LU-TP 22-42
June 2022

Energy Landscapes for Early T Cell Development

Yukun Cheng

Department of Astronomy and Theoretical Physics,
Lund University

Master thesis supervised by Victor Olariu

LUND
UNIVERSITY

Abstract

T cell lineage development from an early thymic progenitor involves programmed shutoff of progenitor gene expression, upregulation of T cell specification genes, and proliferation. This biological system has been deeply studied and a gene regulatory network (GRN) that can accurately describe the system has been presented. Such a GRN offers an excellent opportunity to study lineage commitment from a multipotent progenitor. In this work, we investigated the GRN governing early T cell development by considering cell commitment as movements in an energy landscape. Our energy landscape method provided predictions that are consistent with experimental observations, both with and without simulated gene knockdowns. Moreover, we implemented cell development and proliferation within the energy landscape framework. We were able to show the impact of noise levels on cell developmental speed. The results are coherent with what would be expected for realistic biological systems.

Populärvetenskaplig beskrivning

T cells are one of the most important blood cells in our body that help us defend against outside invaders and diseases. Even though T cells may seem quite different from some other blood cells in terms of both their shapes and functions, they are born from the same kind of stem cells and share the same genetic information, like most blood cells. What differentiates T cells from other blood cells lies in the highly controlled modifications in gene expression. Such differences in gene expression are a result of the path that T cells have taken from stem cell stages during a cellular differentiation process.

Differentiation from a stem cell to a mature cell is a key cellular process in normal tissue development. This process is where a stem cell becomes a more specific type of cell with unique functions. A differentiation process usually involves many genes. Genes work together and interact with each other during this process. The final results of such interactions are that some genes are activated (expressed), but some are not. The information from activated genes are then used to produce functional products such as proteins. The products of activated genes will ultimately affect the properties of the cells.

To help understand the differentiation process better, we now treat an overall state of the genes as a cell state. Following this treatment, if we look at how the cell state develops during a differentiation process, there are patterns to follow. Some cell states develop into other states, but not the other way around. This is like a ball rolling on a road: the ball tends to go downhill but not uphill because the gravitational energy is higher uphill. In a similar vein as this gravitational energy, we can evaluate the energy of different cell states, which leads us to an energy landscape.

In the energy landscape, a state with lower energy would be more favorable. The differentiation process can be viewed as a ball rolling in the energy landscape, and the fate of a cell will depend on the pathway the ball takes and where it settles down. The landscape view of the cellular differentiation process can be a powerful tool for analyzing biological systems. T cell development provides a great example to carry out such studies.

In this work, we tested a landscape method based on a modified binary logic on a gene interaction model for early T cell development. Our landscape method can make predictions that agree with experimental results.

Contents

1 Introduction

In our daily lives, we are frequently exposed to infectious agents such as bacteria, protozoa, and viruses [1]. For our immune system to protect us against outside invaders or diseases, lymphocytes play a vital role. Among lymphocytes, B cells can produce antibodies [2], whereas T cells can help direct the immune response or wipe out the infected cells [3]. Despite the fact that T cells and B cells function differently in our immune system, they share the near-identical genome just as many other types of cells in our body. What separates a T cell from a B cell is which genes are activated (in other words, expressed), a result of taking different developmental pathways from a pluripotent progenitor.

Now we take a close look at the process of early T cell development. Pluripotent progenitor cells originating from hematopoietic stem cells (HSCs) in the bone marrow reach the thymus where they start to develop into functional T cells [4]. The cells that just arrived at the thymus do not express the CD4 or CD8 receptor and are termed double negative (DN). From DN1 cells or early thymic progenitors (ETP), the first step in T cell commitment is towards DN2a cells. At the DN2a state, cells have positive surface CD25 expression. At the next step, DN2a cells develop into DN2b cells. This step is marked by upregulation of Bcl11b. Once the cells have the Bcl11b expression, they will enter Bcl11b-dependent T-lineage commitment [5–7].

Much progress has been made in terms of understanding important regulatory factors controlling the early stages of T cell development [8,9]. A gene regulatory network (GRN) architecture describing the dynamics of T cell specification regulatory genes during DN1/ETP - DN2a transition has been presented [10].

One approach to exploring such a GRN architecture is by building models that treat continuous gene expression levels and using rate equations to carry out deterministic or stochastic simulations [11]. Such an approach has been adopted in previous studies [10]. However, this treatment results in an infinite gene expression space, which may lead to the need of simplification.

To simplify the gene expression level, we can consider it to only be either ON (1) or OFF (0). In this case, the gene expression space is reduced to a finite size of 2^N states (N is the number of genes in the GRN) [12, 13]. In such a Boolean model, it is reasonable to construct an energy function (landscape) based on the interactions between genes and the genes' activation statuses. The landscape can be a discrete function of all possible gene states, and it indicates which gene states are more favorable than others.

Metaphorically, the fate decision process of the cells can then be considered as a real marble rolling in the landscape, going through different paths (figure 1), and finally being captured by an attractor in the landscape [14,15]. Like in physics, a system tends to evolve into a state where the energy is lower. For a GRN controlling cell commitment, minimum energy in the landscape corresponds to a stable cell state. Cell differentiation can indeed be viewed as a process where the system navigates the energy landscape, moving from a

local minimum to another minimum from which it cannot escape.

With an energy landscape based on the GRN, the dynamics of the system can be explored by studying the fate of a system initialized at specified states in the landscape. In this work, we tested an energy landscape method on the GRN governing the transition from the DN1/ETP state to the DN2a state in T cell development. Our model provided results similar to experimental observations with or without gene knockdown. However, we found that to model cell proliferation in the energy landscape, a continuous energy function and a better understanding of cell cycles could be necessary.

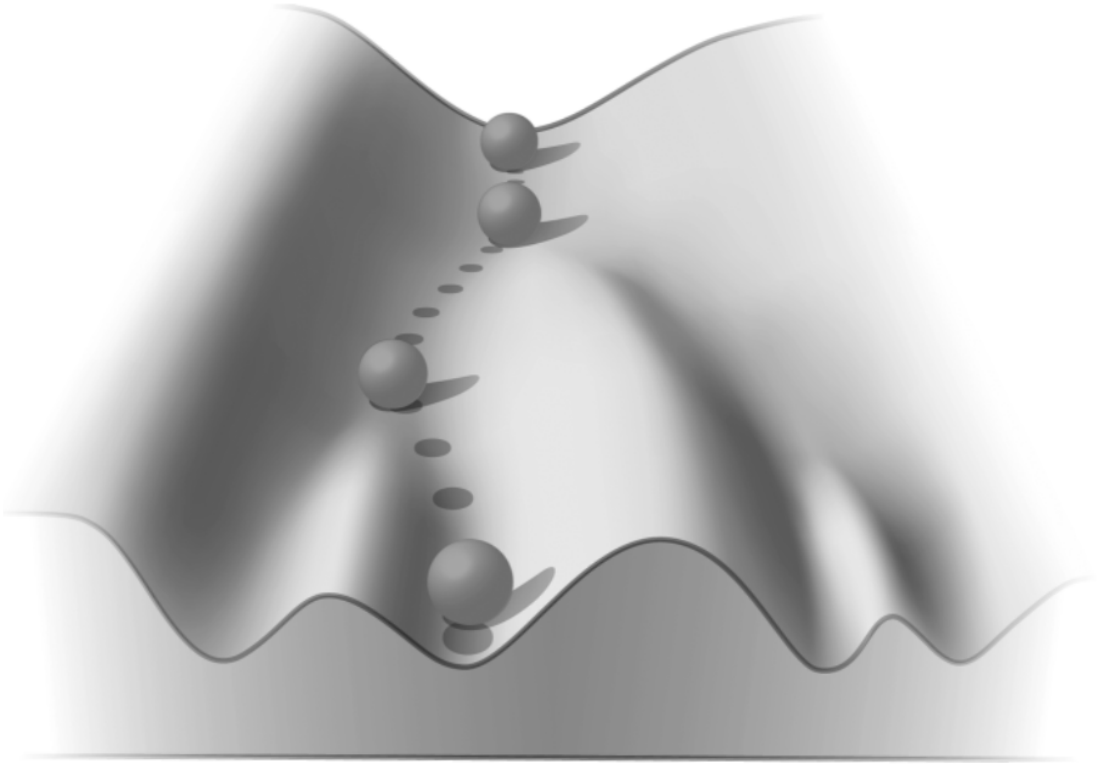

Figure 1: An energy landscape. A ball rolls down the hill with different paths available (modified from `https://figshare.com/articles/_Waddington_s_8220_Epigenetic_Landscape_8221_/620879`, Creative Commons license CC BY 4.0) [16].

2 Background

2.1 Mapping Boolean implementation of GRNs into landscapes

The specific framework we used in this work is CELLoGeNe (Computation of Energy Landscapes of Logical Gene Networks) [17].

2.1.1 Gene interactions and the logic

In CELLoGeNe, the activation state of a gene is distinguished from the gene's action on another gene (the logic is slightly different). To be more precise, a gene's activation state can be either ON (1) or OFF (0). However, a gene can activate (+1), repress (−1), or have no effect (0) on another gene.

With this logic, the energy contribution from a single gene is calculated based on two factors: (1) its activation state and (2) how other genes act on it. This energy measures if a gene is in its favorable state. Thus, a lower energy is given when the state of the gene is in agreement with other genes' action on it (we say this state is stable). A higher energy is assigned if the state is not stable.

Now we consider a gene B repressed by a gene A (figure 2 (a)). Due to the repressive action of A, B will be forced towards being OFF (0). Therefore, as long as A is ON (1), lower energy is assigned to $B = 0$ and higher energy is assigned to $B = 1$. When A is OFF (0) meaning that no action on B, a neutral energy is assigned to B for both $B = 0$ and $B = 1$.

The energy contribution of gene B for (A, B) = (0, 0), (0, 1), (1, 0), and (1, 1) is 0, 0, −1, and 1 respectively. Note that only gene B has an energy contribution here. Gene A is not regulated by any gene, and therefore, its energy contribution is 0.

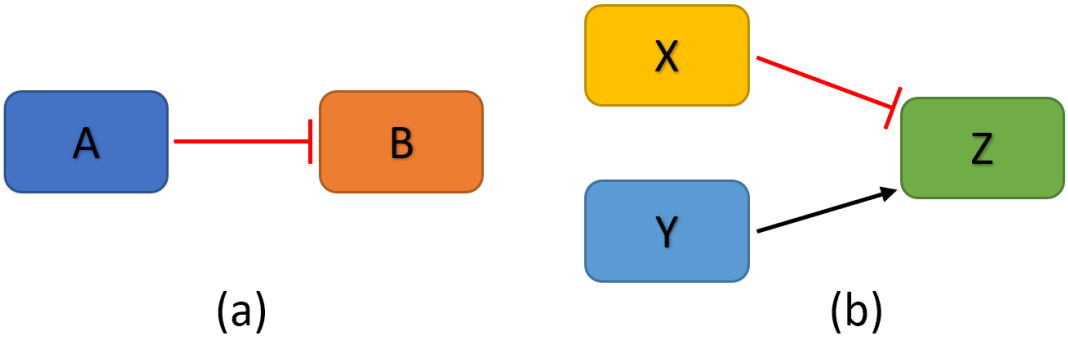

Figure 2: Two very simple networks. The activation effect is depicted by (→) and repression is depicted by (⊣). (a) Gene A represses gene B. (b) Gene Z is repressed by gene X but activated by gene Y.

2.1.2 Combining multiple inputs to a single gene

Now we look at a network where a gene Z is repressed by a gene X but activated by a gene Y (figure 2 (b)). In this case, when X is ON (1), it provides a repressive (-1) input to gene Z, but when X is OFF (0), it provides an input of 0 (no effect) to gene Z. Similarly, when Y is ON (1), it provides an activating (1) input to gene Z, but an input of 0 when OFF.

Logical operators (see Appendix A) are needed to combine the inputs from X and Y to Z. The net effect on a gene being regulated may depend on the operator choice. Some operators could have an activation effect always win against a repression effect, while some could have a repression effect override an activation effect. For example, UBOR behaves the same as normal logical OR, except when the inputs are $+1$ and -1, UBOR outputs -1. In this case, it means that a repressive input (-1) always overcomes an activating input ($+1$).

Sometimes in a GRN, there could be multiple genes taking multiple inputs. This can lead to a very large space of operator combinations. Consequently, the operator choice not only affects the energy of a single gene at a given state, but also alters the whole energy landscape.

2.1.3 Energy landscapes for an entire network

For a given state of genes, the energy of an entire network is calculated by *adding together the energy contributions of all genes in the network*. The energy landscape is then revealed by calculating the energies for the entire state space of the network. Now if we consider the network with genes X, Y, and Z, the entire state space of the network is

$$(X,Y,Z) = (0,0,0),\ (0,0,1),\ (0,1,0),\ (0,1,1),\ (1,0,0),\ (1,0,1),\ (1,1,0),\ (1,1,1)$$

There are eight distinct states for the network, and the energy landscape consists of these eight states with their respective energy. This is similar to what we have shown in figure 1, but with only eight discrete states instead.

As mentioned in the previous section, the energy landscape depends on the operator choices. In order to construct an energy landscape of a given GRN architecture, the operator space needs to be searched by CELLoGeNe. Otherwise, the operator choices should be specified. Once the operators have been assigned, CELLoGeNe will compute the energy landscape. In case the landscape obtained by a set of operator combinations does not agree with binarized experimental data, i.e., the attractors in the landscape do not match with known distinct cell states, CELLoGeNe can search through the operator space exhaustively or stochastically to find desired operator combinations. However, if no suitable operator combinations can be found by an exhaustive search, it suggests that the GRN topology should be further optimized.

When a landscape that matches experimental data is found, we can take further steps to analyze it. CELLoGeNe also provides a visualization tool to plot the landscape with more than three dimensions.

2.2 A GRN governing the process from DN1 to commitment

When T cell progenitors arrive at the thymus, they are exposed to Notch and then enter the T cell developmental program. The very first step of this program is the DN1/ETP - DN2a transition. The GRN that governs this transition consists of: (1) four positive regulators, (2) one antagonistic regulator (possibly PU.1), and (3) a function X (figure 3 (a)) [10].

The four positive regulators are Notch, TCF1 (Tcf7), Gata3, and Runx1. They are all required during the DN1/ETP - DN2a transition. Furthermore, they also control the entry from the DN2a state into the DN2b state, where Bcl11b is activated. However, only Runx1 and Notch are necessary for stable expression of Bcl11b [6].

Function X was originally defined as an epigenetic constraint on Bcl11b, and it is correlated with CD25 expression. In this work, we model only the transition to the DN2a state which does not involve Bcl11b, and therefore, the state of X can be used to identify DN2a cells (positive CD25 expression).

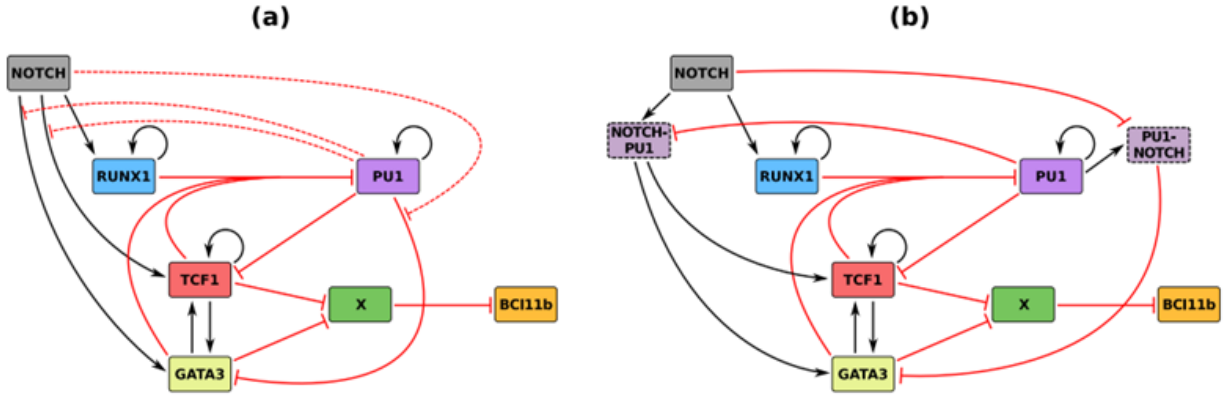

Figure 3: (a) The T-cell GRN considered in this work. The activation effect is depicted by ($\rightarrow$) and repression is depicted by ($\dashv$). Dashed lines show the special repressive effect of NOTCH / PU1 on the actions of PU1 / NOTCH respectively. Not like other interactions between genes, function X has an epigenetic repression effect on Bcl11b. (b) A modified network where the special repressive effects in (a) are replicated by adding normal nodes, in order to make it compatible with CELLoGeNe. The modified network should behave similarly to the original one with a proper operator choice.

2.3 Modified T cell network compatible with CELLoGeNe

In the original T cell network, repressions of the actions of genes are also included which are not standard in CELLoGeNe. To account for such interactions and make it compatible with CELLoGeNe, extra nodes need to be added and a particular operator UBOR (a repressive input always overcomes an activating input) needs to be assigned to their inputs (figure 3 (b)). With these modifications, the network behaves similarly to the original one and is also compatible with CELLoGeNe.

3 Energy landscapes for early T-cell development

In the first part of this work, we applied CELLoGeNe to the GRN governing early T-cell development. We tested the robustness of the GRN by examining the effect of different operator combinations (in other words, configurations) on the resulting energy landscapes. We also looked into the dynamics of systems within the energy landscape by stochastic simulations. The results were connected to the behaviors of realistic biological systems.

3.1 Construction of landscapes and analyses

3.1.1 Mapping the GRN to energy landscapes

With the GRN ready to be mapped to an energy landscape, we need to search or specify the logical operators that are used to combine multiple inputs to one gene. For the network used in this work, there is a large number of possible operator combinations, which can give rise to too many possible landscapes. To reduce the number of landscapes and find landscapes that agree with experimental data, we imposed some constraints.

The T-cell GRN being studied in this work governs the process where progenitor T cells (proT-cells) evolve from the DN1/ETP stage to the DN2a stage. Therefore, we consider the gene expression states at both stages to be stable (minima, attractors in the landscapes). For the DN1/ETP and DN2a stages, the gene activation states (NOTCH, PU1, NOTCH-PU1, PU1-NOTCH, TCF1, GATA3, RUNX1, X) are:

$$\text{DN1/ETP: } (0, 1, 0, 1, 0, 0, 0, 1)$$

$$\text{DN2a: } (1, 0, 1, 0, 1, 1, 1, 0)$$

By setting the constraints that both the DN1/ETP state and the DN2a state are attractors in the energy landscape, CELLoGeNe can search for suitable logical operator combinations (suitable configurations). Furthermore, for simplicity, the allowed operators in the search can be specified as long as we can find desired landscapes. For example, we can search for configurations that only use AND and UBOR.

3.1.2 Stochastic probe of the energy landscapes

At the next step, we conducted a stochastic simulation that corresponds to the marble metaphor of cell fate decisions in the energy landscape. In the simulations, we view the system as a real marble, and it is able to roll in the landscape and take different pathways to settle at a final state. From now on, we refer to such simulations as 'marble' simulations (see Appendix B).

The barriers between different attractors can be overcome with perturbations. If we view the landscape in a biological sense, the attractors in the landscape resemble distinct cellular states and the transitions between attractors represent the cell development process. For example, we can study the DN1/ETP - DN2a transition by the following procedure:

(i) Initializing the system at the DN1/ETP state.

(ii) Following the path of the system in the presence of noise.

(iii) Checking the destination of the system in the landscape.

3.2 Results and discussion

3.2.1 Robustness of the network

We found it not necessary to use too many different logical operators to find a landscape that has both the DN1/ETP and DN2a states as attractors. In fact, with only two or three allowed operators, it is already rather easy to find suitable landscapes, indicating that it is a robust feature of the network itself.

Furthermore, for a fixed set of allowed operators, no matter how the operators are combined or assigned to the inputs, DN1/ETP and DN2a are likely to be attractors. When UBAND and UBOR are the two allowed operators, out of 32768 possible configurations, 32128 of them have the DN2a state as an attractor, and 8024 have the DN1/ETP state as an attractor. This also indicates that the network is robust, especially in terms of maintaining the DN2a state as an attractor.

3.2.2 Visualization and the structure of the landscape

In order to have a closer look at the landscape, we picked one valid landscape obtained by allowing UBOR and UBAND operators. In this landscape, on top of the attractors specified by the constraints, there is a new degenerate state of DN1/ETP (table 1). This comes from the fact that function X only takes inputs from TCF1 and GATA3. However, at the DN1/ETP state, both TCF1 and GATA3 are not expressed and therefore function X contributes the same amount of energy to the network no matter if it is ON or OFF. The two resulting states with the same energy are neighbors and can be treated as one attractor.

The last column in table 1 shows the specific configuration to obtain the landscape. Also, note that the exact same landscape can be obtained with a slightly different configuration.

Table 1: The constraints and the resulting attractors from the configuration in the last column. A filled table cell corresponds to a gene being ON (1), and a blank cell means OFF (0).

	Constraints		CELLoGeNe results		INPUT (denoted as initial letters)
	DN1/ETP	DN2a	DN1/ETP	DN2a	
NOTCH					None
PU1					(T UBAND P) UBOR (R UBOR G)
NOTCH_PU1					N UBOR P
PU1_NOTCH					N UBOR P
TCF1					(G UBAND P) UBOR (T UBOR NP)
GATA3					(PN UBAND NP) UBOR T
RUNX1					R UBOR N
X					G UBAND T

Visualization of the landscape

The GRN has eight nodes, and the resulting full landscape has eight dimensions ($2^8 = 256$ discrete states) which is large and chaotic to plot in 2D. However, by fixing some of the nodes to be ON or OFF, we can still extract some features of the landscape. To observe the DN1/ETP state, we set the NOTCH and NOTCH-PU1 to OFF due to the experimental results that Notch signaling is needed for T-lineage commitment [5]. While on the other hand, to observe the DN2a state, we set NOTCH and NOTCH-PU1 to be ON.

Figure 4(a) and (b) show the hyperplanes containing the DN1/ETP and DN2a attractors respectively for the picked landscape. The color map shows the energy scale of the landscape, and the legends show the correspondence between genes and the pies in the plot. The legends also show which nodes are fixed to what values. The nodes are denoted by their initial letters. A filled slice of pie corresponds to the gene being ON, while a blank slice corresponds to a gene being OFF.

In figure 4 (a), we observe the DN1/ETP state, and its degenerate state is on the top left of the node. They have lower energy compared to their neighboring states (arrows all pointing towards them). The DN2a state has the lowest energy in the plotted hyperplane and also is an attractor (figure 4 (b)).

The DN1/ETP and DN2a states are both attractors in the landscape, but the energy of the DN2a state is lower. In such a landscape, if we initialize the system at the DN1/ETP state, in the presence of some noise, we can expect that the system will overcome the barrier between the DN1/ETP state and the DN2a state, and settle down at the DN2a state where the energy is lower. This is in line with the fact that in a natural situation, cells develop from the DN1/ETP state to the DN2a state, but not the other way around.

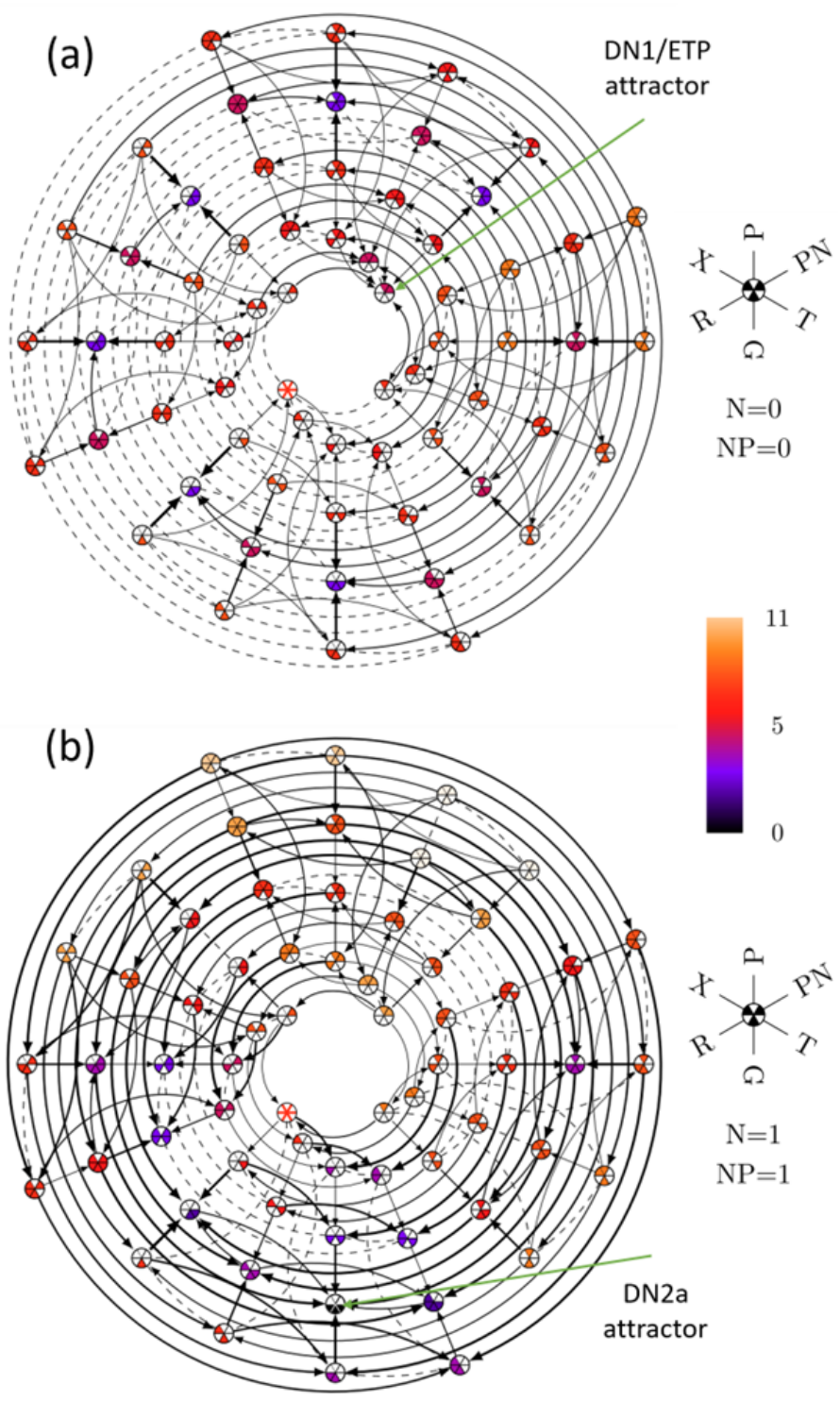

Figure 4: Hyperplanes of a landscape resulting from a picked configuration with two allowed operators, UBOR and UBAND. (a) contains the DN1/ETP attractor and (b) contains the DN2a attractor. A filled slice of pie corresponds to a gene being ON (1), and a blank slice means OFF (0).

3.2.3 Marble simulations

We probed the landscape structure by starting the system in the landscape with different initial states and controlling the noise level. The systems stop evolving after staying at the same state in the landscape for three consecutive steps. Figure 5-7 show simulation results where several noise levels were considered. A larger value of β means a lower level of noise. Plot (a) in figure 5-7 depicts the probability for the system to end up at the DN1/ETP state (orange), the DN2a state (blue), and other states (gray), when the system started from all possible states for the network. Plot (b) and (c) in figure 5-7 show the results when the DN1/ETP and DN2a states were initial states, respectively.

In figure 5, β is 3 which means that the noise level is quite low. No matter where the system started, it was very likely to end up at the DN2a state (plot (a)). For a small portion of simulations, the system was more likely to end up at the DN1/ETP state. Those were the simulations with an initial state close to the DN1/ETP state. At a low noise level, we found the system to be very stable at the DN2a state: when we initialized the system at the DN2a state, it stayed there. For systems starting at the DN1/ETP state, there was already a chance for it to transition to the DN2a state.

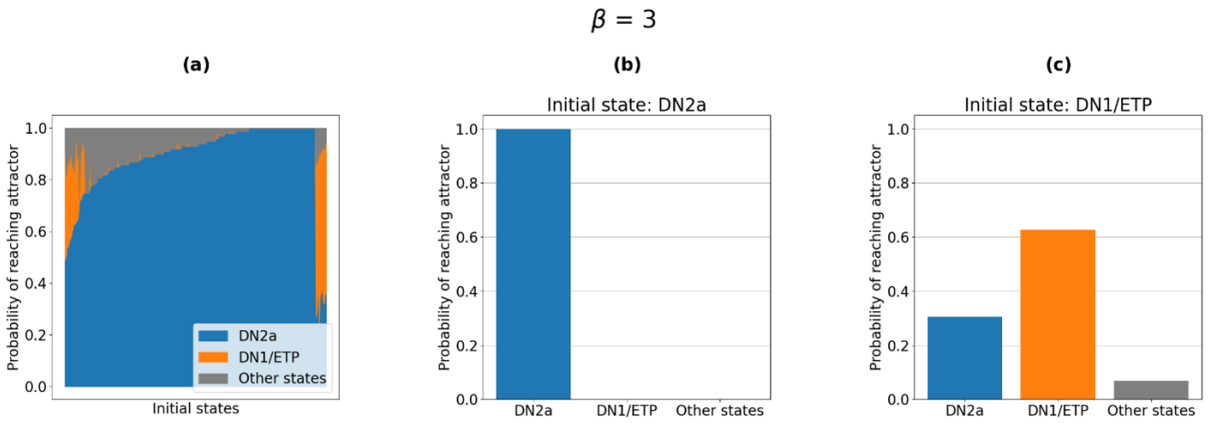

Figure 5: Marble simulations with a low level of noise. (a) The system was very likely to end up at the DN2a state. (b) Systems starting from the DN2a state stayed at the DN2a state. (c) DN1/ETP - DN2a transitions could happen, but the probability was only around 30%.

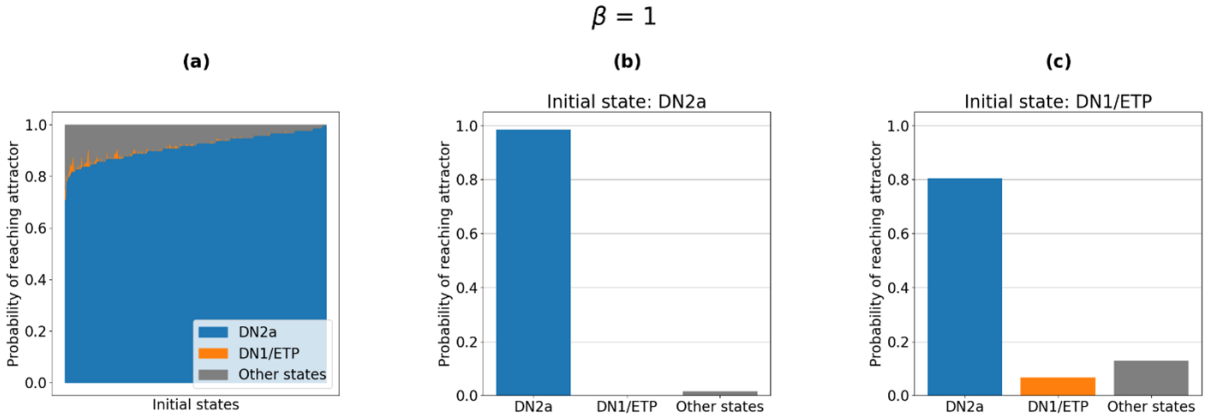

Figure 6: Marble simulations with a medium level of noise. (a) The system was very likely to end up at the DN2a state. (b) Systems starting from the DN2a state almost always stayed at the DN2a state. (c) DN1/ETP - DN2a transitions could happen, and the probability was around 80%.

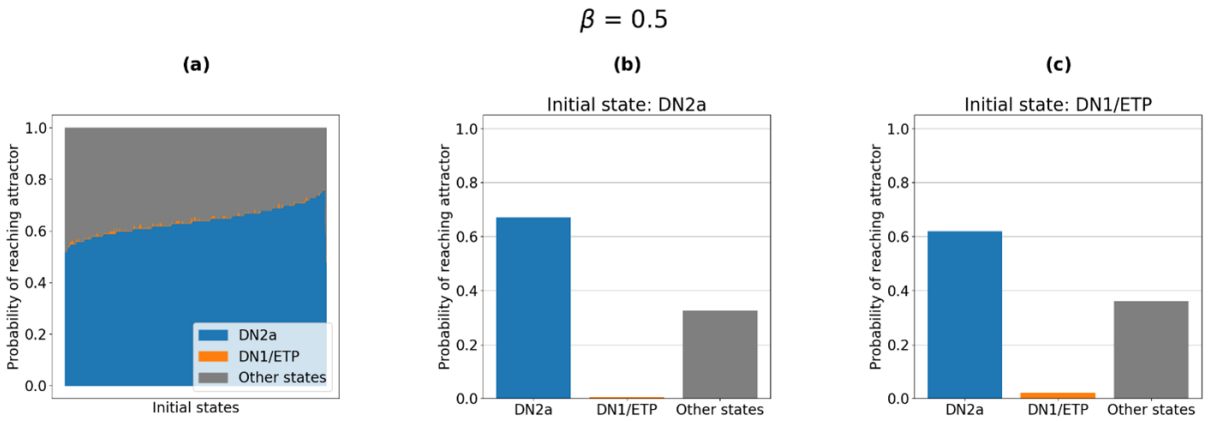

Figure 7: Marble simulations with a high level of noise. (a) The system was likely to end up at the DN2a state. (b) Systems starting from the DN2a state mostly stayed at DN2a states. (c) DN1/ETP - DN2a transitions could happen, and the probability was around 60%.

When we increased the noise level, transitions between different states happened more often, and the system would seldom end up at the DN1/ETP state no matter where it started (figure 6 and figure 7). When the noise level was high enough, the attractors started to lose the ability to trap the system, and the system ended up at other states with a probability of around 40% (figure 7).

The stochastic simulations further connect the landscape for the T cell GRN with real cell differentiation process. For a wide range of noise levels, the system tends to evolve towards the DN2a state regardless of where it starts, and the DN1/ETP - DN2a transition in the landscape does resemble part of the process in T cell development.

4 Gene knockdown effect on the landscapes

In the second part of this work, we looked into the gene knockdown effect on the landscapes with regard to three aspects: (1) likelihood of having a landscape with the DN1/ETP state or the DN2a state as an attractor in terms of configuration choices, (2) predicted attractors in the landscapes, and (3) the probability of DN1/ETP - DN2a transitions.

4.1 Gene knockdown

Gene knockdown is an experimental technique used to reduce the expression level of genes. This technique has been used to study the effects of gene expression levels on cells at different developmental stages [6].

It is also possible to approximate the gene knockdown scenario in CELLoGeNe by excluding the connections to the specified gene or genes in the network. We investigated the new landscapes when gene knockdown is applied in a similar fashion as it was done in section 3. We then compared the results from landscapes after gene knockdown to the results from the original landscape.

4.2 Results and discussion

4.2.1 Knockdown scenarios tested on different sets of operators

When gene knockdown is applied to the network, the connections between nodes in the GRN are reduced. Therefore, there will be fewer possible configurations for the same set of allowed operators. We tested three sets of allowed operators each in four scenarios: without gene knockdown, RUNX1 knockdown, GATA3 knockdown, and TCF1 knockdown. Since TCF1 has the most connections in the network, TCF1 knockdown results in the fewest possible configurations (table 2). Also, allowing three operators will significantly increase the number of configurations.

In all the scenarios, despite the gene knockdown, it was still always possible to find a configuration that has both the DN1/ETP state and the DN2a state act as attractors. However, the proportion of configurations where the DN1/ETP state or the DN2a state is an attractor varied.

In all the cases, the knockdown of either GATA3 or TCF1 increased the chance to have a configuration that makes the DN1/ETP state an attractor, but at the same time decreased the chance to have configurations with the DN2a state as an attractor. This indicates that the knockdown of GATA3 or TCF1 leads to the DN1/ETP state being ‘more of an attractor’ and the DN2a state being ‘less of an attractor’. In such a situation, if we assume the systems to all start at the DN1/ETP state in a landscape, we can expect to have more

Table 2: Four knockdown scenarios with three different sets of operators each. The proportions of configurations that make the DN1/ETP and DN2a states attractors are listed respectively.

Allowed operators	Knockdown gene	Total configurations	DN1 is an attractor (Percentage)	DN2a is an attractor (Percentage)
UBOR UBAND	-	32768	8024 (24.49%)	32128 (98.04%)
	RUNX1	16384	4011 (24.48%)	15423 (94.13%)
	GATA3	4096	2175 (53.10%)	3543 (86.50%)
	TCF1	1024	543 (52.02%)	831 (81.15%)
UBOR OBAND	-	32768	6756 (20.62%)	22128 (67.53%)
	RUNX1	16384	4247 (25.92%)	11087 (67.67%)
	GATA3	4096	1791 (43.72%)	2503 (61.11%)
	TCF1	1024	459 (44.82%)	607 (59.28%)
UBOR UBAND OBOR	-	4084101	1787544 (43.77%)	3639129 (89.10%)
	RUNX1	1361367	595847 (43.77%)	1150786 (84.53%)
	GATA3	194481	110396 (56.76%)	155650 (80.03%)
	TCF1	27783	15770 (56.76%)	18332 (65.98%)

systems staying at the DN1/ETP state and fewer transitions to the DN2a state when GATA3 or TCF1 knockdown is applied. This agrees with the experimental results [6].

RUNX1 knockdown did not seem to have a consistent impact on the proportion of configurations with the DN1/ETP state or the DN2a state being an attractor, but the impact was also less significant than that of GATA3 or TCF1 knockdown. This may come from the fact that RUNX1 has fewer connections in the network than both GATA3 and TCF1. We need to further investigate the effect of RUNX1 knockdown. It is better to look closer into it on specific landscapes.

4.2.2 Attractors in landscapes after gene knockdown

As previously mentioned, after gene knockdown, it was still possible to find landscapes that fulfill the constraints of both the DN1/ETP state and the DN2a state being attractors. However, for some configurations, we found a third attractor in the landscape. Table 3 shows two valid landscapes obtained separately from GATA3 knockdown and RUNX1 knockdown allowing UBOR and UBAND operators.

Table 3: The constraints and resulting attractors for GATA3 knockdown and RUNX1 knockdown scenarios. UBOR and UBAND are the two allowed operators. There exists a third attractor for both cases. A filled table cell corresponds to a gene being ON (1), and a blank cell means OFF (0).

GATA3 Knockdown	Constraints		CELLoGeNe results				
	DN1/ETP	DN2a	DN1/ETP		DN2a	Attractor 3	
NOTCH							
PU1							
NOTCH_PU1							
PU1_NOTCH							
TCF1							
GATA3							
RUNX1							
X							

RUNX1 Knockdown	Constraints		CELLoGeNe results				
	DN1/ETP	DN2a	DN1/ETP		DN2a	Attractor 3	
NOTCH							
PU1							
NOTCH_PU1							
PU1_NOTCH							
TCF1							
GATA3							
RUNX1							
X							

4.2.3 Marble simulations

Effect of a third attractor

We studied the effect of the possible third attractor in the landscape first. It turned out that the third attractor would only mildly impede the transition from the DN1/ETP state to the DN2a state. Figure 8 shows the results from marble simulations in a valid landscape (RUNX1 knockdown) obtained by allowing UBAND and UBOR operators.

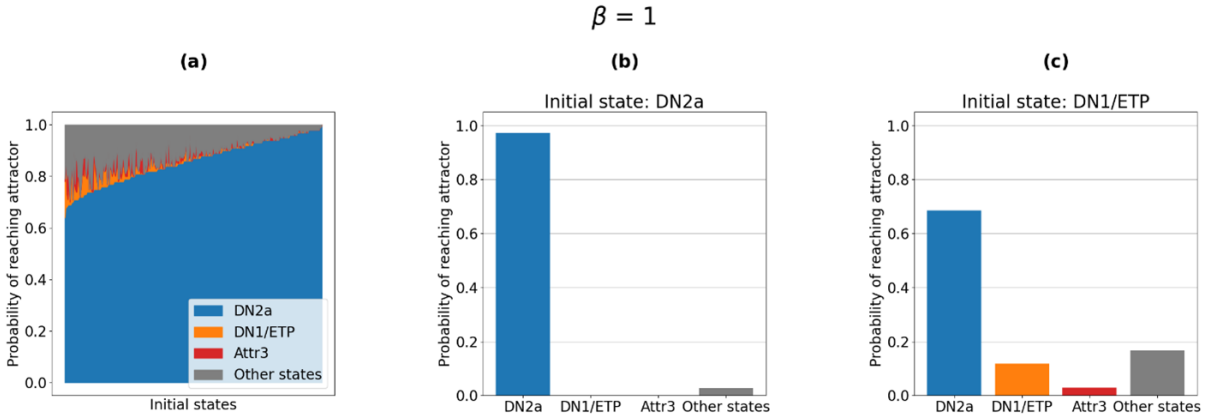

Figure 8: Marble simulations with a medium level of noise in a landscape after RUNX1 knockdown. (a) The system had a small chance to end up at the third attractor (red) depending on the initial state. (b) Systems starting from the DN2a state almost always stayed at the DN2a state. (c) A small portion of systems starting from the DN1/ETP state ended up at the third attractor, but the DN1/ETP - DN2a transition was still more common.

Even after gene knockdown, it was still possible to find landscapes that fulfill the constraints. We found an additional attractor for some landscapes, but it did not have a strong impact on the marble simulations in the landscape. In the presence of a medium amount of noise, regardless of where it started in the landscape, the system was most likely to end up at the DN2a state. The fact that even after gene knockdown, the DN2a attractor can still be maintained also shows that the network is robust.

Gene knockdown effect on DN1/ETP - DN2a transition

None of the TCF1 knockdown, GATA3 knockdown, or RUNX1 knockdown changed the landscape completely. The DN2a state had the lowest energy and stayed to be the most favorable state in the landscape. However, gene knockdown did affect the probability of the DN1/ETP - DN2a transition. We carried out marble simulations for the three knockdown scenarios and compared the results to those from a landscape without gene knockdown. All the landscapes were from configurations only allowing two operators (UBAND and UBOR).

First, we initialized the systems at the DN1/ETP state, and recorded the probability of ending up at the DN1/ETP state. Figure 9 shows the simulation results in the presence of various levels of noise. The transition probability for each scenario with each β value was obtained from 10000 simulations.

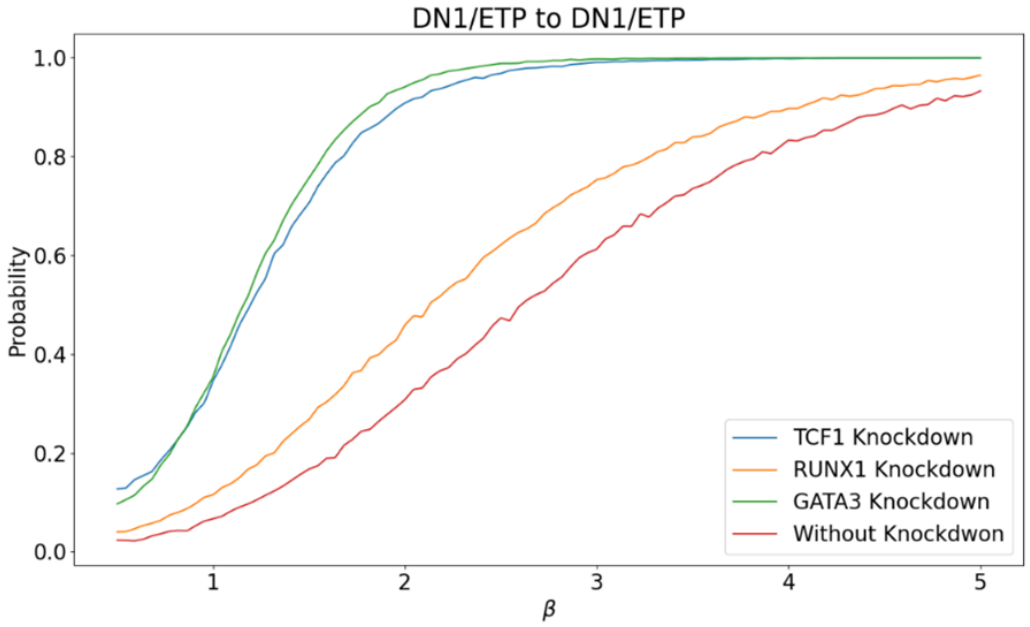

Figure 9: Probability of ending up at the DN1/ETP state when the system was initialized at the DN1/ETP state. β varies from 0.5 (large noise) to 5 (small noise). Landscapes are indicated by the legend.

In a wide range of noise levels, of all simulated scenarios, the landscape obtained from the network without knockdown showed the lowest probability of staying at the DN1/ETP state. This means that the system is more likely to leave the DN1/ETP attractor leading to potential transitions to the DN2a state.

Knocking down either TCF1 or GATA3 increased the probability for the landscape to trap the system at the DN1/ETP state.

RUNX1 knockdown did not seem to affect the chance of having a configuration with the DN1/ETP state as an attractor in a consistent way. In fact, in the case where UBAND and UBOR were the two allowed operators and when RUNX1 knockdown was applied, the chance of the DN1/ETP state being an attractor even decreased by a tiny amount compared to the network without any knockdown (table 2). Yet, on the other hand, the resulting landscape with RUNX1 knockdown was more likely to trap the system at the starting DN1/ETP state than a landscape without any knockdown (figure 9).

Second, we studied the probability of the transition from the DN1/ETP state to the DN2a state. Figure 10 shows the simulation results in the presence of various levels of noise.

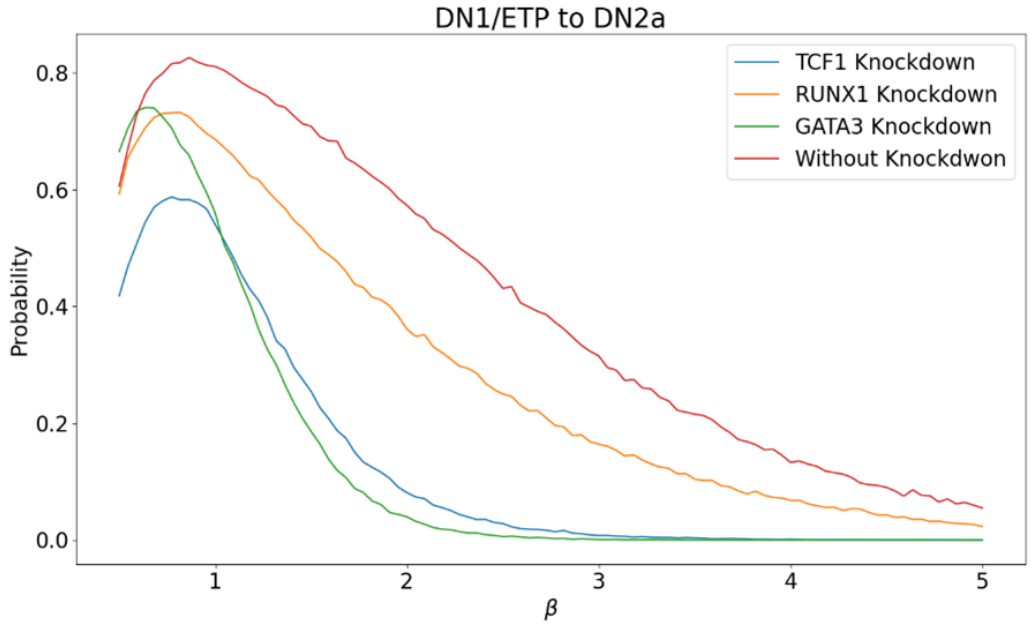

Figure 10: Probability of ending up at the DN2a state when the system was initialized at the DN1/ETP state. β varies from 0.5 (large noise) to 5 (small noise). Landscapes are indicated by the legend.

For all four scenarios, as we were reducing the noise level, the probability of DN1/ETP - DN2a transition increased and peaked at somewhere where β was below 1. In this range, the system had a high chance of getting out of the basin of the DN1/ETP attractor, and the decrease in noise helped the system settle down at the DN2a state rather than some random states in the landscape. As we further reduced the noise, the system started to get trapped in the DN1/ETP state, and hence the probability of DN1/ETP - DN2a transition started to decrease.

In the presence of large noise (β close to 0.5), RUNX1 knockdown and GATA3 knockdown shared similar results with the original landscape, while knocking down TCF1 reduced the DN1/ETP - DN2a transition probability by about 20%. When the level of noise was reduced (β larger than 1), the landscape obtained from the network without gene knockdown started to give the highest probability of DN1/ETP - DN2a transitions. This higher transition probability agrees with the experimental results [6].

In both figure 9 and figure 10, we noticed that in general, the landscape when RUNX1 knockdown was applied gave results closer to the landscape without any knockdown. This may also come from the fact that RUNX1 has the fewest connections in the network among GATA3, TCF1, and RUNX1.

The marble simulations were able to make a prediction of the probability of DN1/ETP - DN2a transition for different scenarios with or without gene knockdown. The results suggest that knocking down any one of RUNX1, GATA3 or TCF1 will impede the transition from the DN1/ETP state to the DN2a state in the process of T cell development.

5 Cell development and proliferation

When DN1/ETP cells are cultured in an experimental condition with Notch signaling, the cells not only go through the T cell developmental pathway, but also proliferate. Therefore, it can help better understand the network and the dynamics of the system to simulate the cell development process while also considering cell divisions in the landscapes.

In the third part of this work, we made an attempt to model cell divisions in the landscape and investigated the time evolution of cell distributions. For this purpose, we first needed to find how to identify cell states and how to implement time in the landscape.

5.1 Cell development and divisions in the energy landscapes

5.1.1 The state of the cells

Since our goal is to study the T cell development process from the DN1/ETP state to the DN2a state while also considering cell divisions, it may seem reasonable to initialize the cells at the DN1/ETP state and investigate the probability of DN1/ETP - DN2a transitions, like in previous marble simulations. However, as we showed in section 3.2.3, if a system was initialized at the DN1/ETP state, regardless of the noise level, the probability of a system ending up at a state other than DN2a was always larger than 20%. In reality, if Notch signaling is present, a DN1/ETP cell is much more likely to enter the DN2a state eventually.

At first glance, it seems that the problem can be solved by letting the cell state evolve in the landscape without ever stopping it. The issue is that the cell will not always stay at the DN2a state after it reaches there for the first time, but instead it has a chance to go to a neighboring state and hop around. The cell then will have a probability p of being at the DN2a state. For a fixed noise level, this probability will approach a fixed value α if the cell has been given enough time to evolve in the landscape.

Therefore, one way to tell if the cell is a DN2a cell could be by checking p and comparing it to α. An alternative approach could be freezing the cell when and only when it stays at the DN2a state for some consecutive steps. After that, the cell is considered to stay as a DN2a cell. We will discuss later that the second approach represents the real cell development process better.

5.1.2 Time in the landscape

Time is essential for the implementation of cell divisions in the landscape. Since the energy landscapes are mapped from a binarized GRN, naturally the energy is a discrete function of the discrete gene states. A unit step inside the landscape is one gene state being turned ON or OFF. The simplest way to treat time is then by connecting it to the number of unit

steps in the landscape. For example, a completely random guess could be that one hour in real time corresponds to 10 unit steps in the landscape. In this manner, we can connect the time elapsed in the landscape directly to the steps a cell has taken in the landscape.

5.1.3 Cell divisions

In this work, we assume the cell divisions to be symmetric, and the daughter cells share the same content as the mother cell, but the cell cycle length for each cell is independent. Each cell also evolves independently in the landscape (each cell has its own gene states).

For simplicity, we implemented the cell cycle length in a way that it only depends on the cell generation and the cells of higher generations proliferate faster. We will later show that with such implementation, cell proliferation does not affect the distribution of DN1/ETP cells and DN2a cells.

5.2 Results and discussion

We carried out all the simulations in a landscape obtained by allowing two operators (UBAND and UBOR).

5.2.1 Noise level and transition time

Probability of a cell being at the DN2a state - time evolution

As described in section 5.1.1, if the cells are released at the DN1/ETP state and keep evolving in the landscape, we expect that the probability of the cells being at the DN2a state should approach α for a fixed noise level. We investigated the time it takes for p to reach α for different noise levels. Figure 11 shows the results when β is 0.5, 1, 2 and 3.

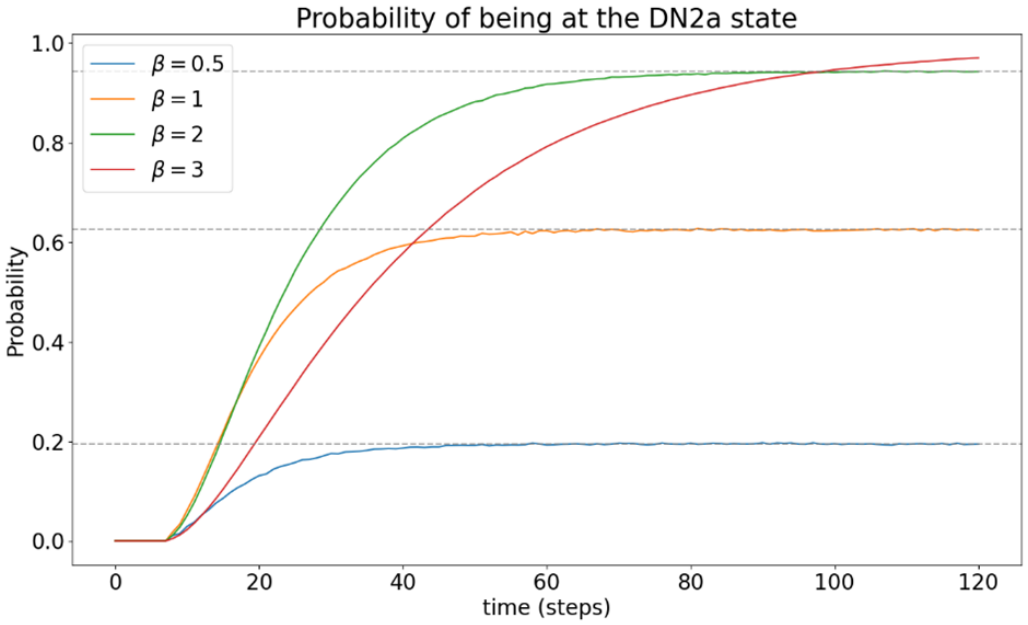

Figure 11: Probability of the cell being at the DN2a state (p) depending on time. The dashed lines from the bottom to the top mark the α for $\beta = 0.5$, 1, and 2 respectively. When $\beta = 0.5$, it took the least time for the probability to reach α, but α was also the lowest. When $\beta = 3$, the probability of the cell being at the DN2a state was yet to reach α at step 120. The curve ($\beta = 3$) suggests that it will eventually reach α which will have the largest value of all four plotted scenarios.

As we showed in the results, the increase in the noise level decreased the time it took for p to reach α. If the cell is considered to be a DN2a cell when p has reached α, such results suggest that the increase in the noise level only makes the transition faster.

There are two crucial issues if we decide the cell state in this manner. First, biologically, some perturbations from either the intrinsic noise in the network or variations in external signals are needed to initiate the transition process. An initial increase in the noise level helps speed up the transition. However, when the noise is increased to an extremely high level, we should not expect that the transition happens even faster. Second, when β was equal to 0.5, α was merely about 20%. It indicates that the cell was not even at a state close to the DN2a state all the time, and it makes little sense to consider such cells to be DN2a cells.

Stopping the cell development when it stays at the DN2a state

The previous approach did not capture the cell development process very accurately. Now, the cell is considered to be a DN2a cell and stops evolving only if it stays at the DN2a

state for a limited number of steps, otherwise, it keeps evolving in the landscape. We investigated the effect of noise levels on the time it takes for DN1/ETP - DN2a transitions by simulations with $\beta = 1, 2$, and 3. Figure 12 (a) shows the probability of a cell initialized at the DN1/ETP state becoming a DN2a cell at step N, and figure 12 (b) shows the cumulative sum. The cell had the highest chance of becoming a DN2a cell at around 25 steps for all three levels of noise.

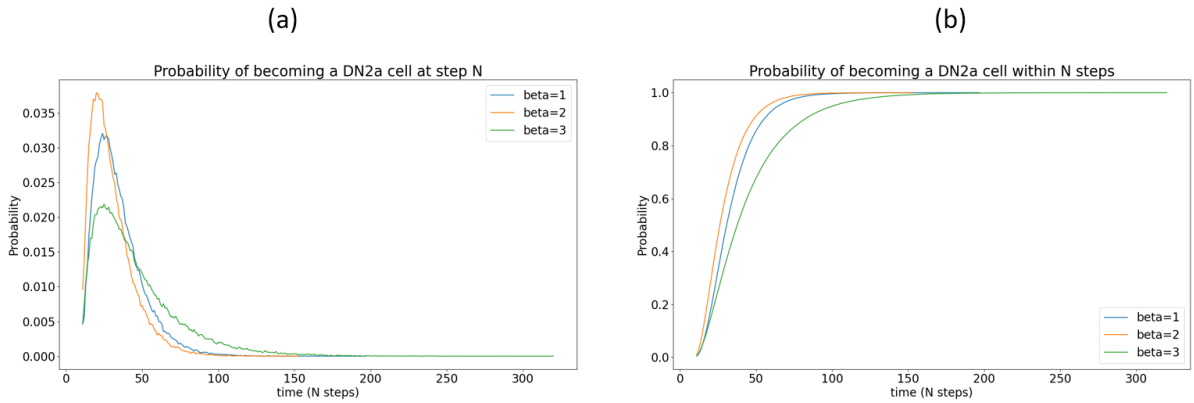

Figure 12: The probability of a cell initialized at the DN1/ETP state becoming a DN2a cell. Three β values (1, 2, and 3) were considered. (a) Probability of becoming a DN2a cell at individual steps. (b) Probability of becoming a DN2a cell within N steps (cumulative sum).

For the three noise levels considered, the highest level of noise no longer gave the shortest transition time to DN2a cells. Instead, it took the least amount of time for the cell to become a DN2a cell when β was equal to 2. The transition from a DN1/ETP cell to a DN2a cell was slower when the noise was either too low or too high, as we would expect for a realistic biological system. It is even more clear in figure 13 where we show the steps it took for the cell to have a 50% chance to become a DN2a cell. In the simulations, we also varied the steps needed to stop the cell from further development when it stays at the DN2a state.

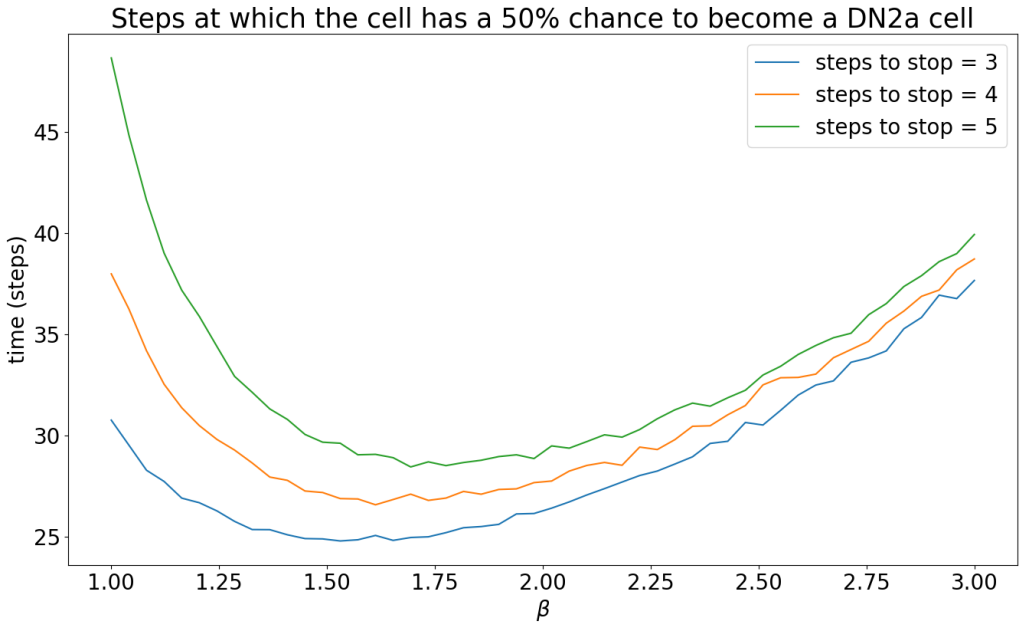

Figure 13: Steps needed for the cell to have a 50% chance to become a DN2a cell depending on β. The legend indicates the steps needed to stop the cell from further development when it stays at the DN2a state.

In the case when the cell development is stopped after the cell staying at the DN2a state for 3 consecutive steps, for it to have a 50% chance of becoming a DN2a cell, at least about 25 steps were needed. The least steps came from a β between 1.25 and 1.75.

Also, increasing the steps needed to stop the cell slightly slowed down the DN1/ETP - DN2a transition, as it took more steps for the cell to settle down at the DN2a state.

5.2.2 Estimation of time elapsed in the energy landscape

Figure 12 (b) shows the probability of a cell initialized at the DN1/ETP state becoming a DN2a cell within N steps. We can also interpret that it shows the proportion of DN2a cells if multiple cells are released from the DN1/ETP state. In the experiment, bulk DN1 cultures demonstrate similar overall dynamics in terms of cell distributions. Therefore, by comparing with the experimental data, it is possible to estimate the relationship between steps in the landscape and real time.

Bulk DN1 cells were cultured and harvested each day from day 2 to day 5, and the percentages of DN2a cells were measured [10]. We fitted such results to the curve of DN2a cell proportion obtained from a landscape with $\beta = 1$ (figure 14).

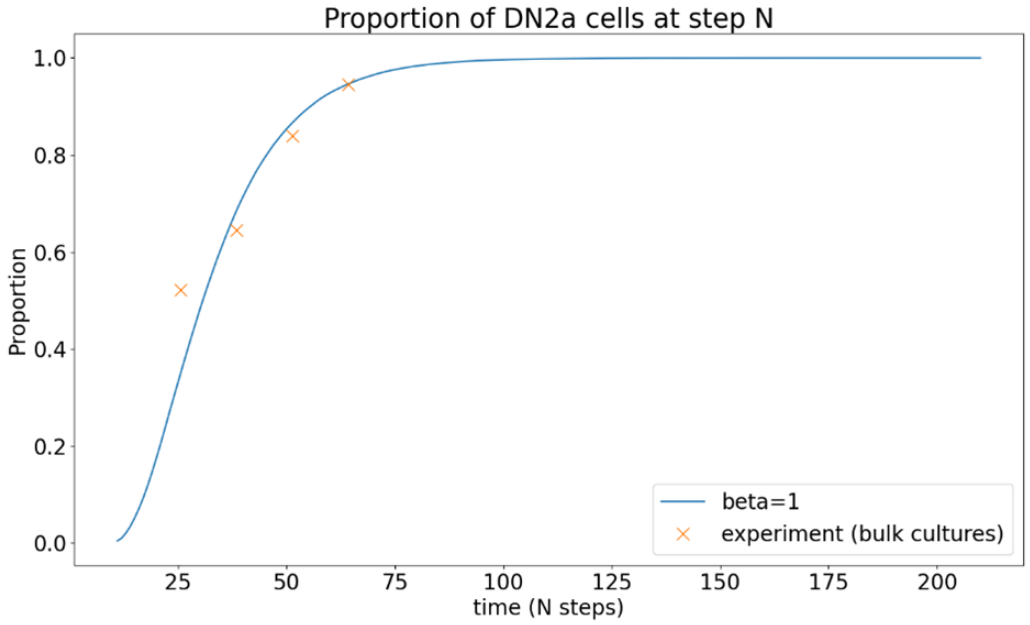

Figure 14: DN2a cell proportion of bulk DN1 cultures fitted to the results from landscape prediction.

The leftmost 'x' mark in the plot represents day 2 of the experiment, and it corresponds to about 25.67 steps in the landscape. For the period of one day (24 hours), we estimated the cells to be able to take about 12.83 steps in the energy landscape.

It can merely be an estimation because the noise level will also affect the dynamics of the system and therefore change the estimated relationship between real time and steps in the landscape. We chose the β value to be 1 here because it resulted in the highest DN1/ETP - DN2a transition probability in marble simulations (figure 10) and could be closer to reality.

5.2.3 Cell development accompanying proliferation

Once the time is defined by steps taken in the landscape, we can convert the cell cycle length from time to steps in the landscape.

In the beginning, one single cell (generation 0) is released at the DN1/ETP state, and its cell cycle length is chosen from a normal distribution. When the cell divides, two daughter cells are created at the same state as their mother cell and their cell cycle lengths are chosen from another normal distribution independently. We used normal distributions with different means and variances to decide the cell cycle length depending on the generations. Generally, a cell of a higher generation divides faster. We compared the DN2a

cell proportions when cell proliferation is considered to the ones without cell proliferation (figure 15).

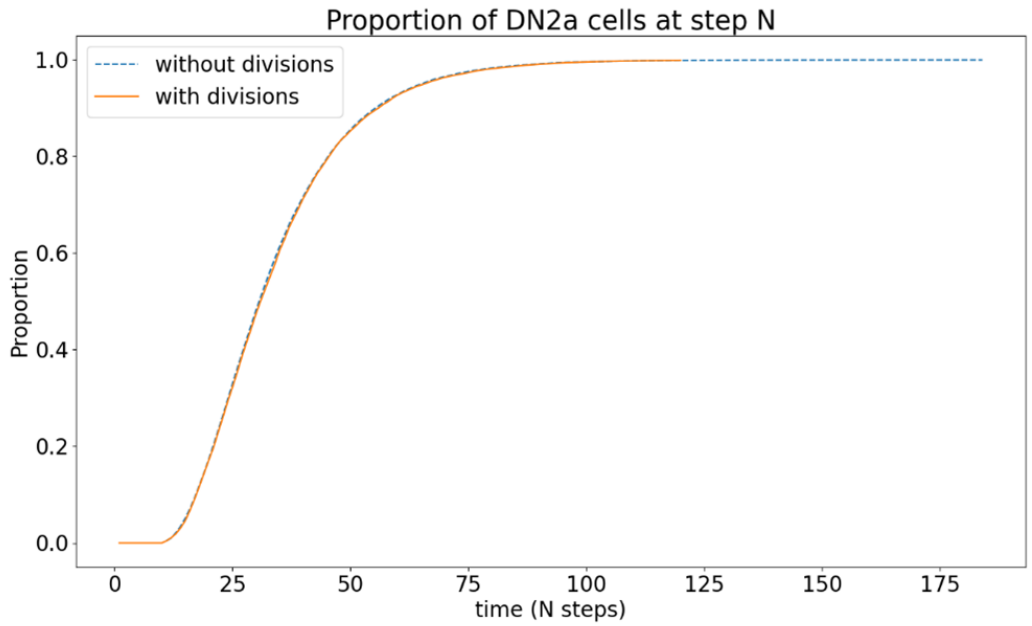

Figure 15: Proportion of DN2a cells depending on time. The orange line and the dashed line show the results with and without considering cell divisions, respectively. The two lines were almost identical.

Cell divisions did not affect the proportion of DN2a cells for the particular implementation. We believe the reason is as follows. The cell cycle length only depended on the cell generation. The cell generation had nothing to do with the cell state at all, which means that the DN1/ETP cells and DN2a cells would have the exact same generation profile. Even though at a later time, most of the population were DN2a cells, as long as DN1/ETP cells and DN2a cells divide at the same rate, the proportion of them stays unaffected.

However, in a more realistic situation, the cell cycle is more complicated, and faster differentiating cells can feed back on other cells by the exchange of proteins. In order to build models that resemble cell proliferation better, further study regarding the cell cycle is needed.

6 Conclusion and outlook

Mapping the GRN of T cell development to energy landscapes

We mapped a real-world gene regulatory network that describes T-cell development from the DN1/ETP state to the DN2a state into energy landscapes. We showed that the network is robust to small variations in gene interactions within the network, i.e., the favorable states of the network model do not depend much on how the inputs from multiple genes to one gene are combined.

When both the DN1/ETP and DN2a states act as attractors in the landscape, the landscape provided further predictions that the DN2a state possesses the least amount of energy in the whole landscape. The transition from the DN1/ETP state to the DN2a state is much more likely than the opposite transition for a wide range of noise levels. Such predictions are in line with experimental observations.

Gene knockdown effect

We showed that the network is robust under the effect of gene knockdowns. The DN1/ETP and DN2a states can still act as attractors in the landscape after gene knockdown. In fact, when GATA3 knockdown or TCF1 knockdown was applied to the network, the DN1/ETP state not only became more likely to act as an attractor but could also become a stronger attractor in terms of trapping the system. On the other hand, RUNX1 knockdown did not have a consistent impact on the likelihood of the DN1/ETP or DN2a state being an attractor, but it did enhance the strength of the DN1/ETP attractor.

In some cases, we found a third attractor in the landscape when GATA3 or RUNX1 knockdown was applied. The potential third attractor mildly impeded DN1/ETP - DN2a transitions in the landscape.

We evaluated the combined effects of the altered landscapes by stochastic simulations and compared the results to the ones from the original landscape. Knocking down any of the RUNX1, GATA3, or TCF1 reduced the probability of the cell transitioning from the DN1/ETP state to the DN2a state, which has also been observed in experiments.

Cell development and proliferation in the landscape

In the last part, we made an attempt to implement the process of cell development and proliferation in the landscape. We showed that it is possible to model cell development in a discrete energy landscape. With the correct implementation, the effect of noise levels on DN1/ETP - DN2a transition time was similar to what would be expected for a realistic biological system. Either an extremely high or low level of noise slows down the transition.

The resulting distribution of DN2a cells from landscape simulations resembles that of bulk DN1 cultures. Even though the relationship between real time and steps in the landscape can be estimated by comparing DN2a cell distributions, we should not forget that such estimation may vary depending on the noise level.

After time in the landscape was defined, we implemented cell divisions in the landscape. Cell divisions based on the assumption that the cell cycle length completely depends on cell generations turned out to not affect DN2a cell distributions at all. It suggests that to include cell proliferation in a discrete energy landscape, it may require modeling the cell cycle length based on the cell states.

Outlook: continuous energy instead of discrete energy

The discrete energy landscapes mapped from the T cell development GRN have been able to give great predictions on probabilities of finding the cells in a certain state and the DN1/ETP - DN2a transition. It is also effective to study the knockdown effect of certain genes.

However, when time is involved in the studies, the landscape method may not perform as well due to its discretized energy. For a discrete energy landscape, the steps in the landscape are restricted to one gene being turned ON or OFF, which may give rise to some issues. One issue is that the finite size of the gene state space depends on the size of the network. Thus, the distance between two states in the landscape will depend on the number of genes in the network (even though some genes may not have any effect on a certain process). Another issue is that in a discrete energy landscape, the energy is used to decide the probability of the system transitioning to a neighboring state, but the transition speed is not defined. The assumption that each step in the landscape takes the same amount of time may not be valid.

Therefore, a continuous energy landscape could be helpful when time is involved. We expect to carry out further studies using a tool in CELLoGeNe that is capable of constructing continuous energy landscapes by interpolation.

References

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Introduction to pathogens*. Garland Science, 2002.

[2] T. W. LeBien and T. F. Tedder, "B lymphocytes: how they develop and function," *Blood, The Journal of the American Society of Hematology*, vol. 112, no. 5, pp. 1570–1580, 2008.

[3] E. L. Reinherz and S. F. Schlossman, "The differentiation and function of human t lymphocytes," 1980.

[4] D. A. Zlotoff and A. Bhandoola, "Hematopoietic progenitor migration to the adult thymus," *Year in Immunology*, vol. 1217, pp. 122–138, 2011.

[5] M. A. Yui, N. Feng, and E. V. Rothenberg, "Fine-scale staging of t cell lineage commitment in adult mouse thymus," *Journal of Immunology*, vol. 185, no. 1, pp. 284–293, 2010.

[6] H. Y. Kueh, M. A. Yui, K. K. H. Ng, S. S. Pease, J. L. A. Zhang, S. S. Damle, G. Freedman, S. Siu, I. D. Bernstein, M. B. Elowitz, and E. V. Rothenberg, "Asynchronous combinatorial action of four regulatory factors activates bcl11b for t cell commitment," *Nature Immunology*, vol. 17, no. 8, pp. 956–965, 2016.

[7] L. Li, M. Leid, and E. V. Rothenberg, "An early t cell lineage commitment checkpoint dependent on the transcription factor bcl11b," *Science*, vol. 329, no. 5987, pp. 89–93, 2010.

[8] M. E. De Obaldia and A. Bhandoola, "Transcriptional regulation of innate and adaptive lymphocyte lineages," *Annual Review of Immunology Vol 33*, vol. 33, pp. 607–642, 2015.

[9] M. A. Yui and E. V. Rothenberg, "Developmental gene networks: a triathlon on the course to t cell identity," *Nature Reviews Immunology*, vol. 14, no. 8, pp. 529–545, 2014.

[10] V. Olariu, M. A. Yui, P. Krupinski, W. Zhou, J. Deichmann, E. Andersson, E. V. Rothenberg, and C. Peterson, "Multi-scale dynamical modeling of t cell development from an early thymic progenitor state to lineage commitment," *Cell Reports*, vol. 34, no. 2, 2021.

[11] G. K. Ackers, A. D. Johnson, and M. A. Shea, "Quantitative model for gene-regulation by lambda-phage repressor," *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, vol. 79, no. 4, pp. 1129–1133, 1982.

[12] I. S. Peter, E. Faure, and E. H. Davidson, "Predictive computation of genomic logic processing functions in embryonic development," *Proceedings of the National Academy of Sciences*, vol. 109, no. 41, pp. 16434–16442, 2012.

[13] E. H. Davidson, "Emerging properties of animal gene regulatory networks," *Nature*, vol. 468, no. 7326, pp. 911–920, 2010.

[14] V. Olariu, E. Manesso, and C. Peterson, "A deterministic method for estimating free energy genetic network landscapes with applications to cell commitment and reprogramming paths," *Royal Society Open Science*, vol. 4, no. 6, 2017.

[15] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of theoretical biology*, vol. 22, no. 3, pp. 437–467, 1969.

[16] K. J Mitchell, "Waddington's "epigenetic landscape"," 2013.

[17] E. Andersson, M. Sjö, K. Kaji, and V. Olariu, "Cellogene-an energy landscape framework for logical networks controlling cell decisions," *bioRxiv*, 2022.

List of Figures

List of Tables

Appendices

A Table of operators

As the inputs to a gene are based on a three-state logic, the operators are slightly more complicated than ordinary Boolean operators (table 4).

Table 4: Truth table for operators that take three-state inputs.

A	B	A AND B	A OBAND B	A UBAND B	A OBOR B	A UBOR B
0	0	0	0	0	0	0
± 1	0	0	0	0	± 1	± 1
0	± 1	0	0	0	± 1	± 1
± 1	± 1	± 1	± 1	± 1	± 1	± 1
± 1	∓ 1	0	$+1$	-1	$+1$	-1

B Marble simulations

B.1 Implementation

In order to investigate the cell fates in the landscape and to simulate the development of cell states in connection with real time, marble simulations have been implemented within the framework of CELLoGeNe. A marble is assumed to walk randomly in the landscape with weighted transition probabilities based on the energy of the states. In each update, the marble is able to either stay where it is or take a step to a neighboring state (turning one gene from OFF/ON to ON/OFF). The marble can be stopped in the landscape when certain criteria are met. By repeating such simulations a large number of times, it is possible to capture some features of the landscape, for example, the relative strength of attractors.

During each update, the probabilities for all possible transitions are weighted with the Boltzmann factor. From an initial state s, a marble is able to transition to one of its N neighbors (N is the size of the GRN) or stay at s. The probability of a transition from s to μ is given by:

$$p_{\mu \leftarrow s} = \frac{e^{-(E_\mu - E_s\}\beta}}{\sum\limits_{k \in \{1,\ldots,n,s\}} e^{-(E_k - E_s\}\beta}}, \quad \mu \in \{1, \ldots, n, s\} \tag{B.1}$$

where β is the parameter of the noise level. The decrease in β increases the chance of a transition to a higher energy state happening. The number of consecutive steps that the marble has stayed at the same state can be used to define the stopping criteria.

B.2 Improvement of simulation speed

The marble simulations are implemented in Python, and therefore it is generally better to include the NumPy library. It can speed up the simulations a lot by using array objects and methods within NumPy over the original Python ones.

However, the calculation of transition probabilities can still be very heavy during marble simulations, especially when the β is small and the stopping criteria rely on the marble staying at the same state for some consecutive steps.

For each step of the marble, N Boltzmann factors are needed to find the transition probabilities. If such probabilities are calculated dynamically, a large number of steps in the landscape will result in very slow simulations. It is, therefore, better to precalculate the possibilities for all possible transitions in the landscape as long as the noise level does not vary during the simulations. By assigning neighbors and the corresponding transition probabilities to each state in the landscape before starting the marble simulation, the simulation can be more than ten times faster than the original implementation.

C Supplementary results

C.1 DN2a – DN2a transition

It has been found in the experiment that the early knockdown of RUNX1, GATA3, or TCF1 will halt the cell progression from the DN1/ETP state. Marble simulations also showed that the knockdown lead to a lower probability of DN1/ETP - DN2a transitions (figure 10).

Experimental observations also tell us that the effects of TCF1 and GATA3 depend on the cell stage and are less important after the cells have reached the DN2 state [6]. We carried out marble simulations that are initialized at the DN2a state to investigate the gene knockdown effect on DN2a cells (figure 16).

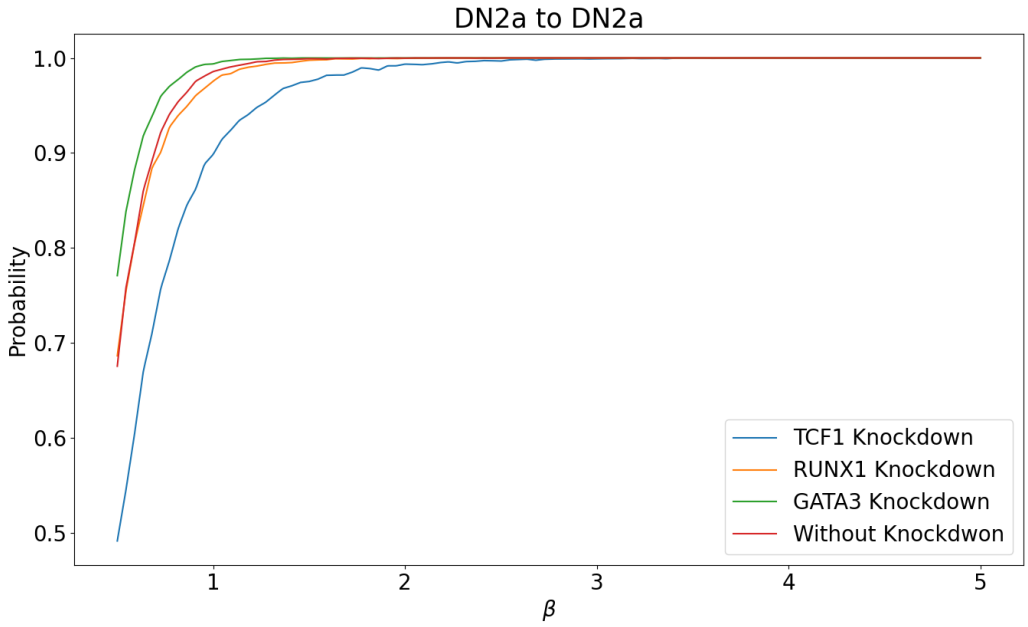

Figure 16: Probability of ending up at the DN2a state when the system was initialized at the DN2a state. β varies from 0.5 (large noise) to 5 (small noise). Landscapes are indicated by the legend.

At low noise levels, the systems were most likely to stay at the DN2a state. When the noise level was raised, TCF1 knockdown showed the strongest impact on the probability of the systems staying at the DN2a state. The probability was decreased by knocking down TCF1, but by contrast, GATA3 knockdown increased the probability of cells staying at the DN2a state. RUNX1 knockdown did not show much impact compared to the landscape without gene knockdown.

The results from marble simulations are not consistent with our expectations. This could come from the fact that the GRN we used to map the landscapes only describes the process of T cell development from the DN1/ETP state to the DN2a state, but does not account for the followed cell states. It is then hard to predict the knockdown effect after DN2a state with such a network. The RUNX1 gene may also be involved in more complicated mechanisms in T cell development, and those effects are not included in the network. So, to conclude, the marble simulations were not able to provide more insight on the stage-dependent role of TCF1 and GATA3. It is also difficult to show the effect of RUNX1 knockdown on DN2a cells.

C.2 Time estimation in the landscape and noise levels

The steps in the landscape were connected to real time by fitting the landscape prediction of DN2a cell proportions to experimental data. Nevertheless, the noise level has an impact on how fast a system evolves inside the landscape. In a situation with a high level of noise, the estimation may vary greatly from the result obtained at a rather low level of noise (figure 17).

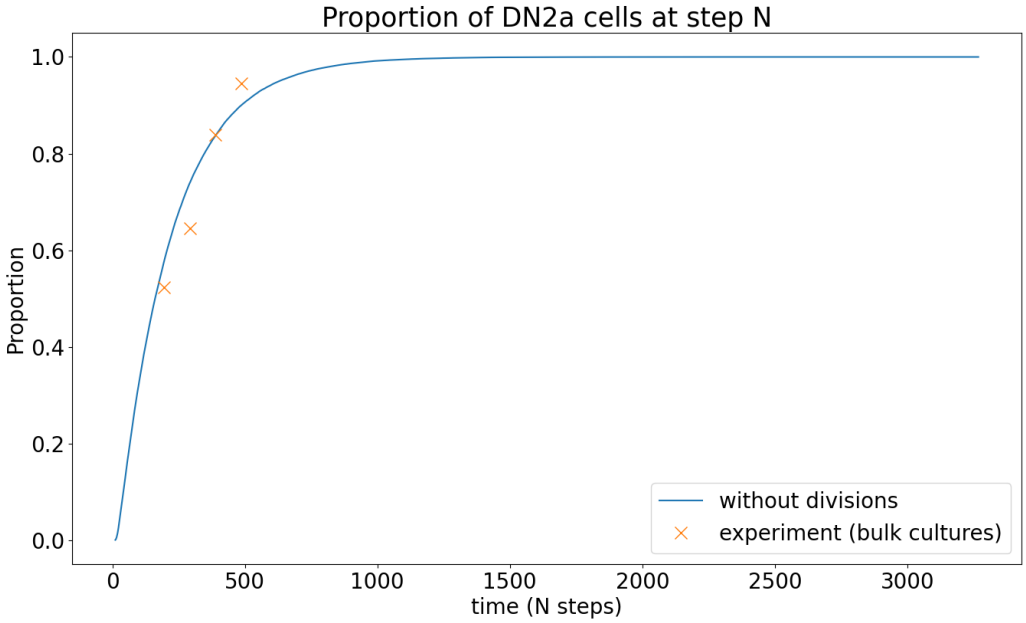

Figure 17: DN2a cell proportion of bulk DN1 cultures fitted to the results from landscape prediction ($\beta = 0.5$).

When β was 0.5, we estimated 24 hours to correspond to almost 100 steps in the landscape, which is much larger compared to the estimation from a landscape with $\beta = 1$ (15.78 steps). In this case, what prevented the cells from entering the DN2a state was the requirement that they must stay at the DN2a state for three consecutive steps. With very high noise levels and enough time passed, the cells are not evolving towards the DN2a state anymore, but instead, they are just wandering around the DN2a attractors. It is therefore not reasonable to assume such high noise levels.