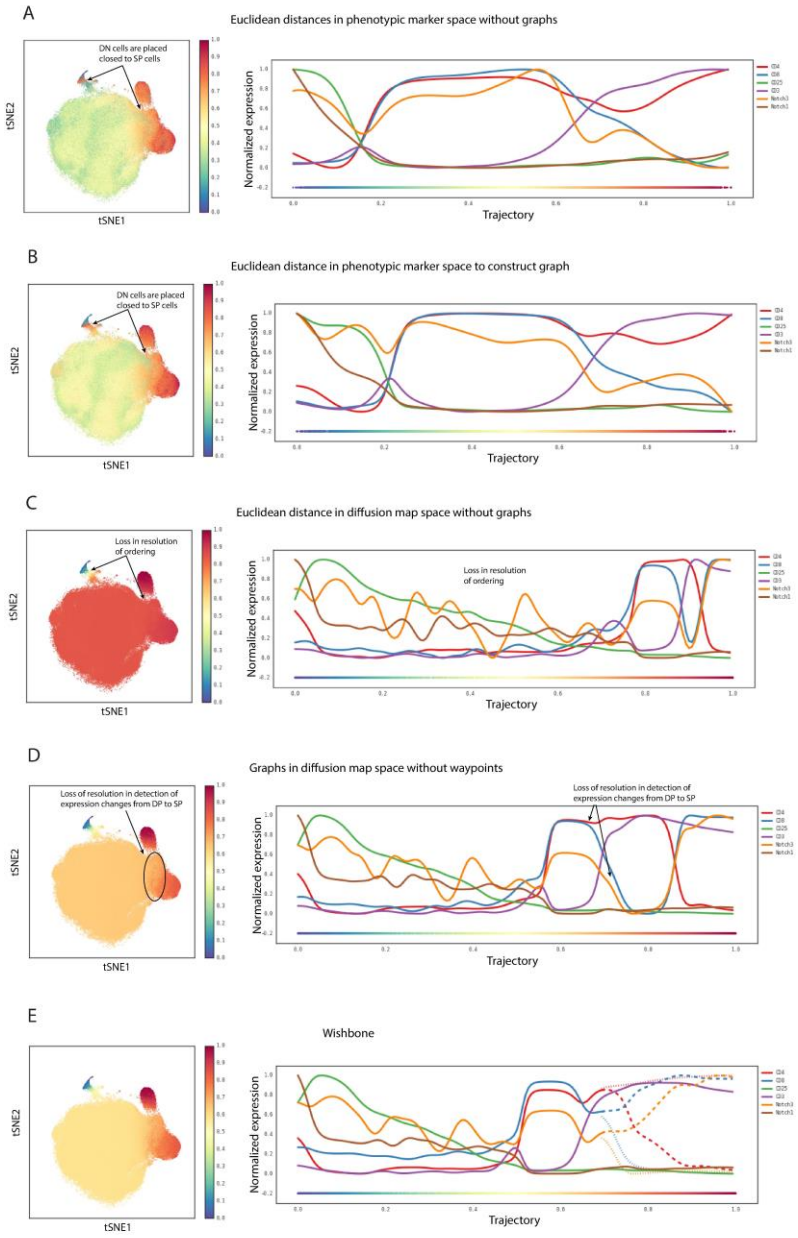


Supplementary Fig. 18: Comparison of Wishbone results to Monocle

(A-C) Plots showing Monocle results for different subsamples of cells. None of the runs yielded the right branching in Monocle with both SP branches invariably grouped together.

(D) Monocle also accurately recovers the ordering in a random sample of human myeloid cells but the branches again do not correspond to the mature populations.

(E) Monocle order a subset of the HSPC cells incorrectly and the same subset of HSPCs are assigned to erythrocyte and monocyte branches.



Supplementary Fig. 19: Principles of Wishbone and their role in achieving accurate, high-resolution trajectories.

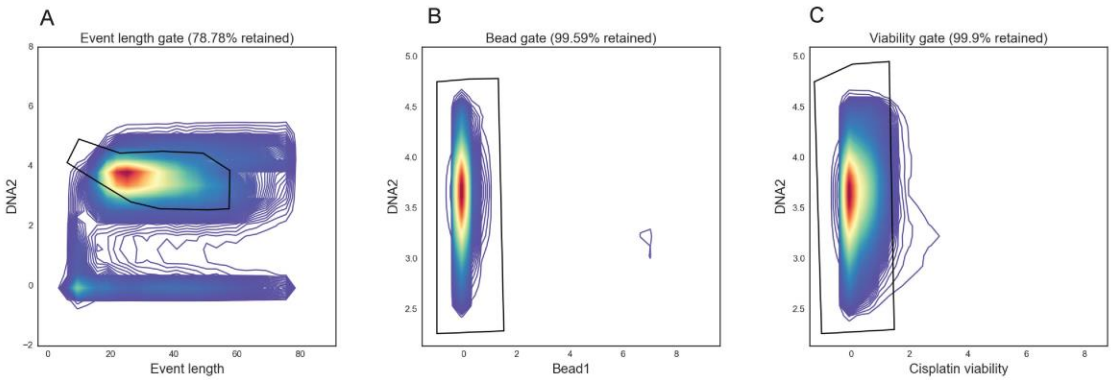
(A) tSNE map showing the trajectory determined by ordering of cells by Euclidean distance from the start cell without using graphs or waypoints. The right panel shows the dynamics of markers along the inferred trajectory. DN cells are placed closed to SP cells since the non-linear relationships are not adequately captured by Euclidean distance.

(B) Use of k-nearest neighbor graph without diffusion maps also leads to similar issues because of short circuits in the data.

(C) Diffusion maps provide a non-linear clean up of the data. Ordering in diffusion space without use of graph correctly recovers the known stages of T cell development but leads to a loss of resolution beyond the SP stage.

(D) Nearest neighbor graphs in diffusion maps help improve the resolution but key subtle expression changes are not inferred when waypoints are not used for local refinement of trajectories.

(E) Wishbone results, derived with nearest neighbor graphs in the diffusion map space with use of waypoints for refinement, are accurate, robust and achieve high-resolution.

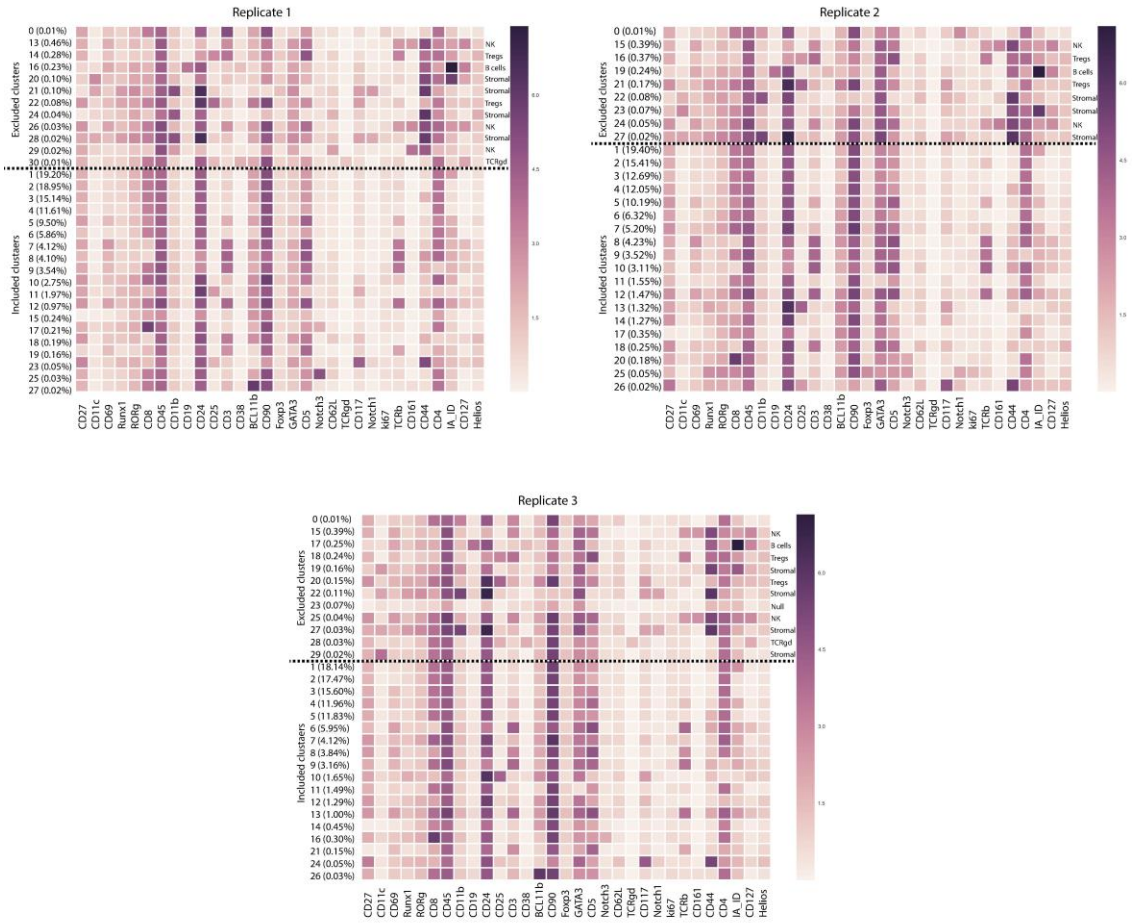


Supplementary Fig. 20: Gating scheme to identify viable cells in the data

(A) Gating on event length and DNA channel was used to remove doublets and debris.

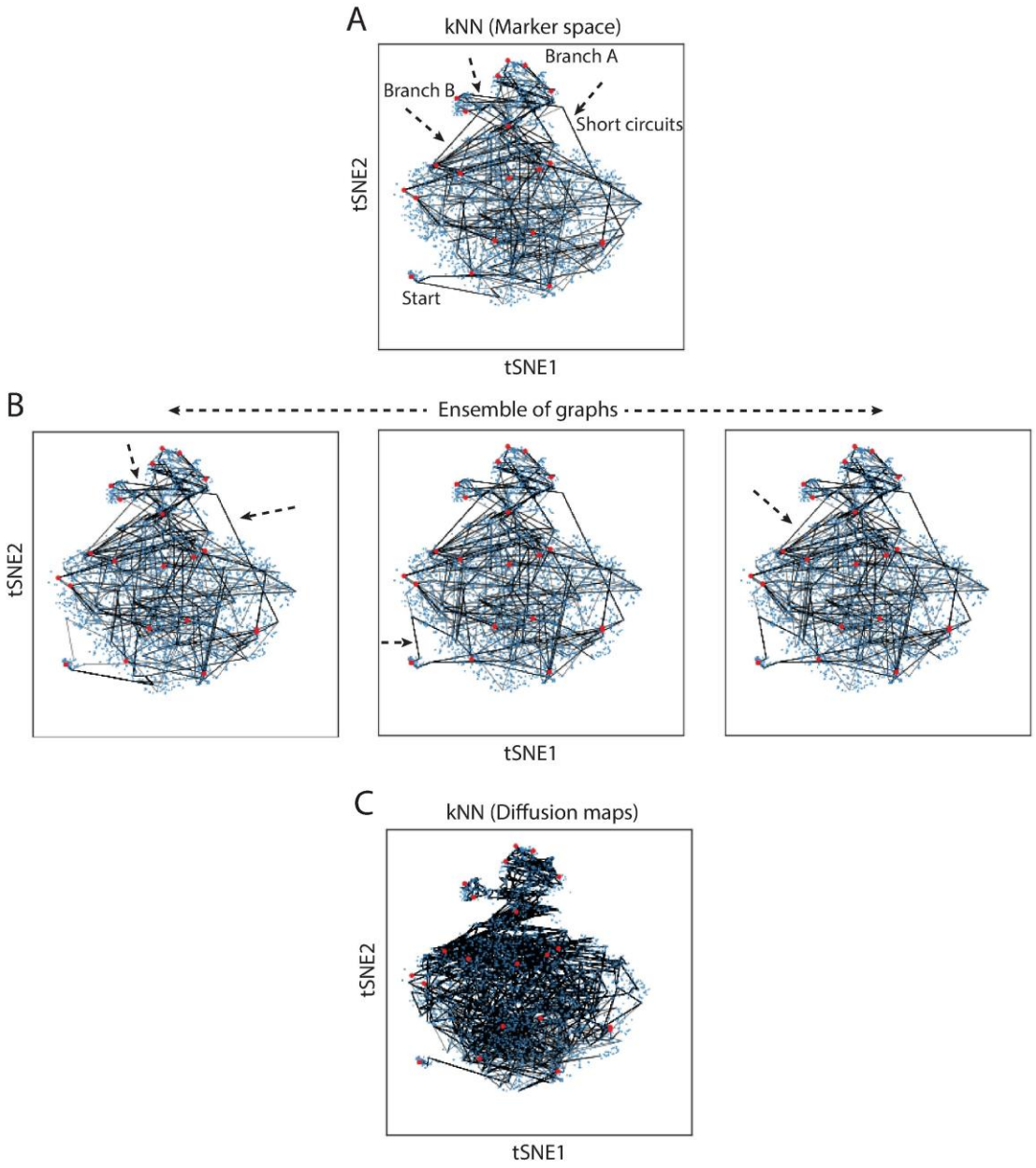
(B) Bead channel is used to remove beads left over from the first gating.

(C) Finally viable cells are selected using the Cisplatin viability and DNA channels.



Supplementary Fig. 21: Mouse thymus data clean up

Phenograph [1] was used to identify the different phenotypic populations or clusters in the data. Heatmaps show the mean expression of markers in each cluster identified by Phenograph. Non T cell clusters were removed based on expression of following markers: CD19 (B cells), CD161 (NK cells), CD11b/CD11c (APCs). In addition, Tregs ($CD4^{+}CD25^{+}$) and TCR$\gamma\delta$ cells were also removed before running Wishbone.

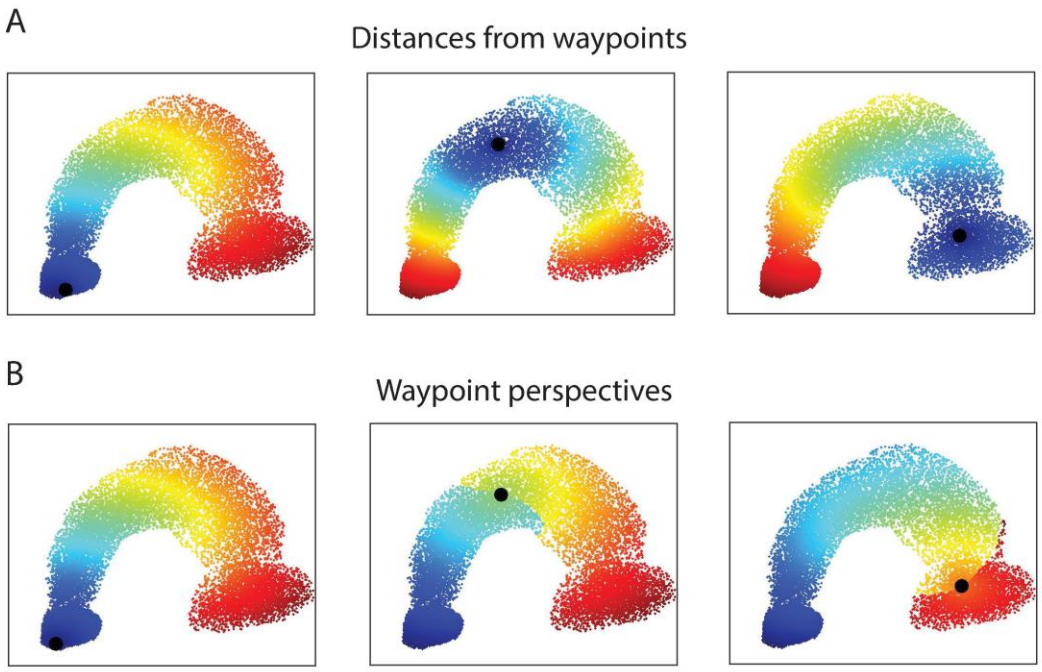


Supplementary Fig. 22: Short circuits in branching datasets.

(A) k -nearest neighbor graph in the phenotypic marker space creates a large number of short circuits in branching datasets.

(B) Ensemble of graphs is not sufficient to clean the data of all the short circuits.

(C) De-noising using diffusion maps removes the short circuits and the graph now faithfully captures developmental relationships between cells.

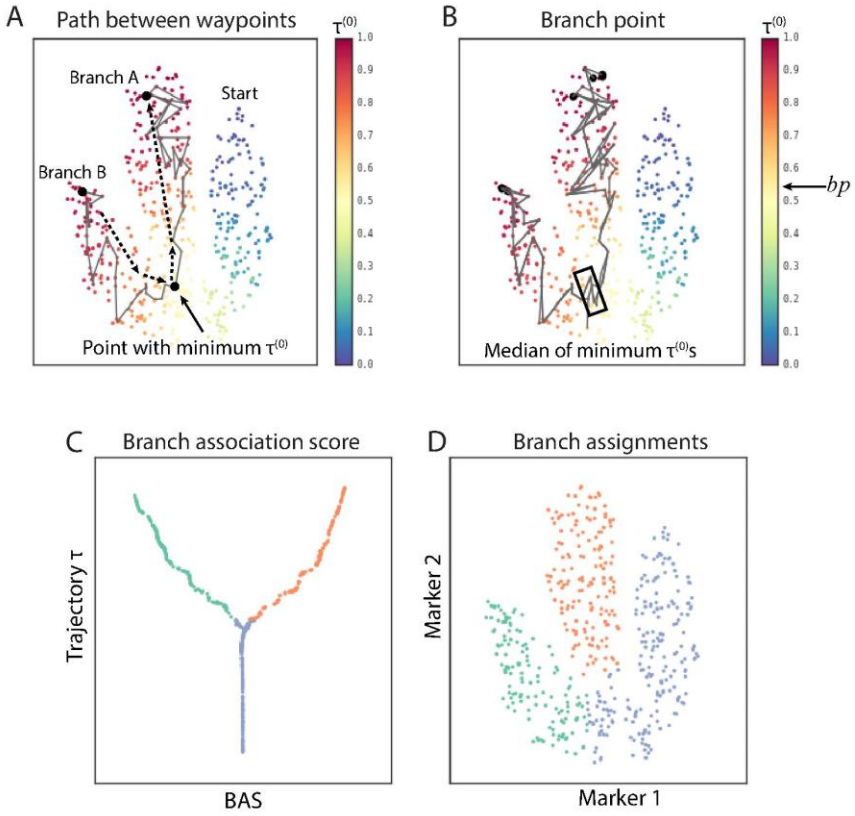


Supplementary Fig. 23: Waypoints and perspectives.

(A) Wishbone uses cells sampled throughout the trajectory called waypoints to overcome the additive noise of the shortest path distances. As with the distances from early cell, the distance of each cell from each waypoint is determined using the shortest path algorithm.

(B) The waypoint distances are then aligned with the early cell distances to determine waypoint perspectives. A weighted average over all the perspectives is used for calculating the refined ordering of cells.

Nature Biotechnology: doi:10.1038/nbt.3569



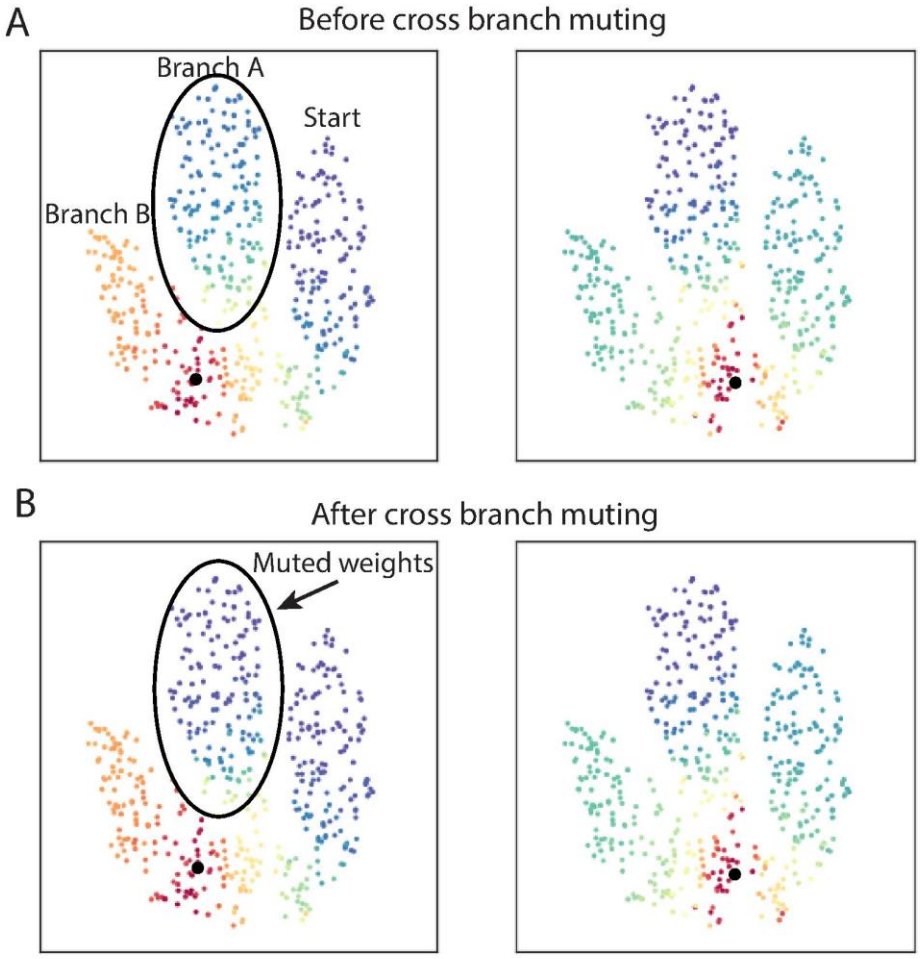
Supplementary Fig. 24: Branch point identification and branch assignments to cells.

(A) Wishbone uses second Eigen vector of Q matrix to identify the 5 most mature waypoints on both branches. A path from such a waypoint on Branch B to another on Branch A traverses through Branch B, changes direction and then traverses through Branch A. The cell with minimum trajectory value on such a path i.e., cell that is closest to the start represents the region with phenotypic properties of both the branches.

(B) Wishbone determines the set of cells with minimum trajectory values on paths between all pairs of waypoints on the two branches and identifies the branch point as the median of these minimum values.

(C-D) Plot showing the BAS or branch association score for each cell calculated from the second Eigen value of the Q matrix. The branch association scores are used to determine branch assignments for all cells. Any cell with $BAS < bp$ is

assigned to the trunk and cells with $BAS > bp$ are assigned to one of the two branches depending on the sign of BAS



Supplementary Fig. 25: Cross branch muting.

(A) Plots showing the weight or degree of influence in ordering of cells by the highlighted waypoints. The left panel shows that a waypoint on Branch B can influence ordering of cells in Branch A. The weights are exponentially reduced for cells on the different branch using the BAS scores.

(B) The muted weights are shown in the left panel with the waypoint in Branch B no longer influencing the cells in Branch A. The weights of waypoints on the trunk remained unaltered (middle panel)

Supplementary Note 1: Wishbone accurately recovers differentiation events with high resolution and expression trends of key TFs

Runx1, Bcl11b and Notch1 provide a test for accuracy of Wishbone trajectory since these markers were not used for learning the trajectory. The abundance of all these markers is consistent with their known roles in DN stages of T cell development: (a) Runx1 is necessary for CD4 suppression in DN stages [2] and is high in the DN cells, (b) Bcll11b knockout leads to arrest of cells in DN2-4 stages with seemingly no effect on DN1 cells [3] and consistent with these studies, Bcl11b is highly expressed starting from DN2 till DN4 and (c) Notch1 mediated signaling is central in DN cells [4] and correspondingly Notch1 is specifically detected in DN cells

Wishbone also accurately recovers known molecular events that punctuate T cell differentiation with high resolution. First, we observe that CD8 is upregulated before CD4 upregulation in the transition from DN to DP (Figure 2E(1)). This upregulation matches prior observations of immature CD8 single positive cells, a small fraction of $CD8^+$ and $CD4^-$ cells that occur between the DN to DP transition [5, 6]. Then the two lineage markers are significantly upregulated in transition to the DP stage (Figure 2E(2)) and subsequently remain relatively stable throughout the DP stage following their initial upregulation (Figure 2E(3)).

We next observe the hallmarks of positive selection and transition from the DP to SP stage with a coordinated change in multiple markers: downregulation of the lineage markers, upregulation of TCRβ, TCR complex component CD3, co stimulatory molecule CD5 and activation marker CD69 (Figure 2E(4)). While individually, the role of each marker has been characterized in the DP to SP transition [7, 8], the coordinated and synchronized change in expression of these markers has not been characterized.

The initial downregulation of the lineage markers is followed by a further downregulation of CD8 (Figure 2E(5)). This is consistent with the prevailing

model of lineage commitment which states that downregulation of CD8 leading to a loss of signaling through MHC Class I molecules, in turn leading to loss in TCR signal, is necessary for CD8 lineage commitment [9, 10]. CD4 selected cells continue along the CD4 lineage with a sustained TCR signal [10], resulting in bifurcation of the two lineages (Figure 2E(6)). The detection of branch point following downregulation of CD8 also indicates that the lineage commitment follows positive selection, consistent with previous studies.

Further along the trajectory, hallmarks such as downregulation of CD69 & CD24 and upregulation of the homing receptor CD62L indicate successful maturation of the two SP populations (Figures 2E(7)) [8]. Thus we demonstrate that Wishbone can recover precise temporal ordering and branching of cells along with the bifurcation point with high resolution and without use of genetic perturbations to study development by leveraging multiplexed, single-cell analysis of a complex primary tissue.

Supplementary Note 2: Robustness analysis

We tested the robustness of Wishbone by varying the following free parameters:

Parameter	Description	Default
k	Number of neighbors for k nearest neighbor graph	15
nW	Number of waypoints	200
$\mathbf{WP}$	Random set of waypoints	

The robustness of Wishbone was tested on the same datasets used in Figures 2-3. The trajectory from each run was scaled to be between 0 to 1 using:

$$\boldsymbol{\tau}_{\mathbf{norm}} = \frac{\boldsymbol{\tau} - \min(\boldsymbol{\tau})}{\max(\boldsymbol{\tau}) - \min(\boldsymbol{\tau})} \quad (1)$$

Normalized trajectories from different runs were compared using Pearson correlation.

The similarity of branch associations is measured using fuzzy rand index, a probabilistic analogue of the adjusted rand index, which is a statistical measure of similarity between two clustering solutions [11]. As a first step, we calculate the probability of each cell belonging to the trunk and the two branches using the same principle for calculating the branch association score. The probability of a cell i belonging to a branch b is calculated as

$$p(i,b) = \frac{\sum_{w \in \mathbf{WP_b}} W_{iw}}{\sum_{p \in} \sum_{w \in \mathbf{WP}_p} aW_{iw}} \quad (2)$$

b is either the trunk or one of the two branches. $\mathbf{WP}_b$ is the set of waypoints assigned to b. W_{iw} is the affinity of cell i relative to waypoint w and is calculated using a Gaussian kernel on the distance (Equation 3). For each cell, the probability of the cell belonging to the associated branch is show in Supplementary Fig. 7A. This probability is lowest in the neighborhood of bifurcation point and increases exponentially away from this point indicating high confidence associations.

Let $\mathbf{C}$ represent the calculated branch probabilities such that for a given cell i,

$$\mathbf{C}(i) = \left(C_1(i), C_2(i), C_3(i)\right) \tag{3}$$

represents the vector of probabilities.

For a pair of cells i and j, define

$$E_{\mathbf{C}}(i,j) = 1 - \|\mathbf{C}(i) - \mathbf{C}(j)\| \tag{20}$$

Then the fuzzy rand index for two branch probabilities for two different clusterings, $\mathbf{C_1}$ and $\mathbf{C_2}$is calculated using

$$FRI(\mathbf{C_1}, \mathbf{C_2}) = 1 - \frac{\sum_{i=2:N} \sum_{j=1:i} \left| E_{\mathbf{C_1}}(i,j) - E_{\mathbf{C_2}}(i,j) \right|}{N(N-2)/2} \tag{21}$$

We first investigated the sensitivity of Wishbone trajectory and branching to the sampling of waypoints. Supplementary Fig. 7B shows a comparison of two runs with different sampling of waypoints for the same k and nW. These runs are highly correlated in both trajectory and branching (Pearson correlation: 0.99 & Fuzzy rand-index: 0.97). This also generalizes for multiple waypoint samplings (Supplementary Fig. 7C-E), with the correlation never falling below 0.99 and fuzzy rand-index never below 0.95. Furthermore, these results are reproduced across the replicates (Supplementary Fig. 7C-E), demonstrating that Wishbone results are robust for different sampling of waypoints.

Next, we investigated the robustness of Wishbone to different choices of k and nW (Supplementary Fig. 7F-H). The trajectory is again extremely stable across the different parameter settings (Pearson correlation > 0.99) and is consistent

with the observations of consistency in Wanderlust [12]. The branching results are consistently significant with nW at least 100 (Fuzzy rand-index > 0.87, with Fuzzy rand-index > 0.95 for nW =250) and inconsistent at lower values. The branch identification through the computation of the "**Q**" matrix relies on consistent observation of contradictions across groups of waypoints (Figure 1) and as such a significant number of waypoints is necessary to achieve a stable result. More over, nW =100 is still <0.1% of all the cells in the data and as such is not computationally expensive. These results are again reproducible across the different replicates (Supplementary Fig. 7F-H) and demonstrate that Wishbone is robust to different parameter choices.

We next compared the branching results of Wishbone derived using DN cell as the input early cell to results from using cells in the two SP populations as the input early cells (Supplementary Fig. 10). The input cell was randomly sampled from the $CD3^+CD4^+$(or $CD8^+$) populations in each of the three replicates and Wishbone was run with the above-mentioned parameters. The branching results from the three start cells were compared using adjusted Rand index. The adjusted rand index across different comparisons and replicates is always > 0.92, again highlighting the robustness of Wishbone. Wishbone results are also stable to removal of individual markers from the analysis (Supplementary Fig. 9).

Nature Biotechnology: doi:10.1038/nbt.3569

Supplementary Note 3: Cell type specific TF dynamics along trajectory

ThPOK, Gata3 and Runx1 are key SP specific transcription factors. While ThPOK, Gata3 and Runx3 have been demonstrated to play a role in the SP lineage commitment, the dynamics of these markers along a continuous trajectory have not been previously characterized. Wishbone proposes distinct expression dynamics of these TFs along the differentiation trajectory.

ThPOK levels are higher in CD4 SP cells compared with CD8 SP. Its level is high in the CD4+/CD8low stage, and drops in CD8 SP cells together with the increase in CD8 levels from intermediate to high, consistent with its role as suppressing CD8 expression (Figure 4C(1-2)). Gata3 is also high in CD4 SP compared to CD8 SP (Figure 4C(1-4)) but not the same extent as ThPOK. Runx3 is higher in CD8 SP compared with CD4 SP (in which it is very low). In addition, Repression of ThPOK in CD8+ cells is at the same time (or maybe slightly after) the increase in Runx3 in these cells, consistent with a role for Runx3 as a suppressor of ThPOK transcription (Figure 4C, D(5)) [10]. The observed dynamics are consistent across replicates (Supplementary Fig. 12).

Supplementary Note 4: Wishbone pseudocode

Inputs

$\mathbf{X} \in \mathbb{R}^{N \times M}$: Dataset of N cells.
k: Number of neighbors of nearest-neighbor graph
s: Early cell
nW: Number of waypoints

Algorithm

1. Determine low dimensional embedding $\mathbf{L} \in \mathbb{R}^{N \times R}$ of dataset $\mathbf{X}$ using diffusion maps

2. Construct the k-NN graph $\mathbf{G}$ using Euclidean distance in the embedded space $\mathbf{L}$. Each cell is a node in the graph and a cell is connected by weighted edges to its k nearest neighbors

$$G_{ij} = \begin{cases} \left\| \mathbf{L_i} - \mathbf{L_j} \right\|_2 & if\ i\ and\ j\ are\ neighbors \\ 0 & otherwise \end{cases}$$

3. Sample $nW - 1$ cells from the N cells and median filter to identify waypoints.

For each sampled cell i,

$$i_{med} = median\{k\ nearest\ neigbhors\ of\ i\ in\ \mathbf{G}\}$$

Find cell j, nearest to median cell i_{med}

Add j to $waypoint\ set$

Add early cell s to $waypoint\ set$

4. Determine shortest path distances from all waypoints to all cells to construct the distance matrix, $\mathbf{D} \in \mathbb{R}^{nW \times N}$ using Dijkstra's algorithm

$$D_{ij} = \min_{\mathbf{P}} \sum_{e \in \mathbf{P}} G_{e_1, e_2}$$

5. Initialize the ordering, $\tau_i^{(0)} = D_{is}$, $t = 0$

6. Repeat until $corr\left(\boldsymbol{\tau}^{(\mathbf{t-1})}, \boldsymbol{\tau}^{(\mathbf{t})}\right) < 0.9999$

7. Compute the perspective matrix, $\mathbf{P} \in \mathbb{R}^{nW \times N}$

$$P_{iw} := \begin{cases} \tau_i^{(t-1)} + D_{iw} & \text{if } \tau_i^{(t-1)} < \tau_w^{(t-1)} \\ \tau_i^{(t-1)} - D_{iw} & \text{otherwise} \end{cases}$$

8. Compute the mutual disagreement matrix, $\mathbf{Q} \in \mathbb{R}^{nW \times nW}$

$$Q_{ij} = \left| P_{ij} - \tau_j^{(t-1)} \right|$$

9. Determine the Eigen vector $\mathbf{v_2}$, corresponding to the second highest Eigen value of $\mathbf{Q}$.

10. Determine the branch association score for each cell i

Calculate the weight matrix

Nature Biotechnology: doi:10.1038/nbt.3569

$$W_{wi} = exp\left(\frac{-D_{wi}^2}{\sigma}\right) \Big/ \sum_{k=1:N} exp\left(\frac{-D_{wk}^2}{\sigma}\right)$$

Normalize $\boldsymbol{v_{2w}}$

$$v_{2w\,norm} = sign(v_{2w}) * \frac{abs(v_{2w})}{max\big(v_{2k} |\, sign(v_{2k}) == sign(v_{2w})\big)}$$

Calculate BAS for each cell i

$$BAS_i = \sum_{w \in W} W_{wi} * v_{2w\,norm}$$

11. Cross branch muting of weights to avoid waypoints of one branch influencing the ordering of cells on the other branch

$$Mut_b = \max\big(\, exp\,(-|BAS_k| \,|\, sign(BAS_k) \;\neq\; sign(BAS_i)\,)\big)\; for\; k \in waypoint\; set$$

$$Wmut_{wi} = \begin{cases} W_{wi}, & if\; sign(BAS_w) == sign(BAS_i) \\ W_{wi} * max(exp(-|BAS_i|), Mut_b), & otherwise \end{cases}$$

12. Calculate trajectory at iteration t

$$\tau_i^{(t)} = \sum_{w \in W} P_{iw} * Wmut_{iw}$$

13. $t = t + 1$, repeat from step 7.

14. Pick the five furthest waypoints on the positive and negative spectrum of $\mathbf{v_2}$. Estimate the branch point, bp as the median of the points with minimum $\boldsymbol{\tau^{(t-1)}}$ along all pairwise shortest paths between the chosen waypoints.

Let **BrA** and **BrB** be the set of five furthest waypoints of the two branches.

$$\mathbf{BrA} = \left\{ i \mid highest\; five\; \tau_i^{(t-1)}\; with\; sign(v_{2i}) > 0 \right\}$$

$$\mathbf{BrB} = \left\{ i \mid highest\; five\; \tau_i^{(t-1)}\; with\; sign(v_{2i}) < 0 \right\}$$

These waypoints are then used to determined the branch point by

$$bp = median\left\{ \tau_k^{(0)} : k = min_p\left\{ \tau_p^{(0)} : p \in path(i,j) \right\} \right\} i \in \mathbf{BrA}\; and\; j \;\in \mathbf{BrB}$$

15. Determine branch assignments for all cells

$$c_i = \begin{cases} 1\ (Trunk)\ \text{if}, \tau_i^{(t)} \leq bp \\ 2\ \text{if}, \tau_i^{(t)} > bp\ and\ BAS_i > 0 \\ 3\ \text{if}, \tau_i^{(t)} > bp\ and\ BAS_i \leq 0 \end{cases}$$

Outputs

$\boldsymbol{\tau}^{(\mathbf{t})}$: Trajectory or ordering of cells
bp: Branch point
BAS: Branch association scores
c: Branch assignments

Table legends

Supplementary Table 1: Table listing the staining panel clones, suppliers, isotope reporter and staining concentration.

Supplementary references

1. Levine, J.H., et al., *Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis.* Cell, 2015. **162**(1): p. 184-97.
2. Collins, A., D.R. Littman, and I. Taniuchi, *RUNX proteins in transcription factor networks that regulate T-cell lineage choice.* Nat Rev Immunol, 2009. **9**(2): p. 106-15.
3. Liu, P., P. Li, and S. Burke, *Critical roles of Bcl11b in T-cell development and maintenance of T-cell identity.* Immunol Rev, 2010. **238**(1): p. 138-49.
4. Radtke, F., H.R. MacDonald, and F. Tacchini-Cottier, *Regulation of innate and adaptive immunity by Notch.* Nat Rev Immunol, 2013. **13**(6): p. 427-37.

Nature Biotechnology: doi:10.1038/nbt.3569

5. Yui, M.A. and E.V. Rothenberg, *Developmental gene networks: a triathlon on the course to T cell identity.* Nat Rev Immunol, 2014. **14**(8): p. 529-45.
6. Guidos, C.J., I.L. Weissman, and B. Adkins, *Intrathymic maturation of murine T lymphocytes from CD8+ precursors.* Proc Natl Acad Sci U S A, 1989. **86**(19): p. 7542-6.
7. Azzam, H.S., et al., *CD5 expression is developmentally regulated by T cell receptor (TCR) signals and TCR avidity.* J Exp Med, 1998. **188**(12): p. 2301-11.
8. Love, P.E. and A. Bhandoola, *Signal integration and crosstalk during thymocyte migration and emigration.* Nat Rev Immunol, 2011. **11**(7): p. 469-77.
9. Germain, R.N., *T-cell development and the CD4-CD8 lineage decision.* Nat Rev Immunol, 2002. **2**(5): p. 309-22.
10. Singer, A., S. Adoro, and J.H. Park, *Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice.* Nat Rev Immunol, 2008. **8**(10): p. 788-801.
11. Campello, R.J.G.B., *A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment.* Pattern Recognition Letters, 2007. **28**(7): p. 833–841.
12. Bendall, S.C., et al., *Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development.* Cell, 2014. **157**(3): p. 714-25.