

RESEARCH ARTICLE

# An enriched network motif family regulates multistep cell fate transitions with restricted reversibility

Yujie Ye<sup>1</sup>, Xin Kang<sup>2,3</sup>, Jordan Bailey<sup>1</sup>, Chunhe Li<sup>2,4\*</sup>, Tian Hong<sup>1,5\*</sup>

**1** Department of Biochemistry & Cellular and Molecular Biology, The University of Tennessee, Knoxville, Tennessee, United States of America, **2** Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China, **3** School of Mathematical Sciences, Fudan University, Shanghai, China, **4** Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China, **5** National Institute for Mathematical and Biological Synthesis, Knoxville, Tennessee, United States of America

\* [chunheli@fudan.edu.cn](mailto:chunheli@fudan.edu.cn) (CL); [hongtian@utk.edu](mailto:hongtian@utk.edu) (TH)



**OPEN ACCESS**

**Citation:** Ye Y, Kang X, Bailey J, Li C, Hong T (2019) An enriched network motif family regulates multistep cell fate transitions with restricted reversibility. *PLoS Comput Biol* 15(3): e1006855. <https://doi.org/10.1371/journal.pcbi.1006855>

**Editor:** Becca Asquith, Imperial College London, UNITED KINGDOM

**Received:** June 15, 2018

**Accepted:** February 7, 2019

**Published:** March 7, 2019

**Copyright:** © 2019 Ye et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** CL is supported by the National Science Foundation of China (11771098), Natural Science Foundation of Shanghai, China (17ZR1444500), and Thousand Youth Talents Program. TH is supported by the startup funds provided by the University of Tennessee, Knoxville. Funding for open access to this research was provided by University of Tennessee's Open Publishing Support Fund to TH. The funders had no role in study

## Abstract

Multistep cell fate transitions with stepwise changes of transcriptional profiles are common to many developmental, regenerative and pathological processes. The multiple intermediate cell lineage states can serve as differentiation checkpoints or branching points for channeling cells to more than one lineages. However, mechanisms underlying these transitions remain elusive. Here, we explored gene regulatory circuits that can generate multiple intermediate cellular states with stepwise modulations of transcription factors. With unbiased searching in the network topology space, we found a motif family containing a large set of networks can give rise to four attractors with the stepwise regulations of transcription factors, which limit the reversibility of three consecutive steps of the lineage transition. We found that there is an enrichment of these motifs in a transcriptional network controlling the early T cell development, and a mathematical model based on this network recapitulates multistep transitions in the early T cell lineage commitment. By calculating the energy landscape and minimum action paths for the T cell model, we quantified the stochastic dynamics of the critical factors in response to the differentiation signal with fluctuations. These results are in good agreement with experimental observations and they suggest the stable characteristics of the intermediate states in the T cell differentiation. These dynamical features may help to direct the cells to correct lineages during development. Our findings provide general design principles for multistep cell lineage transitions and new insights into the early T cell development. The network motifs containing a large family of topologies can be useful for analyzing diverse biological systems with multistep transitions.

## Author summary

The functions of cells are dynamically controlled in many biological processes including development, regeneration and disease progression. Cell fate transition, or the switch of cellular functions, often involves multiple steps. The intermediate stages of the transition

design, data collection and analysis, decision to publish, or preparation of the manuscript.

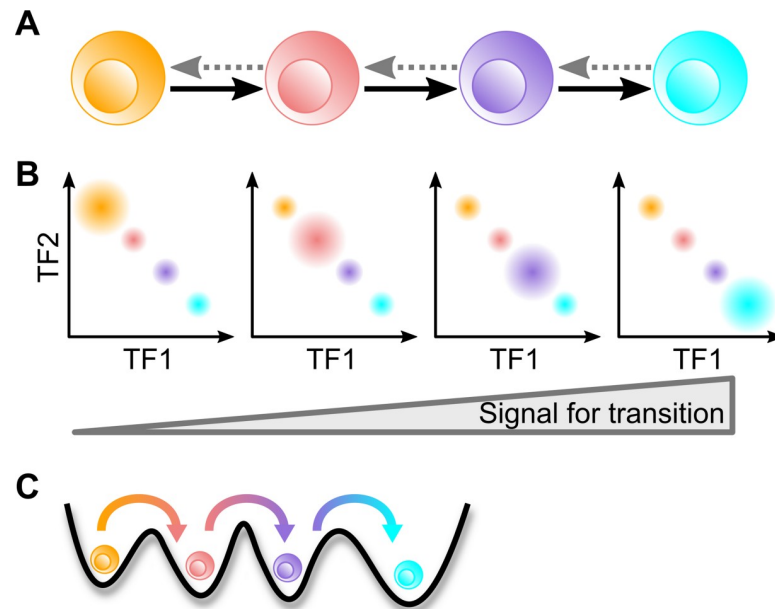
**Competing interests:** The authors have declared that no competing interests exist.

provide the biological systems with the opportunities to regulate the transitions in a precise manner. These transitions are controlled by key regulatory genes of which the expression shows stepwise patterns, but how the interactions of these genes can determine the multistep processes was unclear. Here, we present a comprehensive analysis on the design principles of gene circuits that govern multistep cell fate transition. We found a large network family with common structural features that can generate systems with the ability to control three consecutive steps of the transition. We found that this type of networks is enriched in a gene circuit controlling the development of T lymphocyte, a crucial type of immune cells. We performed mathematical modeling using this gene circuit and we recapitulated the stepwise and irreversible loss of stem cell properties of the developing T lymphocytes. Our findings can be useful to analyze a wide range of gene regulatory networks controlling multistep cell fate transitions.

## Introduction

Cell fate transition, including differentiation, de-differentiation and trans-differentiation, is a fundamental biological process in which the function of a cell gets specialized, reprogrammed or altered. The process often involves significant changes of multiple cellular properties, including the morphology, the self-renewal capacity and the potentials to commit to alternative lineages [1,2]. These changes are controlled by the dynamics of interacting transcription factors (TFs) and the modulation of chromatin structures, which in turn are governed by complex regulatory networks in the cells [3–5]. Interestingly, the fate transitions in many systems are achieved by sequential commitments to a series of cellular states with stepwise changes in their transcriptional profile towards the final stage of the program (Fig 1) [6–11]. The intermediate states between the initial state (e.g. the undifferentiated state in the case of cell differentiation) and the final state may be important for multiple purposes, such as facilitating ‘checkpoints’ that ensure appropriate development of cellular behaviors, or allowing the cells to make correct decisions at the lineage branching points [11–15].

One example of these stepwise cell lineage transitions is the development of T lymphocytes in the thymus. The differentiation from multipotent pre-thymic progenitor cells to committed T cells involves multiple cellular states with stepwise changes of their cellular properties and the transcriptional profiles (Table 1) [16–19]. Several lines of evidence suggest that the transition states at an early phase of the differentiation can serve as stable checkpoints for sequential lineage commitments. The progress through these intermediate states is accompanied by stepwise loss of their potentials to differentiate into other cell types: pre-thymic progenitor cells can be converted to a few types of cells, including B cells, natural killer (NK) cells, dendritic cells (DCs) etc., whereas the multipotency of the intermediate cell types is more limited but not completely lost [20–26]. In addition, the stability of these intermediate states is substantial because the loss of differentiation signals does not result in de-differentiation of some intermediate states [20], suggesting restricted reversibility (or complete irreversibility) of the multiple transitions. In addition, the lymphoid progenitor cells need to divide for a certain number of times at an intermediate state before committing to the T cell lineage, and the stable activities of the lineage defining transcriptional program at the intermediate stages may be important for the proliferations [27]. Finally, the loss of certain transcription factors (e.g. BCL11B) can lead to the termination of the differentiation at some intermediate states, which is often associated with diseases such as leukemia [18,20,28]. This further suggests that the intermediate states are cellular ‘attractors’ between the initial and the final stages of the differentiation (Fig



**Fig 1. Illustration of multistep cell fate transition.** **A.** transition from one cellular state to another via two intermediate states. Dashed arrow indicates the limited reversibility of each transition. **B.** stepwise changes of the levels of two transcription factors during the multistep transitions involving four states. **C.** metaphoric energy landscape depicting the four-attractor system. Colors for cell states and transition arrows in B and C match those in the illustration in A.

<https://doi.org/10.1371/journal.pcbi.1006855.g001>

1, bottom panel). Similar stable intermediate states during cell lineage transitions are observed in other systems, such as the epithelial-mesenchymal transition, and the skin development (Table 1), and those states also serve as regulatory stages for altering cellular properties including self-renewal and migration [10,29–37]. Therefore, the multiple intermediate states are involved in diverse normal development and pathological conditions. Understanding the regulatory programs for the sequential cell lineage commitments is a key step towards the elucidation of mechanisms underlying various biological processes involving multistep lineage transitions. Despite the accumulating data and observations on these stepwise lineage commitments, general mechanisms governing these differentiation processes with multiple intermediate cellular states remain unclear.

**Table 1. Examples of multistep transitions with restricted reversibility.**

Physiological scenario	Cellular phenotypic transition	Key regulators with stepwise modulations	Extracellular signals	Evidence supporting multistep transitions, multiple intermediate states and restricted reversibility
Early T cell development	ETP/DN1 → DN2a → DN2b → DN3	PU.1 TCF-1 GATA3 BCL11B	Notch	[17–20,28]
Skin development	Stem cell → renewable spinous cell → non-renewable spinous cell → granular cell	OVOL1 OVOL2	Calcium ion	[32,33,37]
Epithelial-mesenchymal transition	E → <sup>a</sup> EM1 → <sup>a</sup> EM2 → <sup>a</sup> M	SNAIL1 TWIST ZEB1 miR200	TGF-β	[29–31,34–36]

<sup>a</sup> Reversal transitions were observed, but they occur in a limited subpopulation.

<https://doi.org/10.1371/journal.pcbi.1006855.t001>

In this study, we explored the strategies in terms of the transcriptional network design that gives rise to stepwise transitions during cell differentiation. We first used a generic form of networks containing three interacting TFs to find network motifs that can produce four attractors (the minimum number of attractors in the examples of T cell development, epithelial-mesenchymal transition and skin development) with stepwise changes of transcriptional factor levels. We found two types of network motifs, both involving interconnections of positive feedback loops, which can generate the four-attractor systems. These motifs constitute a large family of gene regulatory networks. We found that there is an enrichment of these motifs in a network controlling the early T cell development. We built a specific model using known interactions among key transcription factors in developing T cells, and the model shows that the transcriptional network governs multistep and irreversible transitions in the development process.

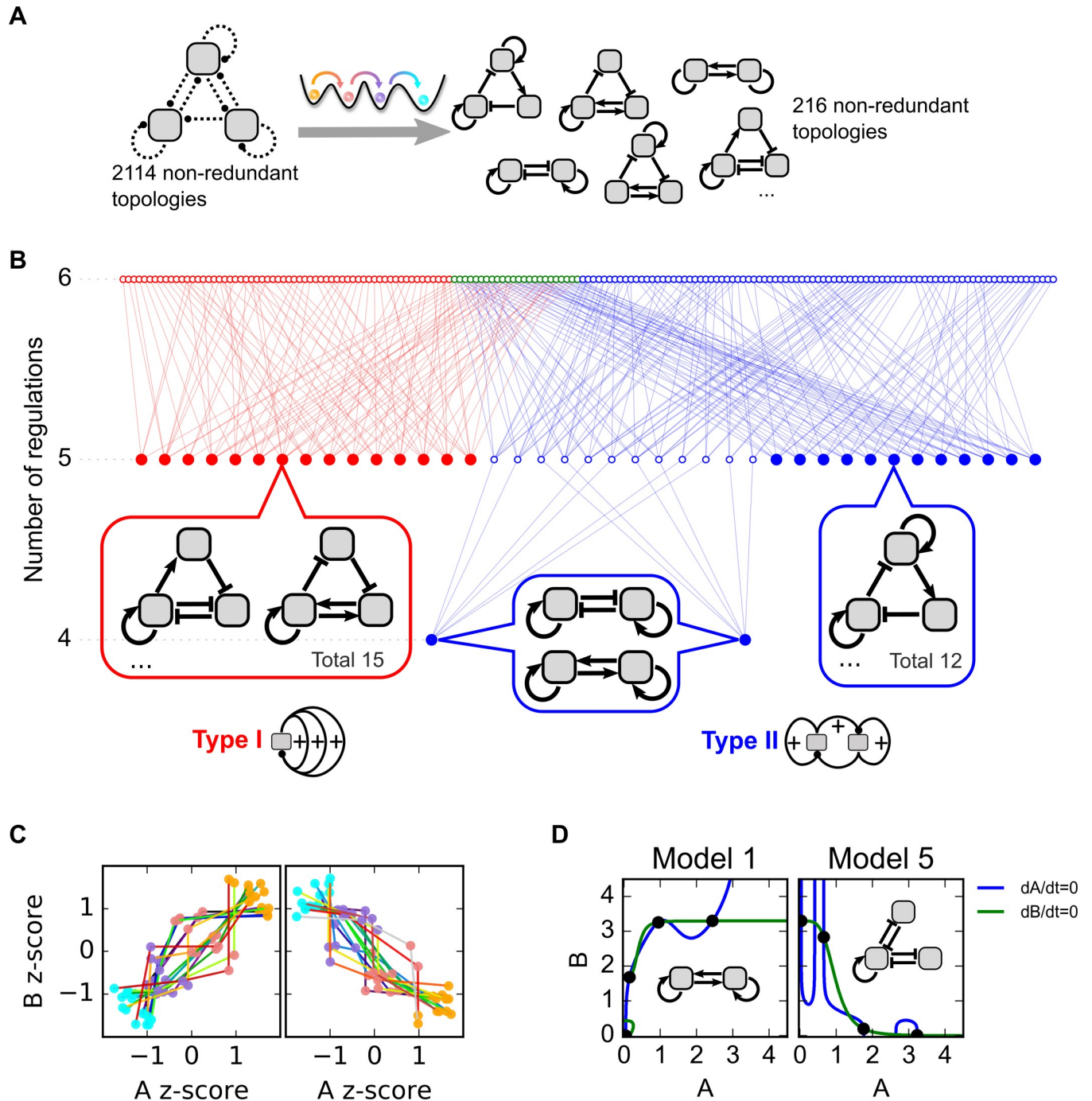
To investigate the stochastic dynamics for early T cell development, we mapped out the quasi-energy landscape for the early T cell development. This landscape characterizes the four attractors representing four stages of early T cell development quantitatively. In addition, by calculating the minimum action paths (MAPs) between different attractors, we quantified the dynamics of the key factors in response to Notch signal with fluctuations, which are in good agreement with experimental observations. Finally, we identified the critical factors influencing T cell development by global sensitivity analysis based on the landscape topography. Overall, our model for early T cell development elucidates the mechanisms underlying the stepwise loss of multipotency and multiple stable checkpoints at various stages of differentiation. The network topologies for multiple attractors found in this study and our motif discovery strategy combined with the landscape methodology can be useful for analyzing a wide range of cell differentiation systems with multiple intermediate states.

## Results

### Networks in a large motif family govern systems with four attractors with stepwise transcriptional modulation

To find transcriptional network topologies that can generate multiple intermediate states during cell fate transition, we first performed random parameter sampling with a network family containing up to 3 nodes (Fig 2A). In this framework of network topology, each node represents a transcription factor (TF) that can potentially influence the transcription levels of other two TFs and itself. Topology searching with a 3-node network was used for motif discovery for various performance objectives in previous studies [38,39]. We performed exhaustive search for topologies with up to 6 regulations from a total of 9 regulations of the network family, and constructed a mathematical model for each topology (see Methods for details). For each model, we performed random sampling in the parameter space from uniformly distributed values (S1 Table). We selected topologies containing at least one parameter set that is able to generate four attractors with stepwise changes of transcriptional levels. We define the system with four attractors with the stepwise changes of transcriptional levels as the scenario in which there are four stable steady states and they can be consistently ordered by the concentrations of any pairs of TFs. In other words, one TF always monotonically increases or decreases with another TF in these four states, and we term these states ‘ordered’ attractors in this paper. Among the 2114 network topologies that we searched, we found 216 topologies that can produce such behavior. In addition, we found 417 topologies that can only produce four unordered steady states (TF concentrations are non-monotonically correlated among the states) (S11 Fig, S12 Fig).

To visualize the relationships among these topologies, we constructed a complexity atlas (Fig 2B), in which the nodes represent the network structures that gave rise to four attractors,



**Fig 2. Network motifs governing four-attractor systems.** **A.** Illustration of the network topology searching. Dashed arrows are regulations sampled. The topologies were screened by the criterion of the four attractors with stepwise changes of TFs. **B.** Complexity atlas for selected topologies. Closed circles denote minimum motifs. Open circles denote topologies containing more regulations than those in the minimum motifs. Each arrow denotes the difference by one regulation in the network. Examples of minimum motifs are shown at the bottom. Red: Type I motif. Blue: Type II motif. Green: Hybrid motif. **C.** Overlaid four attractors for each of the 29 minimum topologies. Factor A denotes the TF on the left of the network diagram. Factor B denotes the TF on the right of the network diagram. In some topologies A and B are positively correlated (left panel), whereas they are negatively correlated in other topologies (right panel). Colored dots denote the stable steady states. Colored lines connect states of their corresponding topologies. The colors of the cell states match the illustration in Fig 1. The colors of the lines denote different representative models. The z-score is calculated by shifting the mean of each four attractors to 0 and then normalizing the four data points to unit variance data. **D.** Example phase planes for two minimum topologies (Type I and Type II respectively). In each case, four out of the seven steady states (intersections denoted by solid dots) are stable. Network structures and phase planes for all 29 minimum motifs are included in S1 and S2 Figs. All models shown in this figure are built with additive form of Hill functions.

<https://doi.org/10.1371/journal.pcbi.1006855.g002>

**Table 2. Definitions and key features of network motifs that generate systems with four ordered attractors.**

	Definition	Minimum number of regulations	Minimum number of positive feedback loops
Type I motifs	Three positive feedback loops that share one or more TFs among all of them.	5	3
Type II motifs	Three connected positive feedback loops. Two of them do not share any TF but are connected via the third loop.	4	3
Hybrid motifs	Motifs containing both Type I and Type II motifs.	6	4

<https://doi.org/10.1371/journal.pcbi.1006855.t002>

and the edges connect pairs of topologies that differ by a single regulation (addition or removal of a transcriptional interaction) [40]. We define the minimum topologies as those of which the reduction of complexity, or the removal of any regulation from the network, will abolish its capability to generate four attractors (solid nodes in Fig 2B and examples in Fig 2C). We found 29 such minimum topologies which represent the non-redundant structures for producing the four-attractor system.

Interestingly, all of the 216 topologies obtained from our search contain three distinct positive feedback loops (including double-negative feedback loops), and they can be categorized into two types of motifs (Fig 2B, bottom panel). The Type I motif contains three positive feedback loops that are closed at a single TF (red nodes and edges in Fig 2B). The Type II motif contains three connected positive feedback loops, two of which do not share any TF but are connected via the third loop (blue nodes and edges in Fig 2B). There is a remarkable diversity of each of the motif types because the interconnected positive feedback loops can share multiple TFs (S1 and S2 Figs). Based on the complexity atlas (Fig 2B), we found that Type II motifs contain 4–6 regulations, and Type I motifs contain 5–6 regulations. Some of the networks with 6 regulations contain subnetworks of both Type I and Type II motifs (Hybrid type, green nodes). The four attractors in the space of two TFs exhibit a variety of patterns of nonlinear monotonic correlations (Fig 2C, S3 Fig), which are governed by intersections of highly nonlinear nullclines in the state space containing the two TFs (Fig 2D, S1 and S2 Figs). The definitions of various types of motifs are listed in Table 2, and the statistics of the topologies discovered are summarized in Table 3 (also see S11 Fig for an illustration). Overall, this motif family represents a large number of networks that can produce a common type of behaviors: multiple stable intermediate states in terms the transcriptional activity.

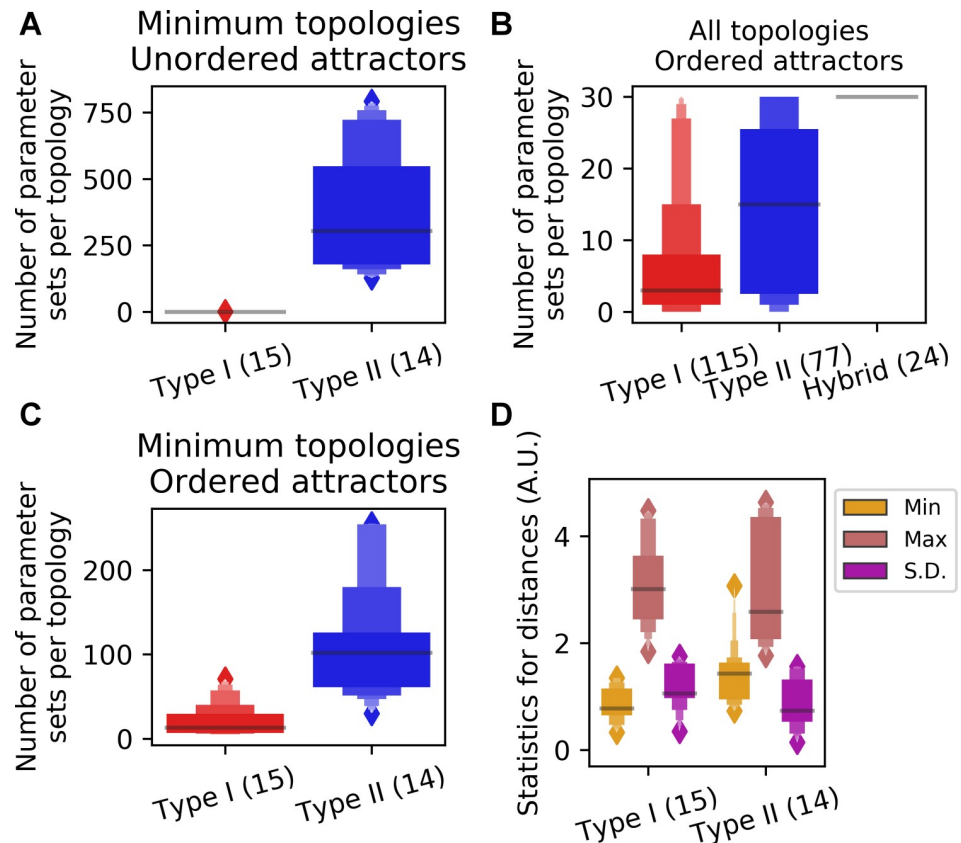
We next asked whether there is a difference between Type I and Type II motifs in terms of their ability to generate systems with four ordered attractors. We found that with the same number of sampled parameter sets, Type II motifs have greater fractions of parameter sets that

**Table 3. Numbers of sampled network structures and discovered motifs<sup>a</sup>.**

	3-node networks	T cell model
Total networks sampled	2114 (12258)	1553 (2047)
Type I motifs	77 (448)	191 (286)
Type II motifs	115 (638)	108 (120)
Hybrid motifs	24 (144)	269 (295)
Minimum Type I motifs	15 (84)	44 (71)
Minimum Type II motifs	14 (78)	22 (26)

<sup>a</sup> In each cell of the table, the first number is the number of non-redundant network topologies. The number in the parentheses is the number of networks (or sub-networks of the T cell model) including the isometric topologies.

<https://doi.org/10.1371/journal.pcbi.1006855.t003>



**Fig 3. Comparison of Type I and Type II motifs for their performance in producing systems with four ordered attractors.** **A.** The numbers of parameter sets per  $10^6$  samples per topology that generate four ordered attractors from two types of minimum motifs. **B.** The numbers of parameter sets per  $10^6$  samples per topology (limited to 30 per topology) that generate four ordered attractors from three types of motifs. **C.** The numbers of parameter sets per  $10^6$  samples per topology that generate four unordered attractors from two types of minimum motifs. **D.** Inter-attractor distances for each parameter set that generates four ordered attractors were calculated and summarized. For each set of parameters associated with the four attractors, the minimum, the maximum and the standard deviation of the distances were analyzed. Minimum Type I and Type II motifs were compared using these statistics.

<https://doi.org/10.1371/journal.pcbi.1006855.g003>

give rise to four ordered attractors than Type I motifs do (Fig 3A and 3B,  $p$ -value  $< 0.0001$ , Mann-Whitney U test). This suggests that Type II motifs may be more robust for governing the four ordered attractors. However, among the 15 Type I and 14 Type II minimum motifs, all Type II motifs are able to generate four unordered attractors (Some TFs are not monotonically correlated. See S12 Fig), whereas there is no Type I motif that has any parameter set that gives rise to four unordered attractors (Fig 3C). This suggests that as compared to Type II motifs, Type I motifs has higher specificity in generating four ordered attractors, which is more relevant to the stepwise cell fate transitions than the unordered ones. Moreover, we observed that the inter-attractor distances between neighboring attractors in the gene expression space were generally more variable with Type I motifs than those with Type II motifs (Fig 3D, magenta boxes). In particular, among the three inter-attractor distances for each model, Type I motifs generated smaller minimum distance than Type II motifs did (Fig 3D, orange boxes.  $p$ -value  $< 0.0001$ , Mann-Whitney U test). We did not observe any significant difference between Type I and Type II motifs in terms of the stabilities of the attractors and the kinetic paths that they generate (S13 Fig See Methods for calculation of quasi-energy landscape and kinetic path). In addition to the effects of motif types, we also asked whether the fractions of

positive or negative regulations in the network can influence the function of generating four-attractor systems. We found that the fraction of positive regulations has a weak but significant correlation with the fraction of successful parameter sets generating four-attractor systems (S14A and S14B Fig). Although negative regulations are slightly less favorable, they might be important to ensure that the levels of some TFs are inversely correlated during multistep cell fate transitions, which is necessary for having at least one highly expressed TFs in each of the initial and the final cell states (S14C Fig).

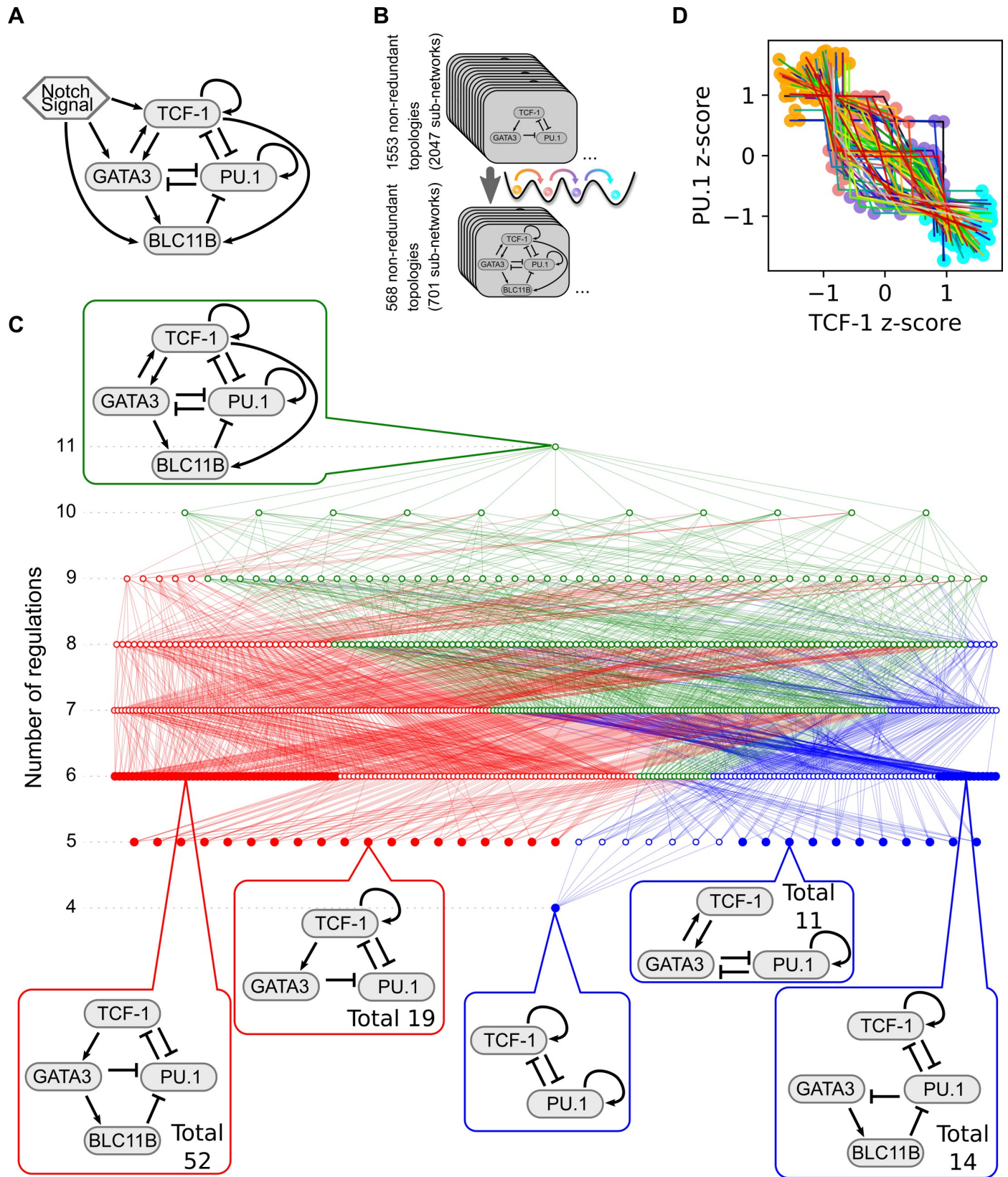
In summary, we found two types of network motifs that generate four attractors with stepwise changes of the transcriptional profile. Two of these attractors represent the multiple intermediate states observed in various biological systems. This exploratory analysis elicits several interesting questions: what are the biological examples of such network motifs? Can the conclusions with respect to the two types of motifs be generalized to networks with more than three TFs? Is there any advantage of combining both types of motifs? How are the transitions among these states triggered deterministically and stochastically? To provide insights into these questions in a more biologically meaningful context, we will use a specific biological system to describe more detailed analysis of these motifs and their underlying gene regulatory networks in the following sections.

### Type I and Type II network motifs are enriched in a transcriptional network controlling early T cell development

We asked whether the motifs that we discovered can be found in any known transcriptional network that potentially control multistep cell differentiation. We used the early T cell differentiation in the thymus as an example to address this question. The differentiation from multipotent lymphoid progenitor cells to unipotent early T cells involves multiple stages at which the cells have differential potentials to commit to non-T lineages and other cellular properties such as proliferation rates. At the early phase of this process, four stages of development T cells (ETP/DN1, DN2a, DN2b, DN3) were identified experimentally, and the progression through these stages is controlled by a myriad of transcription factors including four core factors, TCF-1, PU.1, GATA3 and BCL11B. These TFs form a complex network among themselves (see Fig 4A and supporting experimental observations in S3 Table), and the stepwise changes in the levels of these TFs were observed in the four developmental stages of T cells [20,28]. The interactions involving these core TFs were shown to be critical for the irreversible commitment to the T cell lineage by forming a bistable switch [41]. Among these factors, PU.1 level decreases as the cells commit to later stages, whereas the levels of other three factors increase in this process. It is unclear, however, whether this transcriptional network can serve as a regulatory unit that governs the multistep nature of the T cell differentiation.

We noticed that this T cell transcriptional network contains the motifs that we found in our analysis using the generic form of networks, we therefore hypothesized that the models based on this network can have four attractors with sequential changes of the four TFs. Indeed, using random sampling we were able to find parameter sets that give rise to four-attractor systems similar to what we obtained with the generic 3-node framework. To find the functional components that generate this behavior, we analyzed the subnetworks of the complex T cell regulatory network [42]. We removed the regulations from the network systematically, and we found that out of the non-redundant 1553 topologies (2047 subnetworks), there are 568 topologies (701 subnetworks) that can generate four attractors with stepwise changes of the TFs (Fig 4B). We used a complexity atlas to visualize the relationships among these subnetworks (Fig 4C). We found that the network can be reduced to one of the 66 minimum topologies (97 minimum subnetworks) which retains the four-attractor property (solid nodes in Fig 4C). Notably,





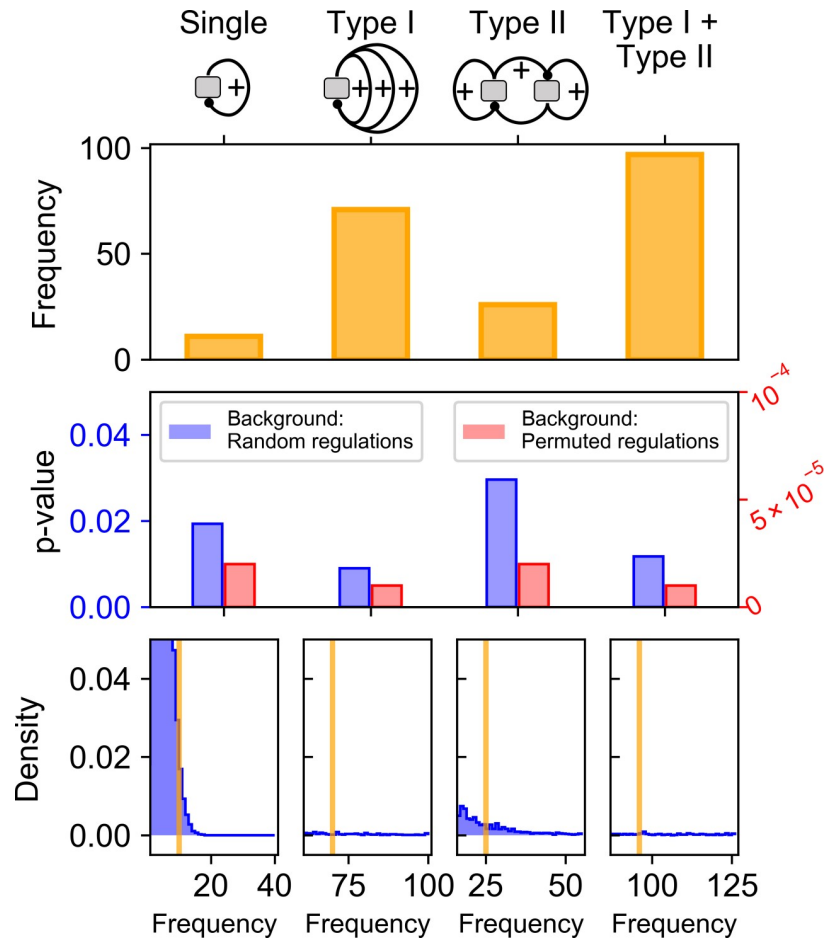
**Fig 4. Four-attractor motifs in the early T cell transcriptional network.** **A.** Influence diagram for transcriptional regulations among four core factors controlling the early T cell development. Arrows represent activations and short bars represent inhibitions. **B.** Functional subnetworks of the T cell network were systematically obtained by removing regulations from the network. These subnetworks were screened by the criterion that four attractors with stepwise changes of TFs exist in the absence of Notch signal. **C.** Complexity atlas showing the relationships of the two four-attractor motifs in the subnetworks of the T cell model. Top callout shows the full network in the absence of Notch. Bottom callouts show examples of the minimum functional subnetworks of the two types with particular numbers of regulations. Red: Type I motif. Blue: Type II motif. Green: Hybrid motif. **D.** Overlaid four attractors for each of the 66 minimum topologies. Colored dots denote the stable steady states. Colored lines connect states of their corresponding topologies. All models shown in this figure are built with the multiplicative form of Hill functions.

<https://doi.org/10.1371/journal.pcbi.1006855.g004>

these networks can be classified into the two types of motifs described earlier (Fig 2B). Similar to the networks that we obtained through the generic framework, the two types of minimum motif have 4–6 regulations. Subnetworks with both types of motifs (green nodes and edges) start to appear when the number of regulations reaches six. The numbers of motifs and subnetworks obtained for the generic framework and the T cell model are summarized in Table 3.

We next quantified the enrichment of the two motif families in the early T cell transcription network. We first generated random networks by perturbing the existing regulations in the network model and computed the empirical p-values for observing the numbers of different types of network motifs. The T cell network contains a large number of positive feedback loops and the two types of motifs that we described earlier (Fig 5, top panel). As expected, the network is significantly enriched with positive feedback loops in general (Fig 5, middle panel, red bars). However, the enrichments of Type I motifs and the combinations of Type I and Type II motifs are even more significant than that of the single positive feedback loops (Fig 5, middle panel, red bars). To exclude the possibility that this differential significance was observed due to the way we generate random networks which gives low p-values ( $<10^{-4}$ ) in general, we used another method to generate random networks with an augmented number of regulations (Fig 5, middle panel, blue bars). Each pair of TFs were assigned with a pair of random regulations (positive, negative or none). Consistent with the previous method, the T cell transcriptional network is enriched with positive feedback loops overall, but the enrichment is more significant for Type I motifs or for the combination of Type I and Type II motifs. Interestingly, motifs that are similar to Type I motif but have higher complexity (more positive feedback loops) do not show more significant enrichment than Type I motif does (S15 Fig). In addition, we found that networks controlling switch-like behaviors, but not multistep cell fate transitions, are not enriched with Type I or Type II motif [43–46]. These results suggest the possibility that the network has been evolved to reach more complex performance objectives than those enabled by simple positive feedback loops alone.

Since the minimum motifs alone can generate the four-attractor system, we asked whether the combination of these motifs enhances the ability of the network to produce the system. We therefore compared a subnetwork containing only one minimum Type I motif with another one containing multiple such motifs in terms of the performance to generate a particular four-attractor system (Fig 6A See Methods and S1 Text for details). We found that the subnetwork with multiple Type I motifs (Network 1) outperforms the one with only one motif (Network 2) in that Network 1 can give a better fit to a hypothetical four-attractor system (Fig 6A–6C). In this hypothetical ‘target’ system, the four attractors are assumed to be determined by dynamics of PU.1 with multiple feedback loops. The assumed degradation (Fig 6B, gray curve) and production rate functions of PU.1 (Fig 6B, green curve) are specified. The curves of these two functions have 7 intersections, four of which represent attractors. The optimized production function obtained from Network 1 (Fig 6B, purple curve) has more robust intersections with the degradation curve than one obtained from Network 2 (Fig 6B, red curve). This difference was observed for production functions of these two categories from multiple runs of optimization (Fig 6C). This suggests the advantage of combining multiple motifs with similar functions

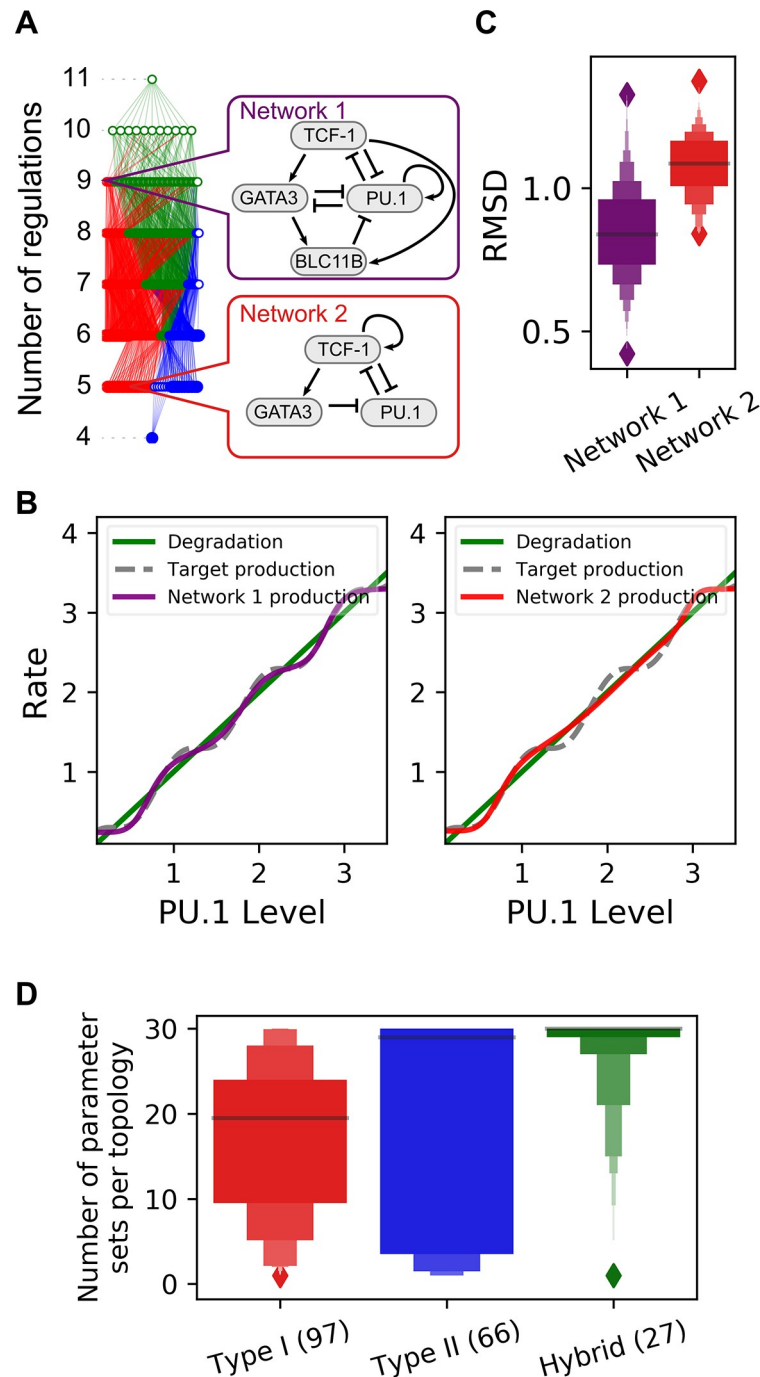


**Fig 5. Enrichment of Type I and Type II motifs in the T cell model.** Top panel: total occurrences of various types of motifs in the T cell network. Middle panel: empirical p-values of the single positive feedback loops and the sum of the two types of motifs. Bottom panel: an illustration of the p-values with the distributions of background population. Random networks were obtained by 1) permuting the regulations in the existing network by randomly assigning their sources and targets (red) and 2) assigning random regulations (positive, negative or none) between each pair of TFs (blue).  $10^5$  random networks were generated with each method. Empirical p-values were obtained by counting the number of the random networks with the number of motifs not less than those in the T cell network. See [Methods](#) for details of the p-value definition. Distributions of motif frequencies obtained from the random networks using the second method are shown in the bottom panel. The yellow vertical bars represent the number of occurrences in the T cell network. The right-tail areas defined by the vertical bars correspond to the p-values shown in the middle panel (blue bars).

<https://doi.org/10.1371/journal.pcbi.1006855.g005>

to enhance its overall performance. We next compared the 66 minimum motifs (Fig 4C bottom solid nodes) and the full topology (Fig 4C top node) in terms of the parameter values that gave rise to the four-attractor systems. We found that the full model contains more parameter sets that have low nonlinearity of the regulations than the minimum motifs do (S16A Fig), and the parameter values are distributed in larger regions with the full model than those with the minimum motifs (S16B Fig).

We next asked whether the topologies that contain both Type I and Type II motifs have greater probabilities to generate the four-attractor system than the topologies with one type of motifs do. When we explored the parameter space randomly for each topology with a fixed number of samples, a larger number of parameter sets that can generate the four-attractor system were found with the topologies containing both motifs than with those containing either



**Fig 6. Comparisons of motifs with different complexity and types.** **A.** Two specific network topologies were selected for comparing models with different complexity. Network 1 contains multiple Type I motifs, whereas Network 2 is a single Type I motif. The color code of the complexity atlas is the same as that in Fig 2 and Fig 4. Red: Type I motif. Blue: Type II motif. Green: Hybrid motif. **B.** Performances of the two subnetworks are compared. Performance was quantified with the sum of squared distance (SSD) from a predefined hypothetical continuous production function (gray curve) of PU.1 level that have 7 intersections with the degradation function, which generates four attractors (see details in supplementary text). The purple and red curves represent the optimized functions fitted to the gray curve. The gray curve is closer to the purple curve than to the red curve, suggesting a better fit with Network 1. **C.** SSD values obtained from 500 optimization runs for each of Network 1 and Network 2. Each value was calculated using the procedure shown in B. **D.** Numbers of parameter sets that generate the four-attractor systems per  $10^6$  random samples per topology. Three types of network motifs with 7 regulations are compared.

<https://doi.org/10.1371/journal.pcbi.1006855.g006>

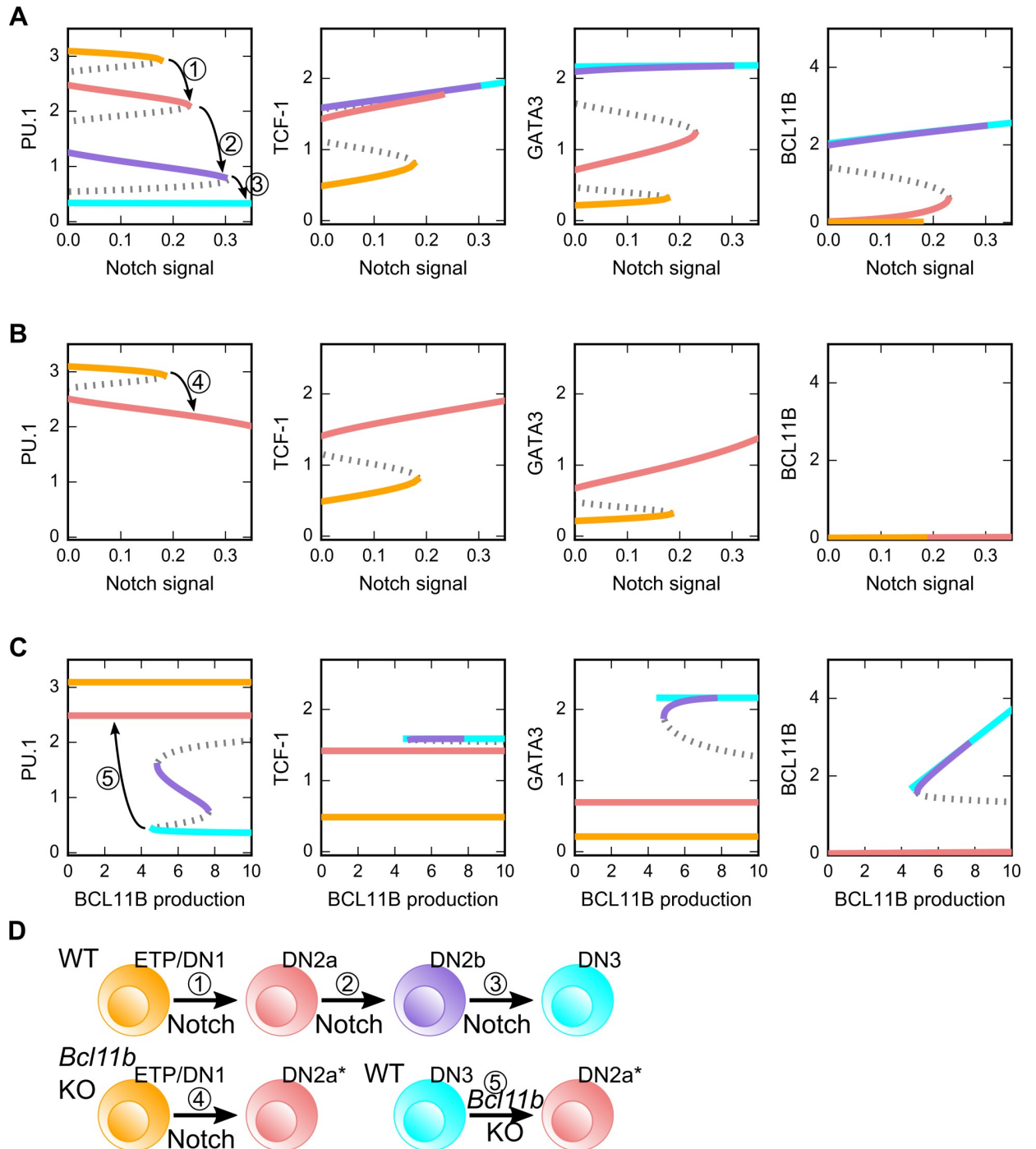
Type I or Type II motifs only (S5 Fig and Fig 6D). This suggests that the combination of both motifs might be a robust strategy to generate the four-attractor system. This pattern was observed for all the topologies in the complexity atlas (S6 Fig) as well as those with the same degree of complexity (Fig 6D, networks with 7 regulations were chosen because they have comparable fractions of the three types of motifs).

In summary, we found that the core transcriptional network controlling early T cell differentiation are enriched with Type I and Type II network motifs. The network composed of these two types of motifs governs a dynamical system containing four attractors, corresponding to four known stages in the early T cell development. The networks with both types of motifs and greater number of such motifs have more robust capability of generating the four-attractor systems than those networks with fewer types of numbers of motifs do.

### Stepwise transitions with restricted reversibility provide robustness to fluctuating differentiation signal to multiple intermediate states

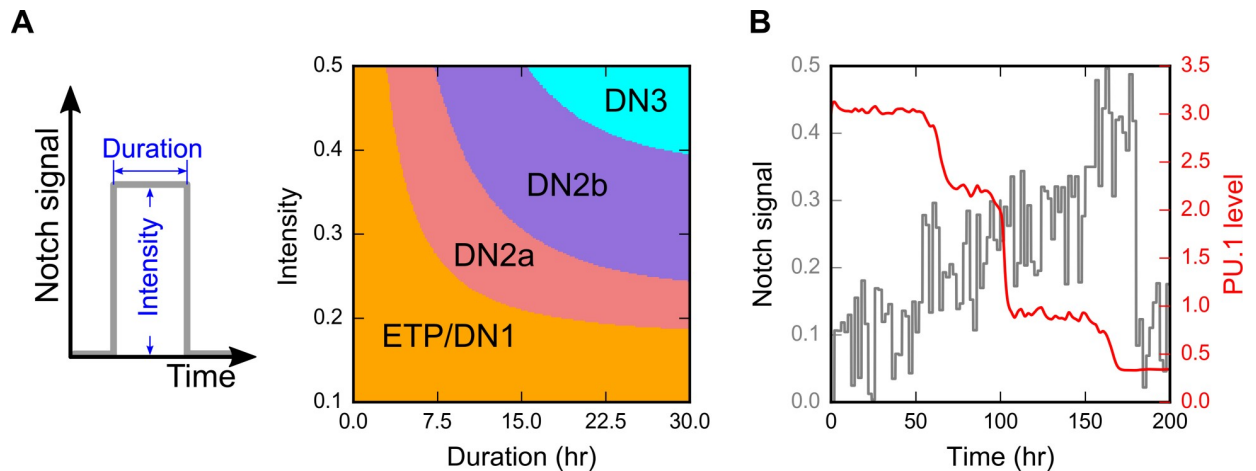
We next characterized the dynamical features of the four-attractor system of the T cell development model in response to differentiation signals. For this and subsequent analysis, we focused on a model describing the network topology shown in Fig 4A (the full model). We first performed bifurcation analysis of the system to the changes of Notch signaling (Fig 7A). With the increasing Notch signal, the system undergoes three saddle-node bifurcations, at which the stability of the proceeding cellular states is lost (Fig 7A, black arrows). These bifurcation points therefore represent the cell state transitions from one stage to the next. The structure of the bifurcation diagram shows a remarkable robust multistep commitment program governed by the T cell transcription network: the commitment to each stage of the program has restricted reversibility in that the attenuation or withdrawal of the Notch signaling does not result in de-differentiation of the developing T cells (i.e. the return of the transcription profile to earlier stages that may have greater multipotency). It was previously shown that the commitment from DN2a to DN2b is an irreversible process with respect to Notch signaling, and this transition eliminates developing T cells' potential to be diverted to any other lineages when Notch signaling is abolished [20,41]. However, simple toggle-switch models do not explain the observation that the multipotency of the early T cells is lost in a stepwise manner. For example, cells at ETP can be differentiated into B cells, macrophages, dendritic cells (DCs), granulocytes, natural killer (NK) cells and Innate lymphoid cells subset 2 (ILC2), whereas the potentials to commit to many of the lineages are blocked even in the absence of Notch signaling at the DN2a stage, at which the cells can only be differentiated into NK cells and ILC2 [20]. Therefore, the stepwise, irreversible transcriptional transitions revealed by our model is consistent with the experimental observations with respect to the loss of multipotency in the stepwise manner.

Although the absence of Notch signal does not allow the reversal of lineage progression, it was previously shown that the absence of BCL11B in lymphoid progenitor cells blocks its ability to progress to DN2b stage, whereas the Cre-controlled knockout of *Bcl11b* in committed T cells (e.g. DN3 cells) reverts its transcriptional profile to DN2a-like cells [28]. Upon blocking the production of BCL11B in our model, we observed the loss of attractors of DN2b and DN3, and the DN2a state is the only stable stage even in the presence of the strong Notch signaling (Fig 7B). As a result, increasing Notch signaling only triggers one saddle-node bifurcation, representing the transition from ETP to DN2a cell (Fig 7B black arrow), whereas the decrease of the BCL11B production triggers the transition back to DN2a instead of ETP (Fig 7C). These results are in agreement with the previous experimental findings [28], and they further support the importance of the multistep differentiation system revealed by our model.



**Fig 7. Stability analysis of the T cell model.** The full model shown in Fig 4A is used for all the analysis. **A.** Bifurcation diagrams for the steady states of the four core factors with respect to the Notch signal. Solid curve: stable steady state. Dashed curve: unstable steady state. **B.** Bifurcation diagram under *Bcl11b* knockout condition with respect to Notch signal. Solid curve: stable steady state. Dashed curve: unstable steady state. **C.** Bifurcation diagram with respect to BCL11B production rate parameter. Solid curve: stable steady state. Dashed curve: unstable steady state. **D.** Illustration of the observed transitions among the four states. Colors of the stable branches of the bifurcation diagrams and the cell icons are matched to the cellular states shown in Fig 1.

<https://doi.org/10.1371/journal.pcbi.1006855.g007>



**Fig 8. Multistep lineage transitions under the influence of varying dynamics Notch signals.** A. Strength and duration of the Notch signal were varied in each simulation. 200X200 combinations of different signal strengths and durations were tested, and the final cellular phenotypes were determined using the levels of the four core factors. B. Dynamics of PU.1 in response to increasing Notch with significant fluctuations. The mean of the Notch signal increases linearly in the first phase, then it is attenuated in the second phase. Fluctuations were simulated with additive noise in small time intervals.

<https://doi.org/10.1371/journal.pcbi.1006855.g008>

The bifurcation analysis shows how the lineage progression is influenced by stably increasing or decreasing Notch signal strengths. We next asked how the duration of Notch signal may control the multistep lineage transition. By inducing the differentiation with varying durations of the Notch signaling, we found that cells experiencing transient Notch signals may only commit to intermediate stages of differentiation (Fig 8A). In addition, the system is able to integrate the information of the signal intensity and duration to make decision on the lineage progression. These results suggest that the multistep lineage transition can be triggered by the increasing strength of the signal, the increasing duration of the signal, or the combination of both types of signal dynamics. Earlier experimental studies have shown that transient Notch signaling can irreversibly drive the cells to an intermediate, but committed stage with a definitive T cell identity (DN2b) [28,41,47]. This is in agreement with our results, and our model further suggests that the commitment to other intermediate states is also irreversible with respect to the lineage progression (note that this irreversibility does not refer to the establishment of T cell identity).

One possible advantage of the multi-stable system is its robustness of response in facing fluctuating signals. We therefore performed numerical simulations of the dynamical system under increasing Notch signaling with significant fluctuations. Under this condition, transient reduction of Notch signaling halted the progress of the lineage commitment but did not trigger the de-differentiation (Fig 8B). Our model suggests that the design of transcriptional network allows the system to stop at intermediate stages before proceeding to the next ones. This strategy has several potential physiological benefits: 1) it protects the cell lineage progression against sporadic fluctuations of Notch signaling; 2) it facilitates the ‘checkpoints’ before lineage commitment in the middle of the developmental process and 3) it allows the stable storage of differentiation intermediates which can be differentiated into mature T cells rapidly when there is an urgent need of new T cells with a diverse T cell receptor repertoire.

### Quantitative analysis of the energy landscapes and minimum action paths delineates the patterns of the multiple-attractor system in T cell differentiation

With the deterministic modeling and bifurcation approaches, we described the local stability for multi-stable T cell model. However, the global stability is less clear from the bifurcation

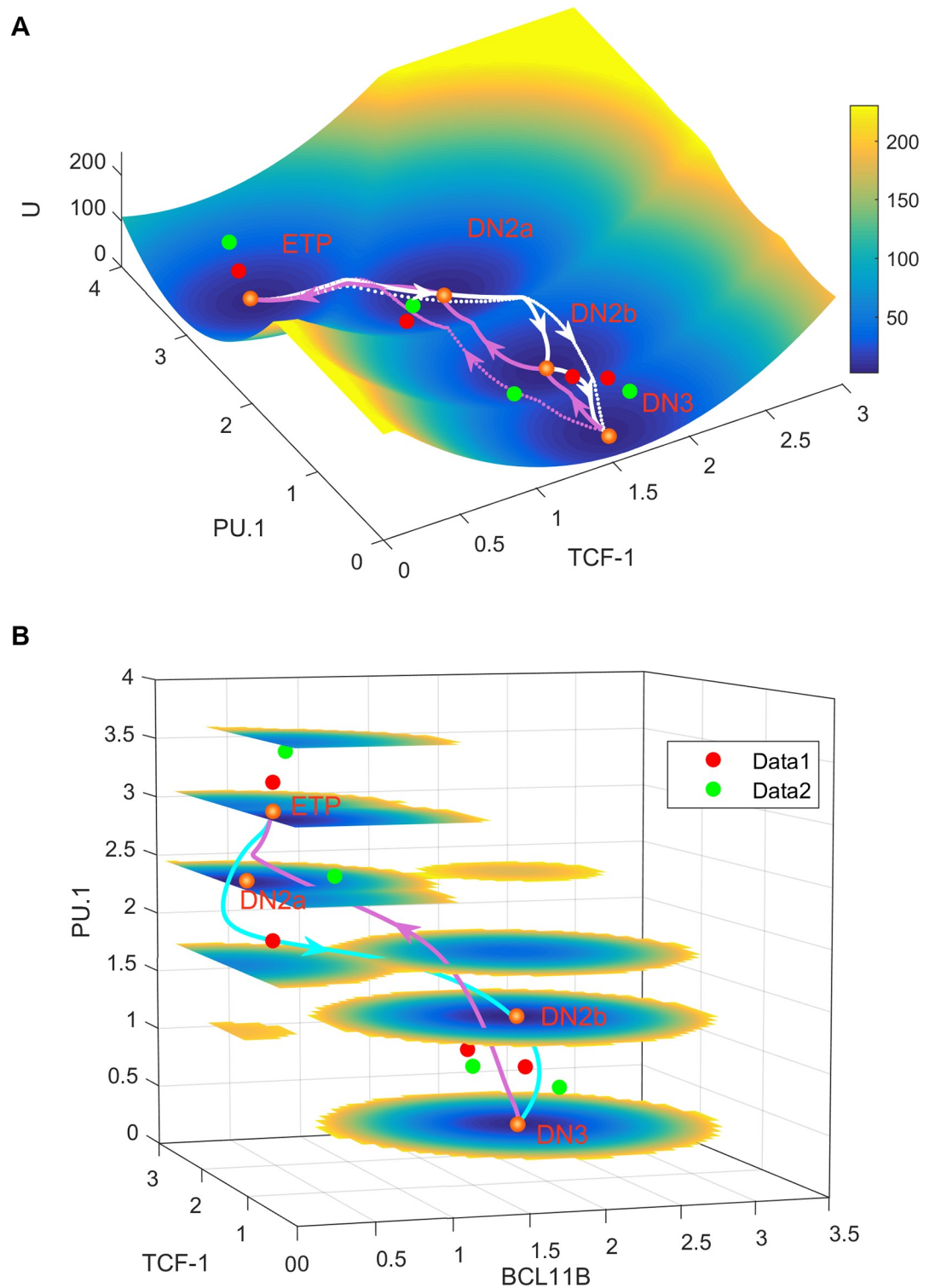
analysis alone. In addition, it is important to consider the stochastic dynamics for T cell development model because the intracellular noise may play crucial roles in cellular behaviors [48,49]. The Waddington landscape has been proposed as a metaphor to explain the development and differentiation of cells [50]. Recently, the Waddington epigenetic landscape for the biological networks has been quantified and employed to investigate the stochastic dynamics of stem cell development and cancer [51–57].

Following a self-consistent approximation approach (see [Methods](#)), we calculated the steady state probability distribution and then obtained the energy landscape for the model of the early T cell development (full model in [Fig 4A](#)). For visualization, we selected two TFs (PU.1 and TCF-1) as the coordinates and projected the 4-dimensional landscape into a two-dimensional space by integrating the other 2 TF variables ([Fig 9A](#)). Here TCF-1 is a representative T cell lineage TF, and PU.1 is a TF for alternative cell fates. We also displayed the landscape in a four-dimensional figure, where the three axes represent three TFs (TCF-1, BCL11B and PU.1), and the color represents the energy  $U$  ([Fig 9B](#)). Note that our major conclusions do not depend on the specific choice of the coordinate (see [S7](#) and [S8](#) Figs for landscapes with PU.1/BCL11B and PU.1/GATA3 as the coordinates).

In the case without Notch signal ( $N = 0$ ), four stable cell states emerge on the landscape for the T cell developmental system ([Fig 9](#)). On the landscape surface, the blue region represents lower potential or higher probability, and the yellow region represents higher potential or lower probability. The four basins of attraction on the landscape represent four different cell states characterized by different TF expression patterns in the 4-dimensional state space ([Fig 9A and 9B](#) provide two types of projections of the whole 4-dimensional landscape). These states separately correspond to ETP/DN1 (high PU.1/low TCF-1/low BCL11B/low GATA3 expression), DN3 state (low PU.1/high TCF-1/high BCL11B/high GATA3 expression), and two intermediate states (DN2a and DN2b, intermediate expression for the four TFs). The existence of four stable attractors is consistent with experiments [16–19]. As the Notch signal ( $N$ ) increases, the landscape changes from a quadristable (four stable states coexist), to a tristable (DN2a, DN2b and DN3), to a bistable (DN2b and DN3) and finally to a monostable DN3 state ([S9 Fig](#)). These results provide a straightforward explanation for the **irreversibility** observed in experiments for the stepwise T cell lineage commitment.

To check whether our modelling results match experimental data quantitatively, we acquired two sets of gene expression data of the four core TFs for T cell development from previous publications [17,47], and mapped the normalized values (see [Methods](#)) to the landscape ([Fig 9A and 9B](#), [S7](#) and [S8](#) Figs). Here, the golden balls represent the four steady states (characterizing four stages of T cell development) from the models, the red balls represent the gene expression data points (Data1) from Zhang et al. [47], and the green balls represent the gene expression data points (Data2) from Mingueneau et al. [17]. We found that these gene expression data agree well with our landscape models in the sense that each data point is almost positioned in the corresponding basin ([Fig 9A and 9B](#), [S7](#) and [S8](#) Figs). We found that the landscapes give a better fit to Data2 (green points), since each green data point can be well positioned in one of the four basins, corresponding to four stages of T cell development. In fact, the two sets of data points are not very close to each other or to the steady states (golden points) from the models. This is reasonable because these two sets of experimental data are measured separately, probably in different conditions, and these data usually delineate the average of multiple measurements from different samples. Also, the gene expression fluctuations are common in biological systems. Therefore, our landscape pictures provide a natural way to reconcile the two different experimental data, i.e. the gene expression data do not have to be at the positions of steady states. Instead, the gene expression data for each individual





**Fig 9. Energy landscape for T cell development.** The landscape and corresponding minimum action paths (MAPs) for the T cell developmental network are shown in 3-dimensional (A) and 4-dimensional (B) figure (the full model shown in Fig 4A is used for all the analysis). **A.** The blue regions represent higher probability or lower potential and the yellow regions represent lower probability or higher potential. White solid lines represent the MAP from ETP state to DN2a, DN2b, and DN3 states. Magenta solid lines represent the MAP from DN3 to DN2b, DN2a, and to ETP state. Dashed lines represent the direct MAP from ETP to DN3 and from DN3 to ETP states, respectively. Here, TCF-1 and PU1 are selected as the two coordinates for landscape visualization. **B.** The landscape and paths are displayed in a 4-dimensional figure, where the three axes represent

three TFs (TCF-1, BCL11B and PU. 1), respectively, and the fourth dimension (color) represents the energy  $U$ . The normalized gene express data (including TCF-1, BCL11B, GATA3, and PU. 1) of four stages for T cell development are mapped to the landscape, where the golden balls represent the four steady states (four stages of T cell development) from the models, the red balls (Data1) represent the data from Zhang et al. [47], and the green balls represent the data from Mingueneau et al. [17]. See S7 Fig and S8 Fig for the landscapes and the gene expression data using other pairs of TFs.

<https://doi.org/10.1371/journal.pcbi.1006855.g009>

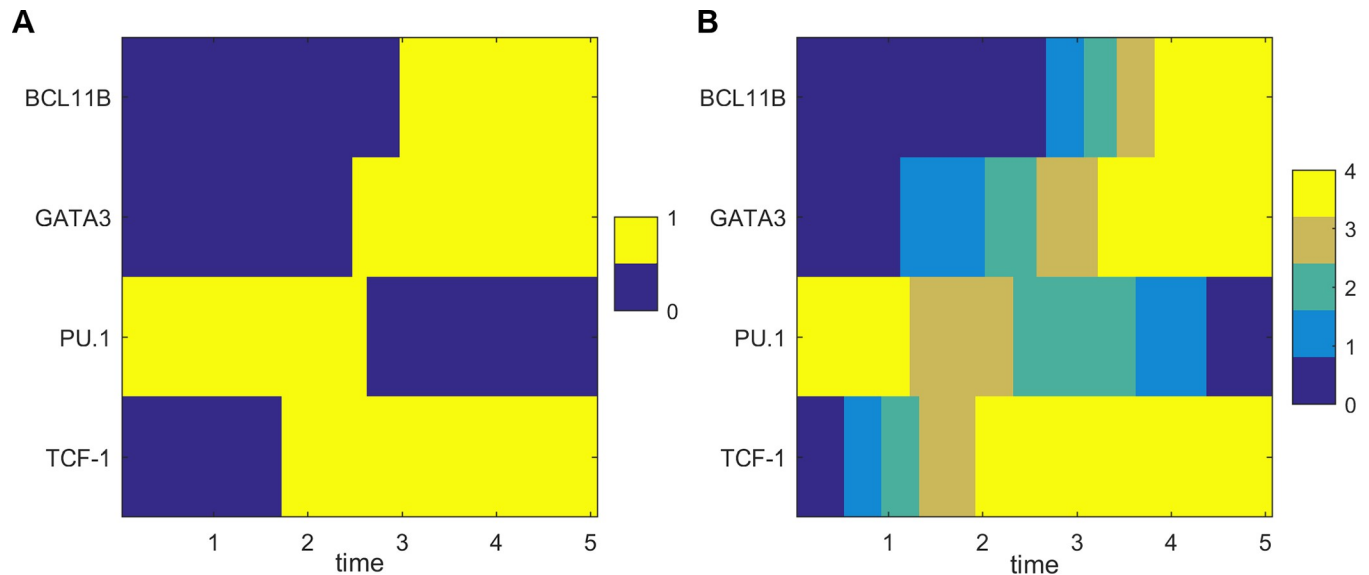
stage could be somewhere around that basin because of the fluctuations. This reflects the original spirit of the classic Waddington developmental landscape.

To examine the transitions among individual cell types, we calculated kinetic transition paths by minimizing the transition actions between attractors [58,59], obtaining minimum action paths (MAPs). The MAPs for different transitions are indicated on the landscape (Fig 9). The white MAPs from the ETP state to the DN3 state, correspond to the T cell developmental process while the magenta MAPs from the DN3 state to the ETP state, correspond to reprogramming process. The lines represent the MAPs, and the arrows denote the directions of the transitions. The MAP for T cell developmental process and the MAP for the backward process are irreversible, since the forward and reverse kinetic paths are not identical. This irreversibility of kinetic transition paths is caused by the non-gradient force, i.e. the curl flux [60,61]. Here, the solid white lines represent three stepwise transitions from ETP to DN2a, DN2a to DN2b, and DN2b to DN3, whereas the dashed white line represents the direct transition path from ETP to DN3. From the MAPs for T cell development, we found that the direct transition path is very similar to the stepwise transition path (the white solid line is similar to the white dashed line, Fig 9, S7 and S8 Figs), which indicates that the T cell developmental process needs to go through the two intermediate states (DN2a and DN2b). This confirms the critical roles of the intermediate states for the T cell differentiation. It is worth noting that the MAPs here quantify the most probable transition paths, which suggest the optimal path (with least transition action) for cells to switch from one state to another. However, in a realistic gene regulatory system, usually a signal is needed to induce rapid cell state transitions (e.g. the Notch signaling is used here to induce T cell development).

To investigate the dynamical developmental process of T cell for multiple TFs, we visualized the 4-dimensional MAP from the ETP to the DN3 state by discretizing the levels of the four TFs. We found that for T cell development, TCF-1 is upregulated first, followed by the activation of GATA3. This leads to the complete inactivation of the alternative fate TF PU.1 and the activation of BCL11B (Fig 10). Interestingly, this temporal order is in good agreement with experimental observations [62] (Also see the gene expression data of four core TFs for four stages in Fig 9, S7 Fig and S8 Fig). These results suggest that the sequence of switching on or off for different TFs can be critical for the lineage commitment of T cell development. Moreover, under the *Bcl11b* knockout condition ( $k_B = 0$ ), the landscape changes from a quadristable (four stable states coexist), to a bistable (ETP and DN2a) state (S10 Fig), which is consistent with the bifurcation analysis (Fig 7) and experimental observations [28].

### Global sensitivity analysis based on landscape topography reveals the critical factors for T cell development

To identify the critical factors (regulations and TFs) which determine T cell development, we performed a global sensitivity analysis based on the landscape topography. Specifically, we use the transition action between attractors as a measure to quantify the feasibility of a transition between different attractors. A smaller transition action, corresponding to a smaller energy barrier, means a more feasible transition from one attractor to another. In this way, by changing the parameters each at a time we can identify the critical parameters for T cell development



**Fig 10. Discrete kinetic transition paths for T cell model.** Transition paths from ETP state to DN3 state in terms of levels of 4 different TFs. **A.** The relative TF levels are discretized to 0 or 1. 1 represents that the corresponding TFs are in the on (activated) state and 0 represents that the corresponding TFs are in the off (repressed) state. **B.** The relative TF levels are discretized to five values from low to high. X axis shows the time along the transition path.

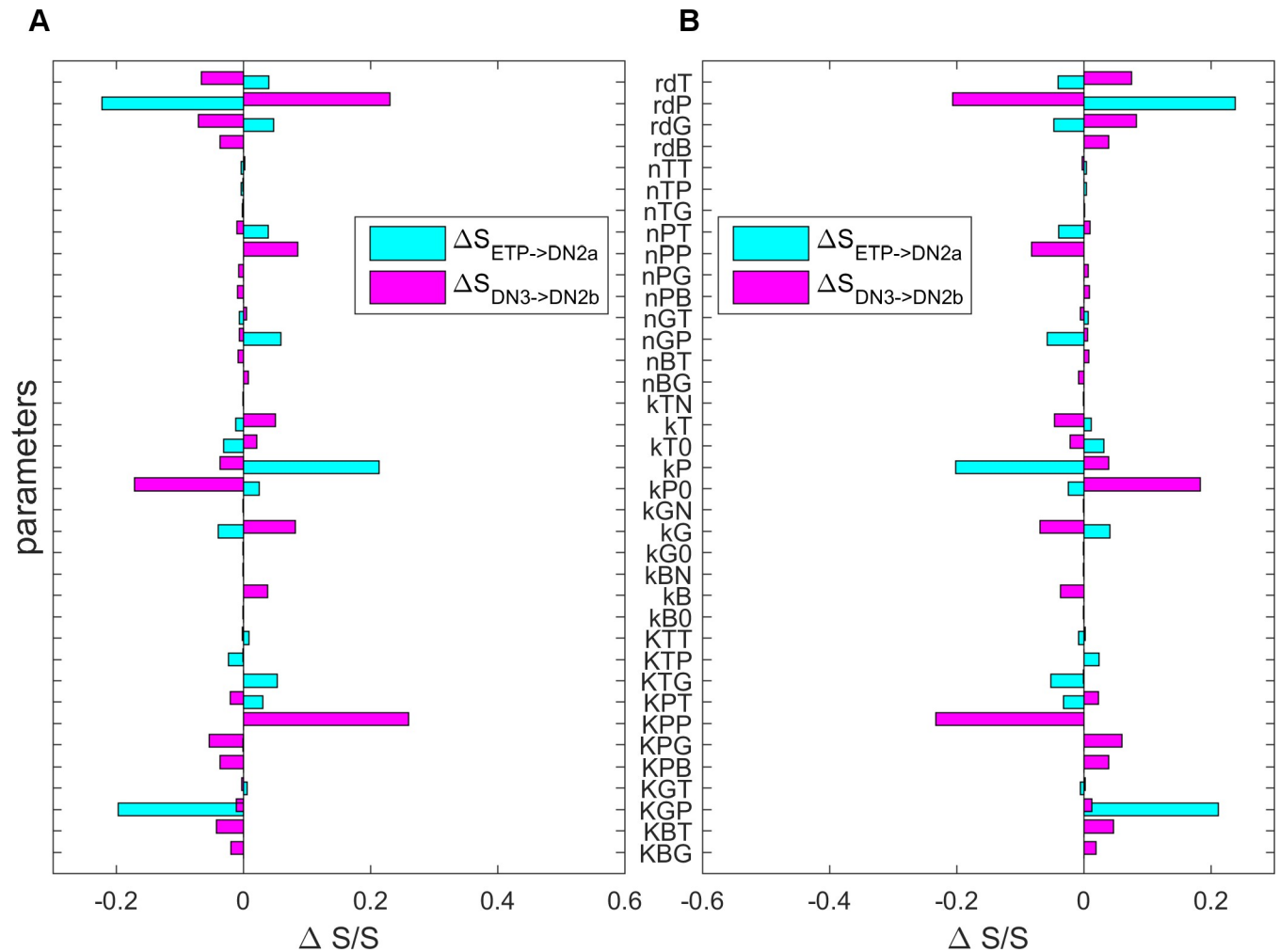
<https://doi.org/10.1371/journal.pcbi.1006855.g010>

(we use the transition from ETP to DN3 as an example). To do this, we constrict the models within the parameter region corresponding to the four-attractor system, so that we can make comparisons for the changes of transition actions as parameters are varied.

We identified some critical parameters of which the variations caused significant changes of transition actions between ETP and DN3 attractor. These parameters include the effective degradation rate of PU.1, (rdP), the regulated production rate of PU.1 (kP), the basal production rate of PU.1 (kP0), the threshold of the self-activation of PU.1 (KPP), and the threshold for the repression of PU.1 on GATA3 (KGP) (Fig 11). In particular, the increase of the self-activation strength of PU.1 (i.e. decreased KPP) reduces the transition action from DN3 to DN2b (Fig 11B), indicating a less stable DN3 state and a more stable ETP state. This is reasonable because the PU.1 is a major TF for alternative cell fates (B-cell, dendritic-cell, and myeloid cell), and silencing of PU.1 is operationally important for T cell commitment [28]. Additionally, the increase of the repression strength of PU.1 on GATA3 (decreased KGP) raises the transition action from ETP to DN2a (Fig 11B), indicating a more stable ETP state and a less stable DN3 state, which is consistent with the observation that GATA3 is a critical TF promoting T cell development. Overall, these results from sensitivity analysis indicate that the PU.1 synthesis/degradation related parameters, the GATA3 synthesis related parameters, and the regulations between PU.1 and GATA3 are critical to the dynamics and the cell fate decisions of T cell development. This indicates that the regulatory circuit between PU.1 and GATA3 plays critical roles for the cell fate determinations during T cell development.

## Discussion

In this study, we identified two types of network motif families that are responsible for generating a four-attractor dynamical system commonly observed in stepwise cell differentiation. Some instances of these motifs were previously described and analyzed in the context of binary or ternary switches during lineage transitions [63–67], but the systematic analysis for these motifs was not performed to our knowledge. In addition, the design principle for multiple



**Fig 11. Global sensitivity analysis for T cell developmental model.** Sensitivity analysis was performed for the 39 parameters in the T cell model. The transition actions between different states ( $S_{ETP \rightarrow DN2a}$  and  $S_{DN3 \rightarrow DN2b}$ ) were calculated to quantify the sensitivity of parameters on the landscape. The Y-Axis represents the 39 parameters. The X-Axis represents the percentage of the transition action (S) changed relative to S without parameter changes. Here,  $S_{ETP \rightarrow DN2a}$  represents the transition action from attractor ETP to attractor DN2a (cyan bars), and  $S_{DN3 \rightarrow DN2b}$  represents the transition action from attractor DN3 to attractor DN2b (magenta bars). **A.** Each parameter is increased by 1%, individually. **B.** Each parameter is decreased by 1%, individually.

<https://doi.org/10.1371/journal.pcbi.1006855.g011>

intermediate states was not clear. Our approach provides a comprehensive framework for analyzing systems with a complex dynamical property, a four-attractor system with stepwise transcriptional modulation, and we illustrate the intricate relationships among these motifs with an intuitive visualization method.

Previous studies on biological circuits governing irreversible transitions focused on the analysis of toggle switches which generate none-or-all type of responses [68,69]. Our work suggests that multistep or graded responses can be associated with irreversible transitions as well. Given the importance of graded response in various biological scenarios [70–72], we expect the design strategy that we found can be useful for discovery of natural-occurring irreversible graded responses or construction of synthetic biological circuits producing these responses. Our work also suggests that the response to signals, or the progression of lineage transition, may be proportional to the intensity and/or the duration of the signal. This is consistent with the previous observations that the duration of the morphogen signal can be critical for cell

lineage choice [73,74]. Of note, when signal strength is converted to digital (none-or-all) response in early phases of signal transduction, its duration can play an essential role in determining the graded response [75].

In our systematic exploration in the network topology space, we took the assumption that network structure is correlated with its function, i.e. assuming the existence of functional motif structure in transcription regulatory networks. The notion of network motifs is very helpful for understanding many complex biological systems [76,77], but the richness of dynamic behaviors of these motifs is beyond their structures—distinct kinetic rates in the same motif can produce diverse responses [78]. Therefore, it is expected that the motifs that we discovered may be able to generate dynamical behaviors different from the four-attractor system (we will discuss some of them in the following paragraphs). We also expect that some of network motifs can be responsible for multiple functions by themselves, and this multifunctionality may explain the diverse motifs that we found for the four-attractor systems in the biological examples. Future work is warranted to examine the distributions of the diverse functions in the parameter space of the motifs that we found. Nonetheless, it is important to understand the capacity of the network motifs in terms of their functional outputs. Our work provides a holistic view of the potential network motif structures governing multistep cell lineage transitions.

Although network motifs with three positive feedback loops closing at a single factor (Type I motifs discussed in this study) were not systematically analyzed in previously studies to our knowledge, some simpler versions of Type I motif, e.g. a pair of interconnected positive feedback loops, have been described in various systems such as the epithelial-mesenchymal transition and the cancer progression [65,79]. These systems typically govern ternary switches with a single intermediate state. These studies and ours suggest a correlation between the number of positive feedback loops and the number of the intermediate states the system may be able to generate. In fact, early studies on multistable systems have shown the requirement of positive feedback loops for generating multiple steady states [80], which was proved mathematically [81]. Intriguingly, an ultrahigh feedback system similar to the Type I motifs was shown to govern irreversible transitions with low differentiation rates for adipocytes [82]. It would be interesting to examine whether controlling the low differentiation rate through cell-to-cell variability and controlling the number of intermediate states suggested by our model can be achieved in the same system. Our findings are consistent with the earlier work in that they highlight the importance of this type of signaling motifs in controlling cell differentiation by preventing the direct and homogeneous transition from the initial state to the final one.

Near symmetrical parameters in models based on a particular instance of the Type II motif class (the one with mutually inhibiting TFs) have been widely used to explain stochastic lineage choice observed in embryonic stem cells, developing hematopoietic cells and CD4<sup>+</sup> T cells [83,84]. Our findings with Type II motifs complement these studies with newly identified functions of these motifs for cell differentiation. Instead of the stochasticity that breaks the symmetry of this motif, the Notch signal may be responsible for switching the system from one side (PU.1 high) to another (PU.1 low) in a stepwise fashion, and the intermediate states mark the stable stages where the system is relatively balanced in terms of two groups of competing TFs.

It was previously suggested that the network consisting of four core transcription factors governs a bistable switch with irreversible transition [41]. Our models based on this network provide explanations for additional experimental observations with respect to the multistep feature of the early T cell development (see [S1 Text](#) for a comparison between our model and the bistable model). Although it is possible that interconnection of multiple positive feedback loops simply enhances the robustness of the bistable switches, the observation that several important irreversible transitions in cell cycle progression are primarily controlled by two

positive feedback loops implies that the enrichment of the positive feedback loops in the T cell transcriptional network is unlikely due to the intrinsic biophysical limits of positive feedback loops in generating bistable switches [46,69]. Instead, other cellular functions, such as generating the multiple intermediate states, might be the performance objectives for the design of this network. Future work with more systematic analysis is warranted to compare the parameter regions corresponding to the bistable, tristable and quadristable systems with ordered and unordered attractors.

Our model of early T cell development suggests that the differentiation program may be stopped at multiple locations in the state space of transcription levels of key factors. These multiple attractors may correspond to the lineage branching points at which the progenitor cells are given opportunities to be converted to T cell as well as other types of lymphocytes. As such, it is possible that this dynamical property is exploited to achieve a better control for the fate determination of the lymphoid progenitor cells at systems level. Given that subpopulations of NK cells and DCs are generated by the thymus [85–87], the multistep lineage transition provides a basis for channeling the lymphoid progenitor to multiple lineages in a precise manner.

Based on the recent landscape-path theory and the T cell gene regulatory network model, we investigated the stochastic dynamics of T cell development. We identified four stable cell states characterized by attractors on the landscape including ETP/DN1, DN3, and two intermediate states (DN2a and DN2b). These attractors (cell states) match two published datasets of gene expression values in a reasonable way [17,47]. We also calculated the kinetic transition paths between different cell states from minimum action path approaches. Importantly, from the MAPs of T cell development, we found that different TFs are switched on or off in different orders. For example, TCF-1 needs to be first activated, and then GATA3 is activated, leading to the inactivation of PU.1 and activation of BCL11B. These predictions agree well with experiments [28,62], which provides further validations for our mathematical model.

In our models, we only considered four core factors based on previous published T cell gene regulatory network for simplicity [41]. In the realistic biological system, there are more factors critical to T cell development [28]. It would be interesting to incorporate other important factors into the network and construct a more realistic model for T cell development. By studying the landscape of more comprehensive T cell development network, we will better understand the underlying regulatory machinery and obtain more insights into the intricate mechanisms for T cell development.

In summary, we identified a large family of network motifs that can generate four attractors that are observed in various biological systems involving cell lineage transition. We built a mathematical model for transcriptional network controlling early T cell development, and we found that the network underlying this developmental process is enriched with the motifs that we identified. The system with the four attractors has a remarkable irreversibility for transitions to multiple intermediate states when the differentiation signal is varied. We suggest that this multistep process may be useful for precise control of the differentiation of lymphoid progenitor cells towards T cell and other cell types. Our T cell model provides new insights into the complex developmental or regeneration processes, and our combined approaches of comprehensive analysis of network motifs for generating multistable systems and landscape-path framework provide a powerful tool for studying a wide range of networks controlling cell lineage transitions.

## Methods

### Framework of mathematical modeling

We used ordinary differential equations (ODEs) to describe the dynamics of the concentrations of transcription factors (TFs). We used Hill function to describe the transcriptional

regulation by TFs. Each ODE has the following form:

$$\dot{X}_i(t) = k_{0,X_i} + k_{X_i} \sum_{j=1}^n \beta_{i,j} \frac{(1 - \theta) + \theta \left(\frac{X_j}{K_{i,j}}\right)^{n_{i,j}}}{1 + \left(\frac{X_j}{K_{i,j}}\right)^{n_{i,j}}} - r_{d,X_i} X_i \quad (1)$$

Here,  $X_i$  represents the concentration of a transcription factor (TF).  $k_{0,X_i}$  is the basal production rate of the TF in the absence of any regulator.  $k_{X_i}$  is the maximum production rate under the control of the transcriptional activators and inhibitors of this TF.  $\beta_{i,j}$  denotes the weight of the influence of the TF  $j$  on  $i$ . The sum of the Hill functions determines the regulation of the production of this TF by other TFs. In each term of the summation,  $\theta = 1$  when the regulating TF ( $X_j$ ) is an activator.  $\theta = 0$  when the regulating TF is an inhibitor.  $K_{i,j}$  is the apparent dissociation constant of the regulating TF binding to its regulatory element of the promoter, and it describes the effectiveness of the regulation in terms of the concentration of the TF.  $n$  is the total number of regulating TFs.  $r_{d,X_i}$  is the effective degradation rate constant. The production rate of the proteins is assumed to be linearly correlated with mRNA production rate. Similar generalized forms of Hill function were previously used for analysis of a variety of gene regulatory networks [52,88]. One time unit of our model corresponds to 6 hours, and all the parameters are dimensionless.

To exclude the possibility that our conclusions are sensitive to the choice of the form of equations, we used an alternative form of ODE to describe the regulatory networks:

$$\dot{X}_i(t) = k_{0,X_i} + k_{X_i} \prod_{j=1}^n \frac{(1 - \theta) + \theta \left(\frac{X_j}{K_{i,j}}\right)^{n_{i,j}}}{1 + \left(\frac{X_j}{K_{i,j}}\right)^{n_{i,j}}} - r_{d,X_i} X_i \quad (2)$$

In these ODEs, multiplication of Hill functions was used instead of addition. Similar forms of Hill function were also used for modeling a variety of gene regulatory networks [66,89]. With this form, the two types of network motifs that generated the four-attractor behavior are the same as those discovered with the additive form of Hill functions (S3 Fig). In fact, using both forms of equations gave rise to the same number of network topologies (216 topologies with the steady states shown in both S3 Fig and S4 Fig). Therefore, our conclusions are robust in terms of the choice of equation form.

During topology searching, random parameters values were chosen from defined ranges (S1 Table, see below).

### Topology searching for four-attractor systems

Network topology searching was first performed for all possible topologies involving up to 3 nodes (TFs) and 6 regulations that are able to generate four-attractor systems with stepwise changes of TF levels. We sampled  $n$  of the 9 arrows in Fig 2A, where  $1 \leq n \leq 6$ . The procedure for the exhaustive search is the following: 1) Choose  $n$  arrows from the 9 arrows in Fig 2A (9 choose  $n$ ), which gives rise to 465 topology templates. 2) For each topology template containing  $n$  regulations, enumerate all  $2^n$  networks (the number 2 represents positive and negative regulations), which gives 12258 networks in total. 3) Remove redundant networks that are different in terms of labels of nodes but are otherwise identical in terms of topology (isometric topologies), which produces 2114 non-redundant topologies. Three-node networks were previously used to explore several types of functional dynamics of network motifs [38,90]. For each topology, we performed random sampling of  $10^6$  parameter sets. For each parameter set, we selected 125 initial conditions in the three-dimensional state space ((0, 3.3) for each variable) using Latin Hypercube sampling, and then solved the ODEs numerically. We stopped

the simulations at time point 500 and checked if the 125 ODE systems are stabilized at four or more distinct steady states. We next checked if the changes of the TFs are monotonically coupled. We first ordered the steady states by the levels of one TF, and then we looked for scenarios in which all other TFs monotonically increase or decrease with the ordered TF (i.e. the attractors with stepwise changes of the TFs). We excluded the scenarios in which one TF is not monotonically correlated with others in terms of their levels at the four attractors. Models that generated oscillations at the final time point were also excluded. The parameter sets which produced the stepwise changes of steady state were accepted and their associated network topologies were analyzed. Parameter values for the minimum topologies are listed in [S2 Table](#).

### Topology classification

To determine whether a network topology contains Type I and/or Type II motifs, we developed a simple algorithm for motif identification. We first enumerated all positive feedback loops (PFLs) in the network. A PFL is defined as any unbranched closed loop structure that has even number of negative regulations. We then followed the scheme described in [Table 2](#) to check if Type I and/or Type II motifs exist in a specific network: we enumerated all combinations of three PFLs in the network. If any three PFLs share one or more TFs, then the network contains a Type I motif. If two of any three PFLs do not share any TF, but they (i.e. at least one TF from PFL 1 and at least one TF from PFL 2) are connected via the third PFL, then the network contains a Type II motif.

Complexity atlas was plotted for the obtained network topologies as described previously by Jiménez et al [40] ([Fig 2B](#) and [Fig 4C](#)).

### Transcriptional network model for early T cell development

We built a model for early T cell development based on the regulations that were previously shown experimentally [91–104]. Information about experimental evidence is described in [S3 Table](#). The form of equations is similar to Eq (2). We chose this multiplicative form of Hill functions because earlier experimental study suggested that regulations of *Bcl11b* gene are combined via an ‘and’ logic gate [105], which favors the use of multiplication. Although similar detailed information is not available for other TFs, we have shown that our main conclusions with respect to the multistep transitions controlled by a network motif family do not depend on the choice of the form of equations ([Fig 2](#) and [S4 Fig](#)). Full list of equations is included in [S1 Text](#). The parameter values were obtained by random searching described above with the four-attractor property as the selection criterion followed by minor manual adjustments. All parameter values are dimensionless. To our knowledge, there is no published experimental measurement that would allow us to directly constrain the ranges of these values except for the degradation rates, which have a unit of the inverse of time. These degradation rates were estimated from a previous study that measured the half-lives (approximately 4 hours) of the transcription factors [41]. The parameter values are listed in [S4 Table](#). To explore the subnetworks of the T cell development model that are essential for the four-stage transition, we performed similar exhaustive search in a set of 1553 non-redundant topologies (2047 subnetworks) to find functional circuit in the model. We obtained 568 topologies (701 topologies) from the search, and we analyzed them with complexity atlas. Isometric topologies were removed in the simulations, but they are included in the complexity atlas so that we do not mix isometric topologies with possibly differential biological meanings specific to certain genes.

In the bifurcation analysis, the value of the parameter  $N$  (Notch signal strength) or  $k_{BCL11B}$  (maximum production rate of BCL11B) is varied and the changes of the steady states of the system were analyzed. We let  $k_{BCL11B} = 0$  to simulate the *Bcl11b* knockout condition.



To simulate the system under various scenarios of Notch signaling, we first varied the strength and/or duration of the Notch signal and checked the steady state distribution of the system under the varying strengths and durations. We tested 200X200 combinations of strengths and durations of Notch signals and obtained the phenotypes of the cells at the steady state. To simulate the fluctuating Notch signals, we divided the time window of the simulation into small intervals (0.1 unites of time). For each interval, we used a random number with a specified mean and an additive noise. The mean of the Notch signal first increased overtime and then became attenuated.

### Enrichment analysis of the four-attractor motifs in the T cell model

To quantify the enrichment of various types of motifs, we used the generic definition of p-value: the p-value for a particular motif is the probability of obtaining at least  $n$  number of motifs from a random network population, where  $n$  is the observed number of such motif in the T cell network. To compute the p-values, we first counted the frequencies of the positive feedback loop, Type I motif and Type II motif in the T cell model (i.e.  $n_1, n_2, n_3, n_4$  representing the numbers of positive feedback loops, Type I motifs, Type II motifs, and the sum of the Type I and Type II motifs respectively). Random networks were generated using two methods: 1) for each regulation in the existing T cell model, we randomly reassign its source and target TFs (referred to as ‘permuted regulations’), and 2) for each pair of TFs from the network, we randomly assign a regulation (positive, negative or none) (referred to as ‘permuted regulations’). For each of the two methods, we generated  $10^5$  networks, and we calculated the empirical p-values by counting the number of the random networks with the numbers of motifs not less than those of respective motifs in the T cell network. The method with permuted regulations is more biologically relevant because the number of the positive and negative regulations are retained in the random networks. We used the second approach as an alternative to exclude the possibility that the conclusion based on the trend of the p-values is due to the low number of networks containing the extreme amount of the motifs.

### Optimization for performance comparison of two subnetworks

Due to the difficulty to compare the performances of regulatory circuits with different complexities in general, we selected two specific instances of Type I network motif for comparison. One of them contains only one Type I motif, whereas the other one contains multiple motifs. For each topology, we reduced the system to one ODE with quasi-steady state assumption and defined a continuous production rate function that can produce four attractors as a surrogate function (see [S1 Text](#)). Multiple runs of optimization using differential evolution algorithm was used, and 500 converged parameter sets for each circuit were used for comparison. This optimization method was previously used for finding optimum parameter sets and for comparing the performances of regulatory circuits [[64,106,107](#)].

### Self-consistent mean field approximation for the quantification of energy landscape

The temporal evolution a dynamical system was determined by a probabilistic diffusion equation (Fokker-Planck equation). Given the system state  $P(X_1, X_2, \dots, X_N, t)$ , where  $X_1, X_2, \dots, X_N$ , represent the concentrations of molecules or gene expression levels, we have N-dimensional partial differential equation, which are difficult to solve because the system has a very large state space. Following a self-consistent mean field approach [[52,61,108](#)], we split the probability into the products of the individual probabilities:  $P(X, t) = P(X_1, X_2, \dots, X_N, t) =$

$\prod_i^N P_i(X_i, t)$  and solve the probability self-consistently. In this way, we effectively reduced the dimensionality of the system from  $M^N$  to  $MN$  ( $M$  is the number of possible states that each gene can have), and thus made the computation of the high-dimensional probability distribution tractable.

Based on the diffusion equations, when the diffusion coefficient  $D$  is small, the moment equations can be approximated to [109,110]:

$$\dot{\bar{x}}(t) = F(\bar{x}(t)) \tag{3}$$

$$\dot{\sigma}(t) = \sigma(t)A^T(t) + A^T(t)\sigma(t) + 2D(\bar{x}(t)) \tag{4}$$

Here,  $\bar{x}(t)$ ,  $\sigma(t)$  and  $A(t)$  are vectors and tensors.  $\sigma(t)$  denotes the covariance matrix and  $A(t)$  is the jacobian matrix of  $F(\bar{x}(t))$ .  $A^T(t)$  is the transpose of  $A(t)$ . The elements of matrix  $A$  are specified as:  $A_{ij} = \frac{\partial F_i(\bar{x}(t))}{\partial x_j(t)}$ . By solving these equations, we can acquire  $\bar{x}(t)$  and  $\sigma(t)$ . Here, we consider only the diagonal elements of  $\sigma(t)$  from the mean field approximation. Then, the evolution of the probability distribution for each variable can be acquired from the Gaussian approximation:

$$P(x, t) = \frac{1}{\sqrt{2\pi\sigma(t)}} e^{-\frac{(x-\bar{x}(t))^2}{2\sigma(t)}} \tag{5}$$

The probability distribution acquired above corresponds to one stable steady state or the basin of attraction. If the system has multiple stable steady states, there should be several probability distributions localized at each basin with different variances. Thus, the total probability is the sum of all these probability distributions with different weights. From the self-consistent approximation, we can extend this formulation to the multi-dimensional case by assuming that the total probability is the product of each individual probability for each variable. Finally, with the total probability, we can construct the potential landscape by:  $U(x) = -\ln P_{ss}(x)$ . Here,  $P_{ss}$  is the steady state probability distribution, and  $U$  is dimensionless potential energy. In this work, we define two quantities based on the landscape theory. One is the energy barrier height, which is defined as the energy difference between the local minimum and the corresponding saddle point. Another quantity is the transition action, which is defined as the minimum action from one attractor to the other. These two quantities both measure the difficulty of the transitions. However, the transition actions are suggested to provide a more accurate description for the barrier crossing between attractors or the transition rate [111]. Therefore, we used the transition actions to quantify the difficulty of the transitions between attractors in this work (see the following section for the approach of calculating minimum action paths).

### Minimum action paths from optimization

Following the approaches based on the [Freidlin-Wentzell](#) theory [58,112,113], for a dynamical system with multistability the most probable transition path from one attractor  $i$  at time 0 to attractor  $j$  at time  $T$ ,  $\phi_{ij}^*(t)$ ,  $t \in [0, T]$ , can be acquired by minimizing the action functional over all possible paths:

$$S_T[\phi_{ij}] = \frac{1}{2} \int_0^T |\dot{\phi}_{ij} - F(\phi_{ij})|^2 dt \tag{6}$$

Here  $F(\phi_{ij})$  is the driving force. This optimal path is called minimized action path (MAP). We calculated MAPs numerically by applying minimum action methods used in [58,112].

## Comparison between simulated attractors and experimental data

Gene expression data for the four core TFs were obtained from two previous studies on T cell development (Zhang et al. and Mingueneau et al. [17,47]). We rescaled these data to the range of the attractors obtained from our models by linearly transforming the expression values so that they match the attractors approximately. This rescaling is necessary because some corresponding expression values from these two datasets differ from each other by more than 10-fold. The corresponding ratios between these two datasets vary significantly as well.

## Supporting information

**S1 Text. Model equations, model reductions and evaluation procedure through optimization.**

(DOCX)

**S1 Table. Ranges of parameter values for sampling 3-node networks.**

(DOCX)

**S2 Table. Parameter values for models of 29 minimum motifs for 3-node networks.**

(XLSX)

**S3 Table. Experimental evidence supporting the regulations in the early T cell development model.**

(DOCX)

**S4 Table. Parameter values for early T cell development model.**

(DOCX)

**S5 Table. Enrichment analysis for three published gene regulatory networks governing systems without four attractors.**

(XLSX)

**S1 Fig. Phase planes for Type I minimum network topologies.** Nullclines for TF A (the node on the left of the network diagram) and TF B (the node on the right of the network diagram) are shown. Stable steady states are shown as black dots. The inset network diagram shows the corresponding network. Random parameter sampling was used to obtain the parameter sets that allows the 4-attractor systems.

(TIF)

**S2 Fig. Phase planes for Type II minimum network topologies.** Nullclines for TF A (the node on the left of the network diagram) and TF B (the node on the right of the network diagram) are shown. Stable steady states are shown as black dots. Random parameter sampling was used to obtain the parameter sets that allows the 4-attractor systems.

(TIF)

**S3 Fig. Overlaid four attractors for each of the 216 topologies from the 3-node network that produce 4-attractor systems.** Factor A denotes the TF on the left of the network diagram. Factor B denotes the TF on the right of the network diagram. In some topologies A and B are positively correlated (left panel), whereas they are negatively correlated in other topologies (right panel). Colored dots denote the stable steady states. Colored lines connect states of their corresponding topologies. The colors of the cell states match the illustration in Fig 1. The colors of the lines denote different representative models. z-score is calculated by shifting the mean of each four attractors to 0 and then normalizing the four data points to unit variance

data. All models shown in this figure are built with additive form of Hill functions.  
(TIF)

**S4 Fig. Four-attractor systems generated with the alternative form of equations.** **A.** Overlaid four attractors for each of the 216 topologies from the 3-node network that produce 4-attractor systems. Factor A is the TF on the left of the network diagram. Factor B is the TF on the right of the network diagram. In some topologies A and B are positively correlated (left panel), whereas they are negatively correlated in other topologies (right panel). Colored dots denote the stable steady states. Colored lines connect states of their corresponding topologies. The colors of the cell states match the illustration in Fig 1. The colors of the lines denote different representative models. z-score is calculated by shifting the mean of each four attractors to 0 and then normalizing the four data points to unit variance data. **B.** Example phase planes for two minimum topologies (Type I and Type II respectively). In each case, four out of the seven steady states (intersections denoted by solid dots) are stable. All models shown in this figure are built with multiplicative form of Hill functions.  
(TIF)

**S5 Fig. Overlaid four attractors for each of the 559 topologies from the T cell network that produce 4-attractor systems.** Colored dots denote the stable steady states. Colored lines connect states of their corresponding topologies. The colors of the cell states match the illustration in Fig 1. The colors of the lines denote different representative models. z-score is calculated by shifting the mean of each four attractors to 0 and then normalizing the four data points to unit variance data. All models shown in this figure are built with multiplicative form of Hill functions.  
(TIF)

**S6 Fig. Comparison of three types of network topologies.** Letter-value plot shows distributions of the numbers of topologies from the entire complexity atlas (Fig 4C) over the space of parameter sets that generate the four-attractor systems per  $10^6$  random parameter sets. Distributions are separately shown for three types of motifs. Red: Type I motif. Blue: Type II motif. Green: Hybrid motif.  
(TIF)

**S7 Fig. Landscape and corresponding minimum action paths (MAPs) for the T cell developmental network in the BCL11B-PU.1 state space.** The blue regions represent higher probability or lower potential and the yellow regions represent lower probability or higher potential. Four basins of attractions characterize four different cell states (ETP, DN2a, DN2b, and DN3). White solid lines represent the MAP from ETP state to DN2a, DN2b, and DN3 states. Magenta solid lines represent the MAP from DN3 to DN2b, DN2a, and to ETP state. Dashed lines represent the direct MAP from ETP to DN3 and the direct MAP from DN3 to ETP states, respectively. The normalized gene express data (BCL11B and PU. 1) of four stages for T cell development are mapped on the landscape, where the golden balls represent the four steady states (four stages of T cell development) from the models, the red balls (Data1) represent the data from Zhang et al. [47], and the green balls represent the data from Mingueneau et al. [17].  
(TIF)

**S8 Fig. Landscape and corresponding minimum action paths (MAPs) for the T cell developmental network in the GATA3-PU.1 state space.** The blue regions represent higher probability or lower potential and the yellow regions represent lower probability or higher potential. Four basins of attractions characterize four different cell states (ETP, DN2a, DN2b, and DN3). White solid lines represent the MAP from ETP state to DN2a, DN2b, and DN3 states. Magenta solid lines represent the MAP from DN3 to DN2b, DN2a, and to ETP state. Dashed lines

represent the direct MAP from ETP to DN3 and the direct MAP from DN3 to ETP states, respectively. The normalized gene express data (GATA3 and PU. 1) of four stages for T cell development are mapped on the landscape, where the golden balls represent the four steady states (four stages of T cell development) from the models, the red balls (Data1) represent the data from Zhang et al. [47], and the green balls represent the data from Mingueau et al. [17].

**S9 Fig. Landscape changes as Notch signal increases.** On the landscape, the blue regions represent higher probability or lower potential and the yellow regions represent lower probability or higher potential. As the Notch signal (N) increases (from 0.1 to 0.2, 0.3, and 0.4), the landscape changes from a quadristable (four stable states coexist), to tristable (DN2a, DN2b and DN3 coexist), to bistable (DN2b and DN3 coexist) and finally to a monostable DN3 state. The landscape change shows how the Notch signal promotes the transition from ETP to DN3, quantitatively.

**S10 Fig. Quasi-energy landscape for the *Bcl11b* knockout condition.** With the *Bcl11b* knockout ( $k_B = 0$ ), the landscape changes from a quadristable (four stable states coexist), to a bistable (ETP and DN2a) state. On the landscape, the blue regions represent higher probability or lower potential and the yellow regions represent lower probability or higher potential. Two basins of attractions characterize two different cell states (ETP and DN2a). The landscape change shows that the *Bcl11b* knockout will make ETP/DN2a states more stable and DN2b/DN3 less stable.

**S11 Fig. Venn diagram of four types of network motifs that can produce four attractors with up to three TFs.** Red and blue areas correspond to Type I and Type II motifs shown in Fig 2B. Green area corresponds to motifs that contain both Type I and Type II networks. Orange area corresponds to motifs that can only produce four unordered attractors, in which the concentrations of the TFs are non-monotonically correlated. Numbers in the diagrams denote the total numbers of non-redundant topologies for each type. The Type II (blue) and Hybrid (green) motifs can produce both ordered and unordered 4-attractor systems, depending on the choice of parameters.

**S12 Fig. A phaseplane for a system containing four unordered attractors.** Model 2 (Type II) is chosen as an example. See S2 Fig for the scenario with ordered attractors. Nullclines for TF A (the node on the left of the network diagram) and TF B (the node on the right of the network diagram) are shown. Stable steady states are shown as black dots. Random parameter sampling was used to obtain the parameter set.

**S13 Fig. The comparisons for the landscape and paths between Type I and Type II minimal motifs.** To make comparisons, two models (A and B) are picked from Type I motifs (Model 11 and Model 17 in S1 Fig) and two models (C and D) are picked from Type II motifs (Model 19 and Model 28 in S2 Fig). On the landscape, the blue regions represent higher probability or lower potential and the yellow regions represent lower probability or higher potential. For each subfigure, four basins of attractions characterize four different cell states (ETP, DN2a, DN2b, and DN3). Both Type I motif (A and B) and Type II motif (C and D) can generate stable four-attractor system.

**S14 Fig. Effects of the fraction of positive regulations in network topologies on the performance of four-attractor systems.** Scatter plots and regression lines show correlations of the numbers of successful parameter sets per topology and the fractions of positive regulations in corresponding topologies. **A.** Sampled 3-node generic topologies. Spearman correlation coefficient: 0.21, p-value = 0.008. **B.** Subnetworks of the T cell development model. Spearman correlation coefficient: 0.31, p-value < 0.0001. **C.** Comparison of networks generating attractors with all positively correlated TFs and those with at least one pair of inversely correlated TFs. Three-node generic topologies were analyzed. (TIF)

**S15 Fig. Enrichment motifs containing varying numbers of positive feedback loops similar to Type I motif.** Top panel: total occurrences of various types of motifs in the T cell network. Middle panel: empirical p-values (middle panel) for these motifs with a background network population. Bottom panel: an illustration of the p-values with the distributions of the background population. Each motif has  $n$  ( $0 < n < 9$ ) positive feedback loops, all of which share at least one TFs in the motif. Type I motif is a special case of such motifs with  $n = 3$ . Random networks were obtained by assigning random regulations (positive, negative or none) between each pair of TFs.  $10^5$  random networks were generated. Empirical p-values were obtained by counting the number of the random networks with the motifs not less than those in the T cell network. See [Methods](#) for details of the p-value definition. Distributions of motif frequencies obtained from the random networks are shown in the bottom panel. The yellow vertical bars represent the number of occurrences in the T cell network. The right-tail areas defined by the vertical bars correspond to the p-values shown in the middle panel (blue bars). (TIF)

**S16 Fig. Comparison of parameter sets from the 66 minimum motifs and the T cell network full model that generated the four-attractor systems.** **A.** Distributions of Hill exponents (steepness of regulations) of 900 parameter sets from each category. The parameter sets from 66 minimum motifs are combined. **B.** Distributions of the activation threshold parameters of the regulation of GATA3 by TCF-1 and the regulation of TCF-1 by GATA3. 440 parameter sets (maximum number of parameter sets from minimum motifs that contain these two parameters) from each category are plotted. (TIF)

## Author Contributions

**Conceptualization:** Chunhe Li, Tian Hong.

**Data curation:** Yujie Ye, Chunhe Li, Tian Hong.

**Funding acquisition:** Chunhe Li, Tian Hong.

**Methodology:** Chunhe Li, Tian Hong.

**Visualization:** Xin Kang, Jordan Bailey, Chunhe Li, Tian Hong.

**Writing – original draft:** Chunhe Li, Tian Hong.

**Writing – review & editing:** Chunhe Li, Tian Hong.

## References

1. Holtzer H, Weintraub H, Mayne R, Mochan B (1972) The cell cycle, cell lineages, and cell differentiation. *Current topics in developmental biology*: Elsevier. pp. 229–256. PMID: [4264706](#)

2. Slack JMW (2009) Essential developmental biology: John Wiley & Sons.
3. Mariani L, Löhning M, Radbruch A, Höfer T (2004) Transcriptional control networks of cell differentiation: insights from helper T lymphocytes. *Progress in biophysics and molecular biology* 86: 45–76. <https://doi.org/10.1016/j.pbiomolbio.2004.02.007> PMID: 15261525
4. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125: 315–326. <https://doi.org/10.1016/j.cell.2006.02.041> PMID: 16630819
5. Ellmeier W, Taniuchi I (2016) Transcriptional control of lineage differentiation in immune cells: Springer.
6. Wang X, Li J (2017) An intermediate cell state allows rerouting of cell fate. *Journal of Biological Chemistry* 292: 19133–19134. <https://doi.org/10.1074/jbc.H117.812537> PMID: 29373866
7. Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, et al. (2016) Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS biology* 14: e1002585. <https://doi.org/10.1371/journal.pbio.1002585> PMID: 28027290
8. Mojtahedi M, Skupin A, Zhou J, Castano IG, Leong-Quong RYY, et al. (2016) Cell fate decision as high-dimensional critical state transition. *PLoS biology* 14: e2000640. <https://doi.org/10.1371/journal.pbio.2000640> PMID: 28027308
9. Lu KT, Kanno Y, Cannons JL, Handon R, Bible P, et al. (2011) Functional and epigenetic studies reveal multistep differentiation and plasticity of in vitro-generated and in vivo-derived follicular T helper cells. *Immunity* 35: 622–632. <https://doi.org/10.1016/j.immuni.2011.07.015> PMID: 22018472
10. Chang HH, Oh PY, Ingber DE, Huang S (2006) Multistable and multistep dynamics in neutrophil differentiation. *BMC cell biology* 7: 11. <https://doi.org/10.1186/1471-2121-7-11> PMID: 16507101
11. MacLean AL, Hong T, Nie Q (2018) Exploring intermediate cell states through the lens of single cells. *Current Opinion in Systems Biology* 9: 32–41. <https://doi.org/10.1016/j.coisb.2018.02.009> PMID: 30450444
12. Wang J, Sun Q, Morita Y, Jiang H, Groß A, et al. (2012) A Differentiation Checkpoint Limits Hematopoietic Stem Cell Self-Renewal in Response to DNA Damage. *Cell* 148: 1001–1014. <https://doi.org/10.1016/j.cell.2012.01.040> PMID: 22385964
13. Berkowska MA, van der Burg M, van Dongen JJM, van Zelm MC (2011) Checkpoints of B cell differentiation: visualizing Ig-centric processes. *Annals of the New York Academy of Sciences* 1246: 11–25. <https://doi.org/10.1111/j.1749-6632.2011.06278.x> PMID: 22236426
14. Bod L, Douguet L, Auffray C, Lengagne R, Bekkat F, et al. (2018) IL-4–Induced Gene 1: A Negative Immune Checkpoint Controlling B Cell Differentiation and Activation. *The Journal of Immunology* 200: 1027–1038. <https://doi.org/10.4049/jimmunol.1601609> PMID: 29288206
15. Ikawa T, Hirose S, Masuda K, Kakugawa K, Satoh R, et al. (2010) An essential developmental checkpoint for production of the T cell lineage. *Science* 329: 93–96. <https://doi.org/10.1126/science.1188995> PMID: 20595615
16. Rothenberg EV, Moore JE, Yui MA (2008) Launching the T-cell-lineage developmental programme. *Nature Reviews Immunology* 8: 9. <https://doi.org/10.1038/nri2232> PMID: 18097446
17. Mingueneau M, Kreslavsky T, Gray D, Heng T, Cruse R, et al. (2013) The transcriptional landscape of αβ T cell differentiation. *Nature immunology* 14: 619. <https://doi.org/10.1038/ni.2590> PMID: 23644507
18. Ha VL, Luong A, Li F, Casero D, Malvar J, et al. (2017) The T-ALL related gene BCL11B regulates the initial stages of human T-cell differentiation. *Leukemia* 31: 2503. <https://doi.org/10.1038/leu.2017.70> PMID: 28232744
19. Rothenberg EV (2011) T cell lineage commitment: identity and renunciation. *The Journal of Immunology* 186: 6649–6655. <https://doi.org/10.4049/jimmunol.1003703> PMID: 21646301
20. Yui MA, Rothenberg EV (2014) Developmental gene networks: a triathlon on the course to T cell identity. *Nature Reviews Immunology* 14: 529–545. <https://doi.org/10.1038/nri3702> PMID: 25060579
21. Taghon T, Yui MA, Rothenberg EV (2007) Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nature immunology* 8: 845. <https://doi.org/10.1038/ni1486> PMID: 17603486
22. Franco CB, Scripture-Adams DD, Proekt I, Taghon T, Weiss AH, et al. (2006) Notch/Delta signaling constrains reengineering of pro-T cells by PU. 1. *Proceedings of the National Academy of Sciences* 103: 11993–11998.
23. Rosenbauer F, Owens BM, Yu L, Tumang JR, Steidl U, et al. (2006) Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU. 1. *Nature genetics* 38: 27. <https://doi.org/10.1038/ng1679> PMID: 16311598

24. Luc S, Luis TC, Boukarabila H, Macaulay IC, Buza-Vidas N, et al. (2012) The earliest thymic T cell progenitors sustain B cell and myeloid lineage potential. *Nature immunology* 13: 412. <https://doi.org/10.1038/ni.2255> PMID: 22344248
25. Li P, Burke S, Wang J, Chen X, Ortiz M, et al. (2010) Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science* 329: 85–89. <https://doi.org/10.1126/science.1188063> PMID: 20538915
26. Li L, Leid M, Rothenberg EV (2010) An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* 329: 89–93. <https://doi.org/10.1126/science.1188989> PMID: 20595614
27. Manesso E, Chickarmane V, Kueh HY, Rothenberg EV, Peterson C (2013) Computational modelling of T-cell formation kinetics: output regulated by initial proliferation-linked deferral of developmental competence. *Journal of The Royal Society Interface* 10: 20120774.
28. Longabaugh WJR, Zeng W, Zhang JA, Hosokawa H, Jansen CS, et al. (2017) Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. *Proceedings of the National Academy of Sciences* 114: 5800–5807.
29. Hong T, Watanabe K, Ta CH, Villarreal-Ponce A, Nie Q, et al. (2015) An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. *PLoS computational biology* 11: e1004569. <https://doi.org/10.1371/journal.pcbi.1004569> PMID: 26554584
30. Jolly MK, Tripathi SC, Jia D, Mooney SM, Celiktas M, et al. (2016) Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* 7: 27067. <https://doi.org/10.18632/oncotarget.8166> PMID: 27008704
31. Mandal M, Ghosh B, Anura A, Mitra P, Pathak T, et al. (2016) Modeling continuum of epithelial mesenchymal transition plasticity. *Integrative Biology* 8: 167–176. <https://doi.org/10.1039/c5ib00219b> PMID: 26762753
32. Blanpain C, Fuchs E (2009) Epidermal homeostasis: a balancing act of stem cells in the skin. *Nature reviews Molecular cell biology* 10: 207. <https://doi.org/10.1038/nrm2636> PMID: 19209183
33. Du H, Wang Y, Haensel D, Lee B, Dai X, et al. (2018) Multiscale modeling of layer formation in epidermis. *PLoS computational biology* 14: e1006006. <https://doi.org/10.1371/journal.pcbi.1006006> PMID: 29481568
34. Huang RY, Wong MK, Tan TZ, Kuay KT, Ng AH, et al. (2013) An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). *Cell death & disease* 4: e915.
35. Zhang J, Tian XJ, Zhang H, Teng Y, Li R, et al. (2014) TGF- $\beta$ -induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Science Signaling* 7: ra91-ra91.
36. Grosse-Wilde A, Fouquier d'Herouel A, McIntosh E, Ertaylan G, Skupin A, et al. (2015) Stemness of the hybrid Epithelial/Mesenchymal State in Breast Cancer and Its Association with Poor Survival. *PLoS one* 10: e0126522. <https://doi.org/10.1371/journal.pone.0126522> PMID: 26020648
37. Koster MI, Roop DR (2007) Mechanisms regulating epithelial stratification. *Annu Rev Cell Dev Biol* 23: 93–113. <https://doi.org/10.1146/annurev.cellbio.23.090506.123357> PMID: 17489688
38. Ma W, Trusina A, El-Samad H, Lim WA, Tang C (2009) Defining network topologies that can achieve biochemical adaptation. *Cell* 138: 760–773. <https://doi.org/10.1016/j.cell.2009.06.013> PMID: 19703401
39. Fu Y, Glaros T, Zhu M, Wang P, Wu Z, et al. (2012) Network topologies and dynamics leading to endotoxin tolerance and priming in innate immune cells. *PLoS computational biology* 8: e1002526. <https://doi.org/10.1371/journal.pcbi.1002526> PMID: 22615556
40. Jiménez A, Cotterell J, Munteanu A, Sharpe J (2017) A spectrum of modularity in multi-functional gene circuits. *Molecular systems biology* 13: 925. <https://doi.org/10.15252/msb.20167347> PMID: 28455348
41. Manesso E, Kueh HY, Freedman G, Rothenberg EV, Peterson C (2016) Irreversibility of T-cell specification: insights from computational modelling of a minimal network architecture. *PLoS one* 11: e0161260. <https://doi.org/10.1371/journal.pone.0161260> PMID: 27551921
42. Conradi C, Flockerzi D, Raisch J, Stelling J (2007) Subnetwork analysis reveals dynamic features of complex (bio) chemical networks. *Proceedings of the National Academy of Sciences* 104: 19175–19180.
43. Albeck JG, Burke JM, Spencer SL, Lauffenburger DA, Sorger PK (2008) Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS biology* 6: e299.



44. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459: 428. <https://doi.org/10.1038/nature08012> PMID: 19363473
45. Zhang X-P, Liu F, Cheng Z, Wang W (2009) Cell fate decision mediated by p53 pulses. *Proceedings of the National Academy of Sciences* 106: 12245–12250.
46. Yao G, Lee TJ, Mori S, Nevins JR, You L (2008) A bistable Rb–E2F switch underlies the restriction point. *Nature cell biology* 10: 476. <https://doi.org/10.1038/ncb1711> PMID: 18364697
47. Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV (2012) Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* 149: 467–482. <https://doi.org/10.1016/j.cell.2012.01.056> PMID: 22500808
48. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297: 1183–1186. <https://doi.org/10.1126/science.1070919> PMID: 12183631
49. Kærn M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics* 6: 451. <https://doi.org/10.1038/nrg1615> PMID: 15883588
50. Waddington CH (1957) The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *The strategy of the genes A discussion of some aspects of theoretical biology With an appendix by H Kacser.*
51. Wang J, Zhang K, Xu L, Wang E (2011) Quantifying the Waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences* 108: 8257–8262.
52. Li C, Wang J (2013) Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. *PLoS computational biology* 9: e1003165. <https://doi.org/10.1371/journal.pcbi.1003165> PMID: 23935477
53. Li C, Wang J (2013) Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *Journal of The Royal Society Interface* 10: 20130787.
54. Li C, Wang J (2015) Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer research* 75: 2607–2618. <https://doi.org/10.1158/0008-5472.CAN-15-0079> PMID: 25972342
55. Liao C, Lu T (2013) A minimal transcriptional controlling network of regulatory T cell development. *The Journal of Physical Chemistry B* 117: 12995–13004. <https://doi.org/10.1021/jp402306g> PMID: 23642089
56. Wang P, Song C, Zhang H, Wu Z, Tian X-J, et al. (2014) Epigenetic state network approach for describing cell phenotypic transitions. *Interface focus* 4: 20130068. <https://doi.org/10.1098/rsfs.2013.0068> PMID: 24904734
57. Margaret JT, Chu BK, Gallivan CP, Read EL (2018) Rare-event sampling of epigenetic landscapes and phenotype transitions. *PLoS Comput Biol* 14: e1006336. <https://doi.org/10.1371/journal.pcbi.1006336> PMID: 30074987
58. Zhou X, Ren W, E W (2008) Adaptive minimum action method for the study of rare events. *The Journal of chemical physics* 128: 104111. <https://doi.org/10.1063/1.2830717> PMID: 18345881
59. Li C (2017) Identifying the optimal anticancer targets from the landscape of a cancer–immunity interaction network. *Physical Chemistry Chemical Physics* 19: 7642–7651. <https://doi.org/10.1039/c6cp07767f> PMID: 28256642
60. Wang J, Xu L, Wang E (2008) Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences* 105: 12271–12276.
61. Li C, Wang J (2014) Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proceedings of the National Academy of Sciences* 111: 14130–14135.
62. Rothenberg EV, Champhekar A, Damle S, Del Real MM, Kueh HY, et al. (2013) Transcriptional establishment of cell-type identity: dynamics and causal mechanisms of T-cell lineage commitment. *Cold Spring Harbor Laboratory Press*. pp. 31–41.
63. Guantes R, Poyatos JF (2008) Multistable Decision Switches for Flexible Control of Epigenetic Differentiation. *PLoS Computational Biology* 4: e1000235. <https://doi.org/10.1371/journal.pcbi.1000235> PMID: 19043543
64. Hong T, Oguz C, Tyson JJ (2015) A Mathematical Framework for Understanding Four-Dimensional Heterogeneous Differentiation of CD4+ T Cells. *Bulletin of mathematical biology* 77: 1046–1064. <https://doi.org/10.1007/s11538-015-0076-6> PMID: 25779890

65. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E (2013) MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proceedings of the National Academy of Sciences of the United States of America* <https://doi.org/10.1073/pnas.1318192110> PMID: 24154725
66. Huang B, Xia Y, Liu F, Wang W (2016) Realization of tristability in a multiplicatively coupled dual-loop genetic network. *Scientific reports* 6: 28096. <https://doi.org/10.1038/srep28096> PMID: 27378101
67. Yu P, Nie Q, Tang C, Zhang L (2018) Nanog induced intermediate state in regulating stem cell differentiation and reprogramming. *BMC systems biology* 12: 22. <https://doi.org/10.1186/s12918-018-0552-3> PMID: 29486740
68. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403: 339. <https://doi.org/10.1038/35002131> PMID: 10659857
69. Novak B, Tyson JJ, Gyorffy B, Csikasz-Nagy A (2007) Irreversible cell-cycle transitions are due to systems-level feedback. *Nature cell biology* 9: 724. <https://doi.org/10.1038/ncb0707-724> PMID: 17603504
70. Stewart-Ornstein J, Nelson C, DeRisi J, Weissman JS, El-Samad H (2013) Msn2 coordinates a stoichiometric gene expression program. *Current Biology* 23: 2336–2345. <https://doi.org/10.1016/j.cub.2013.09.043> PMID: 24210615
71. Rogers KW, Schier AF (2011) Morphogen gradients: from generation to interpretation. *Annual review of cell and developmental biology* 27: 377–407. <https://doi.org/10.1146/annurev-cellbio-092910-154148> PMID: 21801015
72. Giorgetti L, Siggers T, Tiana G, Caprara G, Notarbartolo S, et al. (2010) Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Molecular cell* 37: 418–428. <https://doi.org/10.1016/j.molcel.2010.01.016> PMID: 20159560
73. Hayes L, Ralls S, Wang H, Ahn S (2013) Duration of Shh signaling contributes to mDA neuron diversity. *Developmental biology* 374: 115–126. <https://doi.org/10.1016/j.ydbio.2012.11.016> PMID: 23201023
74. Balaskas N, Ribeiro A, Panovska J, Dessaud E, Sasai N, et al. (2012) Gene regulatory logic for reading the Sonic Hedgehog signaling gradient in the vertebrate neural tube. *Cell* 148: 273–284. <https://doi.org/10.1016/j.cell.2011.10.047> PMID: 22265416
75. Miskov-Zivanov N, Turner MS, Kane LP, Morel PA, Faeder JR (2013) The duration of T cell stimulation is a critical determinant of cell fate and plasticity. *Sci Signal* 6: ra97-ra97.
76. Alon U (2007) Network motifs: theory and experimental approaches. *Nature Reviews Genetics* 8: 450. <https://doi.org/10.1038/nrg2102> PMID: 17510665
77. Hart Y, Alon U (2013) The utility of paradoxical components in biological circuits. *Molecular cell* 49: 213–221. <https://doi.org/10.1016/j.molcel.2013.01.004> PMID: 23352242
78. Ingram PJ, Stumpf MPH, Stark J (2006) Network motifs: structure does not determine function. *BMC genomics* 7: 108. <https://doi.org/10.1186/1471-2164-7-108> PMID: 16677373
79. Lu M, Jolly MK, Gomoto R, Huang B, Onuchic J, et al. (2013) Tristability in cancer-associated micro-RNA-TF chimera toggle switch. *The journal of physical chemistry B* 117: 13164–13174. <https://doi.org/10.1021/jp403156m> PMID: 23679052
80. Thomas R (1981) On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Numerical methods in the study of critical phenomena*: Springer. pp. 180–193.
81. Soulé C (2003) Graphic requirements for multistationarity. *ComplexUs* 1: 123–133.
82. Ahrends R, Ota A, Kovary KM, Kudo T, Park BO, et al. (2014) Controlling low rates of cell differentiation through noise and ultrahigh feedback. *Science* 344: 1384–1389. <https://doi.org/10.1126/science.1252079> PMID: 24948735
83. Huang S (2013) Hybrid T-helper cells: stabilizing the moderate center in a polarized system. *PLoS biology* 11: e1001632. <https://doi.org/10.1371/journal.pbio.1001632> PMID: 23976879
84. Graf T, Enver T (2009) Forcing cells to change lineages. *Nature* 462: 587. <https://doi.org/10.1038/nature08533> PMID: 19956253
85. Ardavin C, Wu L, Li C-L, Shortman K (1993) Thymic dendritic cells and T cells develop simultaneously in the thymus from a common precursor population. *Nature* 362: 761. <https://doi.org/10.1038/362761a0> PMID: 8469288
86. Wu L, Li C-L, Shortman K (1996) Thymic dendritic cell precursors: relationship to the T lymphocyte lineage and phenotype of the dendritic cell progeny. *Journal of Experimental Medicine* 184: 903–911. PMID: 9064350

87. Vosshenrich CAJ, García-Ojeda ME, Samson-Villéger SI, Pasqualetto V, Enault L, et al. (2006) A thymic pathway of mouse natural killer cell development characterized by expression of GATA-3 and CD127. *Nature immunology* 7: 1217. <https://doi.org/10.1038/ni1395> PMID: 17013389
88. Huang B, Lu M, Jia D, Ben-Jacob E, Levine H, et al. (2017) Interrogating the topological robustness of gene regulatory circuits by randomization. *PLoS computational biology* 13: e1005456. <https://doi.org/10.1371/journal.pcbi.1005456> PMID: 28362798
89. Holmes WR, Reyes de Mochel NS, Wang Q, Du H, Peng T, et al. (2017) Gene Expression Noise Enhances Robust Organization of the Early Mammalian Blastocyst. *PLoS Computational Biology* 13: e1005320. <https://doi.org/10.1371/journal.pcbi.1005320> PMID: 28114387
90. Prill RJ, Iglesias PA, Levchenko A (2005) Dynamic properties of network motifs contribute to biological network organization. *PLoS biology* 3: e343. <https://doi.org/10.1371/journal.pbio.0030343> PMID: 16187794
91. Weber BN, Chi AW, Chavez A, Yashiro-Ohtani Y, Yang Q, et al. (2011) A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* 476: 63–68. <https://doi.org/10.1038/nature10279> PMID: 21814277
92. Schmitt TM, Zuniga-Pflucker JC (2002) Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity* 17: 749–756. PMID: 12479821
93. Franco CB, Scripture-Adams DD, Proekt I, Taghon T, Weiss AH, et al. (2006) Notch/Delta signaling constrains reengineering of pro-T cells by PU.1. *Proc Natl Acad Sci U S A* 103: 11993–11998. <https://doi.org/10.1073/pnas.0601188103> PMID: 16880393
94. Del Real MM, Rothenberg EV (2013) Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. *Development* 140: 1207–1219. <https://doi.org/10.1242/dev.088559> PMID: 23444353
95. Garcia-Ojeda ME, Klein Wolterink RG, Lemaitre F, Richard-Le Goff O, Hasan M, et al. (2013) GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice. *Blood* 121: 1749–1759. <https://doi.org/10.1182/blood-2012-06-440065> PMID: 23287858
96. Taghon T, Yui MA, Rothenberg EV (2007) Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nat Immunol* 8: 845–855. <https://doi.org/10.1038/ni1486> PMID: 17603486
97. Yui MA, Feng N, Rothenberg EV (2010) Fine-scale staging of T cell lineage commitment in adult mouse thymus. *J Immunol* 185: 284–293. <https://doi.org/10.4049/jimmunol.1000679> PMID: 20543111
98. Tydell CC, David-Fung ES, Moore JE, Rowen L, Taghon T, et al. (2007) Molecular dissection of pre-thymic progenitor entry into the T lymphocyte developmental pathway. *J Immunol* 179: 421–438. PMID: 17579063
99. Van de Walle I, De Smet G, De Smedt M, Vandekerckhove B, Leclercq G, et al. (2009) An early decrease in Notch activation is required for human TCR-alpha-beta lineage differentiation at the expense of TCR-gammadelta T cells. *Blood* 113: 2988–2998. <https://doi.org/10.1182/blood-2008-06-164871> PMID: 19056690
100. Weerkamp F, Luis TC, Naber BA, Koster EE, Jeannotte L, et al. (2006) Identification of Notch target genes in uncommitted T-cell progenitors: No direct induction of a T-cell specific gene program. *Leukemia* 20: 1967–1977. <https://doi.org/10.1038/sj.leu.2404396> PMID: 16990763
101. Germar K, Dose M, Konstantinou T, Zhang J, Wang H, et al. (2011) T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. *Proc Natl Acad Sci U S A* 108: 20060–20065. <https://doi.org/10.1073/pnas.1110230108> PMID: 22109558
102. Leddin M, Perrod C, Hoogenkamp M, Ghani S, Assi S, et al. (2011) Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. *Blood* 117: 2827–2838. <https://doi.org/10.1182/blood-2010-08-302976> PMID: 21239694
103. Oh P, Lobry C, Gao J, Tikhonova A, Loizou E, et al. (2013) In vivo mapping of notch pathway activity in normal and stress hematopoiesis. *Cell Stem Cell* 13: 190–204. <https://doi.org/10.1016/j.stem.2013.05.015> PMID: 23791481
104. Zarnegar MA, Rothenberg EV (2012) Ikaros represses and activates PU.1 cell-type-specifically through the multifunctional Sspi1 URE and a myeloid specific enhancer. *Oncogene* 31: 4647–4654. <https://doi.org/10.1038/onc.2011.597> PMID: 22231443
105. Kueh HY, Yui MA, Ng KKH, Pease SS, Zhang JA, et al. (2016) Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. *Nature immunology* 17: 956. <https://doi.org/10.1038/ni.3514> PMID: 27376470
106. Ta CH, Nie Q, Hong T (2016) Controlling stochasticity in epithelial-mesenchymal transition through multiple intermediate cellular states. *Discrete & Continuous Dynamical Systems—Series B* 21.

107. Fu Y, Lim S, Urano D, Tunc-Ozdemir M, Phan NG, et al. (2014) Reciprocal encoding of signal intensity and duration in a glucose-sensing circuit. *Cell* 156: 1084–1095. <https://doi.org/10.1016/j.cell.2014.01.013> PMID: 24581502
108. Zhang B, Wolynes PG (2014) Stem cell differentiation as a many-body problem. *Proceedings of the National Academy of Sciences* 111: 10185–10190.
109. Van Kampen NG (1992) *Stochastic processes in physics and chemistry*: Elsevier.
110. Hu G (1994) *Stochastic forces and nonlinear systems*. Shanghai: Shanghai Scientific and Technological Education Publishing House. 184 p.
111. Feng H, Zhang K, Wang J (2014) Non-equilibrium transition state rate theory. *Chemical Science* 5: 3761–3769.
112. Ren W, Vanden-Eijnden E (2004) Minimum action method for the study of rare events. *Communications on pure and applied mathematics* 57: 637–656.
113. Freidlin M, Weber M (2004) Random perturbations of dynamical systems and diffusion processes with conservation laws. *Probability theory and related fields* 128: 441–466.