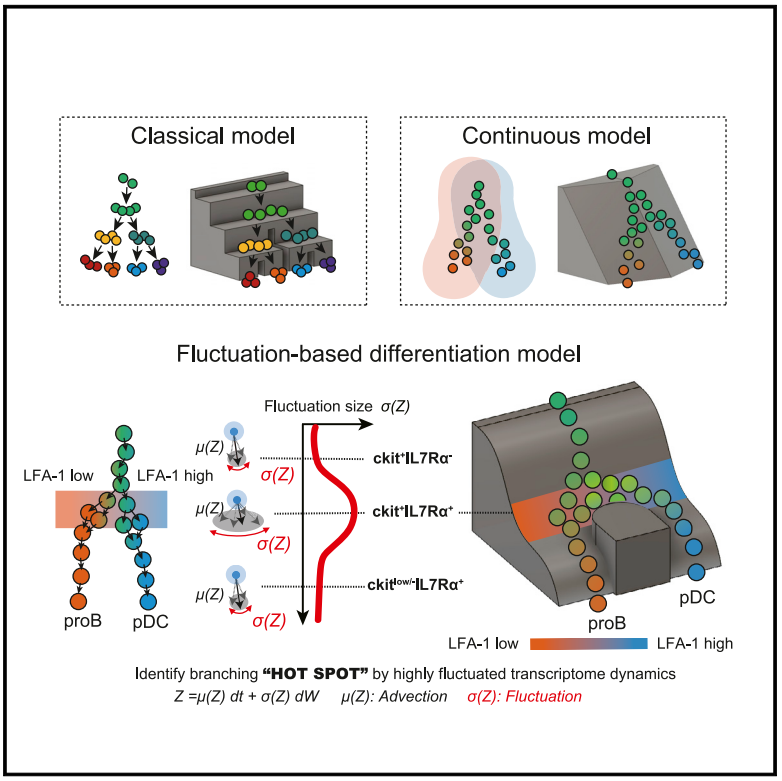
Cell Reports

Article

A bifurcation concept for B-lymphoid/plasmacytoid dendritic cells with largely fluctuating transcriptome dynamics

Graphical abstract



Highlights

- Human B and plasmacytoid DCs have common progenitors
- B/pDC bifurcation is a hotspot of large fluctuating transcriptome dynamics
- LFA-1 is a key molecule associated with pDC-directed cells in dry and wet approaches
- A fluctuation-based concept is proposed for hematopoietic differentiation

Authors

Keiki Nagaharu, Yasuhiro Kojima, Haruka Hirose, ..., Naoyuki Katayama, Hiroyoshi Nishikawa, Kohshi Ohishi

Correspondence

hnishika@ncc.go.jp (H.N.), koishi@clin.medic.mie-u.ac.jp (K.O.)

In brief

Nagaharu et al. show that human B-lymphoid and plasmacytoid DCs have common progenitors with a "hotspot" of large fluctuating transcriptome dynamics, in association of LFA-1 expression with pDC differentiation bias, using dry/wet approaches. They propose a fluctuation-based concept of hematopoietic differentiation that reconciles continuous and discrete models.

Nagaharu et al., 2022, Cell Reports *40*, 111260
August 30, 2022 © 2022 The Author(s).
https://doi.org/10.1016/j.celrep.2022.111260

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

A bifurcation concept for B-lymphoid/plasmacytoid dendritic cells with largely fluctuating transcriptome dynamics

Keiki Nagaharu,[1,10] Yasuhiro Kojima,[2,10] Haruka Hirose,[2] Kodai Minoura,[2] Kunihiko Hinohara,[3,4] Hirohito Minami,[1] Yuki Kageyama,[1] Yuka Sugimoto,[1] Masahiro Masuya,[1] Shigeru Nii,[5] Masahide Seki,[6] Yutaka Suzuki,[6] Isao Tawara,[1] Teppei Shimamura,[2,4] Naoyuki Katayama,[1] Hiroyoshi Nishikawa,[3,4,7,8,*] and Kohshi Ohishi[9,11,*]

[1]Department of Hematology and Oncology, Mie University Graduate School of Medicine, Tsu 514-8507, Japan
[2]Division of Systems Biology, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan
[3]Department of Immunology, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan
[4]Institute for Advanced Research, Nagoya University, Nagoya, Japan
[5]Shiroko Women's Hospital, Suzuka 510-0235, Japan
[6]Department of Computational Biology and Medical Sciences, The University of Tokyo, Kashiwa 277-8561, Japan
[7]Division of Cancer Immunology, Research Institute, National Cancer Center, Tokyo 104-0045, Japan
[8]Division of Cancer Immunology, Exploratory Oncology Research and Clinical Trial Center (EPOC), National Cancer Center, Chiba 277-8577, Japan
[9]Department of Transfusion Medicine and Cell Therapy, Mie University Hospital, Tsu 514-8507, Japan
[10]These authors contributed equally
[11]Lead contact
*Correspondence: hnishika@ncc.go.jp (H.N.), koishi@clin.medic.mie-u.ac.jp (K.O.)
https://doi.org/10.1016/j.celrep.2022.111260

SUMMARY

Hematopoiesis was considered a hierarchical stepwise process but was revised to a continuous process following single-cell RNA sequencing. However, the uncertainty or fluctuation of single-cell transcriptome dynamics during differentiation was not considered, and the dendritic cell (DC) pathway in the lymphoid context remains unclear. Here, we identify human B-plasmacytoid DC (pDC) bifurcation as large fluctuating transcriptome dynamics in the putative B/NK progenitor region by dry and wet methods. By converting splicing kinetics into diffusion dynamics in a deep generative model, our original computational methodology reveals strong fluctuation at B/pDC bifurcation in IL-7Rα^+ regions, and LFA-1 fluctuates positively in the pDC direction at the bifurcation. These expectancies are validated by the presence of B/pDC progenitors in the IL-7Rα^+ fraction and preferential expression of LFA-1 in pDC-biased progenitors with a niche-like culture system. We provide a model of fluctuation-based differentiation, which reconciles continuous and discrete models and is applicable to other developmental systems.

INTRODUCTION

The hematopoietic system has dynamic processes through which hematopoietic stem cells self-renew or generate various lineages of cells by multistep differentiation in response to fluctuating environmental conditions. Hematopoietic stem/progenitor cells (HSPCs) were thought to bifurcate into lineage-committed common myeloid and lymphoid progenitors (CLPs) that subsequently branch into various myeloid or lymphoid lineages (Blom and Spits, 2006; Doulatov et al., 2012; Seita and Weissman, 2010). However, the concept of human lymphopoiesis based on the tree-like hierarchy model of hematopoiesis has been challenged by the following two aspects (Karamitros et al., 2018; Laurenti and Göttgens, 2018; Macaulay et al., 2016; Velten et al., 2017).

First, T-lineage-committed lymphoid progenitors were found to retain the differentiation potential for monocytic cells, which were generally thought to be myeloid-lineage cells (Bell and Bhandoola, 2008; Kawamoto and Katsura, 2009; Wada et al., 2008). Similarly to mice, human CD34$^+$CD38$^-$CD45RA$^+$CD10$^+$ multilymphoid progenitors (MLPs), which correspond to the immature population of CLPs (Galy et al., 1995), possess the differentiation potential for dendritic cells (DCs) and monocytic cells in addition to various lineages of lymphoid cells (Doulatov et al., 2010; Helft et al., 2017). However, the human lymphoid pathway and bifurcation points for DC/monocytic lineage remain incompletely defined.

Second, comprehensive profiling and clustering of hematopoietic progenitors at the single-cell level by single-cell RNA sequencing (scRNA-seq) has revealed the heterogeneity of individual cells among common hematopoietic progenitors and the continuous lineage-priming feature of hematopoietic differentiation (Buenrostro et al., 2018; Haas et al., 2018; Karamitros et al., 2018; Laurenti and Göttgens, 2018; Loughran et al., 2020; Pellin et al., 2019; Velten et al., 2017). However, these concepts were based on scRNA-seq analysis of a static snapshot. Hence, transcriptome dynamics during differentiation and their role in cell fate decision have not been explored from these perspectives.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

La Manno et al. (2018) first focused on splicing kinetics obtained from scRNA-seq data and provided a perspective on lineage direction by recovering the dynamics of the single-cell transcriptome as RNA velocity, and Bergen et al. (2020) developed a methodology estimating RNA velocity without steady-state assumption. Lange et al. (2020) calculated the uncertainty of RNA velocity as a variance of the estimated RNA velocity to appropriately evaluate the cell fate probability, but they did not focus on the uncertainty itself because of the difficulty in distinguishing the biological uncertainty from observation noise. Several studies have revealed fluctuations in transcriptome dynamics during the time course of cell differentiation using a powerful mathematical framework called the "advection-diffusion model," which describes the dynamics of uncertainty by average dynamics (advection) and fluctuation (diffusion) (Fischer et al., 2019; Hashimoto et al., 2016). Fischer et al. (2019) further found that the estimated diffusion dynamics are more dominant than advection dynamics at β selection during T cell maturation. Nevertheless, they did not evaluate or identify molecular factors that underlie the diffusion and did not leverage the dynamics information based on splicing kinetics obtained from scRNA-seq analysis. Thus, existing computational methodologies have difficulties in analyzing the uncertainly or fluctuation of single-cell transcriptome dynamics, which were not well considered in the establishment of the concept that describes hematopoietic differentiation as a continuous process.

To overcome these limitations, we used two originally developed dry/wet approaches. First, we used an advanced lymphoid coculture system that supports the differentiation of human HSPCs into proT cells, proB cells, natural killer (NK) cells, plasmacytoid DCs (pDCs), and monocytic cells under the same culture conditions (Minami et al., 2017; Nakamori et al., 2012). This culture system has allowed us to comprehensively and simultaneously analyze the differentiation potential and tendency of lymphoid progenitors. Second, we developed an original computational methodology that converts single-cell splicing kinetics into an advection-diffusion model in a latent cell state space using deep-learning technology. The variational autoencoder (VAE)-based architecture in our methodology has allowed us to decode the diffusion in the latent cell state space as coordinated fluctuations in gene expression dynamics and to specify the cell state with a large fluctuation in gene expression dynamics. After challenging the expediency of our method in the endocrine development system of the pancreas, we applied this methodology to dissect the human lymphoid DC pathway in the population that was previously thought to be human B/NK progenitors (Doulatov et al., 2012; Galy et al., 1995; Karamitros et al., 2018; Kawamura et al., 2017) and identified a B/pDC bifurcation as a highly fluctuating transcriptome state in association of lymphoid function-associated antigen 1 (LFA-1) expression dynamics with pDC differentiation.

RESULTS

Proliferative and differentiation potentials of lymphoid progenitors change drastically along with an alteration in c-kit expression

To clarify the lineage branching points of lymphoid progenitors, we used an approach to examine the expression profile of cytokine receptors in various lymphoid progenitors and to specify the points at which the proliferation and differentiation potential dynamically change with alterations in receptor expression. We found that the expression profile of c-kit, a receptor for stem cell factor (Rojas-Sutterlin et al., 2014), varied in lymphoid-primed multipotent progenitors (LMPPs), $CD7^+$ MLPs, and $CD7^-$ MLPs in the $CD34^+CD38^-CD45RA^+CD19^-$ immature lymphoid population as well as $CD7^-CD10^-$ double-negative, $CD7^+CD10^-$ single-positive ($CD7^+$ SP), $CD7^-CD10^+$ single-positive ($CD10^+$ SP), and $CD7^+CD10^+$ double-positive progenitors in the $CD34^+CD38^+CD45RA^+CD19^-$ relatively mature lymphoid population. Most proB cells were negative for c-kit (Figure 1A).

To compare the proliferative and differentiation potentials of $c\text{-}kit^+$ and $c\text{-}kit^-$ fractions in $CD7^+$ MLPs, $CD7^-$ MLPs, $CD7^+$ SPs, and $CD10^+$ SPs, each fraction was cultured on stromal cells. After 11–17 days of culture, higher numbers of cells were generated from the $c\text{-}kit^+$ fraction of immature lymphoid progenitor $CD7^+$ or $CD7^-$ MLPs as well as $CD7^+$ or $CD10^+$ SP relatively mature lymphoid progenitors compared with the corresponding $c\text{-}kit^-$ fraction (Figure 1B). Moreover, various lineage of cells, which included proB cells, NK cells, proT, pDCs, monocytic cells, and cDC2, had developed from the $c\text{-}kit^+$ fractions of not only $CD7^+$ and $CD7^-$ MLPs, but also $CD7^+$ and $CD10^+$ SP cells. Low numbers of $CD141^+CLEC9A^+$ cDC1 were also generated, mainly from $c\text{-}kit^+$ fractions (Figures 1C and S1). The $c\text{-}kit^-$ fractions of $CD7^+$ and $CD7^-$ MLPs produced lower percentages of NK cells than $c\text{-}kit^+$ fractions, and $c\text{-}kit^-$ fractions of $CD7^+$ and $CD10^+$ SP cells differentiated to significantly higher percentages of proB, but a lower percentage or absence of NK and proT cells, and pDCs. The generation of cDC2 and monocytic cells was not remarkably different between $c\text{-}kit^+$ and $c\text{-}kit^-$ fractions in all populations, except for a few or no monocytic cells generated from the $c\text{-}kit^-$ fraction of $CD10^+$ SP cells (Figure 1C). Thus, $c\text{-}kit^+$ fractions of $CD38^-$ cells and the $CD38^-$ lymphoid population possessed highly proliferative and multilymphoid, DC, monocytic potentials, while the $c\text{-}kit^-$ fraction of the $CD38^+$ population exhibited a differentiation tendency toward B rather than T, NK, or pDC lineages. An observed association of c-kit negativity with B-lineage differentiation bias in the relatively mature lymphoid population was also suggested, based on the inverse relationship between c-kit and CD10 expression in $CD34^+CD38^+CD45RA^+$ populations and the partial expression of cyCD79a, an early marker of B-lineage cells, and CD19 in their $c\text{-}kit^-$ fraction (Figure 1D). These data revealed that the proliferation and differentiation potentials of lymphoid progenitors change drastically with the alteration of c-kit expression, especially in the $CD34^+CD38^+CD45RA^+$ lymphoid population, and lymphoid progenitors differentiate toward B rather than T, NK, and pDC lineages along with c-kit downregulation.

Development of a computational methodology that quantitatively captures the fluctuation of cell state dynamics during pancreatic endocrine cell differentiation

To capture the fluctuation of transcriptome dynamics during cell differentiation quantitatively, we assumed that the cell state dynamics could be described as the summation of the deterministic transition (advection) and fluctuation (diffusion), which is

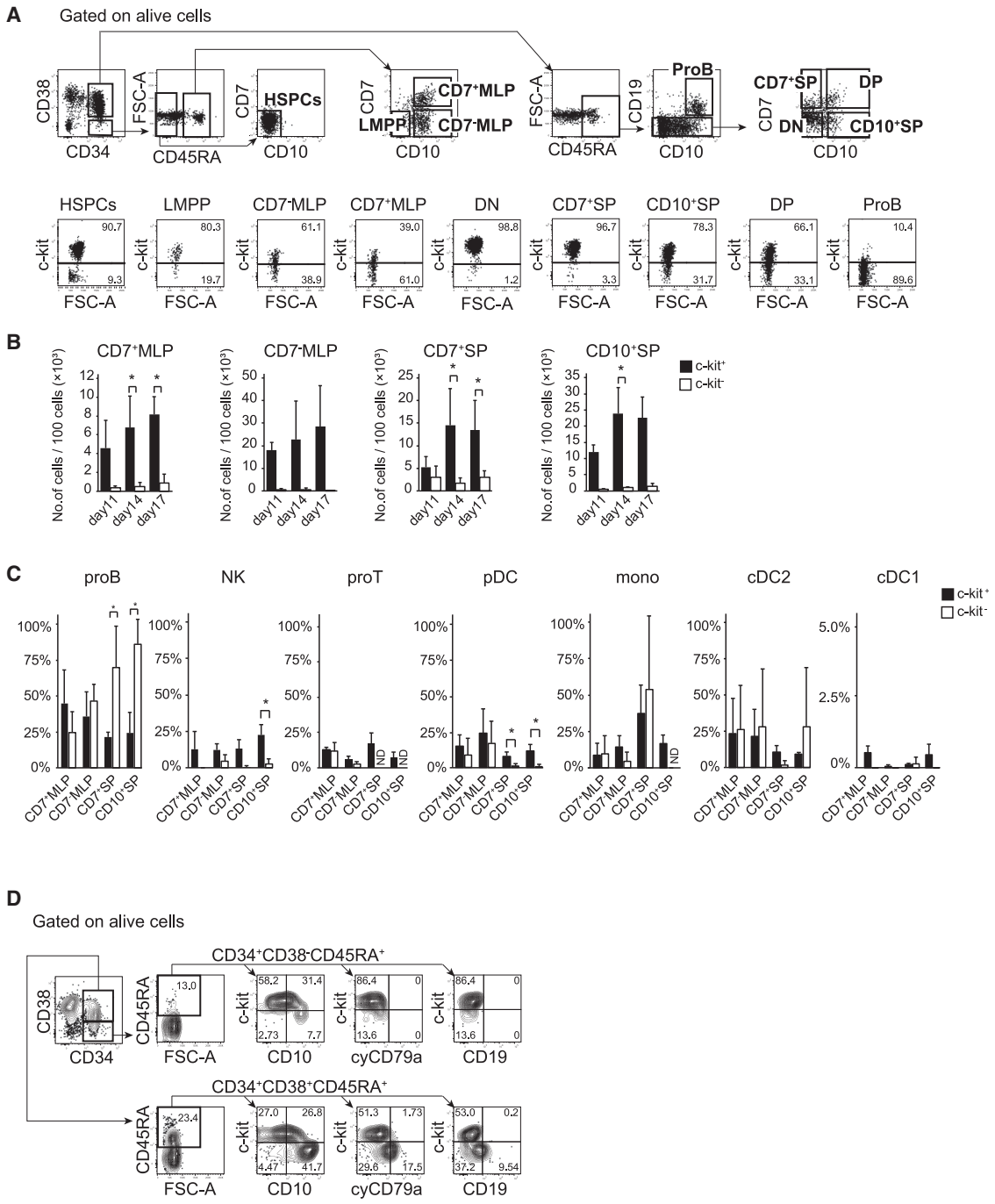


Figure 1. Classification of lymphoid progenitors by c-kit

(A–C) (A) c-kit expression in various lymphoid progenitors. c-kit$^+$ and c-kit$^-$ fractions of CD7$^+$ MLP, CD7$^-$ MLP, CD7$^+$ SP, and CD10$^+$ SP cells (>95% pure) were cultured for 11–17 days. Graphs show (B) the total cell number and (C) percentage of proB cells (day 17), NK cells (day 11), proT cells (day 17), and pDCs among CD45RA$^+$ cells (day 14), and monocytic cells and cDC2 among CD45$^+$ cells (day 14). Error bars represent ±SEM of three independent experiments (unpaired t test; *$p < 0.05$).

(D) Flow cytometry showing the relationship among c-kit, CD10, cyCD79a, and CD19 in CD34$^+$CD38$^-$CD45RA$^+$ and CD34$^+$CD38$^+$CD45RA$^+$ populations.

See also Figure S1.

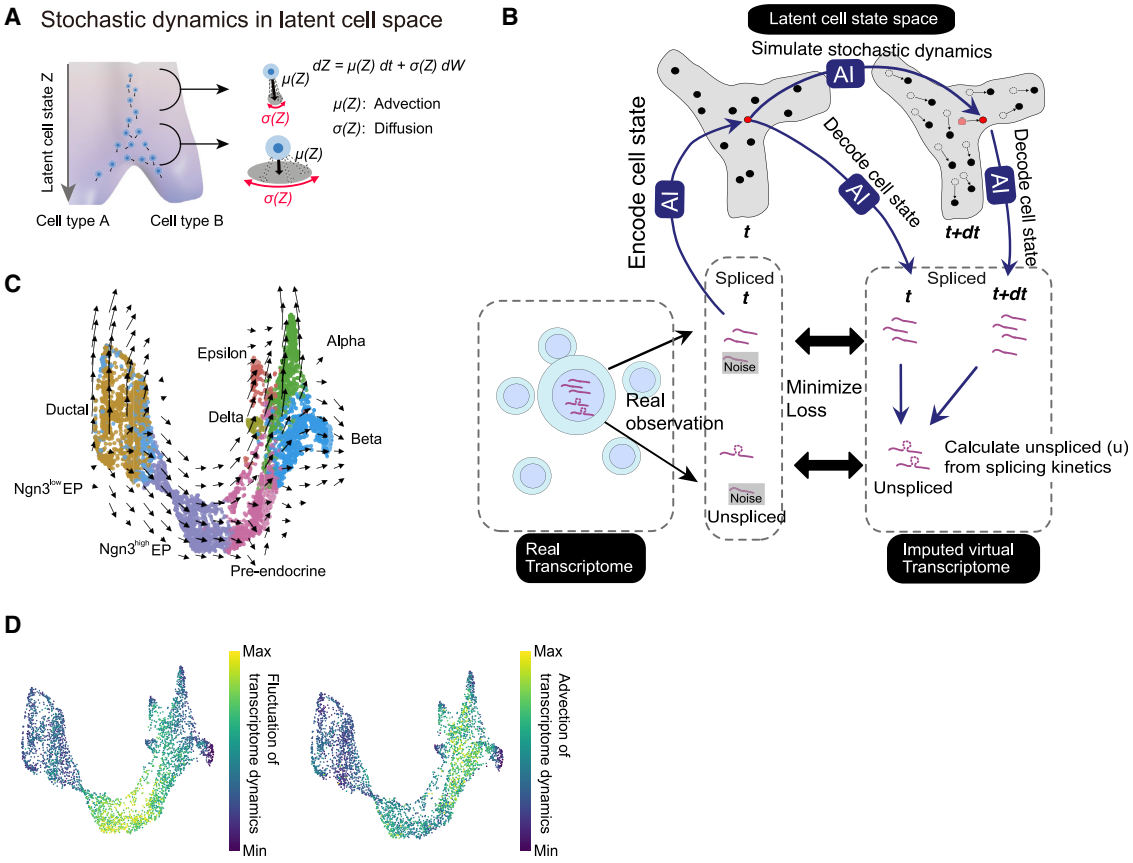


Figure 2. Identification of fluctuating cell state dynamics in pancreas endocrine development

(A) Schematic overview of the advection-diffusion model in the latent cell state space. Transcriptome dynamics in the latent cell state space decomposed into advection (averaged dynamics) and diffusion (fluctuation).

(B) Overview of the AI-based methodology to estimate single-cell transcriptome dynamics with fluctuation from spliced and unspliced transcript abundances.

(C) Flow of estimated transcriptome dynamics. The arrow direction represents the direction of the future latent cell state in UMAP embedding.

(D) Fluctuation and advection of transcriptome dynamics at single-cell resolution.

To prevent the effect of ectopic values from visualization, 1% of the smallest and largest values were forced to be 1% and 99% quantile values. See also Figures S2 and S3.

mathematically termed as the advection-diffusion model (Figure 2A). We developed a computational methodology, variational inference of cell state dynamics with fluctuation (VICDYF), to estimate the advection-diffusion model of the molecular cell state by spliced and unspliced transcript abundances. In this methodology, we optimized the dynamics of the latent cell state, which is encoded from the spliced transcriptome by VAE, so that the time change in the spliced transcriptome decoded from stochastic cell state dynamics is consistent with the unspliced transcriptome under a mathematical model of RNA-splicing kinetics (Figure 2B). To examine the contribution of each gene to the diffusion in the latent cell state space, we decoded the estimated cell state diffusion into the fluctuation of the expression change for each gene using the bidirectional conversion feature of VAE.

To validate the computational methodology, VICDYF, we first applied this methodology to a public scRNA-seq dataset of pancreas endocrine development at embryonic day 15.5 (E15.5) (Bastidas-Ponce et al., 2019; GSE132188). The estimated directionality of cell state dynamics was consistent with the expected flow from endocrine progenitors toward different types of pancreatic endocrine cells such as α, β, δ, γ, or ε cells (Figure 2C). Our methodology revealed that the estimated fluctuation of transcriptome dynamics was the highest in the Ngn3^{high} endocrine progenitor (EP) cluster (Figure 2D) in which cell lineage diversification toward endocrine lineages was expected to occur. The estimated advection of transcriptome dynamics was relatively increased in the population more differentiated into each lineage than the population with a large fluctuation (Figures 2D and S2A).

To clarify the influence of technical factors, such as total unique molecular identifiers per cell and average gene expression levels, on estimating the fluctuation of dynamics, we applied VICDYF to the simulated single-cell transcriptome dataset during the cell differentiation process, which was generated by SERGIO (Dibaeinia and Sinha, 2020). While VICDYF succeeded in reconstructing the correct cell differentiation direction from cell type 0 to 1, we found that the estimated fluctuations in dynamics were large at the boundary between cell type 0 and 1,

and did not exhibit associations with the total counts (Figure S2B). These results suggest that the technical noise due to the observation sparsity had a small contribution to the estimated fluctuation of dynamics.

To confirm whether the unspliced transcriptome, which reflects the future cell state, was indeed relatively variable at the branching point, we evaluated the ratio between the variances of unspliced transcriptomes among k nearest neighbors (k = 30) in the latent space and those of spliced transcriptomes. We found that the relatively higher variance of unspliced transcriptomes was also observed in the Ngn3^{high} EP population (Figure S2C). This finding was consistent with the estimated large fluctuation at the lineage diversification point because unspliced transcriptomes are expected to reflect the later cell state compared with spliced transcriptomes. We confirmed that the fluctuation was more correlated to the variance ratio between spliced and unspliced transcripts (Pearson's r = 0.59) than the mean ratio between them (Pearson's r = 0.45) (Figure S2D). We found that the variance ratio exhibited large values specifically for Ngn3^{high} EP and pre-endocrine cells as well as the fluctuation, while the mean ratio had large values not only for these populations but also for more mature cell populations, including α, β, and δ cells. We also applied VarID (Grün, 2020) to the pancreas dataset and analyzed the differential variability between clusters with large fluctuation and the other clusters (Figure S2E). We found that the variability difference of VarID was consistent with the fluctuation difference of VICDYF (Figure S2F). This suggests that our methodology also captures the fluctuation of gene expression independent from variability-mean dependence.

Computational methodology, VICDYF, extracts lineage drivers based on fluctuation directionality

To clarify the lineage bifurcation and estimated large fluctuation, we estimated cell fate based on the estimated cell state dynamics and compared the estimated fate probabilities with those estimated by other methodologies, CellRank (Lange et al., 2022) and Palantir (Setty et al., 2019). VICDYF assigned the largest fate probability to β cells for Ngn3^{high} EPs, while some fractions were estimated to be committed to other lineages (Figure S3A). These results are consistent with the existing biological knowledge and the estimation of an existing methodology and CellRank (Figure S3B), while Palantir assigned the largest fate probability to ε cells in the population (Figure S3C). However, we found that the downstream cell population with a large fluctuation was estimated to be more committed than those upstream. This suggests that the increase in the fluctuation of the population contributes to the fate decision between these cell lineages, while the majority of the cell population is differentiating into β cells.

To dissect the fluctuation into various lineages, we clustered the population with the largest 20% of fluctuation into five clusters from the fluctuation patterns of genes with high reconstruction accuracy of expression and dynamics, $c_s > 0.3$and $c_r > 0.1$ (Figure S3D). To evaluate the coordinated fluctuation across various genes in fluctuation cluster 2 close to α and β cells (Figure S3D), we calculated the deviation of sampled stochastic gene expression changes from the averaged changes for 50 times iterative sampling of cell state dynamics in the latent cell state of each single cell. By comparing this coordinated fluctuation and the difference in gene expression profiles between α and β cells, we scored the fluctuation directionality of each gene between α and β cells as the α-β score, which is higher when the fluctuation of a gene correlates with that of genes with higher expression in β cells. We found that the aristaless-related homeobox (Arx) gene had the eighth lowest score and pancreas duodenum homeobox 1 (Pdx1) had the sixth highest score in fluctuation cluster 2 (Figure S3E). These results are consistent with previous reports showing that ARX is crucial for α cell identity and survival, that ARX inactivation induces the α-to-β cell reprogramming in pancreatic progenitor cells or mature α cells (Courtney et al., 2011, 2013), and that PDX1 is a critical transcription factor for pancreatic development including β cell maturation and functions (Jara et al., 2020; Zhu et al., 2017). We also compared the lineage driver estimations of VICDYF and CellRank. In contrast to the high scoring of VICDYF for Arx (eighth in α cell fate) and Pdx1 (fifth in β cell fate), CellRank highlighted Nkx6-1 (sixth in β cell fate) and Irx2 (sixth in α cell fates) (Figure S3F), both of which are related to the development of each lineage (Aigha and Abdelalim, 2020). These results showed that both methods captured the lineage determination factors, whereas the extracted factors were largely different. Taken together, these results demonstrated that our methodology had the ability to quantitatively capture the fluctuation in cell state dynamics with deviation of gene expression changes at single-cell levels.

Highly fluctuating transcriptome dynamics at a B/pDC bifurcation with IL-7Rα expression

The culture data described above showed that the CD10$^+$ SP population, which has been thought to comprise B/NK progenitors, was a heterogeneous population in which lineage diversification from multilymphoid progenitors toward lymphoid and DCs was suspected to occur. We therefore focused on this CD10$^+$ SP population and applied our computational methodology to dissect the lymphoid pathway in CD10$^+$ SP cells (n = 2×10^4) to capture the transcriptome dynamics of individual cells during differentiation. We found that each cluster determined by 40-dimensional principal component analysis (PCA) coordinates of scaled expression of highly variable genes was distinctively distributed in uniform manifold approximation and projection (UMAP) (McInnes et al., 2018) embeddings of 10-dimensional latent cell state coordinates (Figure S4), which demonstrated that our VAE-based methodology efficiently embedded the single-cell transcriptome into the latent cell state space as well as the existing VAE-based methodology for dimensional reduction (Lopez et al., 2018). To determine the putative subpopulation toward each cell fate, we evaluated the similarity of denoised expression profiles, which were decoded from latent cell states, to those of all cell types in a public peripheral blood mononuclear cell (PBMC) scRNA-seq dataset (Ding et al., 2020). Notably, the region with similarity to the B lineage was adjacent and overlapped with the region with similarity to pDCs, while partial overlapping was observed between B and monocytic lineages, monocytic and cDC2 lineages, and T and NK lineages (Figures 3A and 3B). The estimated dynamics are directed to the trajectory

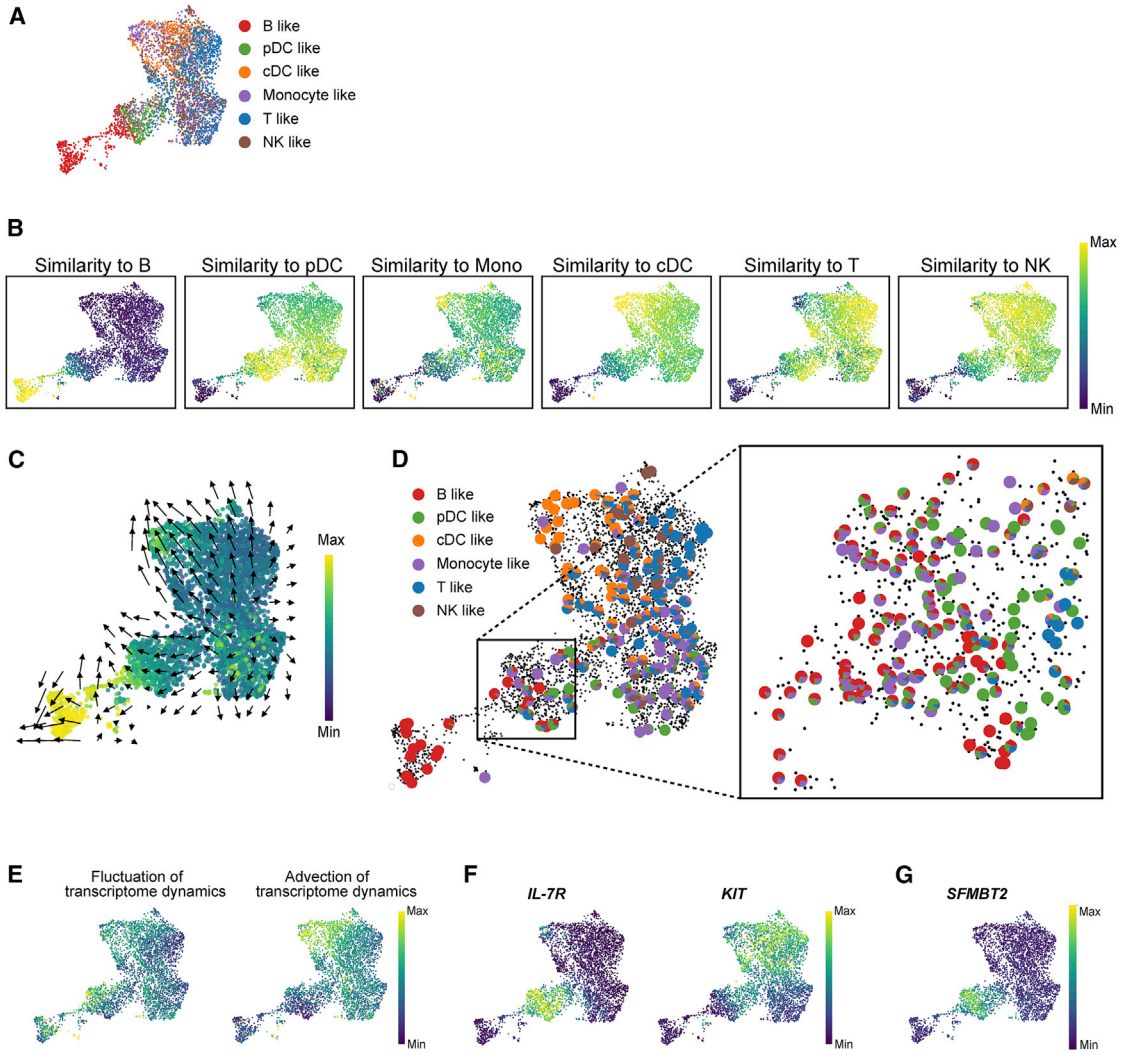


Figure 3. Single-cell RNA sequencing of the CD10$^+$ SP population

(A) Most similar cell types among B cells, NK cells, T cells, pDCs, monocytes, and cDCs in the UMAP representation of latent cell states. We evaluated similarity by comparing the reconstructed expression profile of each cell with the mean expression profile of annotated clusters derived from the peripheral bone marrow cell dataset in the Seurat package.

(B) Similarity of single cells to designated cells.

(C) Direction of transcriptome dynamics in the UMAP representation.

(D) Cell fate probability from estimated cell state dynamics. Each pie chart of each single cell represents the composition of cell fate probabilities for designated cells.

(E) Fluctuation and advection of transcriptome dynamics at single-cell resolution.

(F) Reconstructed expression of IL-7R and KIT.

(G) Reconstructed expression profile of SFMBT2.

To prevent the effect of ectopic values from visualization, 1% of the smallest and largest values were forced to be 1% and 99% quantile values. See also Figures S4 and S5.

toward various lineages such as B/pDC/monocytic, cDC2/monocytic, and T/NK cell lineages (Figure 3C).

To estimate the potential of each cell type, we chained the stochastic dynamics of each cell in latent cell space 1,024 times and estimated the fate probability of each single cell for each cell type, as the probability that the expression profile of the cell type was the most similar to that of the single cell after iterative cell state transition (Figure 3D). This estimation allowed us to identify a cell population with the potential for both B and pDC lineages (Figure 3D, enlarged figure) and that this population was enriched in the cell state region where fluctuation, but not advection of gene expression dynamics, was particularly higher than in surrounding regions (Figure 3E). These data suggested that B/pDC bipotent progenitors exhibited large fluctuation of transcriptome dynamics. We further found that expression of interleukin-7Rα (IL-7Rα) was highly enriched in putative B/pDC

common progenitors, while kit expression was gradually suppressed toward B-lineage progenitors (Figure 3F). Of note, Polycomb group gene SFMBT2, which plays a regulatory role in mediating chromatin remodeling (Miri et al., 2013; Tang et al., 2019), was specifically expressed in the B/pDC bifurcation region (Figure 3G). Single-cell multiome analysis of the CD34$^+$ population revealed that the differentially accessible chromatin regions of B cells and pDCs were simultaneously accessible in the cell population with a large fluctuation of transcriptome dynamics (Figure S5), suggesting that the estimated large fluctuation is associated with transient simultaneous accessibility of loci characterizing the difference between two lineages.

Identification of B/pDC common progenitors at IL-7Rα$^+$ fraction by single-cell culture

On the basis of the transcriptome analysis, we examined the relationship between the c-kit/IL-7Rα expression profile and differentiation potential, especially for B and pDC lineages, in our culture. Expression of IL-7Rα (CD127) was negative in immature lymphoid progenitors as MLPs (data not shown), but positive in a proportion of c-kit$^+$ or c-kit$^-$ fractions of CD10$^+$ SP cells. The majority of the IL-7Rα$^+$ fraction expressed cyCD79a, which indicated that the IL-7Rα$^+$ fraction was biased to the B lineage to some extent (Figure 4A). When c-kit$^+$IL-7Rα$^-$, c-kit$^+$IL-7Rα$^+$, and c-kit$^-$IL-7Rα$^+$ fractions of CD10$^+$ SP cells were cultured, the c-kit$^+$IL-7Rα$^+$ fraction produced a significantly higher percentage of proB cells than the c-kit$^+$IL-7Rα$^-$ fraction. Moreover, the differentiation potential of the c-kit$^+$IL-7Rα$^+$ fraction of pDCs was as high as that of the c-kit$^+$IL-7Rα$^-$ fraction (Figure 4B). Accordingly, the c-kit$^+$IL-7Rα$^+$ fraction gave rise to significantly high percentages of proB cells and pDCs. Relative to the c-kit$^+$IL-7Rα$^+$ fraction, the c-kit$^-$IL-7Rα$^+$ fraction produced an even higher percentage of proB cells but significantly lower percentages or no NK cells, proT cells, NK cells, pDCs, and monocytic cells (Figure 4B). To examine the timing between loss of c-kit and gain of IL-7Rα expression, c-kit$^+$IL-7Rα$^-$CD10$^+$ SP cells were cultured and analyzed for expression of c-kit and IL-7Rα. As shown in Figure 4C, most cells became positive for IL-7Rα, while most remained positive for c-kit$^-$ after 48 h of culture and c-kit expression was downregulated later (data not shown). These data indicated that c-kit$^+$IL-7Rα$^-$CD10$^+$ SP cells differentiated toward c-kit$^-$IL-7Rα$^+$ B-biased progenitors via a c-kit$^+$IL-7Rα$^+$ stage with a high B cell and pDC differentiation potential.

To determine the differentiation potential of individual cells in each fraction more precisely, we performed single-cell culture of cells in each fraction. As shown in Figure 4D, the c-kit$^+$IL-7Rα$^-$ fraction contained various types of progenitors with variable combinations of differentiation potentials for B cells, NK cells, pDCs, cDCs, and/or monocytic cells, which included immature progenitors with a multilineage differentiation potential for B cells, pDCs, and other lineage cells. Compared with the c-kit$^+$IL-7Rα$^-$ fraction, the IL-7Rα$^+$ population, such as c-kit$^+$IL-7Rα$^+$ and c-kit$^-$IL-7Rα$^+$ fractions, was composed mainly of more mature progenitors with a differentiation potential for less than two lineages, and more importantly, B/pDC common progenitors were identified (Figures 4D and 4E). The c-kit$^-$IL-7Rα$^+$ fraction consisted mainly of B-committed progenitors (Figure 4D). Of note, B/NK bipotent progenitors were not detected. These findings demonstrated that the differentiation potential for B cells and pDCs was drastically changed in accordance with alterations in c-kit and IL-7Rα expression.

Association of LFA-1 expression with pDC lineage bias at B/pDC bifurcation revealed by computational prediction and culture

To dissect the molecular mechanisms associated with pDC- or B-lineage differentiation at the B/pDC bifurcation, which was identified as a hotspot with largely fluctuated cell state dynamics by transcriptome analysis, we clustered single cells of the CD10$^+$ SP population with the largest 20% of the fluctuation into eight clusters from the fluctuation patterns of genes with high reconstruction accuracy of expression and dynamics, $c_s > 0.3$ and $c_r > 0.1$ (Figure 5A). We then calculated the B/pDC score of each gene in fluctuation clusters that largely overlapped with the putative B/pDC common (fluctuation clusters 2 and 5) and B progenitors (fluctuation cluster 7) in the same manner as the α-β score in the previous section (Figure S3E). We found that critical B-lineage transcription factor, PAX5 (Cobaleda et al., 2007) and B-lineage-associated gene, CD24 (Fang et al., 2010) exhibited first and third highest score in fluctuation cluster 7, respectively, which had higher similarity to B cells (Figure 5B). Furthermore, B cell antigen receptor complex-associated protein α chain, CD79A, exhibited the second and seventh highest B/pDC score in fluctuation clusters 2 and 5. On the other hand, ID2, which negatively regulates B cell differentiation (Wu and Shao, 2019), had the lowest B/pDC scores in fluctuation clusters 2, 5, and 7. Additionally, interferon regulatory factor 8 (IRF-8), which is a critical transcription factor for human pDC development (Bigley et al., 2018; Collin and Bigley, 2018; Sichien et al., 2016), had the ninth lowest B/pDC score only in fluctuation cluster 5, indicating that IRF8 plays a stage-specific role in pDC development. Additionally, other B-lineage-associated genes (ZCCHC7 [Hystad et al., 2007], POU2AF1 [Levels et al., 2019], and BLNK [Minegishi et al., 1999]) and pDC-associated genes (NFKBIA [Bottero et al., 2006] and RUNX3 [Dicken et al., 2013]) were found in the top ten genes in B/pDC scores of fluctuation clusters 2 and/or 5 (Table S1). Notably, we found that LFA-1 had the fourth and third lowest B/pDC scores in fluctuation clusters 2 and 5. Comparison between expression and fluctuation of genes showed that LFA-1 was broadly expressed except for B-like region-expressing B-lineage-associated gene PAX5, but the fluctuation of LFA-1 was specifically high in the population, which corresponded to the B/pDC progenitor region (Figure 5C). These results demonstrated that the fluctuation-based scoring highlighted the genes associated with lineage commitment and that LFA-1 expression fluctuated positively in pDC-directed cells at the B/pDC bifurcation.

To analyze the expression profile of LFA-1 in accordance with the alteration of lymphoid, DC, and monocytic differentiation potentials, the relationship between LFA-1 and c-kit expression on the surface of various cell populations was examined by fluorescence-activated cell sorting analysis. Compared with HSPCs, the expression levels of LFA-1 were relatively higher in LMPPs and MLPs but did not significantly differ between c-kit$^+$ and c-kit$^-$ fractions. However, a broad range of LFA-1 expression was observed in CD10$^+$ SP cells in which LFA-1 expression was high in the c-kit$^+$ fraction but low or negative in some of

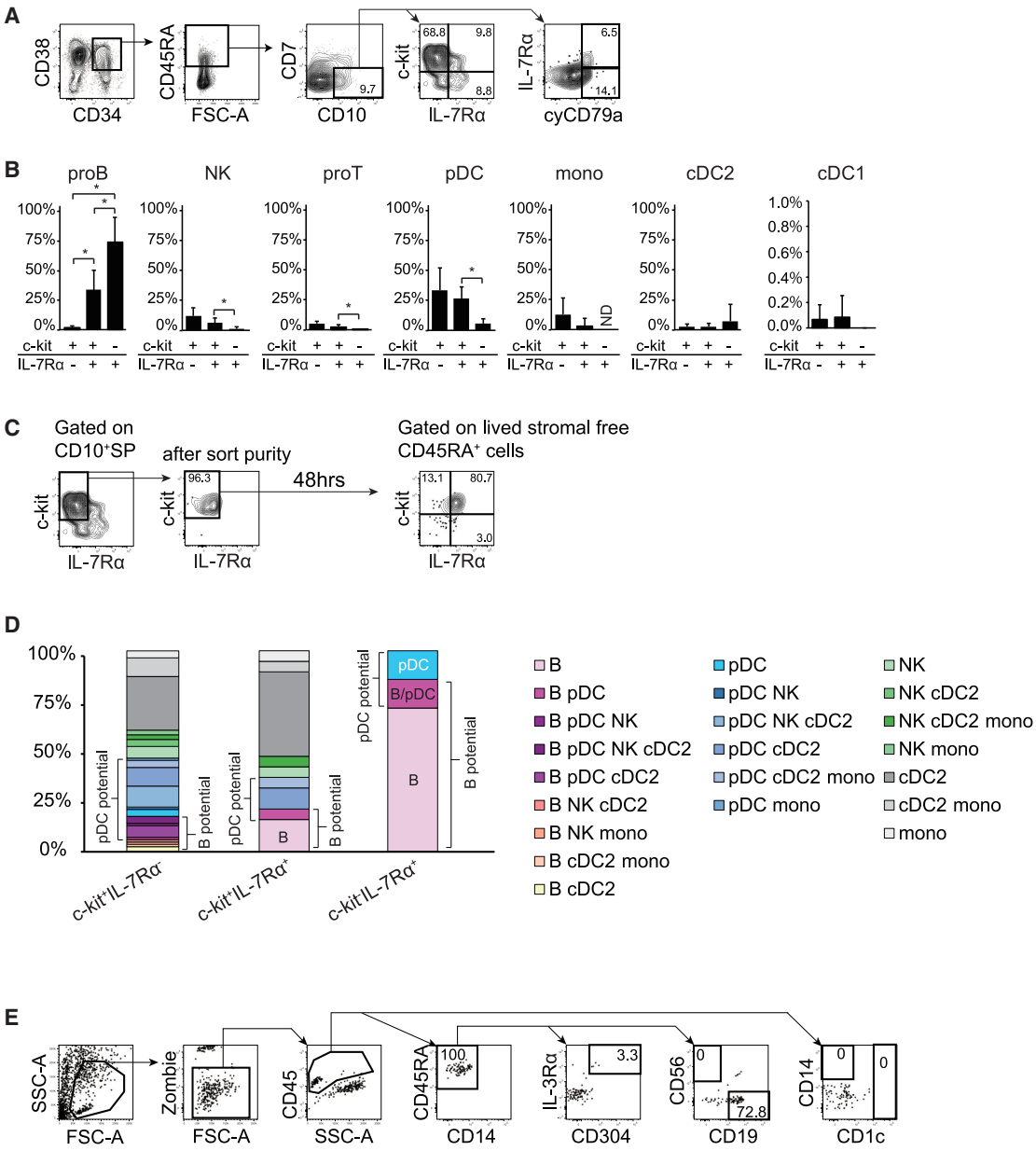


Figure 4. Identification of B/pDC common progenitors in the IL-7Rα$^+$ fraction

(A) Relationship among c-kit, IL-7Rα, and cyCD79a expression in CD10$^+$ SP cells.

(B) c-kit$^+$IL-7Rα$^-$, c-kit$^+$IL-7Rα$^+$, and c-kit$^-$IL-7Rα$^+$ fractions of CD10$^+$ SP cells (>95% pure) were cultured for 14 days and analyzed for the percentages of proB cells, NK cells, proT cells, and pDCs among CD45RA$^+$ cells, and monocytic cells and cDC2 among CD45$^+$ cells. Error bars represent ±SEM of three independent experiments (unpaired t test; *$p < 0.05$).

(C) c-kit$^+$IL-7Rα$^-$ fraction in CD10$^+$ SP cells was cultured on stromal cells and analyzed for expression of c-kit and IL-7Rα at 48 h.

(D) Individual cells of c-kit$^+$IL-7Rα$^-$, c-kit$^+$IL-7Rα$^+$, and c-kit$^-$IL-7Rα$^+$ fractions of CD10$^+$ SP cells were cultured for 14 days, and cultures that contained more than ten non-adherent cells (c-kit$^+$IL-7Rα$^-$ [n = 86], c-kit$^+$IL-7Rα$^+$ [n = 19], c-kit$^-$IL-7Rα$^+$ [n = 7]) were assessed for generation of designated cells.

(E) Representative culture data of single-cell culture of a c-kit$^+$IL-7Rα$^+$ cell, which gave rise to only IL-3Rα$^+$CD304$^+$ pDCs and CD19$^+$ proB cells.

the c-kit$^-$ fraction (Figure 6A). LFA-1 was highly expressed by mature pDCs, NK cells, cDC2, and monocytic cells, but almost negative in proB cells (Figure 6B). To assess the relationship between LFA-1 expression and the B/pDC differentiation potential in the CD10$^+$ SP population, c-kit$^+$IL-7Rα$^-$ and c-kit$^+$IL-7Rα$^+$ fractions were separated into LFA-1-high and -low populations, and the c-kit$^-$IL-7Rα$^+$ fraction was separated into LFA-1-high, -low, and -negative populations (Figure 6C). Each fraction was cultured and a significant difference was not observed between LFA-1^{high} and LFA-1^{low} fractions of the c-kit$^+$IL-7Rα$^-$ population.

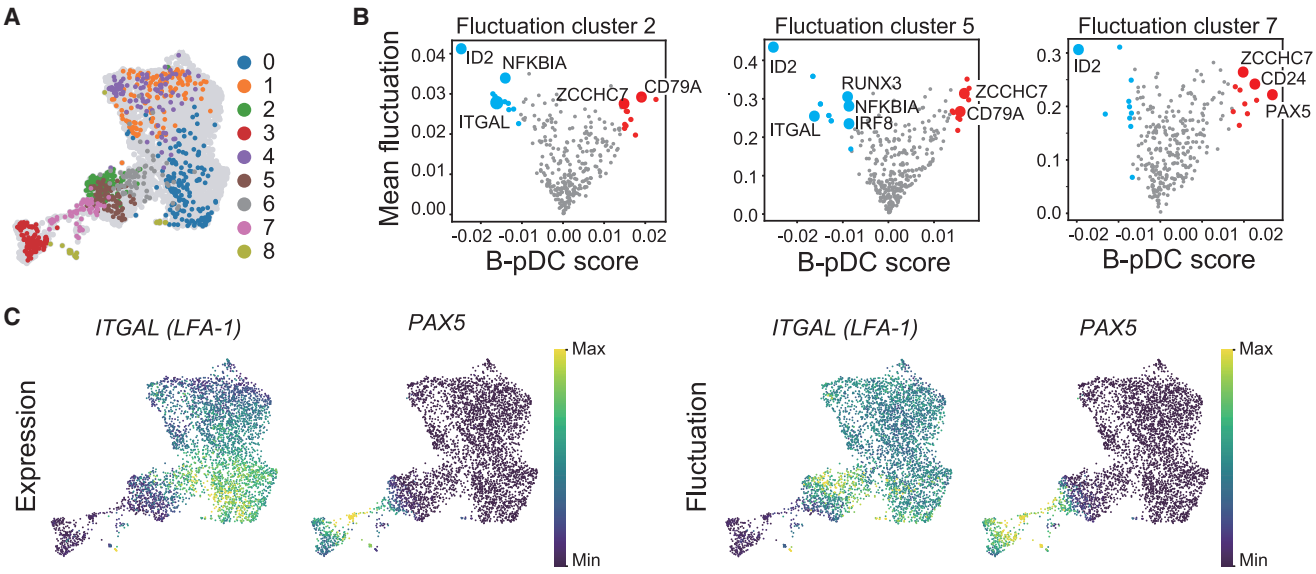


Figure 5. Association of LFA-1 expression dynamics with pDC differentiation

(A) UMAP representation of fluctuation clusters. The fluctuation clusters were derived from clustering cells with the top 20% largest fluctuations. The clustering is based on the fluctuation of selected genes.

(B) B-pDC score and fluctuation of genes with $c_s > 0.3$ and $c_r > 0.1$. A high B-pDC score of a gene represents a high correlation of the fluctuation with those of B-enriched genes, whereas a low B-pDC score of a gene represents a high correlation of the fluctuation with those of pDC-enriched genes. Red dots represent the top ten highest-scoring genes and blue dots represent the top ten lowest-scoring genes.

(C) Reconstructed fluctuation and expression of LFA1 and PAX5.

To prevent the effect of ectopic values from visualization, 1% of the smallest and largest values were forced to be 1% and 99% quantile values.

However, in the c-kit$^+$IL-7Rα^+ fraction, the LFA-1^{high} population had a higher percentage of pDCs than proB cells, while the LFA-1^{low} population produced a higher percentage of proB cells than pDCs (Figure 6D). In the c-kit$^-$IL-7Rα^+ fraction, the LFA-1^{high} population differentiated into pDCs, monocytic cells, and cDC2, but not proB cells. Conversely, the LFA-1$^{low/-}$ population exclusively differentiated into proB cells and lost their pDC differentiation potential (Figure 6D). These data revealed that LFA-1 was broadly expressed in most c-kit$^+$ lymphoid progenitors with a pDC differentiation ability, but in the c-kit$^+$IL-7Rα^+ fraction, higher levels of LFA-1 expression were closely associated with a differentiation bias toward the pDC lineage rather than the B lineage, which was consistent with the findings of scRNA-seq analysis. To further elucidate the relationship between LFA-1 expression and B/pDC differentiation bias in a more B/pDC-enriched fraction, we examined the relationship in the c-kit^{low}IL-7Rα^+ fraction of CD10$^+$ SP because the data in Figure 4B suggested that NK, T, and monocytic potentials were more drastically suppressed than the pDC potential by downregulation of c-kit expression in c-kit$^+$IL-7Rα^+CD10$^+$ SP. We observed that the c-kit^{low}IL-7Rα^+ population mainly gave rise to proB cells and pDCs, and that the LFA-1^{high} fraction had a higher percentage of pDCs and a lower percentage of proB cells than the LFA-1^{low} fraction in the B/pDC-enriched population (Figure S6).

DISCUSSION

An advection-diffusion model is a fundamental model that has been used to describe a wide range of phenomena that includes cell differentiation. The several methods to estimate transcriptome dynamics as the advection-diffusion model are intended to convert the population shift of the scRNA-seq observation to the directionality of cell differentiation (Fischer et al., 2019; Hashimoto et al., 2016). Here, we estimated the advection-diffusion model from spliced and unspliced transcriptome abundances, which is expected to have abundant information about cellular dynamics, and linked the large diffusion in the latent cell space to cell lineage bifurcation. These estimations were validated by knowledge of the pancreatic development system and our advanced culture system.

The VAE-based architecture of artificial intelligence (AI) technology has been used to construct latent cell state space behind a single-cell transcriptome profile and denoise the gene expression profile by decoding compressed cell states (Eraslan et al., 2019; Lopez et al., 2018). Furthermore, recent studies have shown that the additive operation in the latent cell state space corresponds to the additive change in actual cells (i.e., drug treatment) (Lotfollahi et al., 2019). Here, we defined stochastic cell state dynamics as the additive operation in the latent cell state space and fitted the dynamics to the observed unspliced transcriptome in combination with the equation of RNA velocity. This approach enabled us to decode the stochasticity of the dynamics as quantitative transcriptome fluctuation and to estimate the directionality of cell state dynamics. This study estimated stochastic dynamics in the latent cell state space with VAE architecture as an effective approach to analyze the dynamics behind asynchronous observation of numerous single cells.

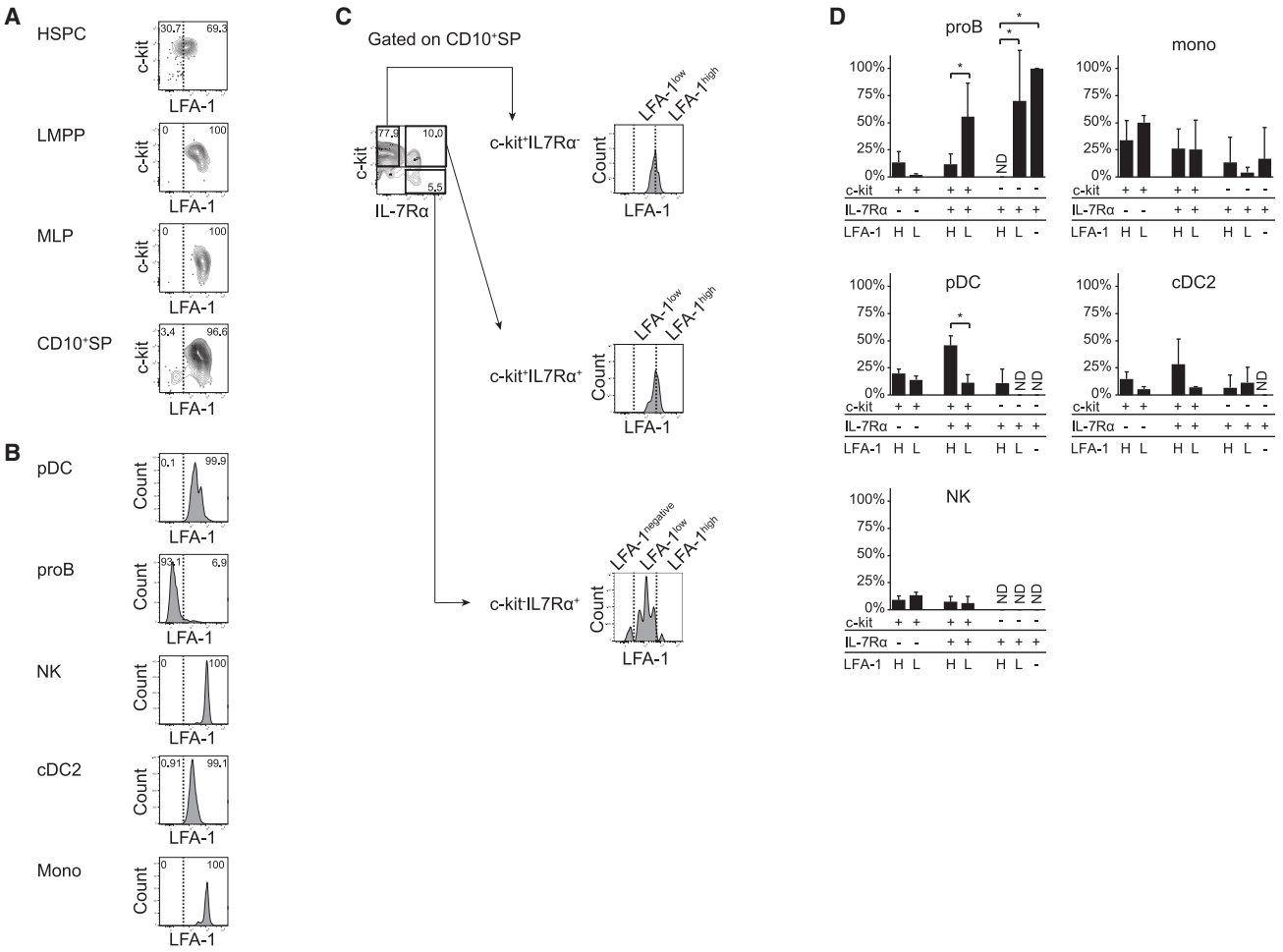


Figure 6. Relationship between LFA-1 expression levels and B versus pDC differentiation

(A) Expression of c-kit and LFA-1 in HSPCs, LMPPs, MLPs, and CD10$^+$ SP cells.

(B) Histograms of LFA-1 expression levels in pDCs, proB cells, NK cells, cDC2, and monocytic cells in cord blood.

(C) Sorting strategy based on LFA-1 expression.

(D) Each fraction of cells (sorting purity: >90%) was cultured for 14 days and then assessed. Error bars represent ±SEM of four independent experiments (unpaired t test; *p < 0.05). See also Figure S6.

Our computational methodology revealed a hotspot of large fluctuation at a B/pDC bifurcation in the human lymphoid pathway. This estimation was consistent with our observations that B cell and pDC differentiation potentials changed drastically with alterations in c-kit and IL-7Rα expression and B/pDC common progenitors were present in the IL-7Rα fraction of CD10$^+$SP cells. These findings are consistent with recent studies in mice, which found that the IL-7Rα$^+$ fraction of CLPs contained B/pDC common progenitors by scRNA-seq (Dress et al., 2019; Herman et al., 2018; Rodrigues et al., 2018). IL-7-deficient mice have low numbers of pDCs and B cells with reduced numbers of IL-7Rα$^+$ CLPs (Vogt et al., 2009). In humans, a case report has described two patients with loss of B cells, NK cells, and pDCs accompanied by a lack of lin$^-$CD34$^+$CD38$^+$CD10$^+$IL-7Rα$^+$ progenitors (Takada et al., 2009). These findings imply that a common pathway of B cells and pDC exists in the IL-7Rα$^+$ fraction and plays a critical role in generating B cells and pDCs in not only mice but also humans. Furthermore, our methodology revealed that LFA-1 expression dynamics fluctuated and were related to pDC differentiation tendency at the B/pDC bifurcation. This assumption was in accordance with culture observations showing that LFA-1 was broadly expressed in c-kit$^+$ lymphoid progenitors with a pDC differentiation ability, but its expression levels changed drastically and were involved in pDC over B-lineage choice in the c-kit$^-$IL-7Rα$^+$ fraction. It has been reported that LFA-1-dependent homotypic cell adhesion and adhesion to various cell types including B cells plays a critical role in interferon-α (IFN-α) production by pDCs (Bencze et al., 2021; Berggren et al., 2012; Reizis, 2019; Saitoh et al., 2017; Tomasello et al., 2018; Ye et al., 2020). Moreover, studies have revealed that crosstalk between pDCs and B cells via cell-to-cell contacts and IFN-α induces subsequent B cell differentiation toward plasma cells (Jego et al., 2003; Menon et al., 2016; Poeck et al., 2004). These findings imply that the B/pDC bifurcation point is susceptive to cellular crosstalk or extrinsic

factors to promptly adjust to immunological conditions, and that LFA-1 is involved in this process. Elucidation of the regulatory mechanism of B/pDC bifurcation will enhance our understanding of the pathophysiology of virus infection and immune-mediated diseases and aid in the development of therapeutic strategies.

Recent scRNA-seq analyses have revealed the heterogeneity of the gene expression profile and continuous and gradual lineage priming at the single-cell level in immature hematopoietic progenitors (Laurenti and Göttgens, 2018). Our data imply that there are critical hotspots in the hematopoietic system, where fluctuated transcriptome dynamics occur and cell lineage choice of hematopoietic progenitors is highly susceptible to intrinsic stochastic events or extrinsic signals. This observation is consistent with single-cell analysis of the dynamic transcriptional state during differentiation (Mojtahedi et al., 2016; Richard et al., 2016). Although our methodology could not resolve the molecular mechanism underlying the fluctuating dynamics, our AI-based analysis of the B/pDC bifurcation population showed that Polycomb group gene SFMBT2 was specifically expressed in the B/pDC bifurcation region. Single-cell multiome analysis further suggested that the estimated large fluctuation is associated with transient simultaneous accessibility of loci characterizing the difference of pDCs and B cells. Further study is required to understand the molecular mechanism of highly fluctuating dynamics at the cell bifurcation point. Thus, by orchestrating biological and computational methods, this study proposes a paradigm of a cell differentiation model based on comprehensive genetic profiling and provides the potential for application to various developmental and differentiation systems.

Limitations of the study

To avoid biological differences and obtain enough target cells, many samples of frozen cord blood were mixed and used for experiments. In particular, more than 30 samples of frozen cord blood were used for scRNA-seq of $CD10^{+}$ SP cells. Yet our experimental designs may come with a limitation because we utilized limited resources on assessment of the $CD10^{+}$ SP population. cDC1 potential was not fully evaluated owing to limited generation in our culture system. Additionally, our computational methodology assumed that the splicing and degradation rate of each transcript was constant across all cells, similarly to other RNA velocity estimation methods (the default estimation in Bergen et al., 2020; La Manno et al., 2018), while existing research suggested varying degradation rates during cell differentiation by combining labeling technology with single-cell transcriptome (Battich et al., 2020). Hence, our estimation of transcriptome dynamics would be improved by considering dynamic splicing and the degradation rate. Another limitation of our computational methodology is the difficulty in decomposing the estimated fluctuation of transcriptome dynamics into the intrinsic stochasticity and the uncertainty of environmental factors. We expect that this will be overcome by extending our computational algorithm to incorporate environmental factors in future.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data code and availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Isolation of hematopoietic precursors
  - Flow cytometric analysis
  - Cultures
  - Single-cell culture assay
  - Single-cell RNA and ATAC sequence analysis
  - Variational inference of single-cell transcriptome dynamics with fluctuation
  - Generative model of spliced and unspliced transcriptomes
  - Variational approximation of the latent variable posterior
  - Evaluation of the accuracy of estimated expression and dynamics for each gene
  - Calculation of the fluctuation in the gene expression change
  - Clustering analysis based on fluctuation
  - Fate prediction based on estimated dynamics
  - Calculation of the transition probability between single cells
  - Fate probability based on long-term transition
  - Scoring fluctuation direction
  - Two-dimensional embedding of velocity in the latent cell space
  - Application of other fate estimation tools
  - Simulation experiment of scRNA-seq
  - Calculation of co-accessibility of B/pDC-differentiating chromatin regions
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111260.

ACKNOWLEDGMENTS

We thank Tetsuo Kon, MD, PhD (Laboratory of Functional Genomics, Graduate School of Bioscience, Nagahama Institute of Bioscience and Technology, Nagahama, Shiga, Japan); Brandon Hadland, MD, PhD (Pediatric Oncology and Stem Cell and Gene Therapy Programs, Fred Hutchinson Cancer Research Center, USA) for technological suggestions regarding scRNA-seq analysis; and Motomu Shimaoka, MD, PhD (Deparment of Molecular Pathobiology and Cell Adhesion Biology, Mie University School of Medicine, Mie, Japan) for critical suggestions. We also thank Mitchell Arico from Edanz (https://jp.edanz.com/ac) for editing a draft of the manuscript. This study was supported by Grants-in-Aid for Scientific Research (C grant nos. 18K08322 and 21K08389 [K.O.] and S grant no. 17H06162 [H.N.]) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Projects for Cancer Research by Therapeutic Evolution (P-CREATE, no. 16cm0106301h0001 [H.N.]), a Development of Technology for Patient Stratification Biomarker Discovery grant (no.19ae0101074s0401 [H.N.]) from the Japan Agency for Medical Research and Development (AMED), the National Cancer Center Research and Development Fund (nos. 28-A-7 and 31-A-7

[H.N.]), JST, ACT-X grant number JPJPMJAX20AB (Y.K.), EKIDEN for Life 2019 (K.O.), and JSPS KAKENHI grant no. 16H06279 (PAGS) (K.O.).

AUTHOR CONTRIBUTIONS

K.N. designed and performed experiments and wrote the manuscript. Y.K. conducted scRNA-seq analysis and wrote the manuscript. H.H., K.M., K.H., H.M., Y.S., M.S., and Y.S. analyzed scRNA-seq data. Y.K., M.M., and I.T. analyzed biological data. N.K. supervised the study. S.N. collected samples. T.S. conducted and supervised scRNA-seq analysis. H.N. and K.O. designed and supervised the study.

DECLARATION OF INTERESTS

K.O. received research funding from Otsuka Pharmaceutical, Sumitomo Dainippon Pharma, Merck & Co, Astellas Pharma, and PharmaEssentia Corporation. H.N. received research funding and honoraria from Chugai Pharmaceutical, Ono Pharmaceutical, Bristol-Myers Squibb, and MSD, and research funding from Taiho Pharmaceutical, Daiichi-Sankyo, Kyowa Kirin, Zenyaku Kogyo, Oncolys BioPharma, Debiopharma, Asahi-Kasei, Sysmex, Fujifilm, SRL, Astellas Pharmaceutical, Sumitomo Dainippon Pharma, and BD Japan outside of this study.

Received: October 19, 2021
Revised: June 2, 2022
Accepted: August 4, 2022
Published: August 30, 2022

REFERENCES

Aigha, I.I., and Abdelalim, E.M. (2020). NKX6.1 transcription factor: a crucial regulator of pancreatic β cell development, identity, and proliferation. Stem Cell Res. Ther. *11*, 459. https://doi.org/10.1186/s13287-020-01977-0.

Bastidas-Ponce, A., Tritschler, S., Dony, L., Scheibner, K., Tarquis-Medina, M., Salinno, C., Schirge, S., Burtscher, I., Böttcher, A., Theis, F.J., et al. (2019). Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. Development *146*. https://doi.org/10.1242/dev.173849.

Battich, N., Beumer, J., de Barbanson, B., Krenning, L., Baron, C.S., Tanenbaum, M.E., Clevers, H., and van Oudenaarden, A. (2020). Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. Science (New York, N.Y.) *367*, 1151–1156. https://doi.org/10.1126/science.aax3072.

Bell, J.J., and Bhandoola, A. (2008). The earliest thymic progenitors for T cells possess myeloid lineage potential. Nature *452*, 764–767. https://doi.org/10.1038/nature06840.

Bencze, D., Fekete, T., and Pázmándi, K. (2021). Type I interferon production of plasmacytoid dendritic cells under control. Int. J. Mol. Sci. *22*, 4190. https://doi.org/10.3390/ijms22084190.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.

Berggren, O., Hagberg, N., Weber, G., Alm, G.V., Rönnblom, L., and Eloranta, M.L. (2012). B lymphocytes enhance interferon-α production by plasmacytoid dendritic cells. Arthritis Rheum. *64*, 3409–3419. https://doi.org/10.1002/art.34599.

Bigley, V., Maisuria, S., Cytlak, U., Jardine, L., Care, M.A., Green, K., Gunawan, M., Milne, P., Dickinson, R., Wiscombe, S., et al. (2018). Biallelic interferon regulatory factor 8 mutation: a complex immunodeficiency syndrome with dendritic cell deficiency, monocytopenia, and immune dysregulation. J. Allergy Clin. Immunol. *141*, 2234–2248. https://doi.org/10.1016/j.jaci.2017.08.044.

Blom, B., and Spits, H. (2006). Development of human lymphoid cells. Annu. Rev. Immunol. *24*, 287–320. https://doi.org/10.1146/annurev.immunol.24.021605.090612.

Bottero, V., Withoff, S., and Verma, I.M. (2006). NF-κB and the regulation of hematopoiesis. Cell Death Differ. *13*, 785–797. https://doi.org/10.1038/sj.cdd.4401888.

Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell *173*, 1535–1548.e16. https://doi.org/10.1016/j.cell.2018.03.074.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420. https://doi.org/10.1038/nbt.4096.

Cobaleda, C., Schebesta, A., Delogu, A., and Busslinger, M. (2007). Pax5: the guardian of B cell identity and function. Nat. Immunol. *8*, 463–470. https://doi.org/10.1038/ni1454.

Collin, M., and Bigley, V. (2018). Human dendritic cell subsets: an update. Immunology *154*, 3–20. https://doi.org/10.1111/imm.12888.

Courtney, M., Gjernes, E., Druelle, N., Ravaud, C., Vieira, A., Ben-Othman, N., Pfeifer, A., Avolio, F., Leuckx, G., Lacas-Gervais, S., et al. (2013). The inactivation of Arx in pancreatic α-cells triggers their neogenesis and conversion into functional β-like cells. PLoS Genet. *9*, e1003934. https://doi.org/10.1371/journal.pgen.1003934.

Courtney, M., Pfeifer, A., Al-Hasani, K., Gjernes, E., Vieira, A., Ben-Othman, N., and Collombat, P. (2011). In vivo conversion of adult α-cells into β-like cells: a new research avenue in the context of type 1 diabetes. Diabetes Obes. Metab. *13*, 47–52. https://doi.org/10.1111/j.1463-1326.2011.01441.x.

Dibaeinia, P., and Sinha, S. (2020). SERGIO: a single-cell expression simulator guided by gene regulatory networks. Cell Syst. *11*, 252–271.e11. https://doi.org/10.1016/j.cels.2020.08.003.

Dicken, J., Mildner, A., Leshkowitz, D., Touw, I.P., Hantisteanu, S., Jung, S., and Groner, Y. (2013). Transcriptional reprogramming of CD11b+Esamhi dendritic cell identity and function by loss of Runx3. PLoS One *8*, e77490. https://doi.org/10.1371/journal.pone.0077490.

Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. *38*, 737–746. https://doi.org/10.1038/s41587-020-0465-8.

Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., and Dick, J.E. (2010). Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nat. Immunol. *11*, 585–593. https://doi.org/10.1038/ni.1889. https://www.nature.com/articles/ni.1889#supplementary-information.

Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: a human perspective. Cell Stem Cell *10*, 120–136. https://doi.org/10.1016/j.stem.2012.01.006.

Dress, R.J., Dutertre, C.A., Giladi, A., Schlitzer, A., Low, I., Shadan, N.B., Tay, A., Lum, J., Kairi, M.F.B.M., Hwang, Y.Y., et al. (2019). Plasmacytoid dendritic cells develop from Ly6D(+) lymphoid progenitors distinct from the myeloid lineage. Nat. Immunol. *20*, 852–864. https://doi.org/10.1038/s41590-019-0420-3.

Eraslan, G., Simon, L.M., Mircea, M., Mueller, N.S., and Theis, F.J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. Nat. Commun. *10*, 390. https://doi.org/10.1038/s41467-018-07931-2.

Fang, X., Zheng, P., Tang, J., and Liu, Y. (2010). CD24: from A to Z. Cell. Mol. Immunol. *7*, 100–103. https://doi.org/10.1038/cmi.2009.119.

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.

Galy, A., Travis, M., Cen, D., and Chen, B. (1995). Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. Immunity *3*, 459–473. https://doi.org/10.1016/1074-7613(95)90175-2.

Grün, D. (2020). Revealing dynamics of gene expression variability in cell state space. Nat. Methods *17*, 45–49. https://doi.org/10.1038/s41592-019-0632-3.

Haas, S., Trumpp, A., and Milsom, M.D. (2018). Causes and consequences of hematopoietic stem cell heterogeneity. Cell Stem Cell *22*, 627–638. https://doi.org/10.1016/j.stem.2018.04.003.

Hashimoto, T., Gifford, D., and Jaakkola, T. (2016). Learning population-level diffusions with generative RNNs. In Proceedings of the 33rd International Conference on Machine Learning, B. Maria Florina and Q.W. Kilian, eds. (PMLR).

Helft, J., Anjos-Afonso, F., van der Veen, A.G., Chakravarty, P., Bonnet, D., and Reis e Sousa, C. (2017). Dendritic cell lineage potential in human early hematopoietic progenitors. Cell Rep. *20*, 529–537. https://doi.org/10.1016/j.celrep.2017.06.075.

Herman, J.S., Sagar, and Grün, D. (2018). FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nat. Methods *15*, 379–386. https://doi.org/10.1038/nmeth.4662.

Hystad, M.E., Myklebust, J.H., Bø, T.H., Sivertsen, E.A., Rian, E., Forfang, L., Munthe, E., Rosenwald, A., Chiorazzi, M., Jonassen, I., et al. (2007). Characterization of early stages of human B cell development by gene expression profiling. J. Immunol. *179*, 3662–3671. https://doi.org/10.4049/jimmunol.179.6.3662.

Jara, M.A., Werneck-De-Castro, J.P., Lubaczeuski, C., Johnson, J.D., and Bernal-Mizrachi, E. (2020). Pancreatic and duodenal homeobox-1 (PDX1) contributes to β-cell mass expansion and proliferation induced by Akt/PKB pathway. Islets *12*, 32–40. https://doi.org/10.1080/19382014.2020.1762471.

Jego, G., Palucka, A.K., Blanck, J.P., Chalouni, C., Pascual, V., and Banchereau, J. (2003). Plasmacytoid dendritic cells induce plasma cell differentiation through type I interferon and interleukin 6. Immunity *19*, 225–234. https://doi.org/10.1016/s1074-7613(03)00208-5.

Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., et al. (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. Nat. Immunol. *19*, 85–97. https://doi.org/10.1038/s41590-017-0001-2.

Kawamoto, H., and Katsura, Y. (2009). A new paradigm for hematopoietic cell lineages: revision of the classical concept of the myeloid-lymphoid dichotomy. Trends Immunol. *30*, 193–200. https://doi.org/10.1016/j.it.2009.03.001.

Kawamura, S., Onai, N., Miya, F., Sato, T., Tsunoda, T., Kurabayashi, K., Yotsumoto, S., Kuroda, S., Takenaka, K., Akashi, K., and Ohteki, T. (2017). Identification of a human clonogenic progenitor with strict monocyte differentiation potential: a counterpart of mouse cMoPs. Immunity *46*, 835–848.e4. https://doi.org/10.1016/j.immuni.2017.04.019.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2020). CellRank for directed single-cell fate mapping. Preprint at bioRxiv. https://doi.org/10.1101/2020.10.19.345983.

Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2022). CellRank for directed single-cell fate mapping. Nat. Methods *19*, 159–170. https://doi.org/10.1038/s41592-021-01346-6.

Laurenti, E., and Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. Nature *553*, 418–426. https://doi.org/10.1038/nature25022.

Levels, M.J., Fehres, C.M., van Baarsen, L.G.M., van Uden, N.O.P., Germar, K., O'Toole, T.G., Blijdorp, I.C.J., Semmelink, J.F., Doorenspleet, M.E., Bakker, A.Q., et al. (2019). BOB.1 controls memory B-cell fate in the germinal center reaction. J. Autoimmun. *101*, 131–144. https://doi.org/10.1016/j.jaut.2019.04.011.

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

Lotfollahi, M., Wolf, F.A., and Theis, F.J. (2019). scGen predicts single-cell perturbation responses. Nat. Methods *16*, 715–721. https://doi.org/10.1038/s41592-019-0494-8.

Loughran, S.J., Haas, S., Wilkinson, A.C., Klein, A.M., and Brand, M. (2020). Lineage commitment of hematopoietic stem cells and progenitors: insights from recent single cell and lineage tracing technologies. Exp. Hematol. *88*, 1–6. https://doi.org/10.1016/j.exphem.2020.07.002.

Macaulay, I.C., Svensson, V., Labalette, C., Ferreira, L., Hamey, F., Voet, T., Teichmann, S.A., and Cvejic, A. (2016). Single-cell RNA-sequencing reveals a continuous spectrum of differentiation in hematopoietic cells. Cell Rep. *14*, 966–977. https://doi.org/10.1016/j.celrep.2015.12.082.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: uniform manifold approximation and projection for dimension reduction. Preprint at arXiv. https://doi.org/10.48550/arXiv.1802.03426.

Melsted, P., Booeshaghi, A.S., Liu, L., Gao, F., Lu, L., Min, K.H.J., da Veiga Beltrame, E., Hjörleifsson, K.E., Gehring, J., and Pachter, L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat. Biotechnol. *39*, 813–818. https://doi.org/10.1038/s41587-021-00870-2.

Menon, M., Blair, P.A., Isenberg, D.A., and Mauri, C. (2016). A regulatory feedback between plasmacytoid dendritic cells and regulatory B cells is aberrant in systemic lupus erythematosus. Immunity *44*, 683–697. https://doi.org/10.1016/j.immuni.2016.02.012.

Minami, H., Nagaharu, K., Nakamori, Y., Ohishi, K., Shimojo, N., Kageyama, Y., Matsumoto, T., Sugimoto, Y., Tawara, I., Masuya, M., et al. (2017). CXCL12-CXCR4 Axis is required for contact-mediated human B lymphoid and plasmacytoid dendritic cell differentiation but not T lymphoid generation. J. Immunol. *199*, 2343–2355. https://doi.org/10.4049/jimmunol.1700054.

Minegishi, Y., Rohrer, J., Coustan-Smith, E., Lederman, H.M., Pappu, R., Campana, D., Chan, A.C., and Conley, M.E. (1999). An essential role for BLNK in human B cell development. Science *286*, 1954–1957. https://doi.org/10.1126/science.286.5446.1954.

Miri, K., Latham, K., Panning, B., Zhong, Z., Andersen, A., and Varmuza, S. (2013). The imprinted polycomb group gene Sfmbt2 is required for trophoblast maintenance and placenta development. Development *140*, 4480–4489. https://doi.org/10.1242/dev.096511.

Mojtahedi, M., Skupin, A., Zhou, J., Castaño, I.G., Leong-Quong, R.Y.Y., Chang, H., Trachana, K., Giuliani, A., and Huang, S. (2016). Cell fate decision as high-dimensional critical state transition. PLoS Biol. *14*, e2000640. https://doi.org/10.1371/journal.pbio.2000640.

Nakamori, Y., Liu, B., Ohishi, K., Suzuki, K., Ino, K., Matsumoto, T., Masuya, M., Nishikawa, H., Shiku, H., Hamada, H., and Katayama, N. (2012). Human bone marrow stromal cells simultaneously support B and T/NK lineage development from human haematopoietic progenitors: a principal role for flt3 ligand in lymphopoiesis. Br. J. Haematol. *157*, 674–686. https://doi.org/10.1111/j.1365-2141.2012.09109.x.

Ohishi, K., Varnum-Finney, B., and Bernstein, I.D. (2002). Delta-1 enhances marrow and thymus repopulating ability of human CD34+CD38– cord blood cells. J. Clin. Invest. *110*, 1165–1174. https://doi.org/10.1172/JCI16167.

Pellin, D., Loperfido, M., Baricordi, C., Wolock, S.L., Montepeloso, A., Weinberg, O.K., Biffi, A., Klein, A.M., and Biasco, L. (2019). A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. Nat. Commun. *10*, 2395. https://doi.org/10.1038/s41467-019-10291-0.

Poeck, H., Wagner, M., Battiany, J., Rothenfusser, S., Wellisch, D., Hornung, V., Jahrsdorfer, B., Giese, T., Endres, S., and Hartmann, G. (2004). Plasmacytoid dendritic cells, antigen, and CpG-C license human B cells for plasma cell differentiation and immunoglobulin production in the absence of T-cell help. Blood *103*, 3058–3064. https://doi.org/10.1182/blood-2003-08-2972.

Reizis, B. (2019). Plasmacytoid dendritic cells: development, regulation, and function. Immunity *50*, 37–50. https://doi.org/10.1016/j.immuni.2018.12.027.

Richard, A., Boullu, L., Herbach, U., Bonnafoux, A., Morin, V., Vallin, E., Guillemin, A., Papili Gao, N., Gunawan, R., Cosette, J., et al. (2016). Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding

irreversible commitment in a differentiation process. PLoS Biol. *14*, e1002585. https://doi.org/10.1371/journal.pbio.1002585.

Rodrigues, P.F., Alberti-Servera, L., Eremin, A., Grajales-Reyes, G.E., Ivanek, R., and Tussiwand, R. (2018). Distinct progenitor lineages contribute to the heterogeneity of plasmacytoid dendritic cells. Nat. Immunol. *19*, 711–722. https://doi.org/10.1038/s41590-018-0136-9.

Rojas-Sutterlin, S., Lecuyer, E., and Hoang, T. (2014). Kit and Scl regulation of hematopoietic stem cells. Curr. Opin. Hematol. *21*, 256–264. https://doi.org/10.1097/moh.0000000000000052.

Saitoh, S.-I., Abe, F., Kanno, A., Tanimura, N., Mori Saitoh, Y., Fukui, R., Shibata, T., Sato, K., Ichinohe, T., Hayashi, M., et al. (2017). TLR7 mediated viral recognition results in focal type I interferon secretion by dendritic cells. Nat. Commun. *8*, 1592. https://doi.org/10.1038/s41467-017-01687-x.

Seita, J., and Weissman, I.L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip. Rev. Syst. Biol. Med. *2*, 640–653. https://doi.org/10.1002/wsbm.86.

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. Nat. Biotechnol. *37*, 451–460. https://doi.org/10.1038/s41587-019-0068-4.

Sichien, D., Scott, C.L., Martens, L., Vanderkerken, M., Van Gassen, S., Plantinga, M., Joeris, T., De Prijck, S., Vanhoutte, L., Vanheerswynghels, M., et al. (2016). IRF8 transcription factor controls survival and function of terminally differentiated conventional and plasmacytoid dendritic cells, respectively. Immunity *45*, 626–640. https://doi.org/10.1016/j.immuni.2016.08.013.

Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods *18*, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5.

Takada, H., Kikushige, Y., Arinobu, Y., Doi, T., Ishimura, M., Ishikawa, F., Akashi, K., and Hara, T. (2009). Novel primary immunodeficiency syndrome with a developmental defect of B cells, NK cells and plasmacytoid dendritic cells. In The 5th Congress of Asian Society for Pediatric Reserch *Hangzhou*.

Tang, P., Miri, K., and Varmuza, S. (2019). Unique trophoblast chromatin environment mediated by the PcG protein SFMBT2. Biol. Open *8*, bio043638. https://doi.org/10.1242/bio.043638.

Tomasello, E., Naciri, K., Chelbi, R., Bessou, G., Fries, A., Gressier, E., Abbas, A., Pollet, E., Pierre, P., Lawrence, T., et al. (2018). Molecular dissection of plasmacytoid dendritic cell activation in vivo during a viral infection. EMBO J. *37*, e98836. https://doi.org/10.15252/embj.201798836.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.

Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. *19*, 271–281. https://doi.org/10.1038/ncb3493.

Vogt, T.K., Link, A., Perrin, J., Finke, D., and Luther, S.A. (2009). Novel function for interleukin-7 in dendritic cell development. Blood *113*, 3961–3968. https://doi.org/10.1182/blood-2008-08-176321.

Wada, H., Masuda, K., Satoh, R., Kakugawa, K., Ikawa, T., Katsura, Y., and Kawamoto, H. (2008). Adult T-cell progenitors retain myeloid potential. Nature *452*, 768–772. https://doi.org/10.1038/nature06839.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

Wu, H., and Shao, Q. (2019). The role of inhibitor of binding or differentiation 2 in the development and differentiation of immune cells. Immunobiology *224*, 142–146. https://doi.org/10.1016/j.imbio.2018.09.006.

Ye, Y., Gaugler, B., Mohty, M., and Malard, F. (2020). Plasmacytoid dendritic cell biology and its role in immune-mediated diseases. Clin. Transl. Immunology *9*, e1139. https://doi.org/10.1002/cti2.1139.

Zhu, Y., Liu, Q., Zhou, Z., and Ikeda, Y. (2017). PDX1, Neurogenin-3, and MAFA: critical transcription regulators for beta cell development and regeneration. Stem Cell Res. Ther. *8*, 240. https://doi.org/10.1186/s13287-017-0694-z.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
PE/Cyanine 7 Mouse anti-CD34	BioLegend	Cat. No. 343515; RRID AB_1877252
APC/Cyanine 7 Mouse anti-CD38	BioLegend	Cat. No. 356615; RRID AB_2562576
BV510 Mouse anti-CD45RA	BioLegend	Cat. No. 304141; RRID AB_2561384
PE/Dazzle Mouse anti-CD10	BioLegend	Cat. No. 312227; RRID AB_2565877
PE Mouse anti-CD127	BioLegend	Cat. No. 351303; RRID AB_10719960
APC Mouse anti-CD117	BioLegend	Cat. No. 313205; RRID AB_314984
BV421 Mouse anti-CD7	BD	Cat. No. 562635; RRID AB_2736907
FITC Mouse anti-LFA1	BD Pharmingen	Cat. No. 555383; RRID AB_395784
FITC Mouse anti-CD14	BioLegend	Cat. No. 367115; RRID AB_2571928
PE Mouse anti-CD56	BioLegend	Cat. No. 318305; RRID AB_604093
APC Mouse anti-CD10	BioLegend	Cat. No. 312209; RRID AB_314920
BV510 Mouse anti-CD45	BioLegend	Cat. No. 304035; RRID AB_2561383
BV605 Mouse anti-CD45RA	BioLegend	Cat. No. 304133; RRID AB_11126164
FITC Mouse anti-HLA-DR	BD Pharmingen	Cat. No. 555811; RRID AB_396145
PE Mouse anti-CD123	BioLegend	Cat. No. 306005; RRID AB_314579
PE/Cyanine 7 Mouse anti-CD304	BioLegend	Cat. No. 354507; RRID AB_2561556
APC Mouse anti-CD303	Miltenyi Biotec	Cat. No. 130-113-190; RRID AB_2726017
Alexa Fluor 488 Mouse anti-CD14	BioLegend	Cat. No. 367129; RRID AB_2721359
BV421 Mouse anti-CD11c	BioLegend	Cat. No. 301627; RRID AB_10898313
FITC Mouse anti-CD66b	BioLegend	Cat. No. 305103; RRID AB_314495
PE Mouse anti-CD141	Miltenyi Biotec	Cat. No. 130-113-880; RRID AB_2726094
BV421 Mouse anti-CD1c	BioLegend	Cat. No. 331525; RRID AB_10933249
PE Mouse anti-CD123	BioLegend	Cat. No. 306013; RRID AB_755989
Chemicals, peptides, and recombinant proteins		
GM-CSF	PeproTech	300–03
FLT3L	PeproTech	300–19
TPO	PeproTech	300–18
SCF	PeproTech	300–07
Critical commercial assays		
Chromium Single Cell 30 Gel Bead and Library Kit	10X Genomics	120235, 120234, 120236, and 120262
Deposited data		
single cell RNA-seq of $CD10^{+}SP$	This study	JGAS000551
Single Cell Multiome ATAC and gene expression data of human umbilical cord blood CD34	This study	JGAS000528
Single cell RNA-seq of pancreas	Bastidas-Ponce et al., 2019	GSE132188
Software and algorithms		
VICDYF	This study	https://github.com/kojikoji/vicdyf https://doi.org/10.5281/zenodo.6890408
Cell Ranger	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cell Ranger ARC	10X Genomics	https://support.10xgenomics.com/single-cell-multiome-atac-gex/software/pipelines/latest/what-is-cell-ranger-arc
kallisto | bustools	Melsted et al. (2021)	https://www.kallistobus.tools
velocyto	La Manno et al. (2018)	http://velocyto.org
Python	Python Software Foundation	https://www.python.org/
Pytorch	PyTorch community	https://pytorch.org/
R	R Development Core Team	https://www.r-project.org/
Seurat	Butler et al., 2018	http://satijalab.org.seurat/
Signac	Stuart et al., (2021)	https://satijalab.org/signac/index.html
scanpy	Wolf et al., 2018	https://scanpy.readthedocs.io
UMAP	McInnes et al., (2018)	https://umap-learn.readthedocs.io/en/latest/
Leiden algorithm	Traag et al., (2019)	https://leidenalg.readthedocs.io/en/stable/
VarID	Grün, 2020	https://github.com/dgrun/VarID_analysis
CELLRANK	Lange et al., (2022)	https://cellrank.readthedocs.io/en/stable/
Palantir	Setty et al., (2019)	https://github.com/dpeerlab/Palantir

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the lead contact Kohshi Ohishi (koishi@clin.medic.mie-u.ac.jp).

Materials availability

This study did not generate new unique reagents.

Data code and availability

Single Cell Multiome ATAC and gene expression data of human umbilical cord blood CD34$^+$ cells and single cell RNA-seq of CD10$^+$SP cells have been deposited in the Japanese Genotype-phenotype Archive (https://www.ddbj.nig.ac.jp/jga) which is hosted by the Bioinformation and DDBJ Center, under accession number JGAS000528 and JGAS000551, respectively. Computational methodology for variational inference of cell state dynamics with fluctuation (vicdyf) is available on GitHub (https://github.com/kojikoji/vicdyf). A version recode for the GitHub repository has been archived at Zenodo. A DOI for the codes is listed in the key resources table.

Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human umbilical cord blood samples were collected from full-term deliveries after obtaining informed consent in accordance with a protocol approved by the Ethics Committee of Mie University Hospital (approval number 341).

METHOD DETAILS

Isolation of hematopoietic precursors

CD34+ cells were separated from mononuclear cells using CD34 immunomagnetic beads (Miltenyi Biotec, Auburn, CA, USA) in accordance with the manufacturer's instructions. To avoid biological differenced and obtain enough target cells, frozen cord blood were mixed and used for experiments. After mixture of frozen cord blood CD34$^+$ cells, cells were stained with antibodies and HSPCs (CD34$^+$CD38$^-$CD45RA$^-$CD7$^-$CD10$^-$), MLPs (CD34$^+$CD38$^-$CD45RA$^+$CD10$^+$), LMPPs (CD34$^+$CD38$^-$CD45RA$^+$CD10$^-$CD7$^-$), CD7$^+$ SP cells (CD34$^+$CD38$^+$CD45RA$^+$CD10$^-$CD7$^+$CD19$^-$), and CD10$^+$ SP cells (CD34$^+$CD38$^+$CD45RA$^+$CD10$^+$CD7$^-$CD19$^-$) were isolated. Cell sorting was performed with a FACSAria II cell sorter (BD Biosciences, San Jose, CA, USA), which excluded Zombie NIR$^+$ dead cells.

Flow cytometric analysis

Surface staining for flow cytometry was performed using the following murine mAbs: anti-CD14 FITC (BioLegend, San Diego, CA, USA), anti-CD45 FITC (BioLegend), anti-HLA-DR FITC (BD Biosciences), anti-CD10 PE (BioLegend), anti-IL-7Rα PE (BioLegend), anti-IL3Rα PE (BD Pharmingen, San Diego, CA, USA), anti-CD19 APC (BioLegend), anti-CD303 (BDCA2) APC (Miltenyi Biotec), CD14 Alexa Fluor 700 (BioLegend), anti-CD7 Brilliant Violet 421, anti-CD11c Brilliant Violet 421, and anti-CD45RA Brilliant Violet 510 (BioLegend). IgG1 FITC, IgG2a FITC, IgG1 PE (all from BioLegend), IgG2a PE (BD Pharmingen), IgG2b PE, IgG1 PerCP-Cy5.5, IgG1 APC, IgG1 APC-Cy7, IgG1 PECy7, IgG1 Alexa Fluor 700, IgG1 Brilliant Violet 421 (all from BioLegend), and IgG2b Brilliant Violet 510 (BD Biosciences) were used as isotype controls. Cells were blocked with FcR Blocking Reagent (Miltenyi Biotec) and then incubated with antibodies for 30 min at 4°C. Dead cells were excluded by staining with 7-aminoactinomycin D (BD Biosciences) or Zombie NIR.

For cytoplasmic staining, cells were incubated with various Abs against surface Ags and Zombie NIR for 30 min at 4°C. After washing, the cells were permeabilized and fixed with PermeaFix (Ortho, Raritan, NJ, USA) for 20 min at room temperature, washed again, and then incubated with anti-CD79a APC (BD Biosciences) for 30 min at 4°C. Dead cells were distinguished by positive staining for Zombie NIR (Ohishi et al., 2002).

Flow cytometric analysis was performed using FACSCanto II and BD LSRFortessa flow cytometers (BD Biosciences). All data were analyzed using BD FACSDiva software (BD Biosciences) or Fortessa flow cytometers (BD Biosciences), and processed by FlowJo software (TreeStar, San Carlos, CA, USA).

Cultures

First, stromal cells were plated in a 25-cm^2 cell culture flask (Corning, Corning, NY, USA) or 24- or 48-well tissue culture plates (Nunc, Roskilde, Denmark) with αMEM (Life Technologies-Invitrogen, Grand Island, NY, USA) that contained 12.5% horse serum, 12.5% FCS (both from Invitrogen, Carlsbad, CA, USA), and 1 μM hydrocortisone (Sigma-Aldrich, St Louis, MO, USA). Confluent stromal cells were washed with αMEM and isolated target cells (95% pure) were seeded on a confluent monolayer of stromal cells in αMEM supplemented with 20% FCS (HyClone Laboratories, Logan, UT, USA), 50 U/mL penicillin, and 50 μg/mL streptomycin in the presence of four growth factors (SCF + Flt3L + TPO + GM-CSF).

Single-cell culture assay

Briefly, hTERT stromal cells were cultured in 96-well plates (Nunc, Roskilde Denmark) with αMEM that contained 20% FCS and growth factors (SCF + Flt3L + TPO + GM-CSF). Each well was assessed at 21 days. Then, flow cytometry was conducted for positive wells.

Single-cell RNA and ATAC sequence analysis

We conducted scRNA-seq of $CD10^+$ SP and single cell gene expression chromatic accessibility analysis of $CD34^+$ cells. $CD10^+$ SP and $CD34^+$ cells were isolated from human umbilical cord blood as described above. All samples had high mean viability of approximately 95% after cell sorting. Single-cell RNA-sequencing (scRNA-seq) and multiome sequencing libraries were generated using the Chromium Single Cell 3' Library & Gel Bead Kit v3 and Single Cell Multiome ATAC + Gene Expression Kit (10X Genomics, Pleasanton, CA, USA) following the manufacturer's instructions. As a result, a cell barcoding sequence and unique molecular identifier were added to individual cDNA molecules. Libraries were constructed and sequenced at a depth of approximately 120,000 reads per cell using the Novaseq 6000 platform (Illumina, San Diego, CA, USA). For mapping of scRNA-seq of the $CD10^+$SP population and single cell multiome analysis of the $CD34^+$ population, sequences obtained from the 10X Genomics scRNA-seq platform were demultiplexed and mapped using Cell Ranger and Cell Ranger ARC package (10X Genomics), respectively. For single cell transcriptome data of $CD10^+$SP and $CD34^+$ population, we also quantified spliced and unspliced transcript counts using kallisto | bustools (Melsted et al., 2021) and velocyto (La Manno et al., 2018) respectively. Cells were removed if they expressed fewer than 200 unique genes, more than 8000 unique genes, or had greater than 15% mitochondrial reads. Genes not detected in any cell were removed from subsequent analyses. Using the scanpy (Wolf et al., 2018) package in Python language, we conducted PCA of expression profiles of the cells (n_comps = 40) and Leiden clustering (Traag et al., 2019) on the PCA coordinates with scanpy.pp.neighbors and scanpy.tl.leiden (n_neighbors = 30, resolution = 1.0). Using Signac R package (Stuart et al., 2021), we filtered out low quality cells of demultiplexed single cell gene expression and ATAC data of the $CD34^+$ population. In particular, the minimum and maximum values of RNA counts were set to 1000 and 20,000, respectively, the minimum and maximum values of ATAC counts were set to 2000 and 40,000, respectively, the maximum value of nucleosome signal was set to 5 and the minimum value of TSS enrichment was set to 1. We applied the RunRTIDF function of the Signac package for normalization.

Variational inference of single-cell transcriptome dynamics with fluctuation

We developed a computational methodology to estimate single-cell transcriptome dynamics with fluctuation. Our methodology stochastically embedded spliced counts of each cell into a latent cell state space using a deep neural network. For each latent cell state, our methodology stochastically simulated the micro-duration change and decoded the time-shifted spliced transcriptome as well as the original spliced transcriptome. The unspliced transcriptome was reconstructed from original and time-shifted spliced transcriptomes, which leveraged the differential equation of splicing kinetics. We assumed probabilistic observation of spliced and unspliced

counts from reconstructed counterparts and optimized parameters for encoding, decoding, stochastic simulation, and the RNA velocity equation. Here, we describe the generative model for spliced and unspliced transcriptomes and the variational approximation of the posterior distributions of latent variables.

Generative model of spliced and unspliced transcriptomes

This is our generative model for spliced and unspliced transcriptomes of a single cell, $s \in R^g$ and $u \in R^g$ where g is the number of genes. This probabilistic model assumes two latent variables, $z \in R^m$ and $d \in R^m$, where m is the dimension of the latent cell state space. z represents the latent cell state and follows a Gaussian prior distribution

$$p(z) = N(z|0, I)$$

d represents the change in the latent cell state during micro-duration δt and follows a Gaussian prior distribution

$$p(d) = N(z|0, I)$$

We modeled the latent cell state after micro-duration δt as $z' = z+\rho d$. We set $\rho<<1$ so that the change in the micro-duration is much smaller than the whole variation of the latent cell state. We set ρ as 0.01 in this study. This generative model assumes that the time evolution of the latent cell state z follows a Wiener process. Given the corresponding latent cell state z,we assume that spliced transcriptome s follows a Poisson distribution

$$p(s|z) = Poisson(s|\bar{s} = l\lambda_\theta(z))$$

where $l \in R$ is the mean expression across all genes in the single cell and $\lambda_\theta(z) \in R^g$ is the decoding neural network of the latent cell state with 50 hidden units, two layers, and layer normalization. We also decoded the latent cell state before and after micro-duration δt, $z-\rho d$ and $z+\rho d$ as $\lambda_\theta(z-\rho d)$and $\lambda_\theta(z+\rho d)$, which corresponded to mean parameters of the spliced transcriptome before and after the micro-duration. Using these time-shifted spliced transcriptomes, we approximately derived the time change of the mean parameter of the spliced transcriptome

$$\frac{d\bar{s}}{dt} \approx \frac{\lambda_\theta(z+\rho d) - \lambda_\theta(z - \rho d)}{2\delta t} \quad \text{(Equation 1)}$$

Here, we assumed that the mean parameter of spliced and unspliced transcriptomes followed the differential equation of splicing kinetics as with existing tools for RNA velocity estimation (Bergen et al., 2020; La Manno et al., 2018):

$$\frac{d\bar{s}}{dt} = \beta \circ \bar{u} - \gamma \circ \bar{s} \quad \text{(Equation 2)}$$

where $\beta \in R^g$ is a vector of gene-specific splicing rates of unspliced transcripts and $\gamma \in R^g$ is a vector of gene-specific degradation rates of spliced transcripts. By combininUg (1) and (2), we approximately derived the mean parameter of the unspliced transcriptome

$$\bar{u} \approx l\frac{\lambda_\theta(z+\rho d) - \lambda_\theta(z - \rho d) + 2\delta t\gamma\lambda_\theta(z)}{2\delta t\beta}$$

We assumed that the unspliced transcriptome u followed a Poisson distribution

$$p(u|z, d) = Poisson(u|\bar{u})$$

Variational approximation of the latent variable posterior

To derive the latent cell state z and corresponding time change d, we used the variational approximation for the posterior distribution of z and d similarly to a variational auto-encoder. We assumed the variational distribution of the latent cell state as a Gaussian distribution dependent on the observed spliced transcriptome as shown below:

$$q(z|s) = N(z|\mu_\varphi(s), \mathrm{diag}(\sigma_\varphi(s)))$$

where $\mu_\varphi(s)$ and $\sigma_\varphi(s)$ are the encoding neural network with 50 hidden units, two layers, and layer normalization. We assumed that the variational distribution of d is was a Gaussian distribution dependent on the latent cell state z as shown below:

$$q(d|z) = N(d|\mu'_\varphi(z), \mathrm{diag}(\sigma'_\varphi(z)))$$

where $\mu'_\varphi(z)$ and $\sigma'_\varphi(z)$ are neural networks with 50 hidden units, two layers, and layer normalization. These formulations corresponded to an assumption that the time evolution of latent cell state z followed an advection diffusion model

$$dz = \rho\,\mu'_{\varphi'}(z)\,dt + \rho\,\sigma'_{\varphi'}(z)\,dW$$

where W is a Wiener process. Hence, $\sigma'_{\varphi'}(z)$ correspondcorresponded to the fluctuation of latent cell state dynamics and $\mu'_{\varphi'}(z)$ corresponded to the average dynamics of latent cell states. For the estimation process of the parameters of the generative models and the variational distribution, we maximized Evidence Lower BOund (ELBO) loss $L(\theta, \varphi, \varphi')$ defined below:

$$L(\theta,\ \varphi,\ \varphi',\ \alpha,\ \beta)\ =\ -\ E_{q(z,d|s)}\left[\log\frac{P(s,\ u,\ z,\ d)}{q(z,\ d|s)}\right]\approx\ -\ \log\ p(s|z')\ -\ \log\ p(u|z',\ d')+D_{KL}(q(z|s)|p(z))+D_{KL}(q(d|z')|p(d))$$

where $z'$and d' are derived through reparametrized sampling form $q(z|s)$ and $q(d|z')$, and $E_{P(x)}[f(x)]$ represents the expectation of $f(x)$ given $x\ \sim P(x)$. For this maximization, we used the Adam optimizer with a learning rate of 0.0001, mini-batch size of 100, and 1000 epochs at total. Only for the first 500 epochs, we fixed φ' and ignored $D_{KL}(q(d|z')|p(d))$ and $-\log\ p(u|z',\ d')$ from ELBO loss, which corresponded to the variational auto encoder of spliced transcriptome. All implementations were based on the pytorch library of Python language.

Evaluation of the accuracy of estimated expression and dynamics for each gene

Owing to the sparse observation for weakly expressed genes or the strong deviation from our constant RNA degradation assumption, the reconstructed expression and dynamics for some genes were unreliable. To quantify the accuracy of estimated expression for each gene, we calculated the Pearson correlation between the observed spliced transcript count and reconstructed mean parameter c_s. To quantify the accuracy of estimated dynamics for each gene, we calculated the Pearson correlation of the ratio of the unspliced transcript count to the spliced transcript count added by one between the observation and reconstruction c_r. We excluded genes with low accuracy c_s<0.3or c_r<0.1, when we scored the fluctuation direction of each gene as described below section.

Calculation of the fluctuation in the gene expression change

The variational distribution of d enableenabled us to evaluate the stochastic dynamics of latent cell states, while each latent cell state was associated with a comprehensive gene expression profile by decoding the neural network. Here, we iteratively sampled N times micro-duration changes $d_{i,n}(n = 1,..,N)$ for the latent state of the cell i, $z_i\ = \mu_\varphi\ (s_i)$, from the variational distribution $q(d\ |z_i)$. We decoded the cell states after micro-transitions $z_i\ +\ \rho d_{i,n}$ and derived multiple spliced transcript abundances after the micro-duration, which reflected the stochastic dynamics in the latent cell state space. Here, we calculated the fluctuation of the gene expression change as their variance among these cells sampled from one original observed cell. We also calculated the advection of the gene expression change as their average.

Clustering analysis based on fluctuation

On the basis of this deviation from the average expression changes described above, we conducted clustering analysis on cells with the top 20% largest fluctuation. Using the scanpy package in Python language, we conducted PCA of the deviation of cells (n_comps = 20) and clustering on PCA coordinates with scanpy.pp.neighbors and scanpy.tl.leiden (n_neighbors = 30, resolution = 1.0).

Fate prediction based on estimated dynamics

To estimate the cell fate probability for each single cell, we calculated the transition probability between single cells, which is consistent with the estimated cell state dynamics, and chained it to simulate the destination as described below.

Calculation of the transition probability between single cells

To estimate fate probability for each lineage, we quantified the probability that the transition between single cells occurs. We calculated the transition probability from cell i based onbased on the iteratively sampled latent cell state change during the micro-duration $d_{i,n}(n = 1,..,N)$. We assumed that the destination of n th transition from cell i is cell j when the cosine similarity between the transition and latent coordinate difference from cell i to cell j is smallest among the neighbor $j \in N(z_i)$.

We defined the transition probability as shown below:

$$T_{ij}\ =\ \begin{cases}\dfrac{1}{N}\ \sum\limits_{n\ =\ 1,\ldots,N} I\left(\operatorname{argmax}_{j\ \in\ Nk(z_i)}\dfrac{d_{in}\cdot\Delta_{ij}}{||d_{in}||\ ||\Delta_{ij}||}\right) & (j\in N(z_i))\\ 0 & (j\notin N(z_i))\end{cases}$$

where $\Delta_{ij}\ =\ z_j\ -\ z_i,\ N\ =\ 30,\ N(z_i)$ is the k-th nearest neighborhoods of z_i.

Fate probability based on long-term transition

We determined destination cells and assumed that the transition from destination cells was unavailable. The destination cells were defined by cells with the top 5% expression profiles similar to those of expected destination cell clusters. The similarity of cell i to cluster k was formulated as shown below:

$$r_{ik}\ =\ (s_i\ -\ \tilde{s})\cdot(f_k\ -\ \tilde{f})$$

where $\tilde{s}$ is the averaged spliced transcriptome across all cells, f_k is the averaged expression profile of cell type k in the PBMC dataset derived from Seurat, and $\tilde{f}$ is the average of that over all cell types. We chained the corrected transition probability 1024 times and estimated the cell fate probability of a specific cell type by summing the long-term transition probability to cells whose expression profile was the most similar to that of the specific cell type among all destination cell types.

Scoring fluctuation direction

As described in the previous section, we simulated the stochastic gene expression change in each cell. Here, we estimated the average expression change from iteratively sampled cells after a micro-duration as shown below:

$$\bar{s}'_i = \frac{1}{N}\sum_n \lambda_\theta(z_i + \rho d_{i,n}) - \lambda_\theta(z_i)$$

We calculated the deviation of the gene expression change sampled once per cell from these averaged expression changes as shown below:

$$\delta_i = \lambda_\theta\left(z_i + \rho d'_i\right) - \bar{s}'_i$$

where d'_i was sampled from $q(d\,|z_i)$. We evaluated the directionality of the fluctuation deviation of each cell based on the similarity to the expression difference between two specific cell types as shown below:

$$p_{ikk'} = \delta_i \cdot (f_k - f_{k'})$$

large value of $p_{ikk'}$ indicates that the dynamics of cell i is directed toward cell type k rather than cell type k'. Furthermore, this cell-specific score can be used to evaluate each gene directionality between the two cell types in specific cell subpopulation Cas shown below:

$$e_{gkk'} = \sum_{i \in C} p_{ikk'}\delta_{ig}$$

A large value of $e_{gkk'}$ indicates that the expression change of gene g fluctuates positively in cells whose dynamics is directed toward cell type k rather than cell type k'.

Two-dimensional embedding of velocity in the latent cell space

We applied the embedding method of Bergen et al. (2020) to our latent cell state dynamics. First, we calculated the transition probability between single cells as shown below:

$$Q_{ij} = \begin{cases} \frac{1}{C'_i}\exp\left(\frac{c(\bar{d}_i, \Delta_{ij})}{\sigma}\right) & (j \in N(i)) \\ 0 & (j \notin N(i)) \end{cases}$$

where $c(x,y)$ is the cosine similarity between x and y, $\bar{d}_i = \mu_\varphi(s_i)$, C'_i is the normalization factor, so that $\sum_j Q_{ij} = 1$ and σ is a constant value (0.05 in this study).

Next, we calculated the change of embedded coordinates by averaging the difference between cells for the transition probability as shown below:

$$d^u_i = \sum_j Q_{ij}(u_j - u_i)$$

where $u_i \in R^2$ is the UMAP embeddings of the latent cell state of cell i, $\mu_\varphi(s_i)$.

Application of other fate estimation tools

We applied CELLRANK and Palantir to the pancreas dataset with default parameters. For Palantir, we determined the initial state as the cell whose PCA coordinate was nearest to the center of Ngn3^{low} endocrine progenitor cells. Terminal states corresponding to Alpha, Beta, Delta, Epsilon, and Ductal cells were determined as the cells whose PCA coordinates were nearest to the center of the corresponding clusters.

Simulation experiment of scRNA-seq

We simulated the spliced and unspliced transcriptome of differentiating cells using SERGIO (Dibaeinia and Sinha, 2020). We assumed three cell types with 100 genes and differentiation from cell type 0 to 1. To heavily capture the initiation of differentiation, we subsampled 1000 cells per cell type from initial 500 steps. We set the parameters as follows: noise_params = 0.3, decays = 0.8, splice_ratio = 1.5, sampling_state = 30, noise_params_splice = 0.1, and noise_type = 'dpd'. To add library size effects, we set the parameters as follows: mean = 8 and scale = 0.4.

Calculation of co-accessibility of B/pDC-differentiating chromatin regions

To explore the natural accessibility of chromatin regions with different accessibilities between B and pDC populations, we scored their co-accessibility at the single cell resolution. First, we examined the differentially accessible chromatin regions of B and pDC clusters identified in the clustering analysis of transcriptome data from the CD34$^+$ population using FindMarkers function

implemented in Signac R package with $p < 0.01$ and $|log2FC| > 0.25$. Next, we calculated the signature scores for each side as B and pDC lineage accessibility. After negative values were converted to 0, we calculated B/pDC co-accessibility by their product.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using R tool 4.1.0. Statistical tests and p values are described in the figure legends. Statistical tests for comparison were two-sided, and $p < 0.05$ was considered significant. Data are presented as means ± SD as described in the figure legends.