博士論文

Dynamical systems theory of cellular differentiation and reprogramming; Interplay between fast gene-expression dynamics and slow epigenetic feedback regulation

（細胞分化とリプログラミングの力学系理論; 速い遺伝子発現ダイナミクスとエピジェネティック制御との相互作用）

松下　優貴

Contents

Chapter 1

Introduction

1.1 Cellular differentiation

Most multicellular organisms start their lives with a single cell. Throughout the developmental process, they acquire tissues consisting of various cell types with different functions. This cellular specialization process, as is schematically shown in Fig. 1.1, is so-called cellular differentiation. Cell differentiation is remarkable in that it generates different cell types even though their DNA sequence, the blueprint for life, is identical.

Cellular differentiation gradually proceeds downward in Fig. 1.1. Totipotent stem cells (the top of Fig. 1.1) differentiate to pluipotent stem cells (the second from top of Fig. 1.1). Then, pluripotent stem cells differentiate to multipotent stem cells. Finally, cells lose their potential of differentiation (as shown in the bottom of Fig. 1.1). Cellular differentiation sequentially proceeds hierarchically [1,2]. In addition, cell differentiation is basically irreversible, i.e., it cannot be reversed upward in Fig. 1.1. For instance, although totipotent stem cells can differentiate to pluripotent stem cells, pluripotent stem cells or further differentiated cells cannot come back to totipotent stem cells.

Even though these properties in cellular differentiation are already remarkable, the most remarkable properties in cellular differentiation will lie in its robustness and its universality among multicellular organisms. As cellular differentiation proceeds via thousands of chemical reactions based on DNA sequences, noise throughout the chemical reaction therein is inevitable. Still, the final cell types, as well as differentiation pathway to them, are robust against such noise [3, 4].

In this thesis, we attempt to understand these universal properties in cellular differentiation, as well as cellular reprogramming to be mentioned below.

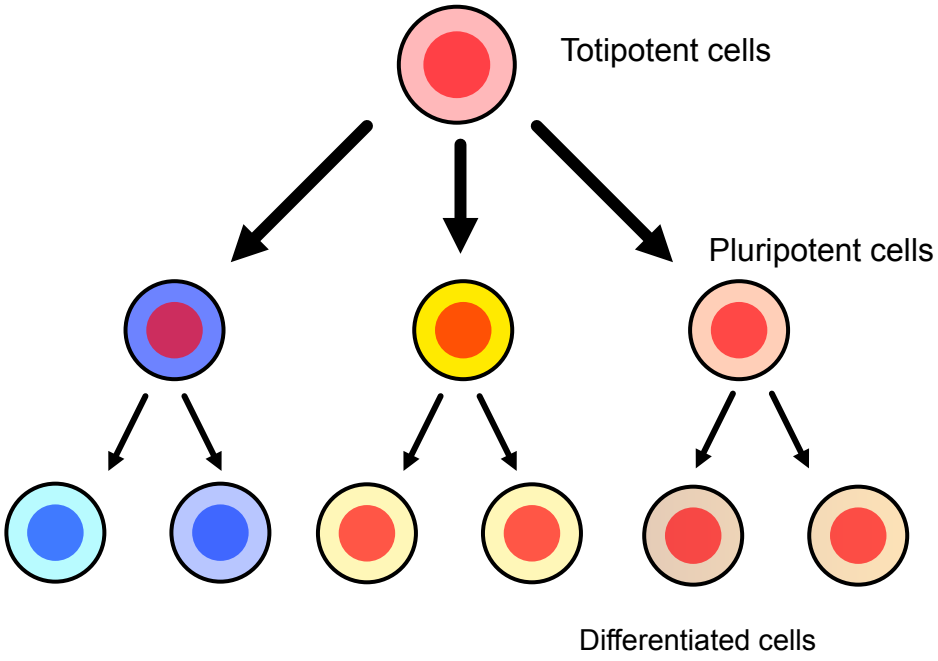


Fig. 1.1: Schematic representation of Cellular differentiation.

1.2 Cellular reprogramming

As mentioned in the previous section, cellular differentiation was believed to be basically irreversible. This common belief was then challenged by seminal studies by Gurdon and Yamanaka.

Gurdon replaced the nuclei of fertilized egg cells of a frog with that of intestinal cells (differentiated somatic cells) of another frog and succeeded in making developmental process from the cells [5]. Replaced nuclei of egg cells have already been inactivated by irradiation, thus development was proceeded based on the genetic information of differentiated, intestinal cells' nuclei of another frog. This result suggests that the nuclei of differentiated cells never lost the information throughout differentiation and pluripotency can be recovered by specific manipulation (which corresponds to transplantation of the nuclei into the cytoplasm of the egg cells).

Yamanaka succeeded in regaining pluripotent cells without the use of egg cells. Takahashi and Yamanaka achieved regaining pluripotency by overexpressing just a few genes (*Oct4*, *Sox2*, *Klf4*, and *c-Myc*) into mouse somatic cells [6, 7], and called these pluripotent cells as induced pluripotent stem cells (iPS cells).

Before the invention of iPS cells, artificial pluripotent cells can be generated by transplantation of egg cells or incubation of early-stage embryonic cells (so-called ES cells). With these methods, application to regenerative medicine is difficult for ethical problems or due to the need of preserving ES cells for all patients. As mentioned above, the production of iPS cells only needs differentiated cells. This opens up the potential of the application of iPS cells to regenerative medicine.

1.3 Regulation of gene expression and multi-stability

Here, let us consider cellular differentiation and reprogramming from the viewpoint of theoretical biology. It is well known that each of the differentiated cell types has identical protein expressions, but distinct by cell type. Here, protein expressions are determined by gene expressions. In this section, we review how gene expressions are regulated.

Fig. 1.2 shows a single pair of promoter and codon sequences in DNA. The promoter part is a starting point of a series of chemical reactions. To synthesize protein, the promoter part responds to the transcriptional input and sends the signal to the transcription start site of the gene. At downstream from the promoter, RNA polymerase produces RNA using Codon sequence as a template. Then, the corresponding protein is synthesized with the aid of the ribosome, an intracellular organelle.

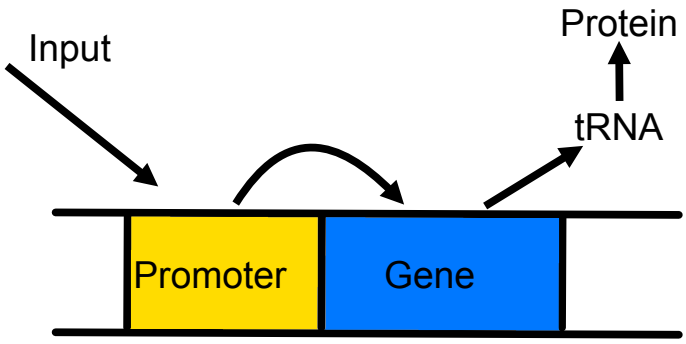


Fig. 1.2: Flow chart of gene expression.

The molecular mechanism of gene expression has been described above. The next problem, however, is the determination of the external input to the promoter. Of course, the environment including cell culture is the candidate of external signaling, but gene expression itself does not need an explicit external input from outside of the cell. Proteins synthesized from the other genes can work as an input. Although Fig. 1.2 shows only a single gene, the DNA sequence consists of thousands of genes. The input for each gene is a result of other genes. Thus, the change of protein expression levels (or concentrations), which follows activation or inhibition among thousands of genes, is represented in terms of dynamical systems theory.

Dynamical systems theory represents the change of state variables in the phase space. It is generally represented by ordinal differential equations. Differenital equations dx_i/dt are introduced for the variables x_i. In dynamical systems theory, the local time evolution of differential equations is given by the flow of each point in phase space $\{x_i\}$. Then, the solution of differential equations is given by trajectory in phase space. Attractor is a stable state to which various trajectories converge.

The fixed point of the dynamical systems is represented by the cross point of plane satisfies $dx_i/dt = 0$ (this plane is so-called nullcline) and it is fixed point attractor if the flows around this cross point are attractive.

If there are multiple attractors in phase space, the system has multi-stability. Multi-stability is often understood by multiple valleys in a landscape. If we put balls onto the landscape, balls are trapped to one of the bottoms of these valleys. Then, which valleys are selected depends on the initial positions of the balls.

The simplest biological example of multi-stability is known as a toggle switch as shown in Fig. 1.3 [8, 9]. The toggle switch consists of two genes that repress each other through synthesized proteins as shown in 1.3a. Here, we introduce variables x_1, x_2 to represent the gene expression (or protein concentration) of gene 1, 2. The dynamics of toggle switch are often described by using Hill function $F^{\mathrm{Hill}}(z) = z^\alpha/(z^\alpha + K^\alpha)$ (for activation) or $F^{\mathrm{Hill}}(z) = 1/(z^\alpha + K^\alpha)$ (for suppression)[*1], or other sigmoidal function. Here, we adopt the following equations which x_1 and x_2 follows as

$$\frac{dx_1}{dt} = F(-x_2) - x_1, \tag{1.1}$$

$$\frac{dx_2}{dt} = F(-x_1) - x_2, \tag{1.2}$$

where $F(z)$ is a sigmoidal function $F(z) = 1/(1 + \exp(-\beta z))$[*2]. In this representation, the toggle switch allows two attractors, State 1 and State 2, in which the gene1, 2, is highly expressed respectively, as shown in Fig. 1.3b. Cellular states are converged to each of the two states, depending on the initial states as shown in Fig 1.3b. Thus, the toggle switch can be represented by a landscape with two valleys as shown in Fig. 1.3c.

Basically, the cell consists of more genes than a toggle switch. To investigate the multi-stability of more complex systems, gene regulatory network (GRN), with a large number of genes with mutual activation (inhibition) regulation, is adopted.

[*1] In Hill function, K is a parameter that represents the threshold where on-off type output changes, whereas α represents the sharpness of Hill function.

[*2] In sigmoidal function $F(z)$, β plays same role as α in F^{Hill}

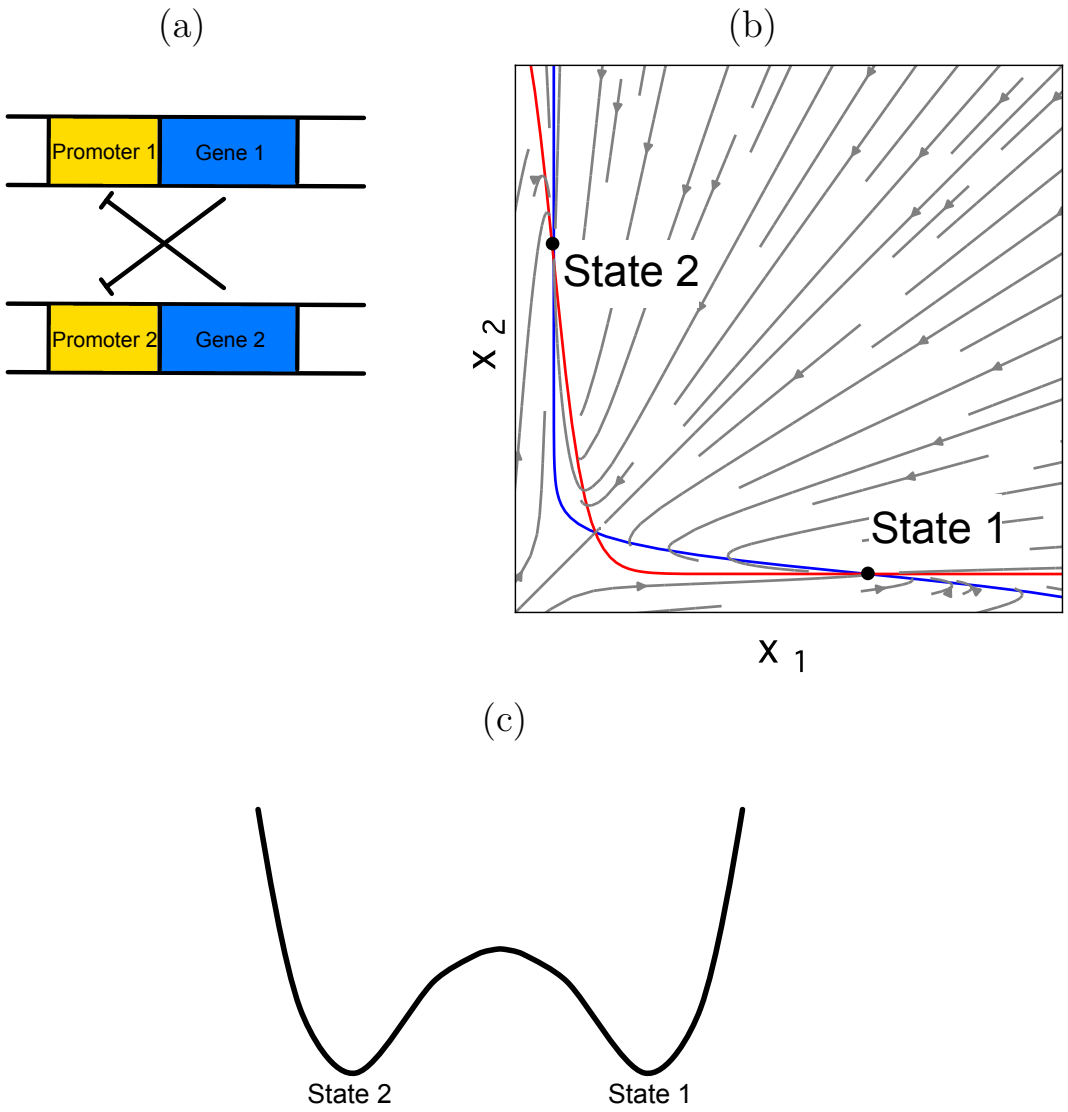


Fig. 1.3: (a) Genetic circuit of toggle switch. Toggle switch consists of two genes that repress each other. (b) $x_1 - x_2$ phase space. x_1 and x_2 are the variables which represents expression state of gene 1 and gene 2 (or protein concentration producted from gene 1 and gene 2). Toggle switch allows two stable states in which gene 1 or gene 2 is highly expressed. (c) A schematic of landscape with two valleys, corresponding to the toggle switch.

Although multi-stability in gene expression dynamics gives us a hint how differentiated cell types emerges from identical DNA sequence, important issues such as hierarchy and irreversibility in cell differentiation, are still unsolved. What kind of valley corresponds to a pluripotent state or to a differentiated state? Or how does the cell differentiation progress, i.e., by jumping over each of the valleys? Another approach, that goes beyond just the analogy of multi-stability, is needed for cellular differentiation and reprogramming.

Cell-cell interaction can be a candidate to make such jumping. In cell population, cells interact with each other via synthesized chemicals. With development,

the number of the cells increases, and the degree of cell-cell interaction changes. For instance, with the increase in the size of cell population, bifurcation from homogeneous to heterogeneous population can occur [10, 11]. In these studies, pluripotent cells are described by the oscillatory state of gene expression dynamics, whereas differentiated cells by the fixed state with oscillation death. In this situation, the population ratio of pluripotent cell and differentiated cells is well-regulated. Besides the cell-cell interaction, however, recent studies suggest that epigenetics which modifies DNA, which is one of the hottest topics in molecular biology, works as another important factor for cellular differentiation and reprogramming.

1.4 Epigenetic landscape and epigenetics

Epigenetics is derived from the important concept of epigenetic landscape proposed by Waddington [3]. Hence, let us start with the introduction of epigenetic landscape.

In 1957, Waddington metaphored developmental process by balls falling down a mountain, as shown in Fig. 1.4a. At the top of the mountain, a single valley exists. Thus, balls are all trapped onto the bottom of the valley. Throughout going down the mountain, the single valley branches to several valleys. With this branching, balls are trapped into one of the different valleys.

Waddington metaphored cell differentiation process by introducing the change in the landscape itself. Here, we need to consider what causes this change. Fig. 1.4b is a picture given by Waddington. Fig. 1.4b shows the backside of epigenetic landscape. According to Waddington, the backside of the landscape is wired by genes, which leads to the landscape change. Hence, Waddington needed to coin the term "epi" + "genetics" for the landscape.

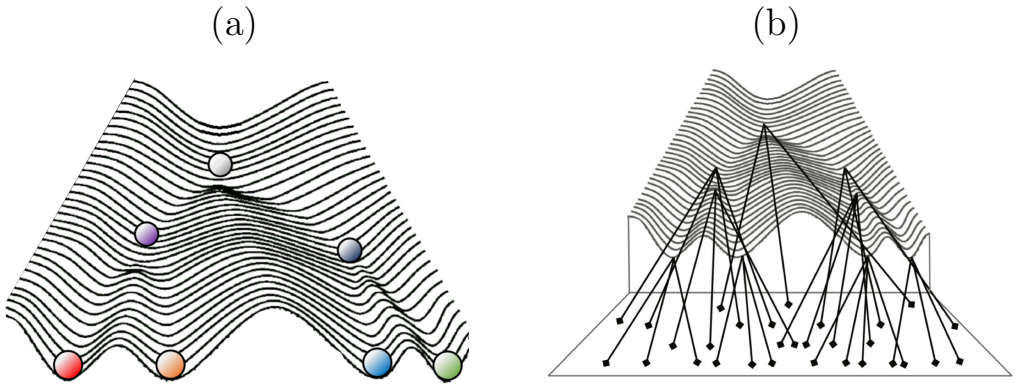


Fig. 1.4: (a) The concept of epigenetic landscape proposed by Waddington. (b) The backside of epigenetic landscape. Both of (a) and (b) are adopted from [3].

Waddington proposed this concept about 70 years ago. Since the discovery of DNA structure by Watson and Crick in 1953, experimental techniques have greatly advanced. Some molecular mechanisms have been identified as the candidates of the change of landscape by Waddington. A series of molecular mechanisms for epigenetic modification alter the feasibility of gene expressions [12–18].

Fig. 1.5 schematically shows DNA methylation, one of the mechanisms known as epigenetics. As shown in Fig. 1.5, the promoter part can be modified by the methyl group. Then, the promoter part becomes insensitive to the input signals and corresponding genes are harder to be expressed. For another example, histone modification changes the openness of chromatin around a promoter of the gene and changes the efficiency of transcription. Besides these numerous epigenetic modifications, novel epigenetic modification mechanisms have been found even now [19, 20]. These modifications change the way to use DNA sequences without changing DNA sequences themselves. Epigenetic modification states are concerned with developmental process; for pluripotent cells, these modification levels are small, whereas differentiated cells have larger modification levels that depend on each cell types [21–24].

Noteworthy, epigenetic modification also depends on the cellular state [25–29]. Thus, cellular differentiation proceeds with complex interaction between gene expression and numerous epigenetic modifications. We need to understand what aspects of the interplay between gene expression and numerous epigenetic modifications are important for cellular differentiation (and reprogramming).

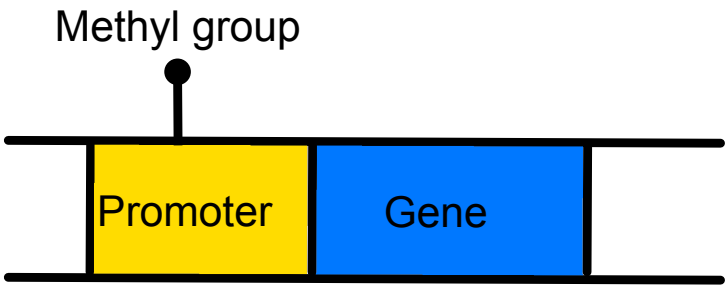


Fig. 1.5: Schematic representation of DNA methylation on the promoter.

1.5 Cellular reprogramming, as climbing epigenetic landscape?

Cellular reprogramming is also discussed with the picture of epigenetic landscape: cellular reprogramming is described as jumping or climbing up epigenetic landscape as shown in Fig. 1.6) [30, 31]. This, however, remains just a metaphor.

As we mentioned, cellular state, however, involves many degrees of freedom (e.g., gene expression levels and epigenetic modification levels, whereas reprogramming manipulation is achieved by overexpressing just a few genes. These two facts are hard to reconcile with each other. Then, we need to understand how few degrees of freedom can reprogram many-dimensional state, and how climbing-up the landscape is represented in terms of dynamical systems theory.

Recent studies reported that reprogramming by overexpression a few genes

works well among a variety of multicellular organisms. Moreover, according to these reports, cellular reprogramming can succeed in using various differentiated cell types and induction sets of genes. This suggests cellular reprogramming is also universal, as well as cellular differentiation. So far, reprogramming has gathered much attention from regenerative medicine, whereas universal properties of cellular reprogramming are less explored. We need to understand how reprogramming manipulation, with a few degrees of freedom, works in terms of dynamical systems theory.

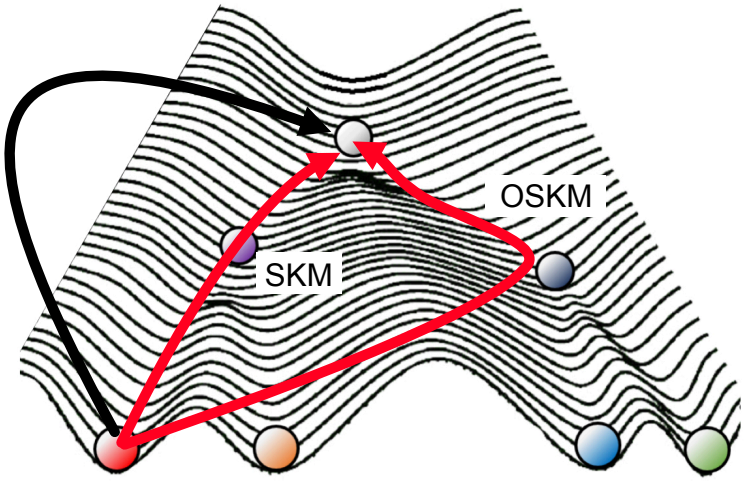


Fig. 1.6: Schematic representation of cellular reprogramming and epigenetic landscape. Cellular reprogramming is metaphored as jumping from the bottom of epigenetic landscape to the top of epigenetic landscape (black arrow) or climbing epigenetic landscape. Each of red arrows shows different reprogramming pathway by different induction sets of genes (each capital represents an element of 4 Yamanaka factors, *Oct4*, *Sox2*, *Klf4*, and *c-Myc*). Adopted from [30, 31].

1.6 About this thesis

In this thesis, we attempt to understand the universal properties of cellular reprogramming and reprogramming. Our goal is to understand the remarkable properties of cellular differentiation, and provide the simplest understanding of epigenetics, which has broad fields and complex mechanisms.

For our goal, we take a phenomenological approach. We construct the cell model with gene regulatory network and epigenetic modification. We introduce epigenetic modification processes as the simplest feedback regulation. Then, we construct dynamical systems theory of cellular differentiation and reprogramming from the viewpoint of the interplay between fast gene expression and slow epigenetic modification.

In chapter 2, we mainly consider a quantitative framework for explaining for cellular differentiation using the proposed cell model. In multicellular organisms, cells differentiate into several distinct types during early development. Determi-

nation of each cellular state, along with the ratio of each cell type, as well as the developmental course during cell differentiation are highly regulated processes that are robust to noise and environmental perturbations throughout development. Waddington metaphorically depicted this robustness as the epigenetic landscape in which the robustness of each cellular state is represented by each valley in the landscape. This robustness is now conceptualized as an approach toward an attractor in a gene-expression dynamical system. However, there is still an incomplete understanding of the origin of landscape change, which is accompanied by branching of valleys that corresponds to the differentiation process. Recent progress in developmental biology has unveiled the molecular processes involved in epigenetic modification, which will be a key to understanding the nature of slow landscape change. Nevertheless, the contribution of the interplay between gene expression and epigenetic modification to robust landscape changes, known as homeorhesis, remains elusive. We demonstrate extensive simulations of the proposed model with dynamical-systems analysis and succeed in extracting Waddington' s epigenetic landscape, which can help to explain how such robustness in the developmental process, which Waddington referred to as "homeorhesis", is generated.

In chapter 3, we elucidate dynamical systems theory of cellular reprogramming. In cellular reprogramming, almost all epigenetic memories of differentiated cells are erased by the overexpression of few genes, regaining pluripotency, the potential for differentiation. Considering the interplay between oscillatory gene expression and slower epigenetic modifications, such reprogramming is perceived as an unintuitive, global attraction to the unstable manifold of a saddle, which represents pluripotency. The universality of this scheme is confirmed by the repressilator model, and by gene regulatory networks randomly generated and those extracted from embryonic stem cells.

Chapter 2

Homeorhesis in Waddington's landscape by epigenetic feedback regulation

In most multicellular organisms, cells differentiate into several types in the course of development, which show distinct gene expression patterns that are robust to external perturbations and internal noise. As a theoretical explanation for this robustness, Waddington introduced the concept of the " epigenetic landscape" more than 60 years ago, as shown in Fig.2.1a. In this concept, a ball falling along the landscape represents the cell differentiation process over time, and each valley corresponds to a differentiated cell type [3]. Although presented visually as a metaphor, Waddington also proposed that this differentiation process can be understood in terms of the dynamical systems of gene expression. Following his insight, each valley is now interpreted as an attractor of an intracellular dynamical system for gene (protein) expression. The cellular state is represented by a set of gene (protein) expression levels reached as a result of such dynamical system, as represented by an attractor of such dynamical system. Then, the cellular state remains in the vicinity of the attractor even under internal noise or external perturbation. In fact, several dynamical-systems models with mutual activation and inhibition of protein expression demonstrated the coexistence of multiple attractors that correspond to distinct cell types [32–35], and supporting experiments have been carried out [36, 37].

According to this dynamical-systems approach, the X axis characterizing the cellular state in Fig.2.1a is represented by the gene expression pattern. However, since there are thousands of genes (or components) in a cell, the state may not be accurately represented by a one-dimensional variable X. Nevertheless, the cellular state can potentially be represented by only a few variables extracted from data reduction of the expression levels of a huge number of components, such as principal component analysis (PCA) [38].

Moreover, the height of the landscape (Z axis) represents changeability of the state. Cellular states are attracted to the bottom of the valley, which, in terms of dynamical systems, are fixed-point attractors at which point no more change will occur.

Along with the dynamics falling onto the bottom of the valley, as represented by

motion along the X axis in Fig.2.1a, the landscape itself is shaped along the other (Y) axis representing the developmental course, in which the valleys are shaped successively and are deepened, in a process known as "canalization" . Therefore, a fundamental question remains: given that the attraction to each valley along the X axis is represented by gene-expression dynamical systems, what does the Y axis representing (slower) landscape change represent?

To address this fundamental question, there are three basic questions to resolve with respect to the postulates of Waddington's landscape itself. First, there is the issue of hierarchical branching. That is, since the valleys are successively generated over developmental time (Fig.2.1a), many valleys (attractors) are not generated independently, but rather the shallower valleys are generated first and are then branched, and these branching processes are repeated [1, 2]. Second, Waddington argued that the developmental process itself, i.e., the motion along the shaping of valleys, is also robust to perturbation, and coined the term "homeorhesis" to represent such path stability [3, 4]. However, the mechanism contributing to the robustness of this shaping process, including successive branchings, remains elusive [34]. Finally, the number ratio of each cell type is also rather robust to perturbations or initial conditions. If we assume that a deeper valley attracts more cells, this robustness implies overall robustness of the landscape, in particular, the depth of each valley [10, 39–44].

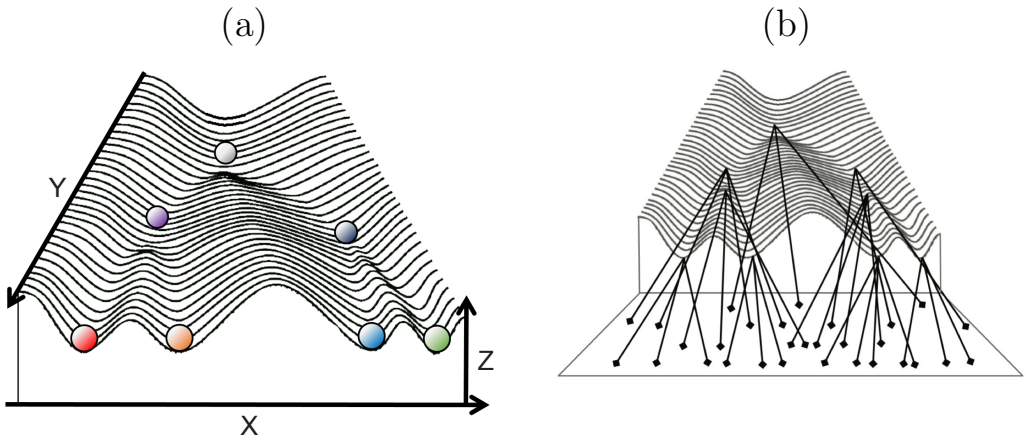


Fig. 2.1: (a) Waddington's epigenetic landscape. The cell differentiation process is conceptually explained as a motion of a ball along the landscape in which valleys correspond to differentiated cell types. Here, the horizontal axis (X) represents a cellular state, the height (Z axis) represents the inverse of the frequency (probability) that a cell takes state X, and the Y axis represents slow developmental change. Adapted from [3]. (b) Shaping the landscape by genes, represented by strings (also adapted from [3]). Waddington entitled the figure as the "complex system interaction underlying the epigenetic landscape", where interactions among genes control the landscape as metaphorically represented by strings.

Considering these three postulates of the landscape, let us now come back to the fundamental question of the nature of the Y axis representing slow developmental change. Waddington drew the diagram Fig. 2.1b to show schematically how genes control the epigenetic landscape, in particular the change of valleys along the de-

velopmental time. After 60 years since the proposal of the epigenetic landscape, we have now possible candidates that could cause such a slow landscape change. One candidate is the cell-cell interactions [10, 43, 44]. As development progresses and the cell number increases, the influence of cell-cell interactions on the intracellular dynamics for each cell-type is stronger. Slow modifications of intracellular expression dynamics can lead to novel attractors or an increase in their robustness.

Another potential source for slow landscape change changes in chromatin structure with epigenetic modification which is currently one of the hottest topics in cell and developmental biology [12–18]. Epigenetics is a field derived from the term "epi(above)-genetic" coined by Waddington, sixty years ago [3, 4]. At that moment it would cover all possible changes beyond the genetic one. Now, epigenetics are mainly studied as a variety of molecular mechanisms that affect feasibility of the expression of each gene at a given time and place without change in DNA sequences, such as the methylation or some other molecular modifications in DNA [45–47]. These modifications change, for example, the openness of chromatin around a promoter of gene, which affects the flows of gene regulation, i.e., the efficiency of transcription binding promoter changes depending on the degree of openness [25,48]. If the chromatin is more open around a gene, it is more feasible to be expressed by the actions of other genes on its promoter. As for modifications to the DNA influencing on the openness over a long time span, novel molecular mechanisms are still being uncovered, with the ongoing development in the field of epigenetics [19].

Note that epigenetic modification generally depends on a given cellular state, i.e., the expression levels of proteins, whereas the epigenetic changes, e.g., the openness of chromatin structures, influence the expression levels. In general, the epigenetic process is slower than the expression dynamics, and the epigenetic change lasts over the time span in the change in protein concentrations [25–29]. The epigenetic change leads to stabilization of cellular states, which can be corresponded to deepening the valleys as schematically represented in Fig. 2.1b. This stabilization may imply the existence of a positive-feedback process between the expression and epigenetic change as suggested experimentally [48, 49], and also theoretically [26–28, 50].

Despite these suggestions and growing reports on epigenetic changes, however, the detailed interplay between epigenetic modification and gene expression dynamics has not been fully explored. In contrast to the extensive body of theoretical and empirical literature on expression dynamics or epigenetic modifications, there is little experimental elucidation of the underlying molecular mechanisms nor theoretical model for the interplay proposed to date. Therefore, at this stage, a simple phenomenological model is needed to investigate how such slow epigenetic change can introduce a novel expression pattern or stabilize the existing expression patterns. Such a model would provide a bridge between epigenetic modification and the epigenetic landscape as Waddington conceptualized.

To formulate the epigenetic process in terms of dynamical systems, we here introduce an epigenetic variable for each expressed gene, represented as a threshold level of the input needed for the gene of concern to be expressed. Using the simplest feedback process, we elucidate the possible conditions for the epigenetic landscape and its properties. Rather than seeking detailed models extracted from

realistic expression dynamics, we instead consider a minimal conceptual model that captures the interplay between the relatively faster gene expression dynamics and slower epigenetic dynamics to address how an epigenetic landscape satisfying the requisites of (1) hierarchical branching, (2) homeorhesis, and (3) robustness in the cell-number ratio is generated. Instead of the simplicity in the model, we have simulated thousands of networks, to extract a universal mechanism and draw a general conclusion, which will hold true in a complicated system with biological reality.

2.1 Model

We consider a cell model with a gene regulatory network (GRN) and epigenetic modification. The cell has N genes and the cellular state is represented by the expression level (or concentration) x_i of each gene i. A schematic representation of the cell model is shown in Fig. 2.2.

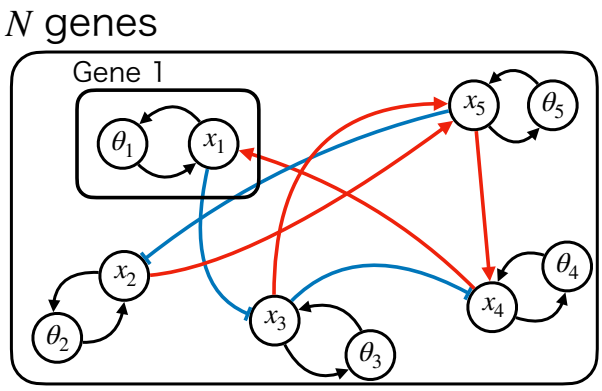


Fig. 2.2: The schematic representation of the cell model.

Here, the GRN represents the mutual control of genes via synthesized proteins. Gene expression typically shows an on-off-type response to the input: a gene is expressed (suppressed) when its input value is above (below) a certain threshold, whereas its expression level is saturated as the input value increases. By normalizing the maximal (i.e., saturated) expression level to unity, we adopt the following gene expression dynamics for simplicity [51–55]:

$$\frac{dx_i}{dt} = F\left(\sum_j^N \frac{J_{ij}}{\sqrt{N}} x_j + \theta_i + c_i\right) - x_i, \tag{2.1}$$

where J_{ij} is the regulatory matrix. If J_{ij} is positive (negative), gene j activates (represses) the expression of gene i, whereas J_{ij} is set to 0 if no regulation exists. To represent the on-off-type expression of genes, we adopt $F(z) = \tanh(\beta z)$ ($\beta = 40$). $x_i = 1$ indicates the full expression of the i th gene and $x_i = -1$ indicates no

expression of the i th gene. c_i is a constant input value, interpreted as an input outside of the N genes (for example, upstream genes) or the natural trend for expression. For most examples, however, c_i is set to 0 unless otherwise noted.

Here, $-\theta_i$ represents the threshold for the input, beyond which the expression is activated. As θ_i is increased (decreased), the i-th gene tends to be expressed (repressed). In the standard GRN model, this threshold is fixed. By contrast, we regard it as a variable by assuming that θ_i represents the epigenetic modification level for each gene i, such as histone modification or openness of chromatin structure.

Further, the epigenetic change depends on the expression level of gene i. In accordance with some theoretical and experimental reports [26,27,48,49,56,57], we adopt a positive-feedback process from gene expression to epigenetic modifications: when a gene is expressed (repressed), it tends to be expressed (repressed), as is consistent with the stabilization by the epigenetic changes. Thus the epigenetic change (openness in the chromatin structure) is given by:

$$\frac{d\theta_i}{dt} = v\left(ax_i - \theta_i\right). \tag{2.2}$$

This assumption is rather natural considering the concept of "canalization", i.e., each valley becomes deeper (or more stable) as development. In Eq. (2.2), the parameter $a(>0)$ represents the strength of the positive-feedback mechanism, and $v(>0)$ gives the rate of change in the epigenetic modification.

In the model, both θ_i that characterizes the degree of openness and $\sum_j J_{ij}x_j$ that represents the actions from other proteins synergetically determine the expression. In the equation, only if $\sum_j J_{ij}x_j + \theta_i$ is larger than the threshold c_i, i.e., if the input is large and the chromatin is sufficiently open, x_i is expressed. As the term $\sum_j J_{ij}x_j$ is bounded by unity, if θ decreases goes to a negative value (towards $-\alpha$), the gene is not expressed. Hence if the chromatin is closed, the gene is not expressed. Further, because θ_i increases as a result of positive feedback from x_i, the gene turns to be expressed only when the chromatin is open and activators bind to the promoter.

Remark: Here we remark on the simplification adopted in our modeling and its limitation. Biologically plausible models often adopt the Hill function such as $F^{Hill}(z) = z^\alpha/(z^\alpha + K^\alpha)$ (or $F_s^{Hill}(z) = 1/(z^\alpha + K^\alpha)$) for the expression dynamics, instead of $F(z)$, where combination of $x_m^\alpha/(x_m^\alpha + K_m^\alpha)$ (for activation) and $1/(x_m^\alpha + K_m^\alpha)$ (for suppression) provides the expression dynamics [58, 59]. The qualitatively same gene-expression dynamics (e.g., on/off expression patterns, multistability, oscillation), however, is observed both for the models with $F(z)$ [51–54] and the Hill-type model.

In the Hill function, the value K corresponds to the threshold for the on/off -type expression dynamics. With the increase in K, the threshold for expression increases. Hence, K and θ in our model play the same role. Now K's change with the affinity with the prompter and proteins, or in other words, with the chromatin openness. Considering that epigenetic change to stabilize each expression pattern, as discussed above, the feedback from x to K in the form of Eq.(2) was introduced in a previous research [58] for a Hill-type model extracted from the real data. Indeed, Eq.(2) is obtained by replacing K by θ.

Here, as long as the threshold-type expression dynamics and slow positive epigenetic feedback process in the type of Eq.(2) is adopted, qualitatively similar behaviors are obtained, with regards to the classes of the expression dynamics and the fixation from oscillatory dynamics, to be discussed below [*1].

Note that the two basic assumptions (i) threshold-type gene expressions via synthesized proteins and (ii) slow epigenetic modification (chromatin openness) that stabilizes the expression pattern by the feedback are essential to draw the conclusion in the present paper. By our approach, however, one cannot make a direct, quantitative prediction on the epigenetic landscape for a specific example, as it requires quantitative information on the specific form of gene expression dynamics and epigenetic feedback. This is a limitation in our modeling.

2.2 Fixed-point analysis

The fixed-point solutions of (2.1) and (2.2) are obtained by setting each term to zero. From the latter, we get $\theta_i = ax_i$ and from the former we obtain

$$\tanh\beta\left(\sum_i^N \frac{J_{ij}}{\sqrt{N}}x_j^* + ax_i^*\right) - x_i^* = 0 \tag{2.3}$$

(note that the case with $c_i = 0$ is considered here). In the large β limit, the tanh function is approximated by the step function, so that the fixed point x_i^* is given by a sequence of $\{-1, 1\}$ that satisfies (2.3). The number of fixed points of (2.3) then increases monotonically with the value of a (Fig. 2.3). If it is large enough (that is, the second term in the brackets in (2.3) is sufficiently larger than the first term), all of the 2^N patterns with any combination of $x_i^* = \pm 1$ (with $\theta_i^* = ax_i^*$) satisfy (2.3). All of these are fixed-point attractors, which are reached by choosing initial conditions close to each $\{-1, 1\}^N$ state. However, for $a = 0$, the number of fixed points satisfying $x_i^* = \tanh\beta(\sum J_{ij}x_j^*/\sqrt{N})$ is much smaller.

[*1] The reason for such agreement comes from the dynamical-systems theory. As shown in the later sections, types of attractors (which genes are expressed, multistability, oscillation) are determined by how nullclines are crossed, whereas a consequence of the epigenetic change is determined by how the nullclines are shifted. Indeed, our model and the Hill-type model have common behaviors as for the crossings and shift of nullclines.

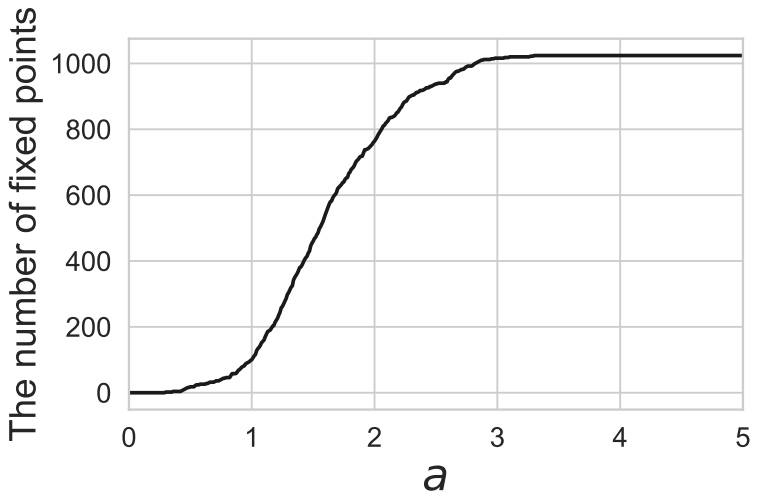


Fig. 2.3: a dependency of the number of fixed points($N = 10$). The number of fixed points monotonically increases for a. At the limit of $a \to \infty$, 2^N fixed points (all gene expression patterns) are allowed.

Here, we focus on a case with sufficiently strong epigenetic feedback, i.e., sufficiently large a, in which all of the possible 2^N states could exist if any value of x_i and θ_i is initially chosen. However, for studying the canalization dynamics, we restrict the initial condition of θ_i as follows: At the initial stages of development, epigenetic modification is not yet introduced [60–64], so that all of θ_is are set to 0. Under this restriction, we investigate which of the 2^N fixed points with $x_i^* = \pm 1$ and $\theta_i^* = ax_i^*$ is reachable through developmental change of the epigenetic modification. As we limit the dynamics to the state of $\theta_i = 0$, we refer to only the final states reached from such initial conditions as attractors throughout the paper (whereas the initial conditions of x_is cover all possible $\{-1, 1\}$ states).

2.3 Attractor generation and pruning

First, we set $N = 10$ and prepare the initial conditions for all gene expression patterns with null epigenetic modification (i.e., 2^N candidates with $x_i = \pm 1, \theta_i = 0$). In the context of the epigenetic landscape, these initial conditions correspond to the balls on the top of the landscape, whereas the valleys are shaped with the change in θ_i and the balls are trapped at the bottoms of the landscape that correspond to the attractor. We then examine which and how many attractors are selected depending on the parameter v.

At the limit of $v \to 0$, i.e., the adiabatic limit in terms of physics, the time scales of the dynamics for x_i and θ_i are well separated. Only after the expression level x_i reaches one of the original attractors with θ_i originally fixed at 0, θ_i begins to show gradual variation. Hence, the number of attractors will be bounded by the expression dynamics when fixing $\theta_i = 0$. At the limit of $v \to \infty$, θ_i reaches $\theta_i = ax_i$ faster, so that all of the 2^N states $x_i^* = \tanh(\beta a x_i^*)$ are attracted depending on the initial x_i values, as long as a is sufficiently large. By considering these two extreme limits, v generally functions as a parameter that limits the final state from all of

the possible 2^N states. Now, from naive estimation based on the above two limits, it might be expected that the number of attractors will monotonically increase with v. Indeed, such monotonic increase could be observed for 80% of randomly chosen networks J_{ij} for $N = 10$.

For $v \sim 0$, the approach to the attractor is completed before epigenetic modification and then θ_i is fixed accordingly. With the introduction of v, θ_i increases or decreases depending on the initial value of x_i. If this process for x_i is fast, x_i is fixed to ± 1 depending on the initial condition; that is, before the approach to the original attractor. Hence, the original basin of attraction is partitioned. With the increase in v, more partitions progress; accordingly, the few attractors that exist at $v = 0$ are successively partitioned toward 2^N states with the increase in v. In this case, for a given v, fixation simply occurs from the neighborhood of each on/off-pattern attractor provided by the initial condition. There exists no hierarchical branching to each attractor over developmental time. Moreover, since only the attractor from the neighborhood of the initial expression state is reached, the final state crucially depends on the initial condition, the final state crucially depends on the initial condition. Neither homeorhesis nor robustness in the cell-number ratio is expected, as will be confirmed later.

However, in the case of $N = 10$, approximately 20 % of the randomly chosen matrix J_{ij} shows non-monotonic dependency of the attractor number on v. Here, different attractors are generated and pruned successively with v in the intermediate range of v. This implies that states separated at smaller v converge again with the increase in v, even though the epigenetic feedback tends to separate each x_i to ± 1. With mutual interference between the fast dynamics of x_i and slower dynamics of θ_i, both the convergence of initial states and divergence to fixed states coexist, as will be discussed below. Further, as will be shown, such convergence of orbits in the initial regime can allow for the creation of an epigenetic landscape that satisfies the three postulates of hierarchical branching, homeorhesis, and robustness in the cell-number ratio.

In this non-monotonic case, the basin volume of each attractor, i.e., the fraction of initial conditions from which each attractor is reached, also changes with v. In particular, dominant attractors successively change with v as shown in Fig.2.4b. This scenario is in stark contrast with the case of a monotonic increase in attractor number, where each basin of attraction is simply partitioned to 2^N successively with the increase in v (Fig. 2.5).

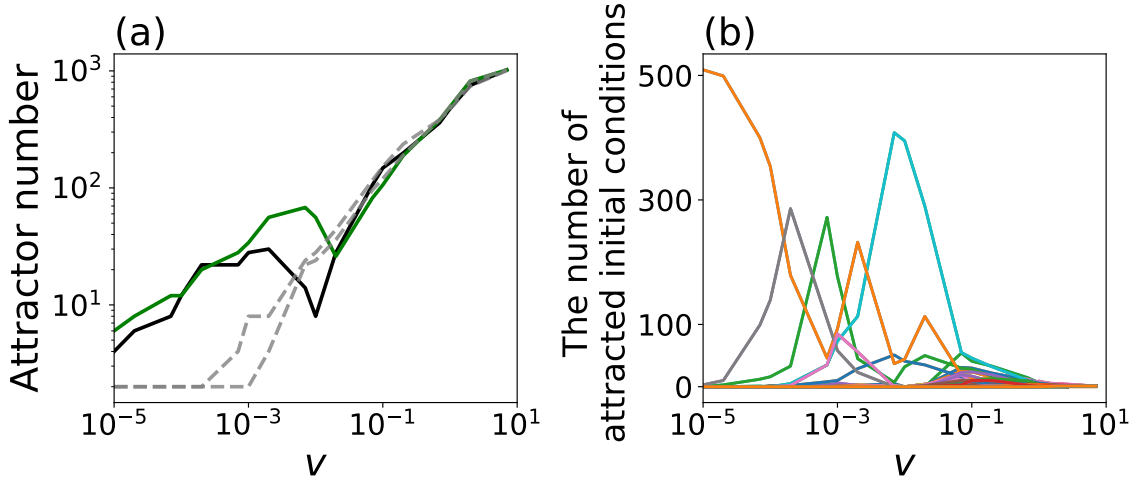


Fig. 2.4: (a) Dependence of the number of attractors (reached states from $\theta_i = 0$) upon v. Grey dotted lines show the case with a monotonic increase of the attractor number against v. The black and green solid lines are examples with non-monotonic dependence on v. Attractors are pruned at $2\times10^{-3} < v < 1\times10^{-2}$. $N = 10$. (b) Dependence of the basin volume of each attractor upon v, for the example of non-monotonic dependence of attractor number shown as the black line in (a). Basin volume is computed by taking 2^N initial conditions of $\{x_i = \pm1\}$ and setting $\theta_i = 0$ initially, and then counting the number of initial conditions reaching each fixed-point state. Each line with a different color shows the basin volume for each different attractor.

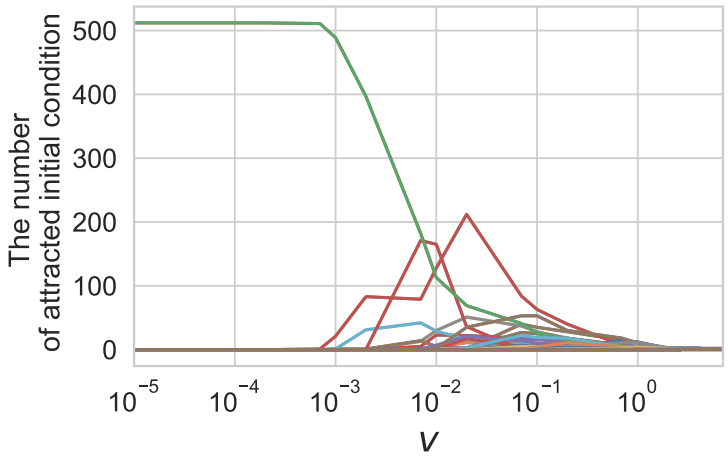


Fig. 2.5: Dependence of basin volume of each attractor upon v for the example in which the attractor number shows monotonic dependency on v.

2.4 Trajectory separation by epigenetic modification: simplest example

To understand how mutual feedback between gene expression and epigenetic modification can lead to the generation and pruning of attractors, we first consider the minimal case with only two genes ($N = 2$). In addition, $c_1, c_2 \neq 0$, which may be also regarded as the inputs from genes other than $i = 1, 2$. We consider the case $J_{11} = J_{22} = 0, J_{12} > 0 > J_{21}$; i.e., one gene activates the other, which then inhibits the first, as shown in Fig.2.6a.

In this simple case, the number of attractors changes as $1 \to 2 \to 1$ with the increase in v over a certain range of parameters c_1, c_2 (Fig. 2.7). For $v < 5.7 \times 10^{-5}$, only trajectories reaching $(-1, 1)$ are realized (Trajectory A)(Fig. 2.7b). By increasing v further, trajectories reaching $(-1, -1)$ then appear (Trajectory B) where $5.7 \times 10^{-5} < v < 7.0 \times 10^{-4}$, and two attractors $(-1, 1), (-1, -1)$ coexist (Fig.2.6b). For larger v ($7.0 \times 10^{-4} < v < 9.1 \times 10^{-3}$), the attractor $(-1, 1)$ disappears completely (Fig. 2.7c). The time course in the development of the two types of trajectories and the change in the basin for each attractor is shown in Fig. 2.8 and Fig. 2.9, respectively.

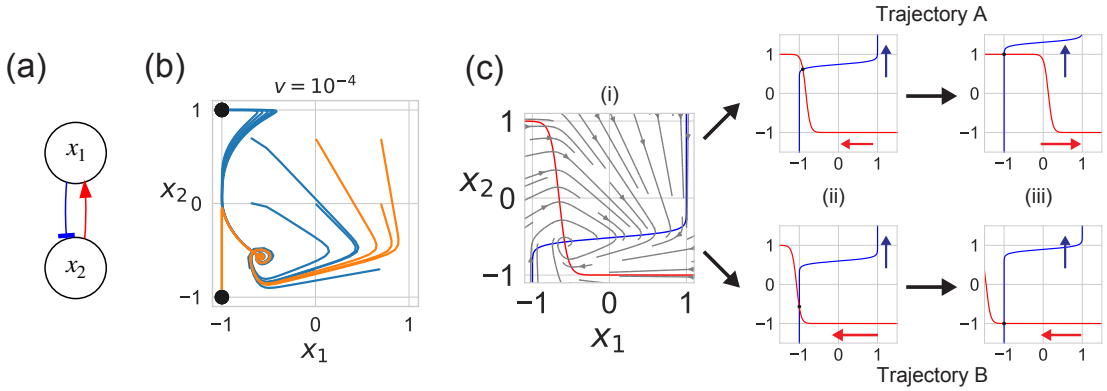


Fig. 2.6: (a) Two-gene system with activation (red arrow) and inhibition (blue arrow): x_2 activates x_1 and x_1 inhibits x_2. The parameter values are chosen as $J_{12} = 0.44, J_{21} = -0.33, J_{11} = J_{22} = 0, c_1 = 0.16, c_2 = -0.15$. (b) Trajectories in x_1, x_2 for $v = 10^{-4}$. Trajectory A reaches the fixed point $(-1, 1)$, whereas Trajectory B reaches $(-1, -1)$. These two types of trajectories coexist (two fixed points as two black dots), depending on the initial condition, for the intermediate value of v. Initial conditions are chosen at even intervals per 0.5 in the phase space of (x_1, x_2). (c) Analyses of the two types of trajectories according to the motion of two nullclines: blue, corresponding to $dx_1/dt = 0$; red, $dx_2/dt = 0$. (i) (x_1, x_2) approaches the crosspoint of the nullclines if θ_i is fixed, whereas the change in θ_i results in a shift of the nullclines. (ii) For both trajectories, x_1-nullcline (blue line) goes up and x_2 nullcline (red line) goes left first, because (x_1, x_2) first approaches the fixed point at $\theta_i = 0$ starting from any initial condition. Upper: As x_2 exceeds 0, the motion of the x_1 nullcline changes its direction, and (x_1, x_2) reaches the fixed point $(-1, 1)$. This gives Trajectory A. Lower: Before x_2 reaches 0, the x_2 nullcline crosses the x_1 nullcline vertically in Trajectory B so that x_2 remains negative and the motion of the nullclines do not change their direction; thus, (x_1, x_2) reaches $(-1, -1)$.

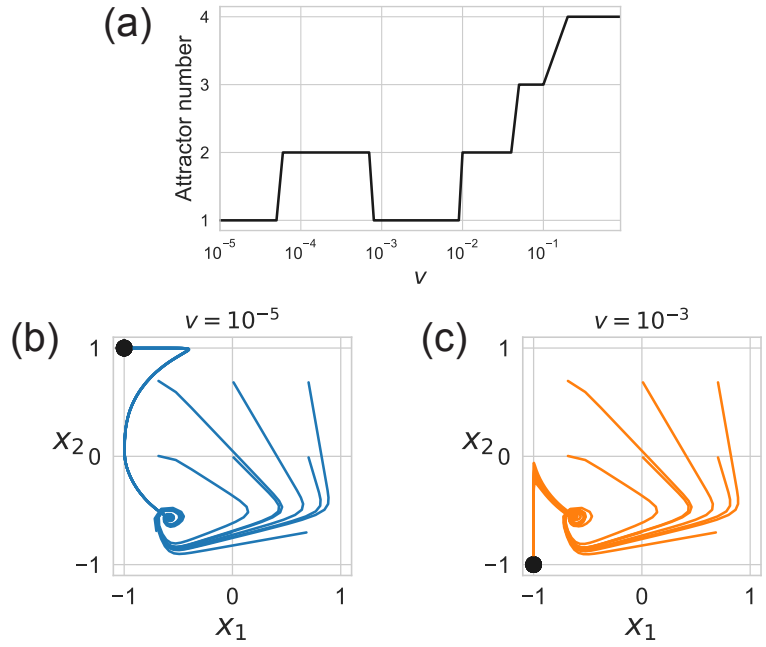


Fig. 2.7: (a) v dependency of attractor number. The attractor number non-monotonically changes as $1 \to 2 \to 1$ for $v < 10^{-2}$. (b) Trajectories realized from several initial conditions for $v = 10^{-5}$. For small v, only Trajectory A (a type of trajectory reaching $(-1, 1)$) is realized. (c) Trajectories for $v = 10^{-3}$. For large v, only Trajectory B (a type of trajectory reaching $(-1, -1)$) is realized. The generation of these trajectories is outlined in Fig. 2.6c.

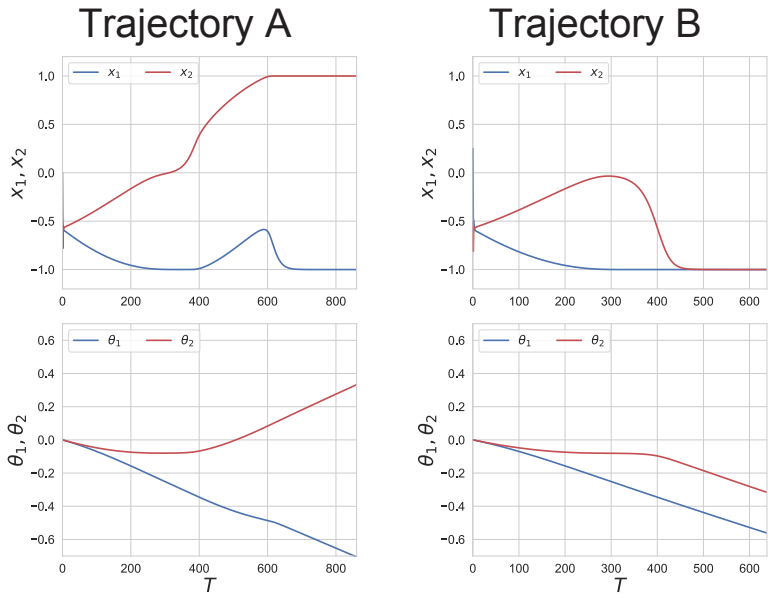


Fig. 2.8: Time-dependent development of Trajectory A and Trajectory B in the minimal two-gene model.

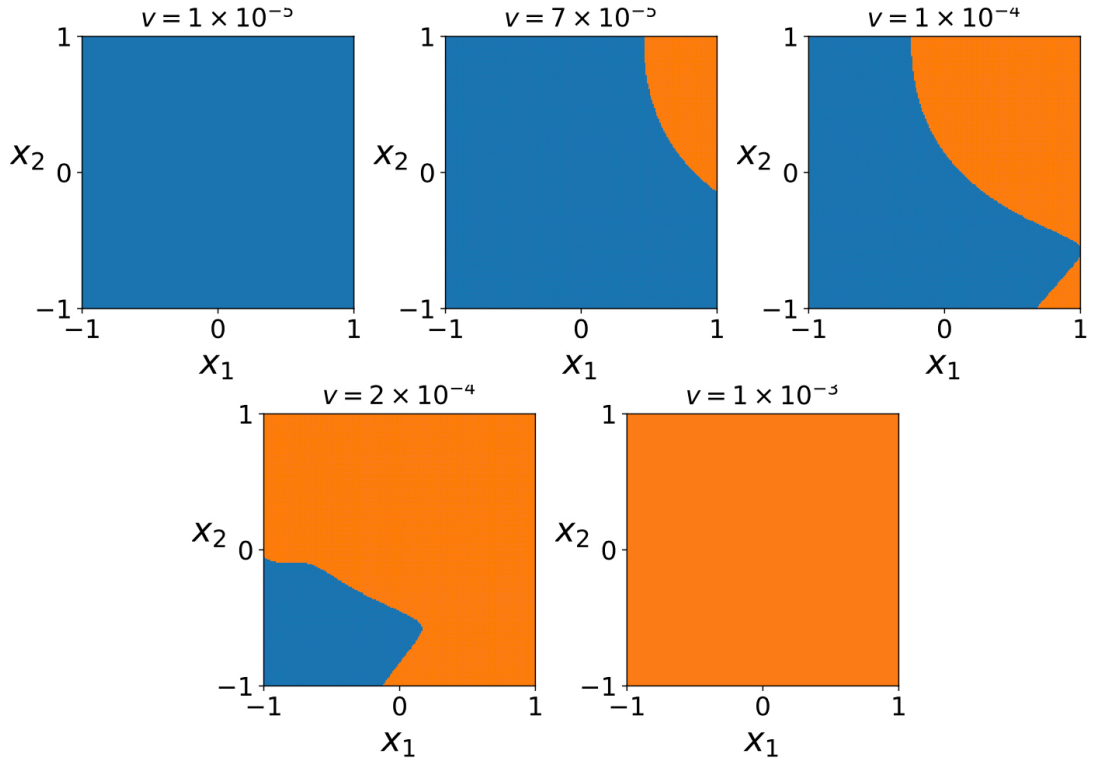


Fig. 2.9: The change of basin for each attractor in the two-gene model. The orange region corresponds to the basin attracted to $(-1, 1)$, and the blue region corresponds to the basin attracted to $(-1, -1)$.

The above v dependency of attractors is explained as follows. When v is small, the dynamics are approximated by the means of "adiabatic elimination"; i.e., x_i reaches the fixed point for a given θ_i, whereas θ_i changes slowly. For given θ_i, the $\{x_i\}$ dynamics are analyzed by the two nullclines, given by

$$dx_1/dt = 0 \;\rightarrow\; x_1 = \tanh\beta\left(\frac{J_{12}}{\sqrt{N}}x_2 + \theta_1 + c_1\right), \tag{2.4}$$

$$dx_2/dt = 0 \;\rightarrow\; x_2 = \tanh\beta\left(\frac{J_{21}}{\sqrt{N}}x_1 + \theta_2 + c_2\right). \tag{2.5}$$

When v is small, while x_i moves towards the crosspoint of the two nullclines, as θ_i slowly changes according to (2.2), the nullclines are slowly shifted.

When this adiabaticity condition is satisfied, only Trajectory A is realized (Fig. 2.7b): at $\theta_i = 0$ (null epigenetic modification), there is a stable fixed point as the crosspoint of the two nullclines at $x_1 < 0$ and $x_2 < 0$ (Fig.2.6c(i)). Then, according to (2.2), each nullcline is shifted as follows: the x_1-nullcline (i.e., $dx_1/dt = 0$ nullcline) goes up, whereas the x_2-nullcline (i.e., $dx_2/dt = 0$ nullcline) goes left. As a result, the crosspoint of the two nullclines itself moves up and left, thus reaching above $x_2 = 0$. Consequently, the shift of the x_2-nullcline changes its direction (as the sign of $d\theta_2/dt$ is approximately given by the sign of x_2). Accordingly, the crosspoint of the nullclines continues to move up, reaches $(-1, 1)$, and then stops.

However, by increasing v, the faster movement of the nullclines generates another trajectory, Trajectory B. First, the crosspoint of the two nullclines moves to the

left and up, in the same way as observed for Trajectory A. However, owing to the faster change in θ, the x_2-nullcline shifts to the left so quickly that the two nullclines cross vertically (see Fig.2.6c (ii) Trajectory B), and the crosspoint does not go above $x_2 = 0$. As a result, the crosspoint moves to the left and down to $(-1, -1)$, where (x_1, x_2) is fixed for some initial conditions. Here, (x_1, x_2) first approaches the fixed point at $\theta_i = 0$ for both Trajectories A and B, and then owing to slight difference in the initial conditions, (x_1, x_2) is directed either to $(-1, 1)$ or $(-1, -1)$.

By increasing v beyond 9.1×10^{-3}, the shift in the nullclines is accelerated, so that the two nullclines cross vertically for all of the initial conditions. In this case, Trajectory A is not realized for any initial condition, and all of the initial conditions are instead attracted to $(-1, -1)$ (Fig. 2.7c).

Hence, the attractor number increases due to the divergence in the motion of the nullclines depending on the initial conditions of $\{x(i)\}$. With a further increase in v, the attractor is pruned because nullclines move faster and no longer split into two directions of motions due to the faster change of θ_i.

2.5 Generation and pruning of attractors from an oscillatory state

The two-gene minimal model described above suggests how the interplay between fast x dynamics and a slow nullcline shift leads to divergence in trajectories, thereby resulting in non-monotonic change in the attractor number. By contrast, for $N = 10$, the non-monotonic behavior of attractor number against v mostly adopts a limit-cycle attractor at $\theta_i = 0$. The frequency of networks showing such behavior is much larger for the limit-cycle case, along with the number of generated and pruned attractors in the intermediate range of v (See Fig.2.10).

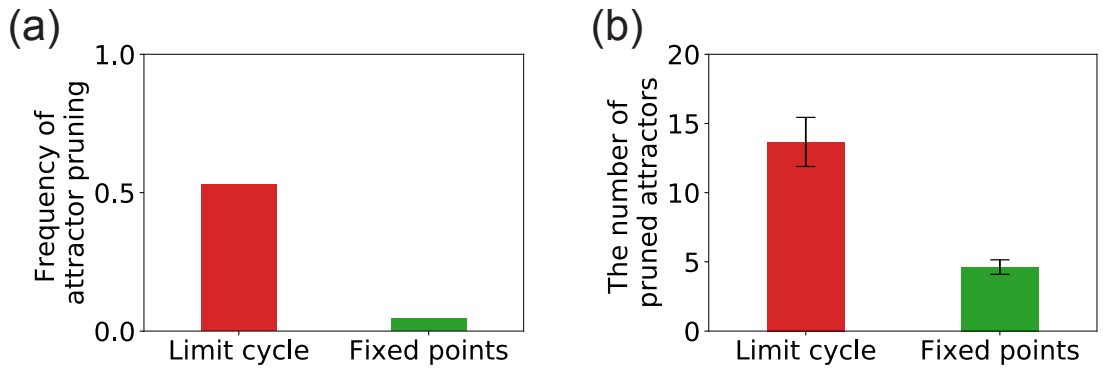


Fig. 2.10: (a) Fraction of the regulatory matrices J_{ij} that exhibit attractor pruning with the increase in v (here defined as a decrease in attractor number of more than 4). The cases with the initial limit-cycle attractors at $\theta_i = 0$ (red) and those with fixed points (green) are sampled separately. (b) Average number of pruned attractors, defined as the difference between the local maximum and local minimum of attractor number against the change in v. Case with a limit cycle (red) and fixed points (green). See Fig. 2.11 for more details.

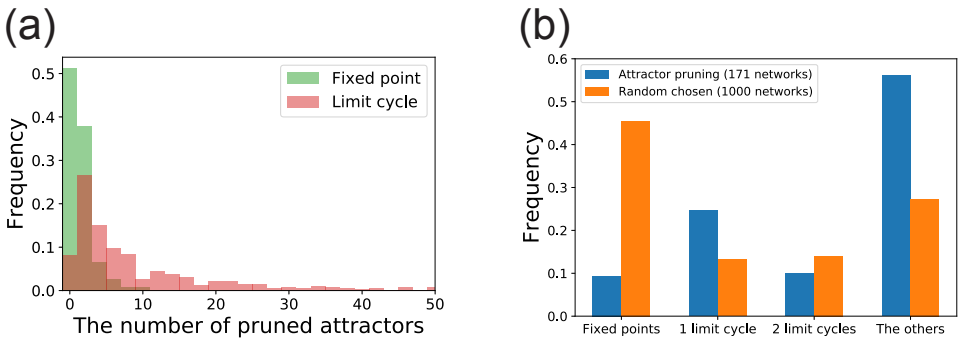


Fig. 2.11: (a) Comparison of the number of pruned attractors for the case in which the original attractor (at $\theta_i = 0$) is the limit cycle and the case in which the original attractor is a fixed point. The number of pruned attractors is defined as the difference between the local maximum and local minimum. If the attractor number monotonically increases with v, the number of pruned attractors is 0. (b) (Blue) Sampling only from the matrices in which attractor pruning occurs. The case in which the original attractor is the limit cycle is most frequently observed. (Orange) Sampling from randomly chosen matrices. The case in which the original attractor is a fixed point is most frequently observed.

This relevance of the limit cycle to the generation and pruning of multiple attractors is explained as follows. First, as the limit cycle travels over a larger portion in the phase space of $\{x_i\}$, the variation in the change in $\{\theta_i\}$ is enhanced so that more attractors can be generated with the increase in v. These attractors are generated hierarchically by branching trajectories successively, stemming from the original limit-cycle orbit. However, with the increase in v, the initial oscillation

is destroyed due to the faster change in θ (shift of nullclines), so that the top of the hierarchy in branching trajectories is destroyed, leading to a drastic decrease in the attractor number.

This hierarchical attractor generation from limit-cycle (HAGL) is illustrated in a simple three-gene system with a limit-cycle attractor (Fig.2.12a). In this three-gene system, only one attractor is reached for small v where the adiabatic condition is satisfied (Fig. 2.13a). With a further increase in v, however, three attractors are reached ($4 \times 10^{-4} < v < 9 \times 10^{-3}$). The trajectories reaching these attractors initially show oscillation around the original limit cycle at $\theta_i = 0$, and then separate into two groups, as shown in Fig.2.12b): two fixed-point attractors are generated from one group, whereas one fixed-point attractor is generated from the other group. Thus, the attractors are generated hierarchically.

In this v region, the branching in gene expression patterns occur in accordance with that in θ_i as shown in Fig.2.12(c), where the first branching occurs in θ_2 at $t \sim 25$, then in θ_1 at $t \sim 50$ and finally at θ_3 at $t \sim 100$. The change in θ_i is triggered by the change in x_i, but it then supports and fixes each of the branching of gene expression patterns. In this sense, the deepening of valleys is guided by θ_i, as is metaphorically represented by strings underlying the Waddington's diagram in Fig. 2.1b.

With the increase in v, the initial limit-cycle orbit is destroyed before the separation into two groups, so that the number of attractors is reduced from three to one ($v \sim 9 \times 10^{-3}$) (see Fig. 2.13 for more details).

Most of the generation-and-pruning of multiple attractors can be understood as HAGL. Note that for much larger N, limit-cycle attractors (or sometimes chaotic attractors) exist more often in the model (2.1) with $\theta_i = 0$, as previously investigated in neural network models [65, 66]. Therefore, the generation-and-pruning of multiple attractors are expected to be ubiquitous. For confirmation, we simulated the model with $N = 100$. Although sampling all 2^N initial conditions $\{x_i = \pm 1; i = 1, \ldots, N\}$ is numerically difficult, simulations with partial sampling showed that non-monotonic change in the attractor number occurred for most of the randomly chosen J_{ij} matrices (Fig. 2.14) where HAGL is commonly observed.

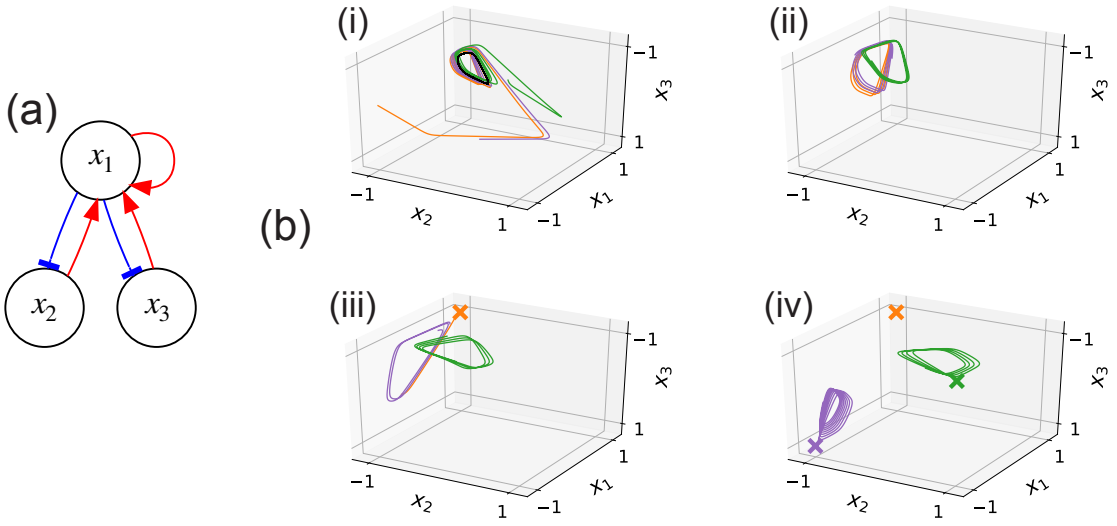


Fig. 2.12: (a) Three-gene system with $J_{11} = 0.26, J_{12} = J_{13} = 0.35, J_{21} = -0.4, J_{31} = -0.36, c_1 = 0.26, c_2 = c_3 = 0.17$. (b) Hierarchical attractor generation from limit cycle (HAGL). $v = 10^{-3}$. (i) Trajectories from different initial conditions, plotted by different colors, approach a limit-cycle attractor at $\theta_i = 0$. $0 < t < 10$. (ii) Trajectories are separated into two groups (green line shows one group, and orange and purple lines show another group), depending on the initial condition. $40 < t < 50$. (iii) Further separation of the group of trajectories is shown in orange and purple. $80 < t < 90$. (iv) Three trajectories reaching distinct fixed points. $120 < t < 150$. (c) Time development of θ_1(i), θ_2(ii), and θ_3(iii). Each line color corresponds to trajectory in (b). HAGL in x space is supported by θ_i, similarity as the strings beyond the epigenetic landscape.

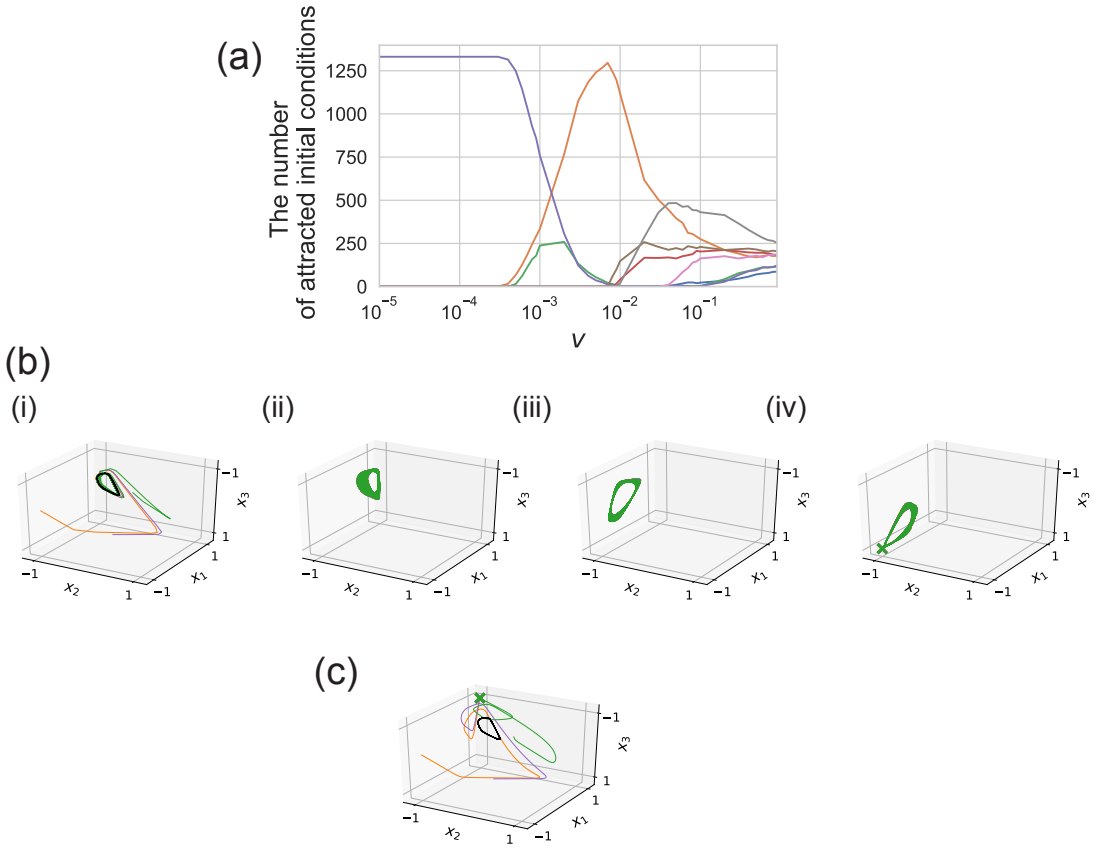


Fig. 2.13: (a) The fraction of basin volume for each attractor. For small v, only one attractor is allowed ($v < 4\times10^{-4}$). For an intermediate range of v, ($4\times10^{-4} < v < 9\times10^{-3}$), three attractors coexist. With an increase in v, certain attractors strongly attract initial conditions due to attractor pruning. (b) Trajectories for $v = 10^{-4}$ (same initial conditions as in Fig. 2.12a). (c) Trajectories for ($v = 9\times10^{-3}$).

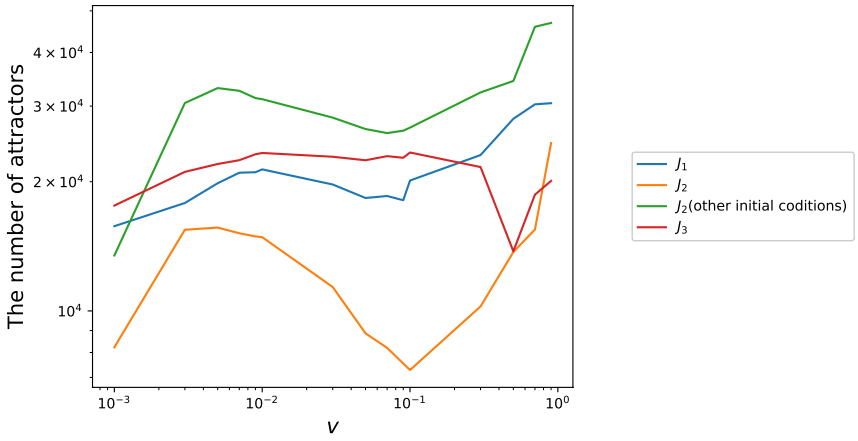


Fig. 2.14: The change of attractor number with much larger N ($N = 100$). Numerical simulation with partial sampling (sampling from randomly chosen 10^6 initial conditions). Blue, orange, and red lines show the result from different matrices. The green line shows the result using the same matrices as used for generation of the orange line but with sampling under different initial conditions.

2.6 Epigenetic landscape and homeorhesis

Thus, HAGL satisfies the first postulate of Waddington's landscape: hierarchical branching. Now, we consider the other two postulates of homeorhesis and robustness in the cell-number ratio. For this purpose, we first need to determine the axes X and Z in the landscape.

As discussed above, the X axis represents the cellular state, which can be extracted from $\{x_i\}$ using PCA. Here, we adopt the 1st PCA mode of $\{x_i\}$ as X. Each valley corresponds to an attractor stabilized by the slow epigenetic change. To explore robustness in the developmental course and generated epigenetic landscape, we introduce noise in (2.1) and (2.2). We adopt the Langevin equation by adding Gaussian white noise $\eta_i(t)$ with $< \eta_i(t)\eta_j(t') >= \sigma\delta_{ij}\delta(t-t')$, with δ_{ij} as Kronecker delta and $\delta(x)$ as a delta function. In general, the specific attractor that is reached depends on the initial condition and perturbation by internal noise. By taking the number of cells under noise, each cell reaches one of the attractors (and stays in its vicinity even under noise). Then, one can compute the number distribution of $P(X)$. As Z is lower, the state with X is more frequently reached. By analogy with the relationship between free energy and probability in thermodynamics, one can adopt $Z = \log(1/P(X))$. Then, the epigenetic landscape can be depicted using the height Z as a function of X.

To compute $P(X)$, we first choose an initial condition of cells (or distribution around a given initial pattern of X). For each initial value, X is computed as a result of time evolution. By starting with a sufficient number of cells, the distribution $P(X)$ is obtained, which may depend on the initial condition of cells. Then, to examine the robustness of the landscape, we explore whether the time evolution of the distribution $P(X)$ is robust against the change in the initial condition of

cells.

First, when v is large, any of the 2^N states is approached from the vicinity of each of the initial expression patterns $\{x_i = \pm 1\}$. In this case, the specific state that is attracted as well as the number distribution of cells for each state crucially depend on the choice of initial conditions. Hence, $P(X)$ is not robust to the change in the initial conditions.

Next, we consider the case with monotonic dependence of attractor number upon v. In this case, if v is not so large, the number of attractors n_A is much smaller. Nevertheless, the specific attractor the cell state reaches is still predetermined by how close the expression state at $\theta_i = 0$ is to the final expression state. The initial x_i state is partitioned into n_A basins, from each of which only one attractor (valley) is generated. Hence, $P(X)$ crucially depends on the initial distribution of x_i's (see Fig. 2.18, 2.19).

In contrast, for HAGL, the postulated robustness is achieved if v is in the intermediate region in which multiple attractors are generated, as shown in Fig. 2.18, 2.19. The obtained $P(X)$ is almost completely independent of the initial conditions of cells(see Fig. 2.19). For most initial conditions, all of the attractors are reached, and the fraction of cells reaching each attractor under noise is quite stable against the change in the initial distribution of $\{x_i\}$ (Fig. 2.18, 2.19). In this case, from any initial conditions, the limit-cycle attractor (at $\theta_i = 0$) is first reached. With the epigenetic feedback, the cells are then distributed to each attractor depending on the phase of the oscillation. Hence, the time course of differentiation to each attractor (cell type), as well as the fraction of each attractor (the number ratio of each cell type) are both independent of the initial distribution of cells.

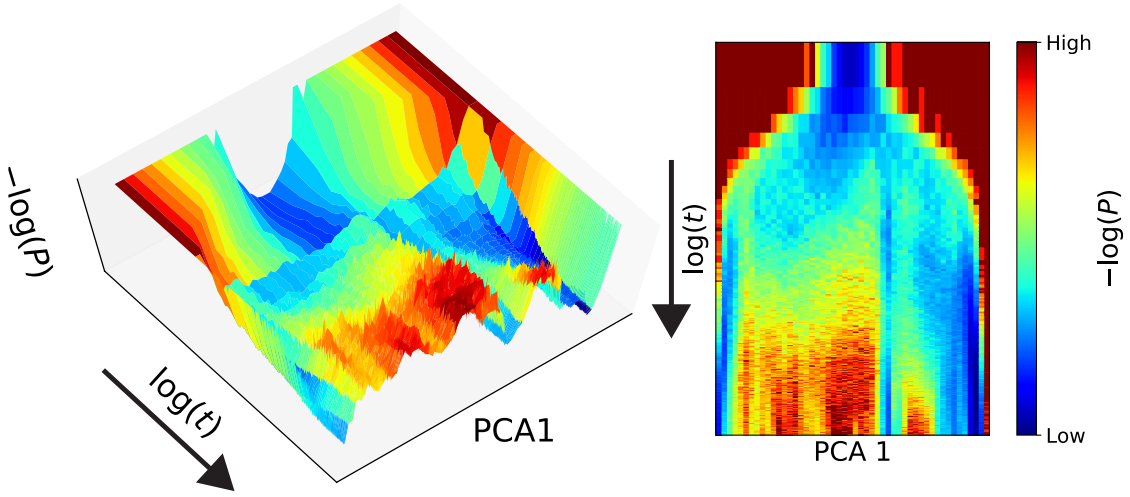


Fig. 2.15: Epigenetic landscape generated from the temporal evolution of cellular states for given J_{ij} that exhibits HAGL. $N = 40$ and c_i is set to a random value sampled from the normal distribution with average 0, variance 0.1. We adopt 1-mode PCA to represent a one-dimensional scalar variable X and $Z = -\log(P(X))$ indicating the depth of valley, where $P(X)$ is the distribution of X over cells developed under noise, and is plotted against time given by $\log(t)$. Red indicates large (i.e., low frequency) and blue indicates small values. The amplitude of Gaussian white noise $\sigma = 0.1$. The right figure shows a one-dimensional representation with the horizontal axis as X and vertical axis as scaled time (from top to bottom), whereas the color represents Z.

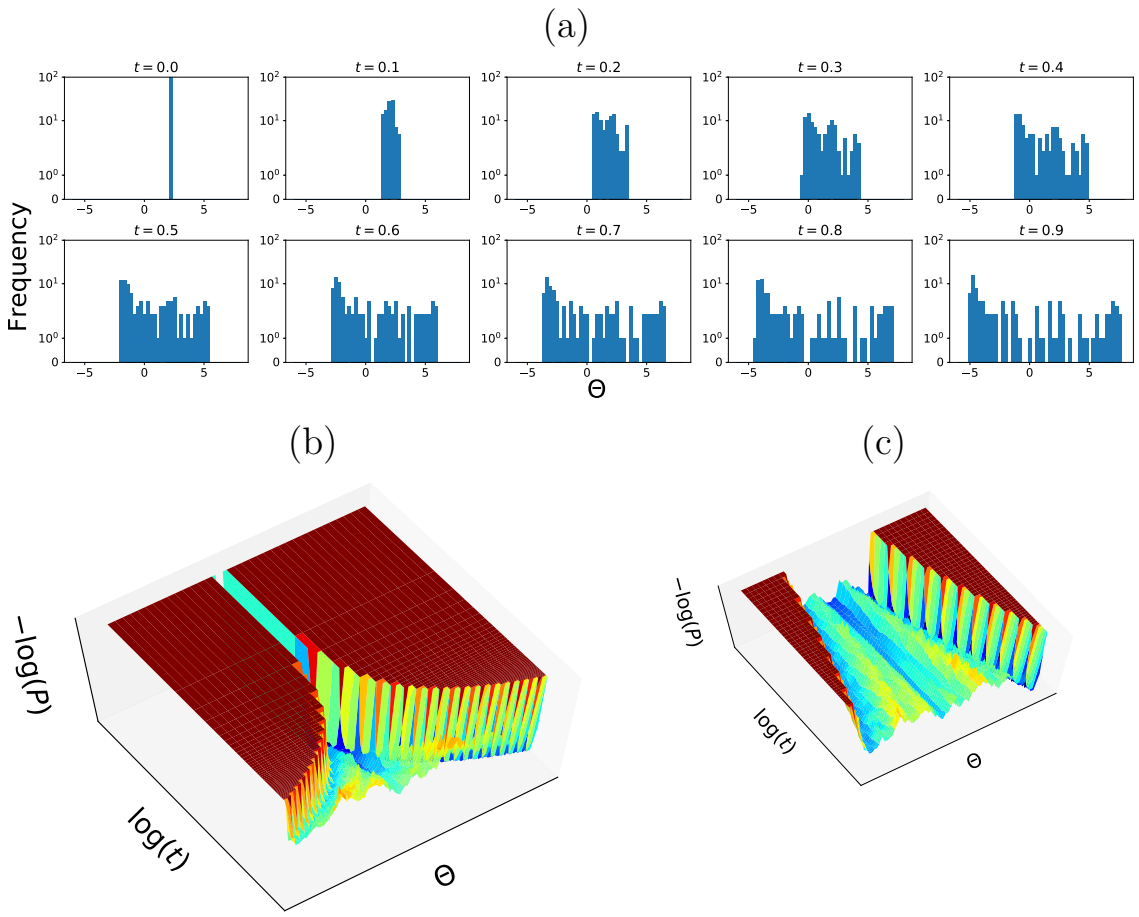


Fig. 2.16: (a) Time development of the histgram of $P(\Theta)$. Each histgram shows distribution of Θ at given (normalized) time. At $t = 0$, distribution of Θ is gathered in one point (reflecting we choose initial θ_i value as $\theta_i = 0$). As time development, distribution of Θ is branched hierarchically, same as the epigenetic landscape in Fig. 2.15. (b) The epigenetic landscape, drawn by Θ, value of θ_i in PCA 1 mode, in Fig. 2.15. (c) Its expansion plot in $0.4 < t < 1$.

We can then depict the epigenetic landscape according to the time evolution of $P(X)$. Here, X (in Fig. 2.1) is given by the 1st PCA mode from $\{x_i\}$ obtained from a distribution of initial conditions. The landscape is depicted by $Z = -\log P(X)$, so that the bottom of the lower valley has a higher population density. The landscape thus depicted is given in Fig.2.15, which shows both the hierarchical branching and robustness to the initial expression or noise. This landscape, as in the case of Fig. 2.12, is shaped with the change in θ_i. Time development of distribution Θ (corresponding to θ_i in the PCA 1 space) is shown in Fig. 2.16a), whereas the landscape thus depicted is given in Fig. 2.16b, c. The figures represent how Θ supports the branching trajectories as in the strings in the epigenetic landscape in Fig. 2.1b.

Finally, we quantitatively characterize the robustness of the final distributions of cellular states reached from different initial distributions. Let us define $P^{\mu}(X)$ as the distribution of X reached from a given initial condition of x_i, indexed by μ (e.g., $x_i(t = 0) = \eta_i^{\mu}$, where $\eta_i^{\mu}(i = 1, \ldots, N)$ is one random sequence in $[-1, 1]$,

whereas $\nu \neq \mu$ denotes a different random sequence). As the measure for the distance between two distributions $P^{\mu}(X), P^{\nu}(X)$ generated from different initial distributions, we adopt the KL divergence $D_{\mathrm{KL}} = \sum_X P^{\mu}(X) \ln\{P^{\mu}(X)/P^{\nu}(X)\}$ for a pair of two distributions $P^{\mu}(X), P^{\nu}(X)$ obtained from two samples μ and ν starting from different initial conditions. If D_{KL} is small, a similar distribution $P(X)$ (i.e., a similar landscape) is obtained, independent of the initial condition, thereby implying robustness at the distribution level. $\overline{D}_{\mathrm{KL}}$ is computed by averaging over the samples μ and ν, which is plotted in Fig.2.17 for the case of monotonic attractor number dependency on v and the non-monotonic HAGL case. As shown in Fig.2.17, the $\overline{D}_{\mathrm{KL}}$ value is kept small up to a large value in v (e.g., $v \leq 10^{-2}$) for the HAGL case. This quantitatively demonstrates that differentiation from the oscillatory state through epigenetic fixation shows higher robustness in the distribution of cellular states.

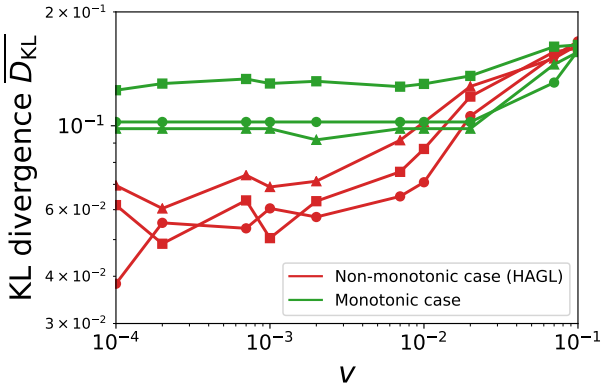


Fig. 2.17: $\overline{D_{\mathrm{KL}}}$ representing the averaged Kullback-Leibler divergence between two distributions of cellular states developed under noise ($\sigma = 0.1$). First, $P(X)$ is computed from 500 cells developed from a given initial condition and J_{ij}. The distribution $P^{\mu}(X)$ is computed over $\mu = 1, 2, \ldots, 10$ starting from different initial conditions. The Kullback-Leibler divergences are then computed over all pairs of 90 distributions and averaged to get $\overline{D_{\mathrm{KL}}}$ (see also Fig. 2.16 for each distribution form). For HAGL, $\overline{D_{\mathrm{KL}}}$ remains low up to large $v \sim 10^{-3}$ (red lines), whereas in the monotonic case (green lines), it takes on a large value over the full range of v.

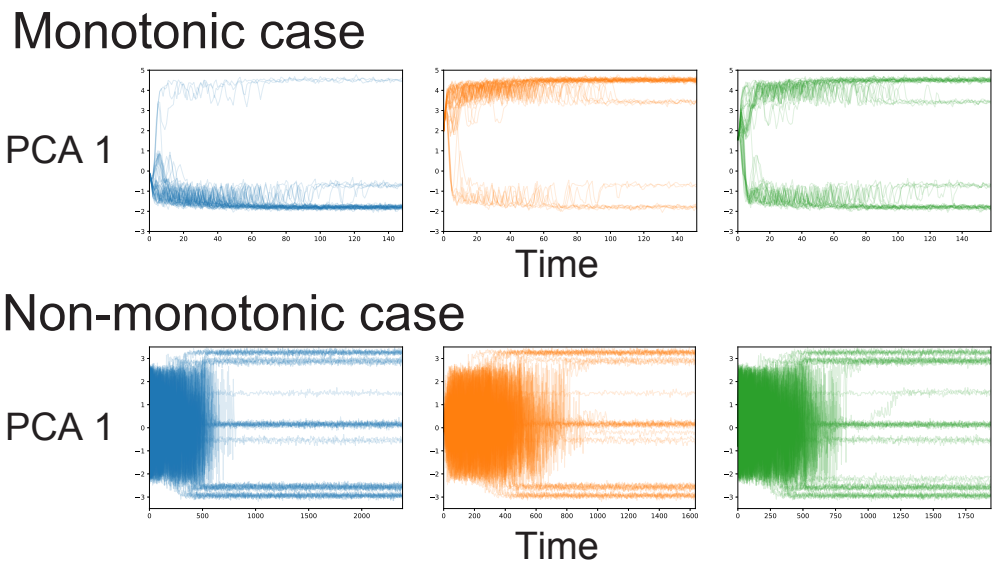


Fig. 2.18: Stochastic simulation. When trajectory separation from the limit cycle is realized [i.e., attractor pruning (AP) occurs], the distribution of cell types has reproducibility even if the initial gene expression pattern is different and gene expression levels change under perturbation. However, when trajectory separation from the limit cycle is not realized, the distribution of cell types strongly depends on the initial condition or shows stochasticity.

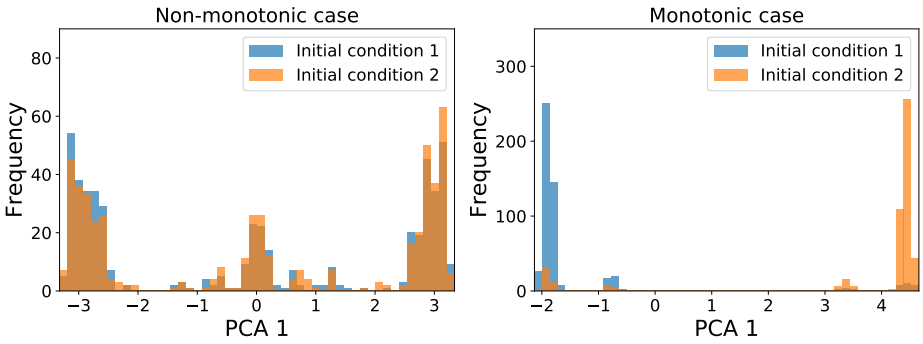


Fig. 2.19: Representative example of the comparison of two distributions simulated from two different initial conditions.

2.7 Discussion

We have introduced a model involving mutual interactions between the expression dynamics controlled by a GRN and epigenetic modification. With more efficient execution of the epigenetic feedback regulation, more attractors with different expression patterns, i.e., more cell types, are generated. In some networks, the initial expression levels are simply embedded into epigenetic modifications, whereas for other networks, mutual feedback between expression levels and modifications

bring about hierarchically ordered attractors from an oscillatory state. In such a case, the attractor number shows non-monotonic change against the rate of epigenetic feedback regulation v. The mechanism of non-monotonic dependency on v, i.e., the attractor generation and pruning, is explained in terms of dynamical systems theory.

By using the change in expression dynamics under the slow epigenetic modification process, Waddington's epigenetic landscape is explicitly depicted, in which the landscape axis (X axis in Fig.2.1) is given by the principal component of the expression pattern; the depth, Y axis, is given by the developmental time with slow epigenetic modification; and the height is given by $-\log(P(X))$ with $P(X)$ as the cell-number distribution of X. In particular, when the original attractor in the absence of epigenetic modification is a limit cycle, the timing of branching to different cell types, number of differentiated cell types, and number fraction of each cell type are all robust to perturbations during the course of development and to the variation of initial conditions. Hence, the generated landscape satisfies the three postulates implicitly assumed in Waddington's landscape: (1) hierarchical branching is supported by the hierarchical attractor generation from the limit cycle; (2) homeorhesis is supported since this branching process is independent of initial conditions and robust to noise; and (3) the cell-number robustness is demonstrated since $P(X)$ is also independent of initial conditions and robust to noise. This robustness in the path and in the cell-number distribution to perturbation is an essential requirement for the development of multicellular organisms [42].

Our theoretical model assumes epigenetic feedback regulation. Although the transient modification in epigenetic factors has been experimentally confirmed [67, 68], the extent to which this modification depends on gene expression is not yet clearly elucidated. Considering that epigenetic change stabilizes the cellular states, it is rather natural to assume positive feedback from the expression level to modification, i.e., if expressed (repressed), it is easier (harder) to be expressed, whereas some molecular mechanisms for such positive feedback have been suggested [26, 27]. However, direct evidence, as well as quantitative estimates for the time scale of epigenetic change, require further experimental elucidation in the future.

The significance of oscillation in the cellular state for the differentiation process was previously discussed [10]. Indeed, the cell state is not fixed but rather involves several oscillatory modes, including circadian and cell-division cycles. Furthermore, oscillatory expression has recently been uncovered for embryonic stem cells [69–72], which is ultimately lost in cells committed to differentiation. Note that the relevance of an oscillatory state to pluripotency was previously discussed in the context of an alternative approach to the epigenetic landscape with respect to inclusion of cell-cell interactions [58]. In this case, the initial oscillation in expression levels is lost with an increase in the cell number and resulting amplification of cell-cell interactions accordingly. Hence, the two approaches, i.e., cell-cell interactions and epigenetic modification, are compatible. Indeed, a model that includes both approaches was previously investigated, in which epigenetic modification of several genes such as Oct4 and Nanog leads to the commitment of cells from an undifferentiated state, which is consistent with experimental observations [73–75].

The canalization in Waddington's landscape is valid for the normal developmen-

tal process. However, through certain external operations, the path of committed cells can be reversed to an undifferentiated state in a process known as reprogramming [6, 76–78]. In the present model, by externally overexpressing some genes for a given time span, the threshold $-\theta_i$ that was initially increased can be decreased so that the expression level recovers, which matches the experimental procedure used to create induced pluripotent stem cells. In the future, it will be important to elucidate the condition required for such reprogramming by identifying the specific genes in the network that need to be overexpressed so as to climb up to the most upstream location in the landscape under the present theoretical framework.

The generation and pruning of attractors that depend on the epigenetic feedback rate is itself an interesting phenomenon in terms of dynamical systems of both fast and slow elements, which requires an analysis beyond the breadth of adiabatic elimination [66]. That is, if the time scales are clearly separated, the change in fast expression would be represented as an attractor change against the slow epigenetic state as a control parameter. In contrast, mutual feedback between the two is important, as shown in the present study with regard to the interaction between the nullclines and the variables. Therefore, an appropriate analytical method that is capable of capturing such feedback dynamics needs to be developed.

Homeostasis, robustness of a steady-state in biological systems has gathered much attention over decades. This, for instance, has been discussed as the stability of the final state (attractor) against perturbations. On the other hand, homeorhesis is concerned with the stability of the time course of a state, against the change in the initial conditions or perturbations. So far, studies on homeorhesis are rather limited: Few examples include relaxation dynamics in signal transduction process independent of the initial condition [79], robust developmental process with cell-cell interaction [10, 43], and robust ecological dynamics in an experiment consisting of algae and ciliates [80]. For homeorhesis to work, the existence of slower time scale and buffering of initial variation will be needed. The hierarchical attractor generation by slower epigenetic feedback after attraction to a limit cycle will provide one general mechanism for the homeorhesis.

Finally, we assume that cellular states start from no epigenetic modification state (that is, initial conditions with $\theta_i = 0$ for all i) in this chapter. We should, however, consider this assumption. Why does development start from these states robustly? We propose the possible scenario in terms of dynamical systems theory by considering cellular reprogramming in the next chapter.

Chapter 3

Dynamical-systems theory of cellular reprogramming

In the development of multicellular organisms, cells with identical genomes differentiate into distinct cell types. This cellular differentiation process has often been explained as balls falling down the epigenetic landscape, as originally proposed by Waddington [3]: balls start from the top of the landscape, and as development progresses, they fall into distinct valleys, which correspond to differentiated cell types. In modern biology, such landscapes are believed to be formed by epigenetic regulation, including DNA and chromatin modifications [12, 13, 15, 17]. For pluripotent cells, these modifications are small, whereas each differentiated cell type has a different epigenetic modification pattern [21–24]. Cells with pluripotency, such as embryonic stem (ES) cells, are located in the vicinity of the first branching point into the valleys, because they can easily differentiate into different types of cells with just slight stimuli [81].

In 2006, a seminal study by Takahashi and Yamanaka reported that differentiated cells can regain pluripotency only by overexpressing few genes (so-called the four *Yamanaka factors*) without direct manipulations of epigenetic modifications. This was termed as reprogramming of induced pluripotent stem (iPS) cells [6]. The reprogramming is often described as "climbing" the epigenetic landscape [30, 31, 82]. This hypothesis, however, has two problems that still need to be addressed: (1) cells have many degrees of freedom, with expression and epigenetic modifications of many genes, whereas reprogramming manipulation involves only few degrees of freedom. How is it possible?; moreover, (2) if the initial pluripotent state is represented by the top of the landscape, it is not a stable point. Thus, how can reprogramming robustly make the cells head toward such an "unstable" state?

Theoretically, these issues should be resolved based on dynamical systems theory. The interplay between fast gene regulation and slow epigenetic dynamics shapes the epigenetic landscape, and differentiated cells are represented by different attractors [32, 33, 35, 35]. Therefore, upon the reprogramming operation, cellular states starting from different attractors first converge into a unique pluripotent state, which is not stable, from which states move toward various attractors afterwards. At first glance, these requirements seem to be incompatible; an unstable state (e.g., repeller) is not attracted from different initial conditions. Hence, to

satisfy these requirements, the pluripotent cell is expected to be represented at least by a saddle that is attracted from many directions and departs only along unstable directions (manifolds), which represent the cell differentiation process, leading to attractors of different destinations. To regain pluripotency by reprogramming, cellular states must be placed on the stable manifold of the saddle by common manipulations from different attractors. Such manipulation, however, would require fine-tuned control. In contrast, reprogramming is mediated by the overexpression of a few common genes across a variety of differentiated cell types. Therefore, some dynamical systems concept beyond just a saddle is needed.

A recent experiment provides some clues on this subject. Temporal oscillations in DNA methylation and corresponding gene expression levels are observed during cellular differentiation [67,68]. In fact, gene expression oscillations have also been reported during somitogenesis and in embryonic stem cells [69, 72, 83], whereas its relevance to cell differentiation has been theoretically investigated for decades [10, 11, 34, 58, 84–86]. Recalling the possible significance of oscillatory dynamics, it is reasonable to consider that if there is an oscillation of fast gene expression around the saddle point of the slow epigenetic dynamics, global attraction to it from broad initial conditions may be attained beyond its stable manifold. As the oscillation dynamics are extended beyond the stable manifold of a saddle, global attraction to the vicinity of the saddle may be possible by taking advantage of the interplay between fast expression and slow modification dynamics.

3.1 Model

Herein, we verified this possibility by using a dynamical system model with a gene regulatory network (GRN) and epigenetic modification. We consider a cell model in which the cellular state was represented by the expression x_i and epigenetic modification level θ_i for each gene i, with $i = 1, 2, \ldots, N$. Gene expression dynamics, with faster time scales, are governed by GRN with mutual activation or inhibition by transcription factors [38,51–54], whereas slower epigenetic dynamics change the feasibility of gene expression, which follows the gene expression patterns. We assumed the epigenetic feedback reinforcement, meaning that as more a gene is expressed (silences), the more feasible (harder) to express. This hypothesis was based on the experimental observations on the Trithorax (TrxG) and Polycomb (PcG) group proteins, two of the essential epigenetic factors for cellular differentiation [87]. Specifically, we adopted:

$$\frac{dx_i}{dt} = F\left(\sum_j J_{ij}x_j + \theta_i + I_i(t)\right) - x_i, \tag{3.1a}$$

$$\frac{d\theta_i}{dt} = \frac{1}{\tau}(x_i - \theta_i). \tag{3.1b}$$

In Eq. (3.1a), gene expression shows an on-off response to the input by adopting the function $F(z) = \tanh(\beta z)$, whereas $\beta = 40$ [*1]. If J_{ij} is positive (negative),

[*1] Although we adopted a symmetric function, the result to be discussed is not changed if asymmetry functions, including the Hill function, are introduced.

gene j activates (inhibits) gene i, whereas J_{ij} is set to 0 if no regulation exists. External input $I_i(t)$ is applied only during reprogramming manipulation to flip the expression of the gene i. For simplicity, $I_i(t)$ takes a constant non-zero value when gene i is overexpressed for reprogramming manipulation and zero otherwise.

In Eq. (3.1a), $-\theta_i$ works as the threshold of the expression of the gene i, which represents the epigenetic modification status (when there is no epigenetic modification, it takes zero). Eq. (3.1b) represents the epigenetic feedback regulation. Following the experimental observation of positive epigenetic feedback [26, 27, 48, 49, 56, 57, 87], we adopted this simple form as its specific form has not yet been confirmed [50, 58, 88, 89]. Here, τ denotes the characteristic timescale for epigenetic modifications, which is assumed to be sufficiently larger than 1; the change in epigenetic modification is much slower than that of gene regulatory dynamics [28, 29, 90].

3.2 3D-Repressilator

Recalling the relevance of oscillatory dynamics, we chose a GRN in which oscillatory dynamics were generated for appropriate θ_i values (specifically at $\theta_i \sim 0$). First, we adopted a repressilator model as a minimal model (see Fig. 3.1), consisting of three genes that repress the expression of the next gene in a cyclic manner [8]. Specifically, we chose $J_{21} = J_{32} = J_{13} = -g = -0.4$ in Eq. (3.1a).

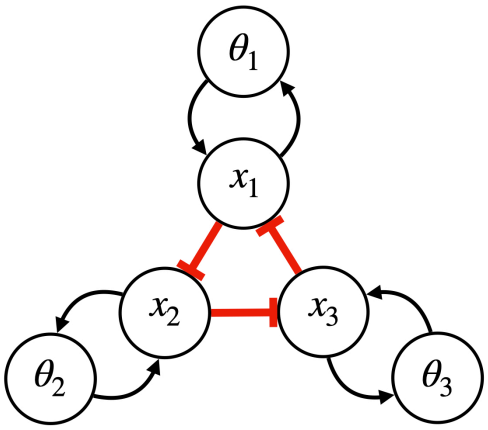


Fig. 3.1: Schematic representation of the cell model, which consisted of N genes, which state is represented by the gene expression x_i and epigenetic modification status θ_i. Red (blue) arrow shows the activation (inhibition) relation according to the gene regulatory matrix J_{ij}. (a) The Repressilator Model comprised three genes that repress each other in a cyclic manner. $J_{21} = J_{32} = J_{13} = -g(= -0.4)$. Black arrow represents positive feedback relation between gene expression and epigenetic modification status.

3.2.1 Dynamics of cellular differentiation and reprogramming in repressilator model

The expression of x_i in this model showed a limit-cycle oscillation when θ_i was close to zero. Thus, for the epigenetic modification to change θ_i following Eq. (3.1b), the states were differentiated into three fixed-point attractors $\{\theta_1, \theta_2, \theta_3\} = \{-1, 1, 1\}, \{1, -1, 1\}, \{1, 1, -1\}$, after first approaching a straight line $\theta_1 = \theta_2 = \theta_3$, as shown in Fig. 3.2a [*2] (see also Fig. 3.3a). In these fixed points, $d\theta_i/dt = 0$ were satisfied; that is, the differentiation of expression x_i was embedded into the epigenetic modification θ_i.

Now, we considered "reprogramming." Starting from one of the differentiated fixed points, we added external input $I_i(t)$ to invoke the transient oscillation again (black dotted line in Fig. 3.2b). Later, $I_i(t)$ was set at zero. After reprogramming manipulation, they approached a line with $\theta_1 = \theta_2 = \theta_3$ around the origin, and then deviated from the line to one of the three fixed points (Fig. 3.2b), in the same manner as the differentiation process. During this reprogramming process, memory of the differentiated states was erased. Once the oscillation in x was recovered, the approach to the straight line and deviation from it always followed (Fig. 3.2c).

[*2] Considering positive/negative symmetry, six attractors exist in the whole space. In this chapter we considered only the side $\sum \theta_i > 0$.

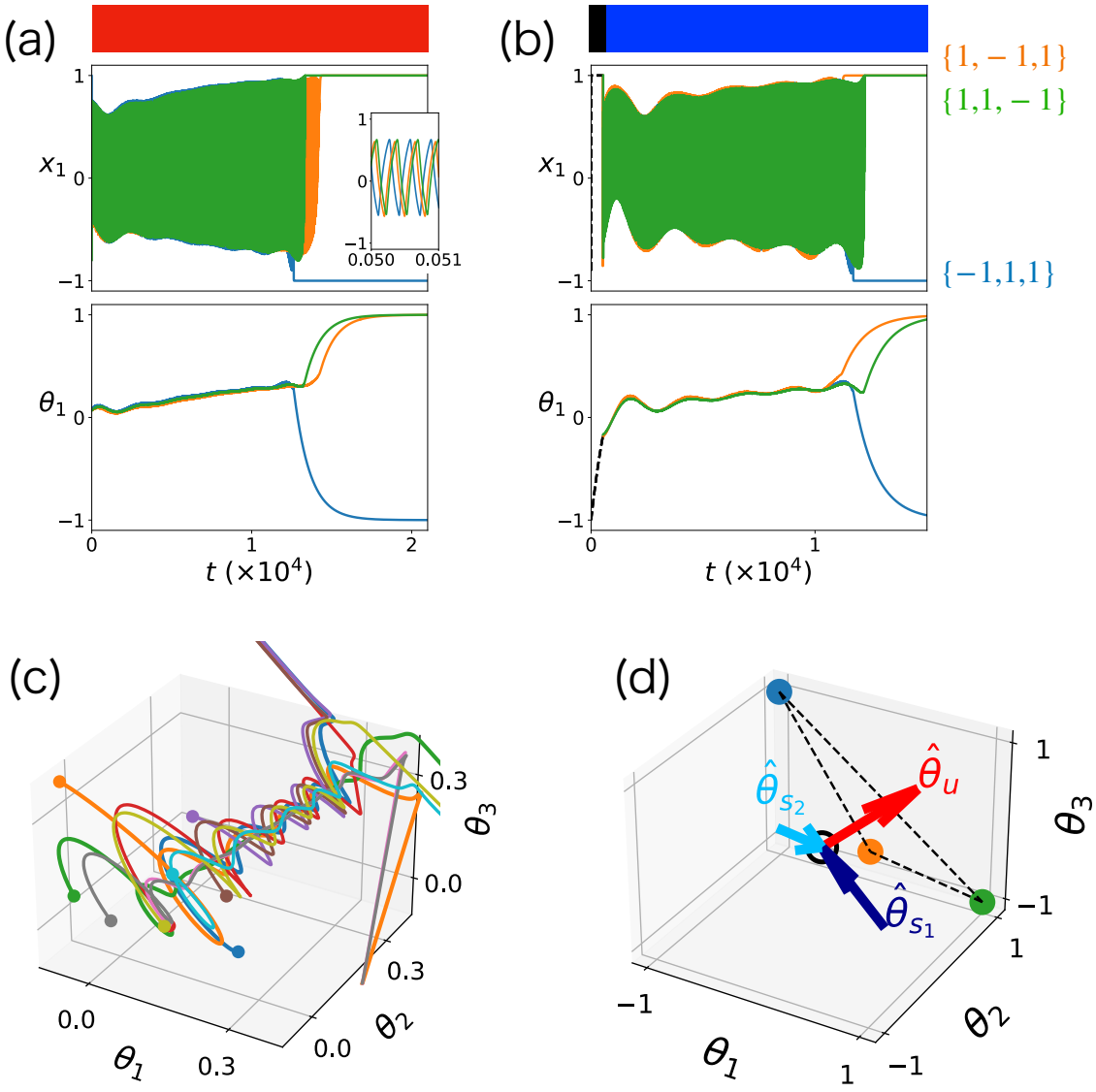


Fig. 3.2: (a) Cell differentiation and (b) reprogramming of the repressilator model with $\tau = 10^3$. Upper bar indicates differentiation (red), reprogramming manipulation (black), and later process (blue) without it ($I_i = 0$). We plotted the time development of x_1, θ_1 (see Fig. 3.3ab for the time series of all variables). (a) Three trajectories are sampled from slightly different initial conditions near $\theta_i = 0$. (b) From the fixed point $\{-1, 1, 1\}$, we tested three slightly different time spans to add external input (520, 530, 540). After reprogramming manipulation, the cellular state first approached $\theta_1 = \theta_2 = \theta_3$, and then differentiated to three fixed points again. (c) Trajectories through reprogramming and differentiation plotted in the $(\theta_1, \theta_2, \theta_3)$ space. Ten attempts were overlaid by considering the initial conditions in $x_i \in [-1, 1], \theta_i \in [-0.25, 0.25]$, which allowed oscillation. Attraction for the straight line $\theta_1 = \theta_2 = \theta_3$ and departure from it was discernible. (d) Eigenvector $\{\boldsymbol{v}_k\}$ and variables $\{\hat{\theta}_k\}$ (see text). Three colored points represent the fixed points $\{-1, 1, 1\}$ (blue), $\{1, -1, 1\}$ (orange), and $\{1, 1, -1\}$ (green).

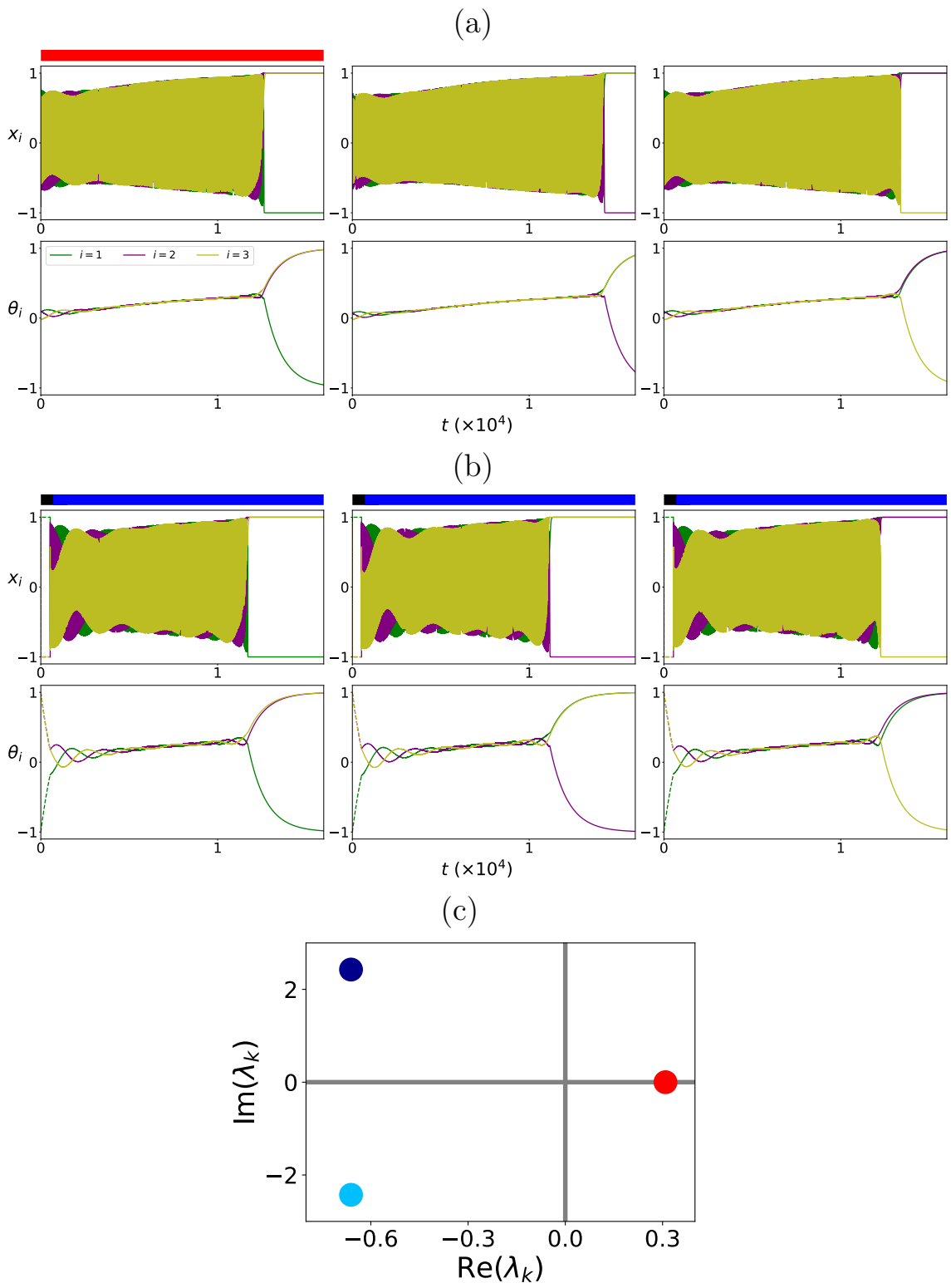


Fig. 3.3: Time development of all variables x_i and θ_i of the (a) cell differentiation process and (b) reprogramming process in the repressilator model. (c) The eigenvalues λ_k of the $\partial\Theta_i/\partial\theta_j$ matrix given by the fixed-point analysis of Eq. (2) around $\{\theta_1, \theta_2, \theta_3\} = \{0, 0, 0\}$.

3.2.2 Adiabatic elimination

Next, we studied how the attraction to the straight line first occurs, followed by differentiation progression. For it, we considered the adiabatic limit of $\tau \to \infty$. For fixed θ_i, we first obtained the attractor x_i. Then, the evolution of θ_i was obtained by replacing x_i in Eq. (3.1b) by its time average $\bar{x}_i$ for a given θ_i, as follows:

$$\frac{d\theta_i}{dt} = \bar{x}_i(\theta_i) - \theta_i (\equiv \Theta_i). \tag{3.2}$$

In the three-variable Eq. (3.2), $\{\theta_1, \theta_2, \theta_3\} = \{0, 0, 0\}$ is a fixed point solution because $x_i(t)$ showed a symmetric limit-cycle oscillation, such that $\bar{x}_i = 0$ for all i therein for $\{\theta_1, \theta_2, \theta_3\} = \{0, 0, 0\}$. By slightly perturbing θ_i as a parameter, $\bar{x}_j$ changed accordingly. From $\partial \bar{x}_i / \partial \theta_j$, we obtained the Jacobi matrix $\partial \Theta_i / \partial \theta_j$ with eigenvalues $\{\lambda_k\}$ and eigenvectors $\{\boldsymbol{v}_k\}$.

3.2.3 Fixed point analysis of Eq. (3.2)

We considered the Jacobi matrix of Eq. (3.2) in the repressilator model as

$$\frac{\partial \Theta_i}{\partial \theta_j} = \frac{\partial \bar{x}_i}{\partial \theta_j} - \delta_{ij} (i, j = 1, 2, 3), \tag{3.3}$$

and fixed-point analysis around $\{\theta_1, \theta_2, \theta_3\} = \{0, 0, 0\}$. The repressilator model obviously has symmetry against a transformation in $1 \to 2 \to 3 \to 1$ around $\{\theta_1, \theta_2, \theta_3\} = \{0, 0, 0\}$.

$$\frac{\partial \bar{x}_1}{\partial \theta_1} = \frac{\partial \bar{x}_2}{\partial \theta_2} = \frac{\partial \bar{x}_3}{\partial \theta_3} (= a_0), \tag{3.4}$$

$$\frac{\partial \bar{x}_2}{\partial \theta_1} = \frac{\partial \bar{x}_3}{\partial \theta_2} = \frac{\partial \bar{x}_1}{\partial \theta_3} (= a_{+1}), \tag{3.5}$$

$$\frac{\partial \bar{x}_3}{\partial \theta_1} = \frac{\partial \bar{x}_1}{\partial \theta_2} = \frac{\partial \bar{x}_2}{\partial \theta_3} (= a_{-1}). \tag{3.6}$$

Eq. (3.3) can be written as the following circulant matrix:

$$\frac{\partial \Theta}{\partial \theta} = \begin{bmatrix} a_0 - 1 & a_{-1} & a_{+1} \\ a_{+1} & a_0 - 1 & a_{-1} \\ a_{-1} & a_{+1} & a_0 - 1 \end{bmatrix}. \tag{3.7}$$

Here, we considered the following matrix C:

$$C = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{3.8}$$

$\partial \Theta / \partial \theta$ can be expanded as:

$$\frac{\partial \Theta}{\partial \theta} = (a_0 - 1) C^0 + a_{+1} C + a_{-1} C^2. \tag{3.9}$$

Matrix C can be diagonalized by a following 3×3 discrete Fourier transform (DFT) matrix F:

$$F = \begin{bmatrix} (\omega_0)^0 & (\omega_1)^0 & (\omega_2)^0 \\ (\omega_0)^1 & (\omega_1)^1 & (\omega_2)^1 \\ (\omega_0)^2 & (\omega_1)^2 & (\omega_2)^2 \end{bmatrix}, \tag{3.10}$$

where $\omega_j = \exp(2\pi ij/3)$ with $j = 0, 1, 2$ (i, used here, is an imaginary unit). Thus, the circulant matrix $\partial\Theta/\partial\theta$ can be diagonalized using the DFT matrix F.

The jth eigenvector $\boldsymbol{v}_j$ is now written as:

$$\boldsymbol{v}_j = \{(\omega_j)^0, (\omega_j)^1, (\omega_j)^2\}. \tag{3.11}$$

Each eigenvalue λ_j corresponding to the eigenvector $\boldsymbol{v}_j$ is

$$\lambda_j = (a_0 - 1) + a_{-1}\omega_j^1 + a_{+1}\omega_j^2. \tag{3.12}$$

Accordingly, we obtained

$$\lambda_0 = (a_0 - 1) + a_{+1} + a_{-1}, \tag{3.13}$$

$$\lambda_1, \lambda_2 = (a_0 - 1) - \frac{a_{+1} + a_{-1}}{2} \pm \frac{\sqrt{3}i}{2}(a_{-1} - a_{+1}). \tag{3.14}$$

Now, we estimated the value of a_0, a_{+1}, a_{-1}. $a_0 = \partial\bar{x}_i/\partial\theta_i$ is given by the change in gene expression against the change in epigenetic modification of the same gene (see Fig. 3.1a). In our model, we adopted a positive feedback relationship between gene expression and epigenetic modification of the gene i; a_0 is positive. In contrast, a_{+1} is given by the change in gene expression against the change in epigenetic modification of the repressing gene; a_{+1} is negative. Similarly, a_{-1} is positive. In addition, the change in gene expression was amplified throughout the negative repressilator loop; thus, a_0, a_{+1}, a_{-1} should satisfy

$$|a_0| < |a_{+1}| < |a_{-1}|. \tag{3.15}$$

Indeed, from numerical calculations, the values of a_0, a_{+1}, a_{-1} followed these estimations ($a_0 = 0.66, a_{+1} = -1.08, a_{-1} = 1.72$ for $g = -0.4$). From these estimations, we obtained the relation of eigenvalues as:

$$\mathrm{Re}(\lambda_1) = \mathrm{Re}(\lambda_2) < \mathrm{Re}(\lambda_0) \tag{3.16}$$

In the present case ($a_0 = 0.66, a_{+1} = -1.08, a_{-1} = 1.72$ for $g = -0.4$), λ_0 is positive, and λ_1, λ_2 is negative:

$$\mathrm{Re}(\lambda_1) = \mathrm{Re}(\lambda_2) < 0 < \mathrm{Re}(\lambda_0) \tag{3.17}$$

Thus, $\boldsymbol{\theta} = \mathbf{0}$-saddle point consists of an eigenvector $\boldsymbol{v}_0$ with $\lambda_0 > 0$ and eigenvectors $\boldsymbol{v}_0, \boldsymbol{v}_1$ with $\mathrm{Re}(\lambda_1) = \mathrm{Re}(\lambda_2) < 0$. To analyze and plot in real θ space, we adopted $\boldsymbol{v}_u \propto \boldsymbol{v}_0, \boldsymbol{v}_{s_1} \propto (\boldsymbol{v}_1 + \boldsymbol{v}_2)/2, \boldsymbol{v}_{s_2} \propto (\boldsymbol{v}_1 - \boldsymbol{v}_2)/2i$ with a normalization factor. The symmetry of the repressilator against the transformation in $1 \to 2 \to 3 \to 1$ is not broken as long as $\theta_1 = \theta_2 = \theta_3$ is satisfied. Therefore, the direction of the eigenvector $\boldsymbol{v}_u$ always coincides with that of the unstable manifold.

3.2.4 Non-linear region of repressilator model

As shown in Fig. 3.2d, $\{\theta_1, \theta_2, \theta_3\} = \{0, 0, 0\}$-fixed point was a saddle, with the eigenvector $\boldsymbol{v}_u = \{1, 1, 1\}/\sqrt{3}$ corresponding to $\lambda_u > 0$ (unstable axis), and $\boldsymbol{v}_{s_1} = \{2, -1, -1\}/\sqrt{6}, \boldsymbol{v}_{s_2} = \{0, -1, 1\}/\sqrt{2}$ for $\mathrm{Re}(\lambda_{s_1}) = \mathrm{Re}(\lambda_{s_2}) < 0$. To investigate θ dynamics along each eigenvectors $\boldsymbol{v}_k$ $(k = u, s_1, s_2)$, we then introduced the variable $\hat{\theta}_k$, projection of $\boldsymbol{\theta}$ on $\boldsymbol{v}_k$ (that is, $\hat{\theta}_k = \boldsymbol{\theta} \cdot \boldsymbol{v}_k$, with $\boldsymbol{v}_k$ normalized). Noteworthy, owing to the symmetry of the repressilator, the unstable manifold was in line with the eigenvector $\boldsymbol{v}_u$.

As shown in Fig. 3.2cd, the straight line to which all trajectories converged agreed with the unstable manifold $\boldsymbol{v}_u$ (Fig. 3.4a). Of course, attraction to the $\boldsymbol{v}_u$ axis was natural if the initial conditions were restricted to the stable manifold for $\{\theta_1, \theta_2, \theta_3\} = \{0, 0, 0\}$. However, we observed an attraction toward the unstable axis over a wide range of initial conditions for $\{\theta_i\}$, which supports the oscillation of x_i. Furthermore, the magnitudes of the eigenvalues for the stable and unstable eigenvectors were of the same order ($\mathrm{Re}(\lambda_u) = 0.31, \mathrm{Re}(\lambda_{s_1}) = \mathrm{Re}(\lambda_{s_2}) = -0.66$, see Fig. 3.3c). Thus, the reprogramming dynamics shown in Fig. 3.2b could not be explained just by the linear stability.

To elucidate whether the nonlinear effect suppresses the instability along the $\boldsymbol{v}_u$ axis, we computed $d\hat{\theta}_u/dt$. As shown in Fig. 3.4c, $d\hat{\theta}_u/dt$ was drastically reduced from the linear case. We also computed $(d\hat{\theta}_{s_1}/dt, d\hat{\theta}_{s_2}/dt)$ for a certain $\hat{\theta}_u$ value (that is, the flow structure in the $(\hat{\theta}_{s_1}, \hat{\theta}_{s_2})$ plane, sliced along the $\hat{\theta}_u$ axis), which showed that $\hat{\theta}_{s_1} = 0$ changed from stable to unstable at $\hat{\theta}_u = \hat{\theta}_u^{\mathrm{th}} (\sim 0.4)$ (see Fig. 3.4d and Fig. 3.5). Up to $\hat{\theta}_u < \hat{\theta}_u^{\mathrm{th}} (\sim 0.4)$, θ_i in $(\hat{\theta}_{s_1}, \hat{\theta}_{s_2})$ plane was attracted to the $\hat{\theta}_u$ axis. By further increasing $\hat{\theta}_u$ beyond $\hat{\theta}_u^{\mathrm{th}}$, θ_i departured from the $\hat{\theta}_u$ axis rotating in the $(\hat{\theta}_{s_1}, \hat{\theta}_{s_2})$ plane, leading to differentiation into three distinct fixed points.

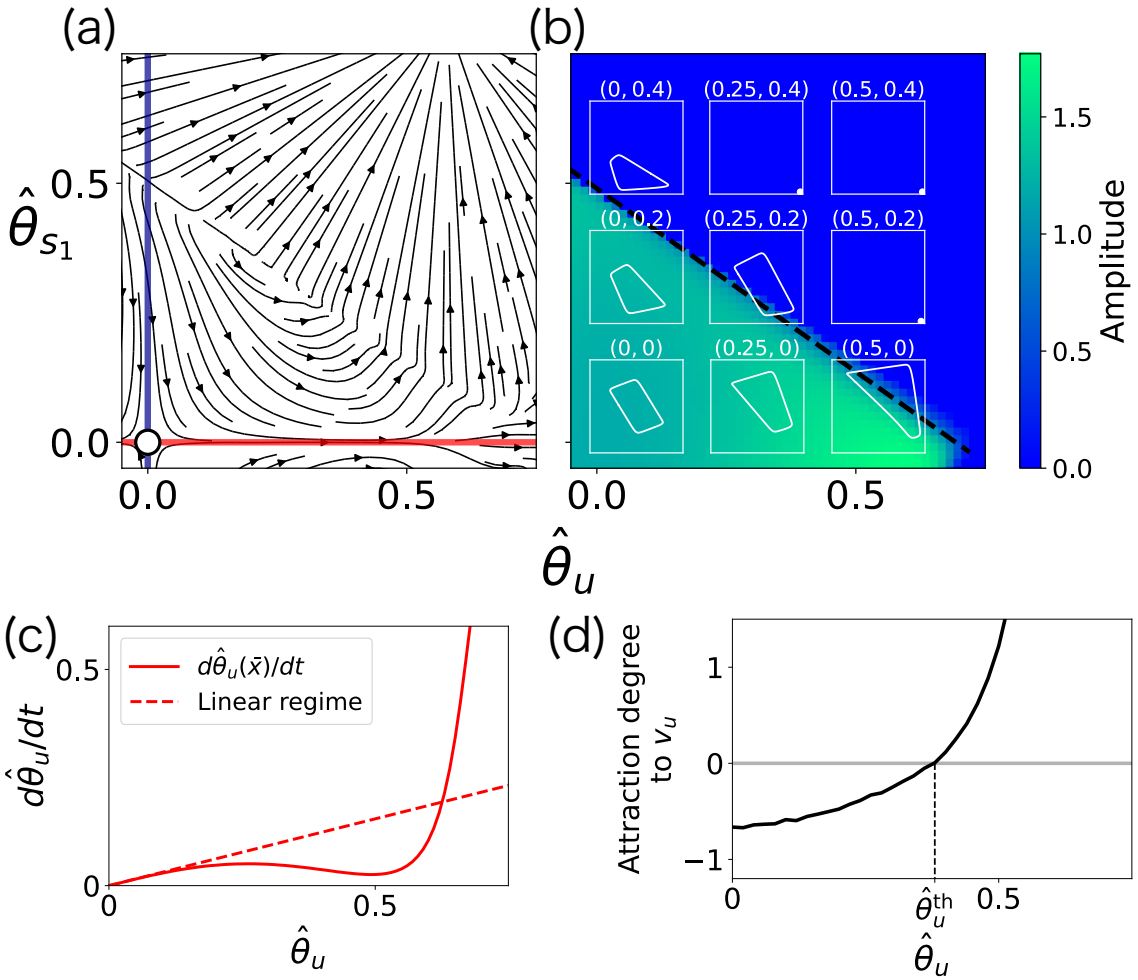


Fig. 3.4: (a) Stream plot of $(d\hat{\theta}_u/dt, d\hat{\theta}_{s_1}/dt)$ in $(\hat{\theta}_u, \hat{\theta}_{s_1})$ space according to Eq. (3.2). Red (blue) line represents the direction of the eigenvector $\boldsymbol{v}_u(\boldsymbol{v}_{s_1})$. (b) Attractors in the x-space for each fixed θ value. For the green (blue)-colored region, the attractor in the x space was the limit cycle (fixed point). Each trajectory shows the dynamics in the $x_1 - x_2$ space for $(\hat{\theta}_u, \hat{\theta}_{s_1})$ (both x_1 and x_2 axes for all figures are ranged within $\{-1.05, 1.05\}$). (c) $d\hat{\theta}_u/dt$ plotted as a function of $\hat{\theta}_u$. For comparison, we plotted $d\hat{\theta}_u = \lambda_u \hat{\theta}_u$(red dotted line). (d) Degree of attraction to $\boldsymbol{v}_u$. ν, in the text, is plotted as a function of $\hat{\theta}_u$. If it is negative (positive), $\{\theta\}$ was attracted to (departed from) $\boldsymbol{v}_u$. See Fig. 3.5 for more detail.

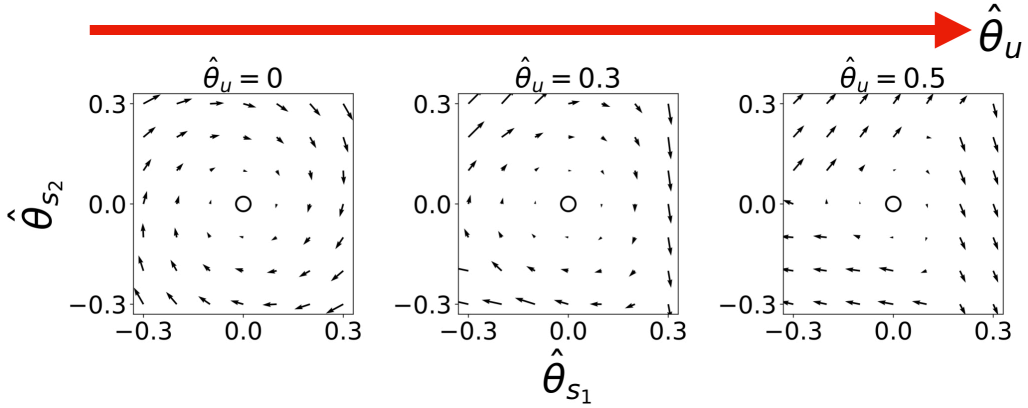


Fig. 3.5: The flow $(d\hat{\theta}_{s_1}/dt, d\hat{\theta}_{s_2}/dt)$ is plotted in the $(\hat{\theta}_{s_1}, \hat{\theta}_{s_2})$ plane along $\hat{\theta}_u$. Each center point $(0, 0)$ sits on the $\hat{\theta}_u$-axis.

To unveil how the slow motion along $\hat{\theta}_u$ and the attraction to $\hat{\theta}_u$ occurred, we first fixed θ_i and studied the change in the x attractor, as shown in Fig. 3.4b. In the green (blue) region, the x attractor was a limit cycle (fixed point) for $(\hat{\theta}_u, \hat{\theta}_{s_1})$. At the line $\hat{\theta}_{s_1} = -\hat{\theta}_u/\sqrt{2} + \sqrt{6}/5$ (as discussed in subsection 3.2.5), x-dynamics exhibited bifurcation from the limit cycle to a fixed point $\{1, -1, 1\}$ (see Fig. 3.9 for more details). Considering the symmetry of the repressilator, bifurcations to three fixed points $\{-1, 1, 1\}, \{1, -1, 1\}, \{1, 1, -1\}$ coexisted in the $(\hat{\theta}_{s_1}, \hat{\theta}_{s_2})$ plane. With the increase of $\hat{\theta}_u$, the limit cycle approached the three fixed points.

Now, we discussed the mechanism of slow-motion along $\hat{\theta}_u$. In the repressilator model, we could estimate the stagnation point, which slowed down, that is, $d\hat{\theta}_u/dt \simeq 0$. From Eq. (3.2), movement along $\boldsymbol{v}_u$ followed $d\hat{\theta}_u/dt = \bar{x}_u(\hat{\theta}_u) - \hat{\theta}_u$ (we defined x_k as a projection on $\boldsymbol{v}_k$). When $\hat{\theta}_u$ increased, the limit cycle orbit moved around differentiated fixed points $\{-1, 1, 1\}, \{1, -1, 1\}, \{1, 1, -1\}$ and was close to a triangular orbit with corners $\{-1, 1, 1\}, \{1, -1, 1\}, \{1, 1, -1\}$, as shown in Fig. 3.6. Thus, we approximated the limit cycle orbit to an equilateral triangle. Fig. 3.7 schematically shows this orbit ($x_1 = x_2$ plane was adopted for plotting).

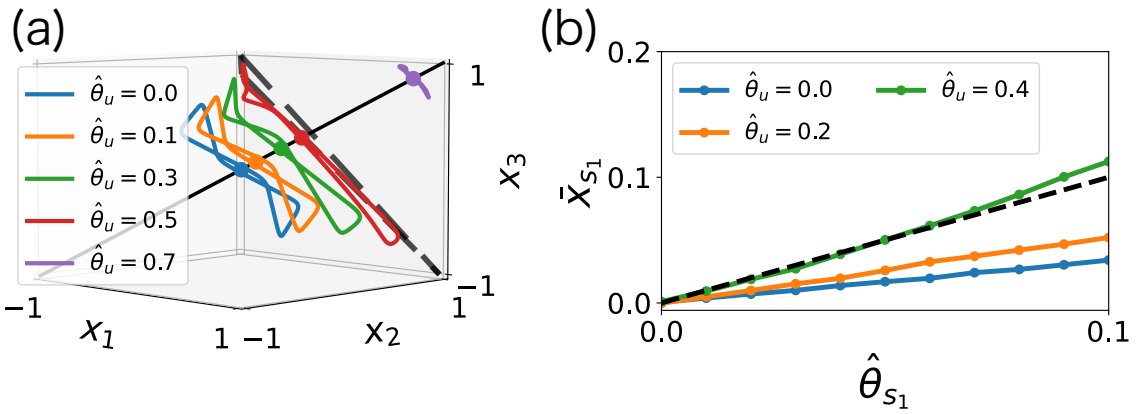


Fig. 3.6: (a) Change in the limit cycle $x(t)$ (closed line) and $\bar{x}$ (point) along $\hat{\theta}_u$. Black dotted line shows an equilateral triangle with corners $\{-1,1,1\},\{1,-1,1\},\{1,1,-1\}$ (see SI 1D). (b) $\bar{x}_{s_1}$ as a function of $\hat{\theta}_{s_1}$. The slope of each line corresponds to $\partial\bar{x}_{s_1}/\partial\hat{\theta}_{s_1}$ for $\hat{\theta}_u$. From $\partial\Theta_{s_1}/d\hat{\theta}_{s_1}$, if the slope is less than one, the orbits are attracted toward $\hat{\theta}_u$. For comparison, we plotted $\bar{x}_{s_1} = \hat{\theta}_{s_1}$ as a black dotted line. At $\hat{\theta}_u \sim 0.4$, $\partial\bar{x}_{s_1}/\partial\hat{\theta}_{s_1}$ exceeds one.

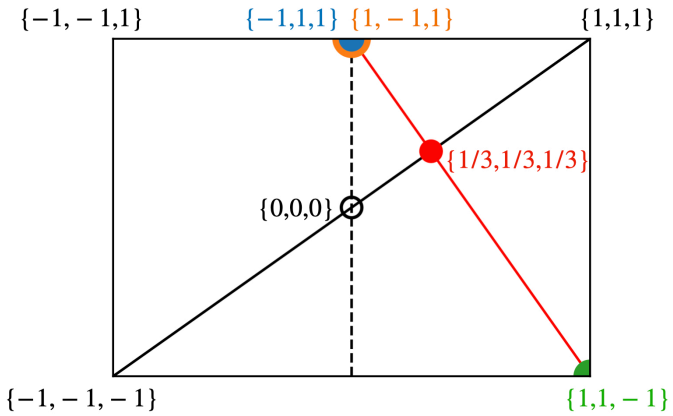


Fig. 3.7: Geometric representation for estimating stagnation point along $\hat{\theta}_u$. x space was sliced on the $x_1 = x_2$ plane.

Then, from the symmetry, the time average of $x_i(t)$ should be located at the red point in Fig. 3.7. Considering geometry, the position of the red point is given by $\{x_1, x_2, x_3\} = \{1/3, 1/3, 1/3\}$. In the geometric representation shown in Fig. 3.7, the value of $\hat{\theta}_u$ agreed with the distance from the origin to the point on the line $x_1 = x_2 = x_3$. Hence, as $\hat{\theta}_u$ approaches $1/\sqrt{3}$, $d\hat{\theta}_u/dt$ was minimized, as shown in Fig. 3.4c.

Next, we considered how attraction to the $\boldsymbol{v}_u$ from the $(\hat{\theta}_{s_1}, \hat{\theta}_{s_2})$ plane was lost at the $\hat{\theta}_u = \hat{\theta}_u^{th}$. By considering $\hat{\theta}_u$ as a parameter, the direction of flow in the $(\hat{\theta}_{s_1}, \hat{\theta}_{s_2})$ plane toward the $\boldsymbol{v}_u$ was determined by the sign of $\partial\Theta_{s_1}/\partial\hat{\theta}_{s_1} = \partial\bar{x}_{s_1}/\partial\hat{\theta}_{s_1} - 1 \equiv \nu\Theta_{s_1}$ (we defined Θ_{s_1} as a projection on $\boldsymbol{v}_{s_1}$). As shown in Fig 3.4b, with the increase in $\hat{\theta}_u$, the bifurcation point from the limit cycle to the fixed point approached the $\hat{\theta}_u = 0$ line. Hence, by slightly changing $\hat{\theta}_{s_1}$, $\bar{x}$ reached fixed

points. Accordingly, $\partial \bar{x}_{s_1}/\partial \hat{\theta}_{s_1}$ increase beyond one, so that $\partial \Theta_{s_1}/\partial \hat{\theta}_{s_1}$ became positive at $\hat{\theta}_u$, approaching $\hat{\theta}_u^{\rm th} \sim 0.4$, as shown in Figs. 3.4d and 3.6b.

3.2.5 Linear stability analysis for x dynamics with fixed θ

We discussed x dynamics of repressilator model with fixed θ as parameter. Full model with fixed θ can be written as

$$\frac{dx_1}{dt} = \tanh\{-\beta g(x_3 - \theta_1/g)\} - x_1 (= X_1) \tag{3.18a}$$

$$\frac{dx_2}{dt} = \tanh\{-\beta g(x_1 - \theta_2/g)\} - x_2 (= X_2) \tag{3.18b}$$

$$\frac{dx_3}{dt} = \tanh\{-\beta g(x_2 - \theta_3/g)\} - x_3 (= X_3). \tag{3.18c}$$

Thus, Jacobi Matrix $\partial X_i/\partial x_j$ can also be written as

$$\frac{\partial X}{\partial x} = \begin{bmatrix} -1 & 0 & \frac{-\beta g}{\cosh^2(-\beta g(x_3-\theta_1/g))} \\ \frac{-\beta g}{\cosh^2(-\beta g(x_1-\theta_2/g))} & -1 & 0 \\ 0 & \frac{-\beta g}{\cosh^2(-\beta g(x_2-\theta_3/g))} & -1 \end{bmatrix}. \tag{3.19}$$

We considered the bifurcation from the limit cycle to a fixed point by changing the parameter θ. From Eq. (3.19), the eigenvalues of the Jacobi matrix $\{\lambda\}$ are given by the solution of the following equation:

$$(-1-\lambda)^3 + \frac{-\beta g}{\cosh^2(-\beta g(x_3 - \theta_1/g)} \cdot \frac{-\beta g}{\cosh^2(-\beta g(x_1 - \theta_2/g)} \cdot \frac{-\beta g}{\cosh^2(-\beta g(x_2 - \theta_3/g)} = 0. \tag{3.20}$$

Here, we plotted $-\beta g/\cosh^2(-\beta g(x-z))$ as a function of x in Fig. 3.8. The $1/\cosh^2$ function has a tall spike at $x = z$ for a sufficiently large β. Considering that we adopted the tanh function, sigmoidal function with range$\{-1, 1\}$, the value of x_i is restricted to $\{-1, 1\}$.

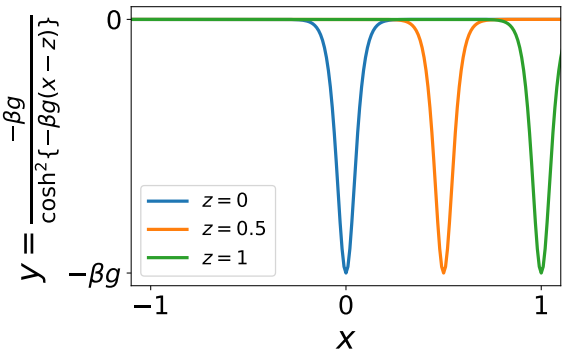


Fig. 3.8: $-\beta g/\cosh^2(\beta g(x-z))$ as a function of x ($\beta = 40, g = 0.4$). This function has a tall spike at $x = z$. Considering the range of the tanh function, this term could be regarded as zero if $|z| > 1$ for a sufficiently large β.

This suggests that if even one of $\theta_i/g (i = 1, 2, 3)$ in Eq. (3.20) exceeds one, and the second term in Eq. (3.20) is close to zero, and the solution to Eq. (3.20) is only determined by $(-1-\lambda)^3$, then $\lambda = -1$, corresponding to a stable fixed point. Considering that Eq. (3.27) gives the solution corresponding to an unstable fixed point, the bifurcation point from the limit cycle to the fixed point was determined as follows:

$$\max(\frac{\theta_1}{g}, \frac{\theta_2}{g}, \frac{\theta_3}{g}) = 1. \tag{3.21}$$

Here, we consider a set of Hopf bifurcation points in $\hat{\theta}_u - \hat{\theta}_{s_1}$ space. Each normalized eigenvector $\boldsymbol{v}_u, \boldsymbol{v}_{s_1}$ is given by $v_u = \{1, 1, 1\}/\sqrt{3}$ and $v_{s_1} = \{2, -1, -1\}/\sqrt{6}$. Hence, the full model with fixed $\hat{\theta}_u, \hat{\theta}_{s_1}$ can be written as

$$\frac{dx_1}{dt} = \tanh\{-\beta g(x_3 - \frac{\hat{\theta}_u/\sqrt{3} + 2\hat{\theta}_{s_1}/\sqrt{6}}{g})\} - x_1 (= X_1), \tag{3.22a}$$

$$\frac{dx_2}{dt} = \tanh\{-\beta g(x_1 - \frac{\hat{\theta}_u/\sqrt{3} - \hat{\theta}_{s_1}/\sqrt{6}}{g})\} - x_2 (= X_2), \tag{3.22b}$$

$$\frac{dx_3}{dt} = \tanh\{-\beta g(x_2 - \frac{\hat{\theta}_u/\sqrt{3} - \hat{\theta}_{s_1}/\sqrt{6}}{g})\} - x_3 (= X_3). \tag{3.22c}$$

From Eq. (3.21), the condition of the bifurcation from the limit cycle to the fixed point is given by

$$\max(\frac{\hat{\theta}_u/\sqrt{3} + 2\hat{\theta}_{s_1}/\sqrt{6}}{g}, \frac{\hat{\theta}_u/\sqrt{3} - \hat{\theta}_{s_1}/\sqrt{6}}{g}, \frac{\hat{\theta}_u/\sqrt{3} - \hat{\theta}_{s_1}/\sqrt{6}}{g}) = 1. \tag{3.23}$$

We considered a case in which $\hat{\theta}_{s_1} > 0$ in Fig. 3.4b, then Eq. (3.23) can be written as

$$\frac{\hat{\theta}_u/\sqrt{3} + 2\hat{\theta}_{s_1}/\sqrt{6}}{g} = 1. \tag{3.24}$$

Hence, the border line for $g = 0.4$ is given by

$$\hat{\theta}_{s_1} = -\frac{1}{\sqrt{2}}\hat{\theta}_u + \frac{\sqrt{6}}{5}. \tag{3.25}$$

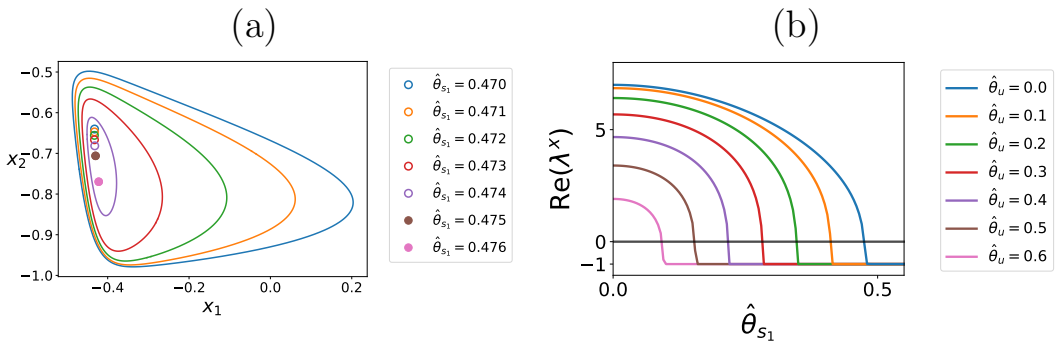


Fig. 3.9: (a) Change in the x attractor in the vicinity of the bifurcation point against the change in $\hat{\theta}_{s_1}$ by fixing $\hat{\theta}_u = 0$. (b) Real part of eigenvalue in Eq. (1a) plotted as a function of $\hat{\theta}_{s_1}$ for various $\hat{\theta}_u$. Each line shows different value of $\hat{\theta}_u$. Bifurcation from the limit cycle to the fixed point occurs when the curve crosses the line $\mathrm{Re}(\lambda^x) = 0$.

Thus, we have unveiled how attraction to the unstable manifold is achieved by slow epigenetic fixation of the oscillation of fast gene expression in the repressilator model. Following this picture, reprogramming is possible by forcing the cells to return to the oscillatory state. Then, the cell is attracted to a pluripotent state with low epigenetic modification $\theta_i \sim 0$, from which differentiation to distinct cell types with specific θ values follows.

3.2.6 The conditions of reprogramming dynamics for Repressilator model

As above analysis, we can impose 2 conditions of reprogramming dynamics in repressilator model: (1) Repressilator model must have limit cycle attractor in x dynamics (2) limit cycle attractor in x space must remain at a stagnation point in θ space.

First, we consider the condition (1). We studied the stability of the fixed point for $\{\theta_1, \theta_2, \theta_3\} = \{0, 0, 0\}$ to elucidate the condition for oscillation in x dynamics with no epigenetic modification. In this case, $\{x_1, x_2, x_3\} = \{0, 0, 0\}$ satisfies the fixed-point condition.

By inserting $x_i = 0, \theta_i = 0$ for all i in Eq. (3.19), we obtained:

$$\frac{\partial X}{\partial x} = \begin{bmatrix} -1 & 0 & -\beta g \\ -\beta g & -1 & 0 \\ 0 & -\beta g & -1 \end{bmatrix}. \tag{3.26}$$

From Eq. (3.26), eigenvalue λ is given by

$$\lambda = -\beta g - 1, (\frac{1}{2}\beta g - 1) \pm \frac{\sqrt{3}}{2}\beta g i. \tag{3.27}$$

In the repressilator model($g > 0$), the first value of (3.27) is always negative. The fixed point is unstable for

$$g > \frac{2}{\beta}, \tag{3.28}$$

where the limit-cycle attractor appeared as a result of Hopf bifurcation, to specific $g > 0.05$ for $\beta = 40$.

Condition (2) can be estimated as follows. Reprogramming dynamics required slowing down along $\boldsymbol{v}_u$. According to our analysis, oscillation in x was necessary for this process. Thus, we imposed the parameter g, the strength of the negative feedback loop in the repressilator, the condition that the oscillation dynamics remained at the stagnation point $\{\theta_1, \theta_2, \theta_3\} = \{1/3, 1/3, 1/3\}$. Then, the Hopf bifurcation should occur beyond the stagnation point with an increase in $\hat{\theta}_u$. From Eq. (3.21), we obtained

$$g > \frac{1}{3} \tag{3.29}$$

As shown in Fig. 3.10 and Fig. 3.2c, the global attraction occurred for $g = 0.35, 0.4$, whereas for $g = 0.2$ and 0.3, θs went to fixed points without passing the stagnation point. Then, reprogramming occurs for $g > 1/3$.

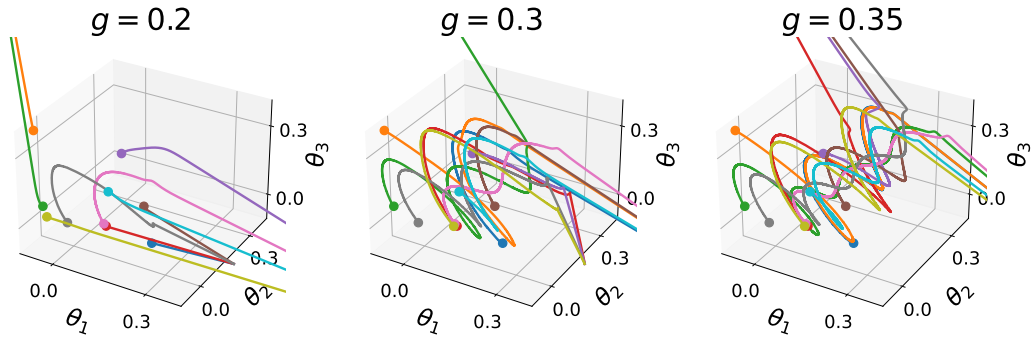


Fig. 3.10: Trajectories in the θ space, starting from the same initial conditions as in Fig. 3.2c, with $g = 0.2, 0.3$, and 0.35.

3.3 10-gene model with randomly generated GRN

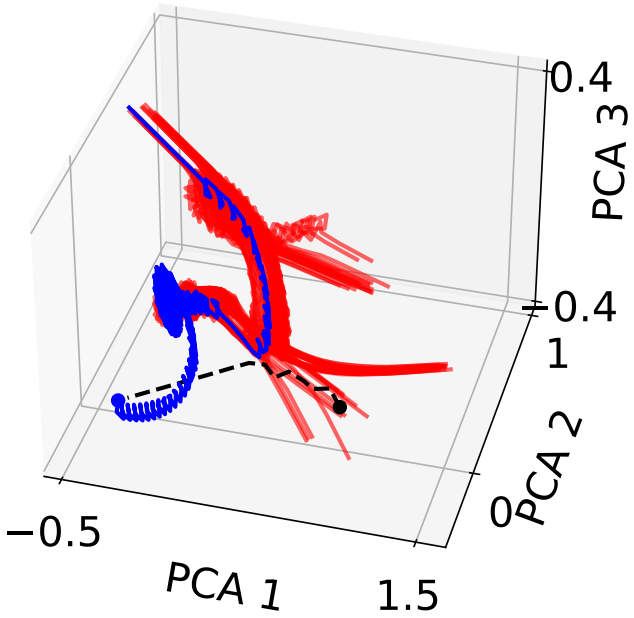
To verify the generality of this reprogramming scheme, we examined several GRN models with more degrees of freedom. As discussed in the previous section, differentiation from oscillatory states is often observed in GRNs (e.g., 20% of randomly generated GRNs show oscillatory dynamics for $N = 10$). An example is shown in Fig. 3.3.1a. From a differentiated state, we overexpressed three genes to regain oscillatory expression (black line in Fig. 3.3.1a). Later, global attraction to unstable manifold also occurred as discussed above.

3.3.1 Dynamics of cellular differentiation and reprogramming in 10-gene model

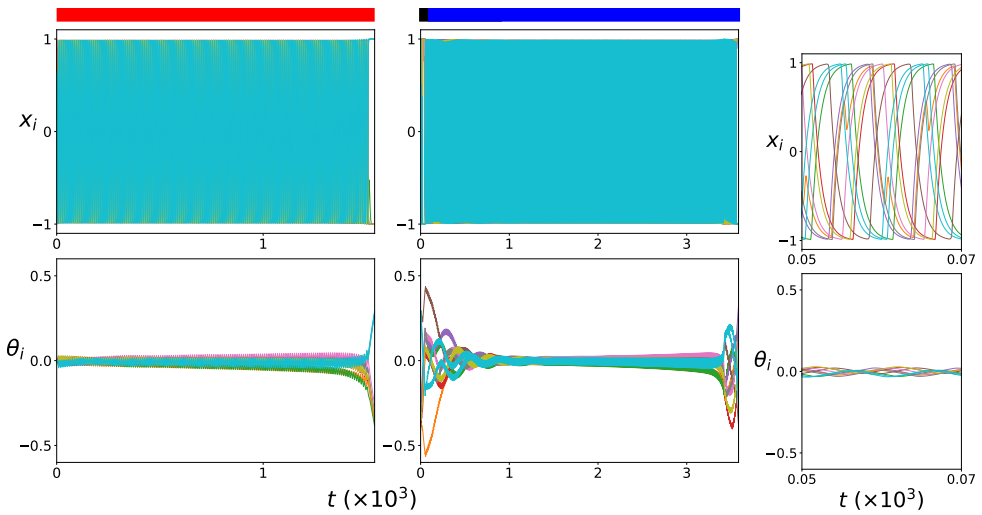
Then, the cell states branch to distinct fixed point states again (blue line in Fig. 3.3.1a). In these cases, the original pluripotent state with $\theta = 0$ was an unstable fixed point, with one positive eigenvalue for the Jacobi matrix of θ dynamics

(Fig. 3.3.1d), as in the repressilator model. Even though the degrees of freedom increased, the unstable manifold is one-dimensional, and the attraction to the manifold occurred from a higher-dimensional state space. This implies that reprogramming manipulation requires only partial degrees of freedom compared with the total number of genes. In fact, overexpression of three genes is sufficient for reprogramming in GRN models with $N = 10$, as far as we have investigated.

(a)


0.4
PCA 3
0.4
1
PCA 2
0
−0.5
PCA 1
1.5

(b)


x_i
θ_i
t (×10^3)
x_i
θ_i
t (×10^3)

(c)

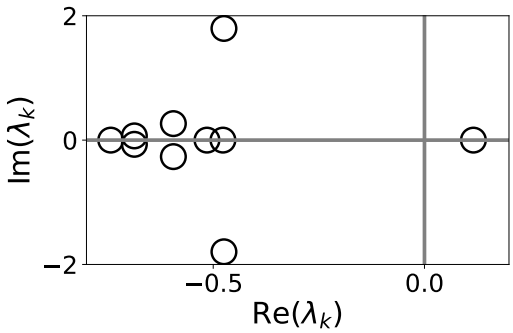

Im(λ_k)
Re(λ_k)

Fig. 3.10: Random gene regulatory network model with $N = 10$. We randomly generated the gene regulatory matrix J_{ij} and selected one sample that showed oscillation dynamics for $x_i(t)$ around $\theta_i = 0$. J_{ij} was generated according to the normal distribution, with an average of zero and a standard deviation of $1/\sqrt{N}$. (a) Cell differentiation and reprogramming. For plotting, we adopted a principal component analysis (PCA) of $\{\theta_i\}(i = 1, 2, \ldots, 10)$ over 1,000 trajectories starting from $\theta_i \sim 0$. The trajectory $\{\theta_i(t)\}$ was plotted by adopting the first, second, and third PCA modes. (Red lines) Trajectories starting from $\theta_i \sim 0$, five different fixed points are reached after transient oscillation dynamics. (Black dotted line) From one of the differentiated cell types, we added external input $I_i(t)$ to three genes as reprogramming manipulation. (Blue line) After reprogramming manipulation, the cell state was attracted to the unstable manifold and differentiated again. (b) Time development of cell differentiation and reprogramming (Fig. 3.3.1a. The upper bar shows the differentiation and reprogramming process, similar to Fig. 3.2ab. (c) Time development of cell differentiation (red region in Fig. 3.3.1b) for a short time window. (d) The eigenvalues λ_k of $\partial\Theta_i/\partial\theta_j$ matrix given by fixed point analysis of Eq. (2) around $\theta_i = 0$ for all i.

3.3.2 Exploration of possible sets of overexpresssion

In $N = 10$ model, various sets of overexpression genes are possible. From experimental reports, which genes are expressed is deeply related to whether the reprogramming is possible or not [30, 31]. In this section, using numerical simulation, we consider the property of reprogrammable sets of overexpression. Here, we assume the number of overexpression genes is three. In this assumption, the number of possible sets of overexpression is $10 \times 9 \times 8/3 \times 2 \times 1 = 120$. Moreover, whether reprogramming is possible or not may depend on how long genes are overexpressed and from where cellular state is overexpressed. Here, we consider 4 time spans $100, 300, 500$,and 700 and 5 differentiated cellular states to overexpress (shown in 3.3.1). To sum up, the number of all overexpression sets is $2,400$. In this trial, based on the previous section, we assume that the reprogramming is successful when the oscillation is recovered.

Among 2,400 overexpression sets, we succeeded in reprogramming 108 sets. And We count the number of overexpression of each gene, as shown in Fig. 3.11a. In Fig. 3.11a, we found an important gene (corresponding to gene 2) even in a randomly chosen network.

Here, we elucidate the mechanism of the appearance of important genes from the viewpoint of dynamical systems theory. For this purpose, we consider the limit that each θ_i value is independent and treated as parameter.

$$\frac{dx_i}{dt} = F(\sum J_{ij}x_j + \theta_i) - x_i \tag{3.30}$$

In this limit, we can draw a bifurcation diagram for each θ_i value, as in Fig. 3.11b. As shown in Fig. 3.11b, whether x dynamics is limit cycle or fixed point strongly depends on gene 2, most frequently overexpressed gene. Fig. 3.11b suggests that important gene for reprogramming corresponds to the center of oscillatory dy-

namics. We roughly check this suggestion to depict the main components of a randomly-chosen network as shown in Fig. 3.12. As shown in Fig. 3.12, network structure shows that some feedback loop is constructed around the gene 2.

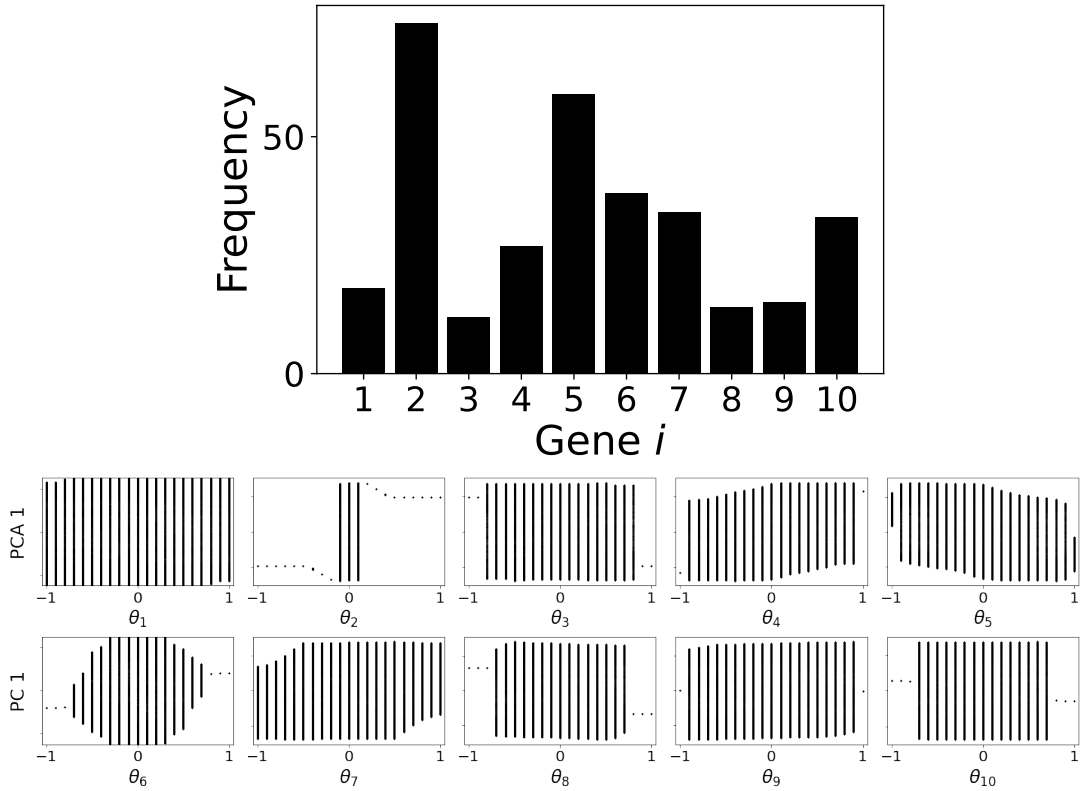


Fig. 3.11: (a) Frequency of each genes among 108 reprogrammable overexpression sets. (b) Bifurcation diagram of each θ_i value. Acconrding to Eq. (3.30), we treat each θ_i as independent parameter and draw bifurcation diagram. When we focus on the change of certain θ_i, the other $\theta_j (j \neq i)$ set to be 0. For plotting 1-dimensional bifurcation diagram, we adopt PCA among limit cycle at $\theta_i = 0$ for all i.

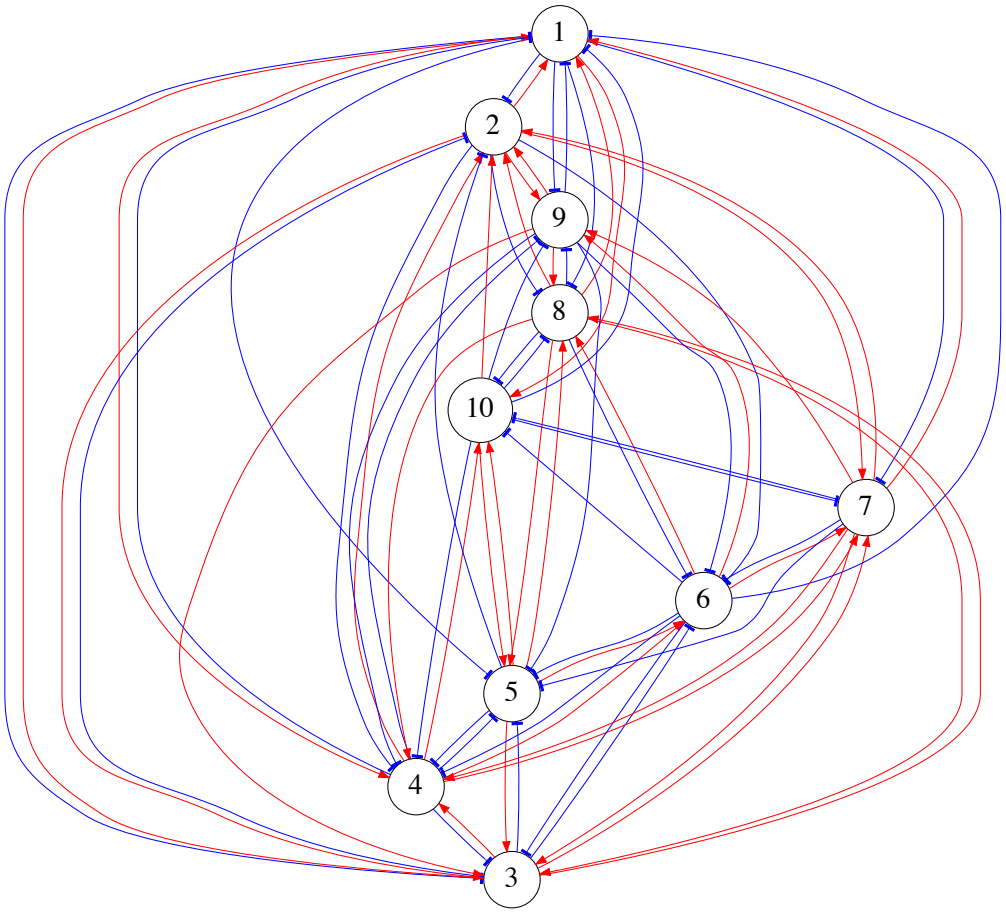


Fig. 3.12: Components of randomly-chosen network. Although network elements J_{ij} are generated according to the normal distribution with an average of zeros and a standard deviation of $1/\sqrt{N}$ (as in the previous sub-section), J_{ij}, satisfies $|J_{ij}| > 0.05$ are depicted for simplicity. Feedback loops are constructed around gene 2, the important gene for reprogramming.

3.4 5-gene model extracted from realistic gene regulatory network

The present mechanism also works in a model extracted from GRN for an ES cell [75], as a core network with five genes (*Nanog, Oct4, Gata6, Gata4,* and *Klf4*) [58] (see Fig. 3.13). *Oct4, Sox2*, and *Klf4* are known as factors to induce reprogramming. The model involves a negative feedback loop, as in the repressilator, in addition to positive feedback regulation.

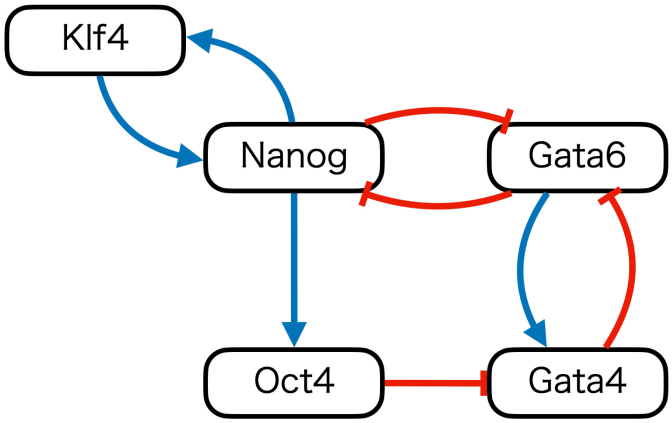


Fig. 3.13: Five-gene model. The θ_i nodes are not represented, but mutual positive feedback exists with the corresponding x_i. *Nanog, Oct4, Gata6, Gata4*, and *Klf4* correspond to x_1, x_2, x_3, x_4, and x_5 with the following parameters: $J_{21} = 0.21, J_{42} = -1.1, J_{34} = -1.0, J_{43} = 0.46, J_{13} = -0.93, J_{31} = -0.16, J_{51} = 0.37, J_{15} = 0.64$, and the other $J_{ij} = 0$.

In this five-gene model, x_i and θ_i oscillate in the region near the origin, and then differentiation to three fixed points progresses as in the repressilator case (three lines in Fig. 3.14a), whereas $\theta_i = 0$ for all i represents a saddle point with one unstable manifold and four stable manifolds, as shown in Fig.3.15b. After overexpression of *Oct4*, *Nanog*, and *Klf4* in one of the differentiated cell types for a certain time span (black dotted line in Fig. 3.14b)[*3], the epigenetic state θ_i approached the unstable manifold for the unstable fixed point $\theta_i = 0$, leading to recovery of pluripotency (blue line in Fig. 3.14b).

[*3] *Sox2* is reduced into Nanog in the five-gene model

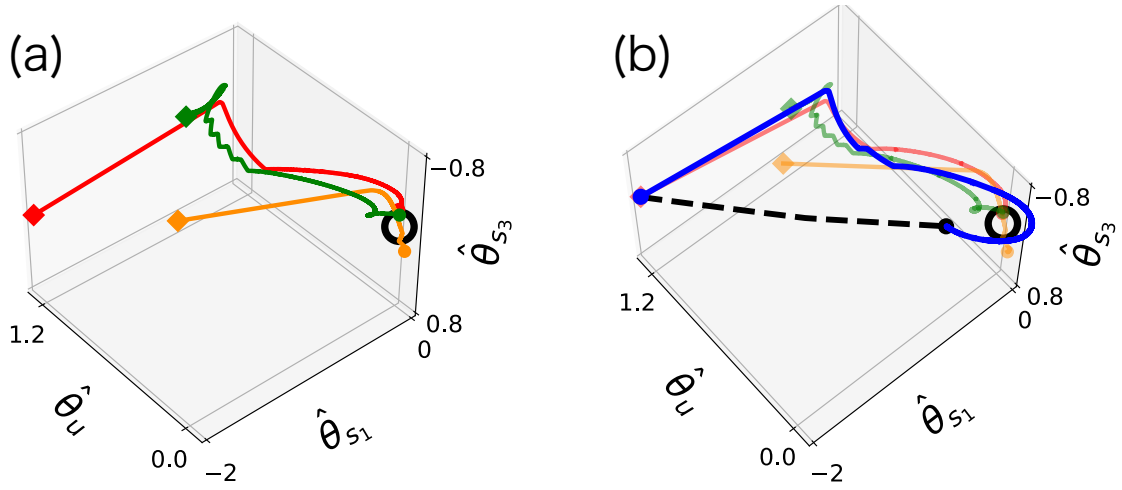


Fig. 3.14: (a) Cell differentiation and (b) reprogramming in the five-gene model (a) Three orbits starting from the vicinity of the saddle point $\theta_i = 0$ for all i (black dotted point), reached three distinct cell types. (b) From differentiated cell types (red point), we added external input $I_i(t)$ to Nanog, Oct4, and Klf4 for a certain time span (black dotted line). After such reprogramming manipulation, we set $I_i(t) = 0$. The cell state then spontaneously approached the saddle point, and reinitiated the differentiation progression again (blue line). $\tau = 10^3$. See Fig. 3.15 for more detail.

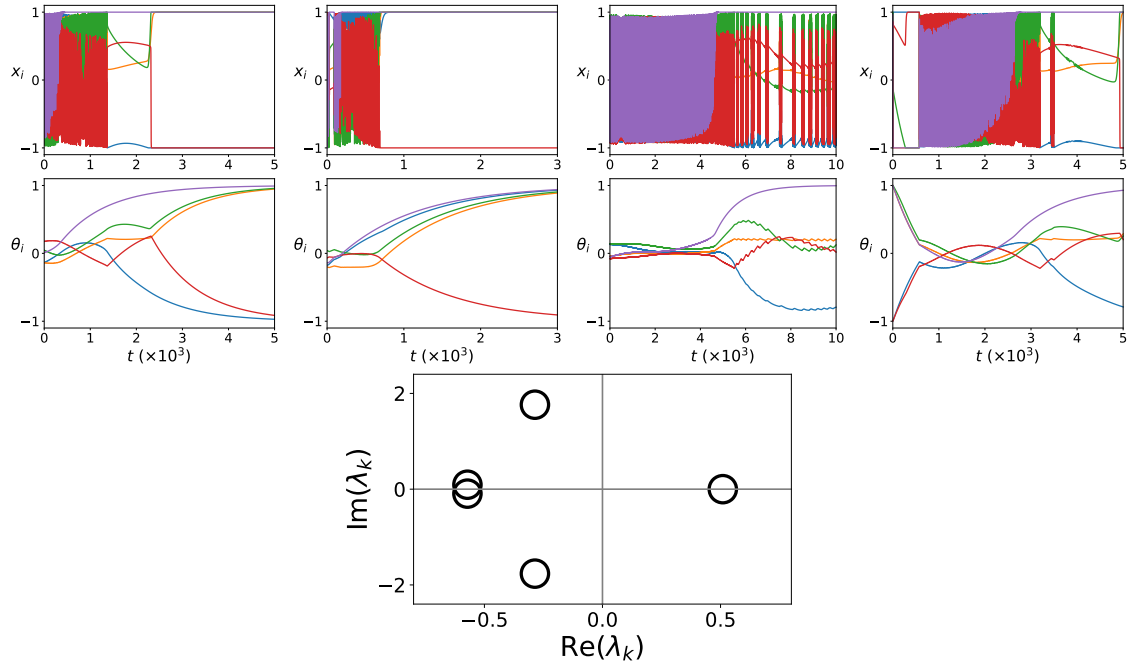


Fig. 3.15: (a) Time development for each trajectory of cell differentiation (yellow, green, and red line in Fig. 3.14a) and reprogramming (black dotted and blue lines in Fig. 3.14b). (Yellow, green, and red) Three trajectories starting from $\theta_i \sim 0$ are plotted, which reach three different attractors. (Black) From the differentiated state, we added an external input $I_i(t)$ to *Nanog*(x_1), *Oct4*(x_2), and *Klf4*(x_5) as reprogramming manipulation. (Blue) After reprogramming manipulation, the cellular state was first attracted to the unstable manifold, and then differentiated again. (b) The eigenvalues λ_k of the $\partial\Theta_i/\partial\theta_j$ matrix given by the fixed-point analysis of Eq. (2) around $\theta_i = 0$ for all i. In Fig. 3.14, we adopted eigenvectors $\boldsymbol{v}_{s_1}, \boldsymbol{v}_{s_3}$ with different Re(λ_k).

In this chapter, we have shown that oscillatory gene expression dynamics with slow epigenetic modifications lead to cellular reprogramming by overexpression of only few genes. The global attraction to the unstable manifold of the saddle point explains the reprogramming process. Now, the return to the top of the landscape by reprogramming, which is seemingly unstable, is explained by the strong attraction toward the unstable manifold of the saddle, and suppressed instability along with the unstable manifold, owing to the approach of limit-cycle of bifurcation to fixed points. The memory of the cellular state before reprogramming manipulation was erased through this reprogramming process.

Moreover, regain of oscillation was found to be the main requirement for reprogramming, whereas elaborate manipulations to induce a cellular state into specific states is not necessary. This explains the role of oscillations in gene expression in pluripotent cells [69] and epigenetic modification through the differentiation process [68], as well as it explains how reprogramming is possible by overexpressing just a few genes among thousand of that [6,31]. The timescale separation between fast expression dynamics and slow epigenetic modification feedback required is also consistent with previous observations [29,90]. In future studies, experimental support is necessary, as well as theoretical analysis of slow-fast dynamical sys-

tems [66, 91].

Chapter 4

Summary and Discussion

4.1 Summary

Here, let us summarize the whole of this thesis. We started this thesis from setting the goal to elucidate the universal properties of cellular differentiation and reprogramming. Both cellular differentiation and reprogramming are remarkable, as well as are universal among multicellular organisms. For a phenomenological approach to achieve this goal, we have constructed a cell model with gene regulatory network and epigenetic modification, as the simplest form. Then, we analyzed the cell model in terms of dynamical systems theory.

In chapter 2, we considered how cellular differentiation can be represented in a cell model with random gene regulatory matrices. Then, we also elucidated how Waddington's landscape emerges, with three salient properties as follows; (1) hierarchy branching of cell types (2) robustness of differentiation process as homeorhesis (3) robustness in the ratio of differentiated cell types. Here, the interplay between oscillation in fast gene expression dynamics and slower epigenetic modification plays an important role. Attractor pruning, non-monotonical change of attractor number against the rate of epigenetic feedback regulation, also emerges. Then, we analyzed the minimal model that consists of two genes, where the change in the epigenetic variables θ shifts the nullclines of expression levels x.

In chapter 3, we presented the mechanism of cellular reprogramming, i.e., how low-dimensional manipulation leads the cellular state to a pluripotent state, the top of Waddington's landscape. Considering the result in chapter 2, we assumed that the interplay between oscillation in fast gene expression dynamics and slower epigenetic modification also plays important role in cellular reprogramming. First, we constructed a minimal model with a three-genes repressilator to verify the hypothesis and analyzed the in terms of dynamical systems theory. We analyzed the dynamics in θ space by assuming adiabatic conditions. In the adiabatic limit, dynamics in θ are determined by the time average of x, whereas the slow change in θ influences the dynamics of x. The global attraction to a saddle point resulted by non-linear slowing down, whereas the bifurcation leads to differentiation. Moreover, global attraction and differentiation are also confirmed by a realistic five-gene model extracted from realistic data of embryonic cells as well as a random network model with more genes.

Let us now come back to epigenetic landscape, as mentioned in chapter 1, in terms of the present theory. Based on the result in chapter 2, cellular is represented

by hierarchical branching of oscillation state by epigenetic modification. It is also be represented by the motion along with the unstable manifold of a saddle and the bifurcations, in chapter 3 (consistency of the result in chapter 2 and chapter 3 are discussed in the next section). The mechanism of cellular reprogramming is explained as a global attraction to the unstable manifold of a saddle.

Based on our results, we can update epigenetic landscape as shown in Fig. 4.1. The left of Fig. 4.1 shows the landscape for the change in gene expression dynamics x, whereas the right of Fig. 4.1 shows a pathway of slower epigenetic modification θ. Cellular differentiation proceeds downward in Fig. 4.1. According to our result in chapter 2, the landscape changes with the hierarchical fixation of oscillation. Oscillation dynamics exist at the bottom of each valley, except for the end of the change, where the fixed point state exists. Epigenetic variables support its change from the backside of each landscape, as in the wires in Waddington's picture (Fig. 1.4). According to the result in chapter 3, the pathway of the cellular development in the epigenetic variables θ is represented by the motion along the unstable manifold and its bifurcation process. These dynamics, on the other hand, are driven by the oscillation in fast gene expression. The interplay between fast gene expression dynamics and slower epigenetic modification is important for the cellular differentiation process.

In our updated picture, cellular reprogramming can be represented by the black arrow in Fig. 4.1. First, differentiated cells move to certain points corresponding to states with oscillatory gene expression, following the reprogramming manipulation. Then, cellular states are strongly attracted to the pluripotent state represented by an unstable manifold in θ space. Throughout the strong attraction, the cellular state no longer has memories of the previous differentiated state. We note that strong attraction to the unstable manifold is also established by the interplay between fast gene expression dynamics and slower epigenetic modification, according to the result in chapter 3. The deletion on most epigenetic memories is completed. Finally, the cellular state follows the developmental pathway from the pluripotent state again. This picture of cellular reprogramming is in strong contrast to the picture provided in Fig. 1.6 in chapter 1, where the reprogramming climbs back again to the same valley as that adopted for the differentiation.

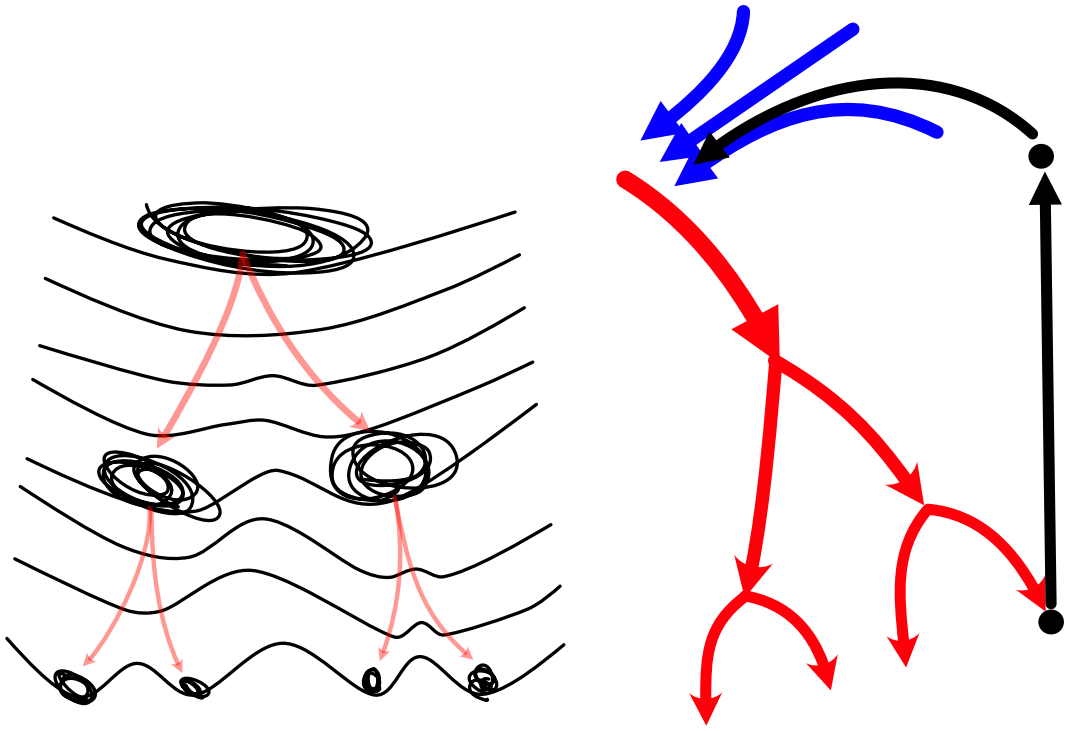


Fig. 4.1: Schematic representation of updated epigenetic landscape. The left figure shows the change in the landscape for gene expression dynamics, whereas the right figure shows the change in epigenetic modifications. Cellular differentiation proceeds downward in the figures. The valleys in the branch are hierarchically in the left figure, whereas bifurcation occurs along the unstable manifold represented by red arrows in the right figure. The blue arrows show the attraction to the pluripotent state. As discussed in chapter 3, due to slowing down in the flow along the unstable manifold attraction to it is stronger. The developmental pathway is thus represented. Black arrows show the trajectory of cellular reprogramming. By reprogramming manipulation, cellular states departed from the differentiation state. Then, the states are attracted to the same developmental pathway as the differentiation process.

4.2 Discussion

Here, we also discuss some points about the whole of this thesis.

4.2.1 Consitstency between the results in chapter 2 and 3

The cell models used in chapter 2 and chapter 3 are identical, although the different notation is adoted (by comparing with Eq. (2.2) and Eq. (3.1b), v corresponds to $1/\tau$).

In chapter 3, we assumed the adiabatic condition, i.e., $\tau \to \infty (v \to 0)$, and hence focused on the dynamics in θ space for understanding the mechanism of cellular reprogramming. In the slow-motion in θ space, the time average of x dynamics for given θ, rather than x dynamics itself, is important. We note the adiabatic condi-

tion is only needed for the simplest analysis, and the result does not change as long as sufficiently large τ (small v) (e.g., $\tau \sim 10^3$). Thus, the mechanism of cellular reprogramming is consistent with the mechanism of differentiating to various cell types as discussed in chapter 2 (the compatibility of cellular differentiation and reprogramming is supported by the results as shown in Fig. 3.3.1 and Fig. 3.14).

In chapter 2, we have also adopted a different approach from that in chapter 3. Focusing on faster dynamics in x space, slower θ variables can be treated as parameters. Then, they give nullclines' shift as discussed in the minimal model in chapter 2. This approach is relevant to understanding the mechanism of cellular differentiation. Different approaches between chapter 2 and chapter 3 are only reflected by different purposes in each chapter. We also note that actual dynamics, both of x and θ, consist of $2N$ variables, and in general, understanding the dynamics with the whole of $2N$-dimension space is required.

4.2.2 Minimal model for each chapter 2 and 3

In each of chapter2 and 3, we mainly analyze the minimal model for understanding in terms of dynamical systems theory. In chapter 2, we consider a minimal model which consists of two genes, whereas we consider a minimal model which consists of 3 genes in chapter 3. In the first place, how the minimal model is determined?

In this section, we discuss the requirements for the establishment of Fig. 4.1 (especially, aspect of dimension). For cellular reprogramming, a saddle attractor in θ space is required as discussed in chapter 3. The saddle has two types of directions. One is the unstable direction that corresponds to the progress for differentiation. The other is a stable direction adopted to return from differentiated states to the pluripotent state. The normal saddle attractor, however, can not satisfy the attraction as in Fig. 4.1. The global attraction to the unstable manifold is needed. In our analysis in chapter 3, oscillatory gene expression, which influences epigenetic modification, is needed for global attraction. Oscillatory gene expression is also for the differentiation to various cell types as also studied in chapter 2. Thus, the second requirement is the oscillation of gene expression at the pluripotent state.

Next, let us characterize these requirements, by focusing on the eigenvalues in each of x and θ space.

According to the above discussion, the first requirement is saddle attractor in θ space. Here, we introduce λ_i^θ as the i th eigenvalue given by the fixed point analysis in θ space. By using λ_i^θ, the first requirement can be written as

$$\mathrm{Re}(\lambda_1^\theta) < \mathrm{Re}(\lambda_2^\theta) < \cdots < \mathrm{Re}(\lambda_{N-1}^\theta) < 0 < \mathrm{Re}(\lambda_N^\theta). \tag{4.1}$$

In fact, the number of positive eigenvalues is unknown, and could be plural. Fewer positive eigenvalues, however, are desirable to give a deterministic and robust differentiation process. Hence, we consider the case with only one positive eigenvalue.

Next, we introduce λ_i^x as the i th eigenvalue for the fixed point analysis in $\boldsymbol{x}$ space. Then, the second requirement can be also written down as

$$\mathrm{Re}(\lambda_1^x) < \mathrm{Re}(\lambda_2^x) < \cdots < 0 < \mathrm{Re}(\lambda_{N-1}^x) = \mathrm{Re}(\lambda_N^x). \tag{4.2}$$

We note that λ^x_{N-1} and λ^x_N are complex conjugate of each other. For oscillation, at least two real parts of all eigenvalues must be positive.

To satisfy the requirements, three dimensions are at least necessary. Thus, the following is minimum requirement to satisfy Eq. (4.1) and Eq. (4.2);

$$\mathrm{Re}(\lambda^\theta_1) = \mathrm{Re}(\lambda^\theta_2) < 0 < \mathrm{Re}(\lambda^\theta_3) \tag{4.3}$$

$$\mathrm{Re}(\lambda^x_1) < 0 < \mathrm{Re}(\lambda^x_2) = \mathrm{Re}(\lambda^x_3). \tag{4.4}$$

Indeed, Eq. (4.4) is satisfied as in the analytic eigenvalues in Eq. (3.27). Then, Eq. (4.3) is satisfied as the eigenvalues obtained numerically in Fig. 3.3c. Thus, the minimal model that satisfies Fig. 4.1 consists of three genes.

We note that Eq. (4.1) and Eq. (4.2) can not be satisfied independently. As shown in Fig. 4.1, however, the relation between x variables and θ variables is much complex. Here, as a rough estimation, let us adopt linearized model of Eq. (3.1a) and Eq. (3.1b) as

$$\frac{d\boldsymbol{x}}{dt} = \beta J\boldsymbol{x} + \boldsymbol{\theta} - \boldsymbol{x}, \tag{4.5}$$

$$\frac{d\boldsymbol{\theta}}{dt} = \boldsymbol{x} - \boldsymbol{\theta}, \tag{4.6}$$

We also assume the adiabatic condition as in chapter 3, i.e., $\tau \to \infty$. It is then obvious that $\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{\theta} = \boldsymbol{0}$ satisfies the fixed point condition. Thus, we consider the fixed-pont analysis in each of x and θ space around $\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{\theta} = \boldsymbol{0}$.

For the second requirement Eq. (4.1),we consider fixed point analysis of Eq. (4.5) under $\boldsymbol{\theta} = \boldsymbol{0}$. $\{\lambda^x_i\}$ are given by the fixed point analysis of following equation around $\boldsymbol{x} = \boldsymbol{0}$;

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}, \tag{4.7}$$

where $A = \beta J - I$ (I is an identity matrix).

On the other hand, $\{\lambda^\theta_i\}$ are given by fixed point analysis of Eq. (4.6). By setting Eq. (4.5) to be 0 under the adiabatic condtion, we obtain the following relation;

$$\boldsymbol{x} = -A^{-1}\boldsymbol{\theta}, \tag{4.8}$$

where A^{-1} is the inverse matrix of A. Inserting this to Eq. (4.6), we also obtain

$$\frac{d\boldsymbol{\theta}}{dt} = -(A^{-1} + I)\boldsymbol{\theta}. \tag{4.9}$$

The eigenvalues, given by the fixed point analysis of Eq. (4.9) around $\boldsymbol{\theta}$, would correspond to $\{\lambda^\theta_i\}$. Comparing Eq. (4.7) and Eq. (4.9), we obtain the relation between $\{\lambda^x_i\}$ and $\{\lambda^\theta_i\}$ as

$$\lambda^\theta_i = -\frac{1}{\lambda^x_i} - 1. \tag{4.10}$$

This estimation, however, would not be valid. Actual dynamics $d\theta_i/dt$ are not given by the unstable fixed point of $\boldsymbol{x}^* = \mathbf{0}$, but by the time average of the limit-cycle solution of $\boldsymbol{x}$, given as Eq. 3.2. Thus, the theory, beyond the above rough estimation, is necessary for future studies of the condition for cellular reprogramming.

Based on the above discussion, we recall the minimal model in chapter 2. The minimal model in chapter 2 needs a robust differentiation process. For this requirement, unstable direction (and bifurcation along with it) is only needed, without the requirement for robust reprogramming process/ Thus, the minimal model in chapter 2 is reduced from the minimal model in chapter 3, i.e., the total number of genes is reduced from three to two.

4.2.3 Regaining pluripotency

Cellular reprogramming is a process to regain the capacity of differentiation by overexpressing just a few genes and thus has been believed remarkable. Regaining the capacity of differentiation itself, however, will be also universal. For instance, a plant callus is given by a mass of undifferentiated cells that are produced to recover wounds. The callus is also produced by grafting. A recent study reports that epigenetic modification, especially histone modification, is related to producing callus [92]. Regaining pluripotency is rather than common mechanism in plants. In the first place, multicellular organisms repeat the developmental process throughout generations (this universality could be metaphorically represented by the concept "epigenetic landscape as Klein bottle" in [93]). Repeating the developmental process, as regaining pluripotency, is also caused by a specific process, such as sexual reproduction, as well as cellular reprogramming.

In the above, we illustrated the universality of reprogramming as the capacity of to regain the pluripotency for differentiation. Most of these examples involve complex procurement, and it is difficult to know how we can manipulate them. Cellular reprogramming, on the other hand, can be induced by explicit and tunable manipulations. Moreover, this manipulation consists of just a few genes. The basic studies about cellular reprogramming will be relevant to understanding the universality of the whole class of phenomena in regaining the capacity for differentiation.

4.2.4 Cellular differentiation and reprogramming

In this thesis, we have presented a coherent view for cellular differentiation and reprogramming based on the interplay between fast oscillatory gene expression dynamics and slower epigenetic modification. Our study suggests that cellular differentiation and cellular reprogramming are two faces of the same coin. As mentioned in the previous section, we suggested a possibility that cellular reprogramming can be a general property for regaining the capacity for differentiation.

For instance, we here consider the relationship between the hierarchy of cell differentiation and the possibility of reprogramming. In chapter 2, we highlight the attractor pruning as a non-monotonic dependency of attractor number with increasing v (the time scale of epigenetic modification). The analysis in chap-

ter 2 shows that this attractor pruning is caused by the collapse of hierarchically branching of the differentiation process in slower epigenetic modification by the increase of v, i.e., the degree of non-monotonicity indicates the degree of the hierarchy of differentiation process in slower epigenetic modification. In section 2.5, we discussed the relation of the non-monotonicity of attractor number and the "globalness" of dynamics in x space. Here, "globalness" means that the limit cycle attractor covers a larger portion in the phase space. The larger limit cycle that travels in x phase space exists, the more hierarchical cell differentiation process progresses. That is, the capacity of differentiation and the "globalness" of oscillatory dynamics are positively correlated.

We sought the requirements for a global attraction of cellular reprogramming in the three-gene repressilator model in chapter 3. The global attraction is necessary for cellular reprogramming, and the global attraction itself needs stronger negative feedback loops in the repressilator (i.e., larger value of g in the repressilator model). Such global attraction, according to chapter 3, implies that oscillation remains beyond the stagnation point in θ space. The strength of negative feedback, on the other hand, is positively correlated to the size of oscillation dynamics ("globalness" of the limit cycle attractor). Thus, the correlation between the global attraction and the "globalness" of oscillatory dynamics is also indicated.

According to the above discussion, the capacity of differentiation and the degree of global attraction should be positively correlated. This suggests that the more cells have the capacity of differentiation, the easier to return by specific manipulation. This hypothesis may give some interpretation about iPS cells, the first instance of artificial pluripotent cells for regaining pluripotency from various differentiated cell types. So far, cellular differentiation and reprogramming are only linked by the metaphor of Waddington's landscape. Deep consideration between cellular differentiation and reprogramming both from theory, as well as from experiments, is necessary.

4.2.5 Other sources for cell differentiation and reprogramming

For simplicity, we did not consider other sources for cellular differentiation and reprogramming, than gene expression dynamics and epigenetic modification. In this section, we briefly discussed the possible connection consisting of our results and other sources such as environments and cell-cell interaction.

It is well-known that cell-cultures influence the maintenance of iPS cells. Based on our results, this can be explained as follows. According to our theory, undifferentiated cellular states follow the slow-motion along the unstable manifold in epigenetic variables. The time to maintain the undifferentiated states is determined by the inverse of the speed of the motion along the unstable manifold. The speed in epigenetic variables is influenced by the dynamics in gene expression. It is natural that cell culture affects gene expression. Thus, we can write down the flow of the effect of cell culture as cell culture $\rightarrow$ gene expression dynamics $\rightarrow$ epigenetic modification. Thus, the culture condition is also an important clue for the improvement of the incubation of iPS cells. Note that, at present, the time to maintain the undifferentiated states is mainly identified by the pluripotent markers. Our theory suggests that tracking dynamics of gene expression or epigenetic

modification can be a clue for characterizing such findings.

Next, we consider the effect of cell-cell interaction. The previous study reported the robust differentiation process with gene expression dynamics and epigenetic modification under cell-cell interaction, whereas the detailed process for it is unknown. Epigenetic fixation of differentiation states induced by cell-cell interaction is discussed by [58]. Following the theory in the present thesis, we should reconsider this issue.

4.2.6 Low dimensionality in developmental process

In chapter 2, we tracked the trajectories of cellular differentiation by adopting principal components analysis (PCA). It is noteworthy that we could track developmental trajectories and depict epigenetic landscape by adopting just a few freedoms (dimensions), even though cellular states can be actually represented N genes. Of course, cellular states can be completely represented by adopting all of the cellular components as variables. In our result, however, most of the differentiation process and differentiated cell types are characterized on just a few degrees of freedom, implying the low-dimensional representation of the cellular differentiation process is valid. Low dimensionality is also suggested by the result in chapter 3, where cellular states through differentiation can follow on the unstable manifold, since cellular states are strongly attracted to it. Experimental reports also suggest that cellular states of various cell types can be mapped to just a few dimensions (e.g., 2D plane) [2].

According to our result, low dimensionality in the cellular differentiation process can be briefly explained as follows. Cellular differentiation is represented by the interplay between oscillatory gene expression dynamics and slower epigenetic modification, where transient dynamics to reach attractors are important. The total dimension of the whole cellular differentiation process can be derived from the pluripotent states as the oscillatory state (limit cycle attractor or chaos), and the dimensionality would not be so much increased from it. In this thesis, we adopted principal components analysis (PCA) to depict the developmental trajectories. PCA is the simplest analysis to extract dimensions for dominant variations. Based on the picture with oscillatory gene expression and slow epigenetic modification, we should consider the possibility for better plotting methods, such as t-Distributed Stochastic Neighbor Embedding (t-SNE) or the viewpoint from the oscillation phase. Moreover, we should also consider the significance of trajectories, as well as experimental studies.

In evolutional biology, the dimensional reduction from high-dimensional phenotypes has recently been discussed [94]. For instance, it is suggested that the changes in phenotypes of Escherichia coli (E.coli), as a response to environmental changes, are strongly restricted in lower-dimensional space [95]. Of course, E.coli originally have a thousand of proteins, and other components (as well as factors for epigenetic gene regulation [96]). The dominant changes of phenotypes, however, are restricted in a few-dimensional space, and this evolutionary dimension reduction is explained by the robust steady growth [97, 98]. The developmental process, however, is not necessary at a steady growth. Hence, we need to the low-dimensional structure in the developmental, dynamical process. Evolu-

tion and development, however, are inseparable, as discussed in genetic assimilation [99, 100] and evolution-development congruence [101–103]. Thus, a comprehensive approach to low-dimensional structure both in development and evolution has to be explored.

4.2.7 Epigenentic spin glass

In this thesis, we adopted the cell model consisting of the gene regulatory network and epigenetic modification. Here, it should be noted that gene expression dynamics with interaction $\sum_j J_{ij}x_j$ has some similarity with spin-glass. Indeed, spin-glass can provide an abstract model for such complex systems [104]. In fact, the analysis discussed in spin-glass has been applied to gene regulation networks, say, Kauffman's Boolean network [105]. Here, x_i comes to up/down of each spin variable. Then, we can also introduce the epigenetic modification process to spin-glass as an external (magnetic) field, this field, however, is influenced by spin variables themselves.

For simplicity, in our cell model, each gene has a variable x_i that represents gene expression levels (or protein concentration), whereas θ_i that represents epigenetic modification levels. According to experimental reports, however, certain epigenetic modifications will affect gene expressions globally. For instance, chromatin openness, i.e., compaction and expansion process of DNA, is changed by epigenetic modification. Then, the change of compaction and expansion process affects the efficiency of a series of genes. Thus, chromatin openness, rather than, seems to be more global variable than local variable (field) θ_i. We should consider the local or global effects of epigenetic modification. In spin-glass, the temperature or external field is a global parameter and various studies have revealed the dependency of thermodynamical properties upon the temperature or external field. Then, the study of "epigenetic spin glass", spin-glass by external field coupled with the spin variables, may be important for our purpose.

4.2.8 Epigenetic adaptation in unicellular organisms

In this thesis, we discussed cellular differentiation and reprogramming, remarkable phenomena commonly observed in multicellular organisms. These are issues how various cell types are generated and how cellular states return to the pluripotent state. The definition of multicellular organisms itself is, however, not so simple. They would be different from just cell aggregates of homogenous cell types. Some organisms, however, decide to be homogeneous or heterogeneous states depending on the environment. Thus, the definition of multicellular organisms or unicellular organisms is generally ambiguous [106, 107] If the border of the unicellular organisms and the multicellular organisms is not so clear as mentioned above, the next issue to be addressed concerns if and how the interplay between fast gene expression dynamics and slower epigenetic modification is significant in unicellular organisms. It is well-known that bacteria, not only multicellular organisms, have epigenetic regulation [96]. In this section, we discuss a simple adaptation mechanism as an extension of our theory that consists of fast gene expression dynamics and epigenetic modification.

Adaptation via oscillatory state, as an example

To consider an adaptation mechanism with fast gene expression dynamics and slower epigenetic modification, let us consider the following model for cellular adaptation, as shown in Fig. 4.2

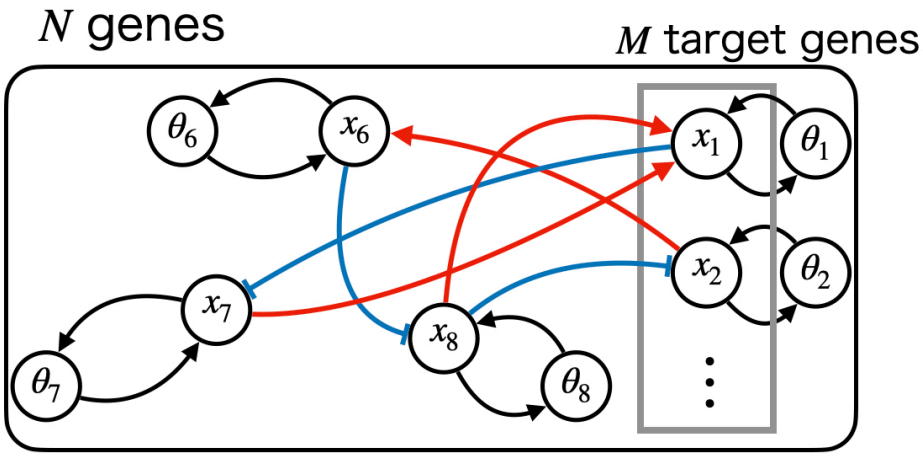


Fig. 4.2: The schematic representation of adaptation cell model. The model consists of N genes and cellular state is represented by gene expression level $x_i (i = 1, 2, \dots N)$ and epigenetic modification level θ_i for gene i. There are M target genes, a subset of N genes. In the cell model, target genes $x_m (m = 1, 2, \dots M)$ converge to X_m^k if the cellular state is adaptable $k (k = 1, 2, \dots, K)$ th environment.

As in the model adopted in chapters 2 and 3, the cell model here consists of N genes, and the cellular state is represented by gene expression levels (or protein concentrations) $x_i (i = 1, 2, \dots, N)$, whereas epigenetic modification levels are given by θ_i. The cell model follows same dynamics as in Eq. (2.1) and Eq. (2.2), except for v^k (as mentioned below),

$$\frac{dx_i}{dt} = F(\sum_j J_{ij} + \theta_i) - x_i, \tag{4.11}$$

$$\frac{d\theta_i}{dt} = v^k(ax_i - \theta_i), \tag{4.12}$$

where $F(z) = \tanh(\beta z)$ $(\beta = 40)$, and $a = 3$.

We consider the adaptation to various environments by referring to the attractor selection mechanism discussed in [88, 108, 109]. Here, the timescale of epigenetic modification changes by cellular activities such as cell growth or metabolism, and such activities depend on the subset of the whole of N genes, i.e., depend on $x_m (m = 1, 2, \dots, M)$. The timescale of epigenetic modification, v^k is thus a variable that represents the timescale of epigenetic modification and the $k (k =$

$1, 2, \ldots, K$) th environmental condition. We adopted the following form

$$v^k = V \exp\left\{ -\frac{b}{\sqrt{M}} \sqrt{\sum_{m=1}^{M} (x_m - X_m^k)^2} \right\}, \tag{4.13}$$

where $b = 3$ and $V = 10^{-2}$. $\boldsymbol{X}^k$ is the k th target expression pattern that indicates the environmental condition. If $x_m \sim X_m^k$ (x_m far from X_m^k), the cell is (not) adapted to k th environment, respectively. Here, we adopt $M = 5$. Then, the total number of possible M-bit patterns with $-1, 1$, which described as $\{1,1,1,1,1\}, \{-1,1,1,1,1\}, \ldots, \{-1,-1,-1,-1,-1\}$, is $2^M = 32$. We introduce these M-bit patterns as the set of target gene expression patterns $\{\boldsymbol{X}^k\}$. We note that the cell model has symmetry $-\boldsymbol{x} \leftrightarrow \boldsymbol{x}$, as discussed in chapter 2 and chapter 3. Considering this symmetry, $-\boldsymbol{X}^k$ and $\boldsymbol{X}^k$ are assumed to be identical. Thus, we adopt $2^M/2 (= 16)$ independent M-bit patterns as target gene expression patterns $\{\boldsymbol{X}^k\}$.

By using the cell model mentioned above, let us first study the adaptation with a random gene regulatory network with oscillatory dynamics. For simplicity, we start adaptation from initial conditions where x_i is randomly chosen, whereas θ_i is set to be 0. This setting is similar to that in chapter 2. We generate 500 random gene regulatory networks that show oscillatory gene expression dynamics at $\theta_i = 0$. Then, we counted the number of adaptable environments for each network. Fig. 4.3 shows the distribution of the number of environments that the cell has adapted. As shown in Fig. 4.3, most networks can adapt to various environments, even though the networks are randomly generated. In contrast, the distribution of the number of adapted environments for networks with fixed-point attractor is also plotted in Fig. 4.3. The number is peaked at 2. This shows the relevance of oscillation dynamics.

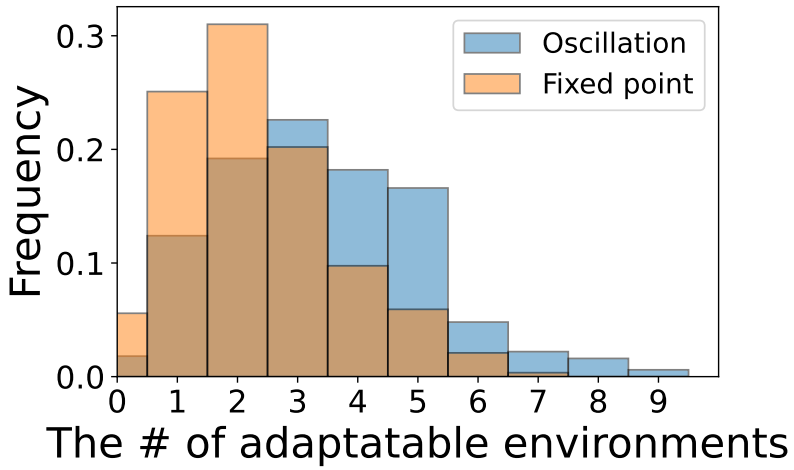


Fig. 4.3: Distribution of the number of adaptable environments. (Blue) The distribution of networks with oscillation dynamics. We generate 500 random gene regulatory networks that show oscillation with $\theta_i = 0$ and count the number of adaptable environments for each network. (Orange) The distribution of networks with fixed-point attractors. We generate 287 networks, as well as networks with oscillation dynamics. The number of all environments is $2^M/2 = 16$.

Fig. 4.4 shows an example of adaptation dynamics. First, the cellular state converges to the oscillatory state, and then oscillation is replaced by a fixed point by epigenetic modification, in which, the cellular state has reached the desired gene expression pattern through the transient oscillatory dynamics. These adaptation dynamics are similar to the dynamics in chapters 2 and 3. We also plotted the oscillatory dynamics at $\theta_i = 0$, as shown in Fig. 4.5. In Fig. 4.5ab, chaotic attractor exists at $\theta_i = 0$. We also plotted the trajectory of x_m in M dimensional subspace by using the PC space as shown in Fig. 4.5c. In the x_m space, the target gene expression patterns $\{\boldsymbol{X}^k\}$ are also plotted as shown in Fig. 4.5c. In Fig. 4.5c, oscilaltion dynamics travels various taret gene expression patterns. Thus, the "globalness", which is discussed in the previous sections, may be also important for adaptation mechanisms.

We also carried out an evolution simulation to select the networks that can adapt to more environments. Then, fitness is defined as the number of success trials among the multiple adaptations to each environment. Throughout mutation and selection of the network J_{ij}, the fitness increase as shown in Fig. 4.6, and an individual with oscillation dynamics is selected.

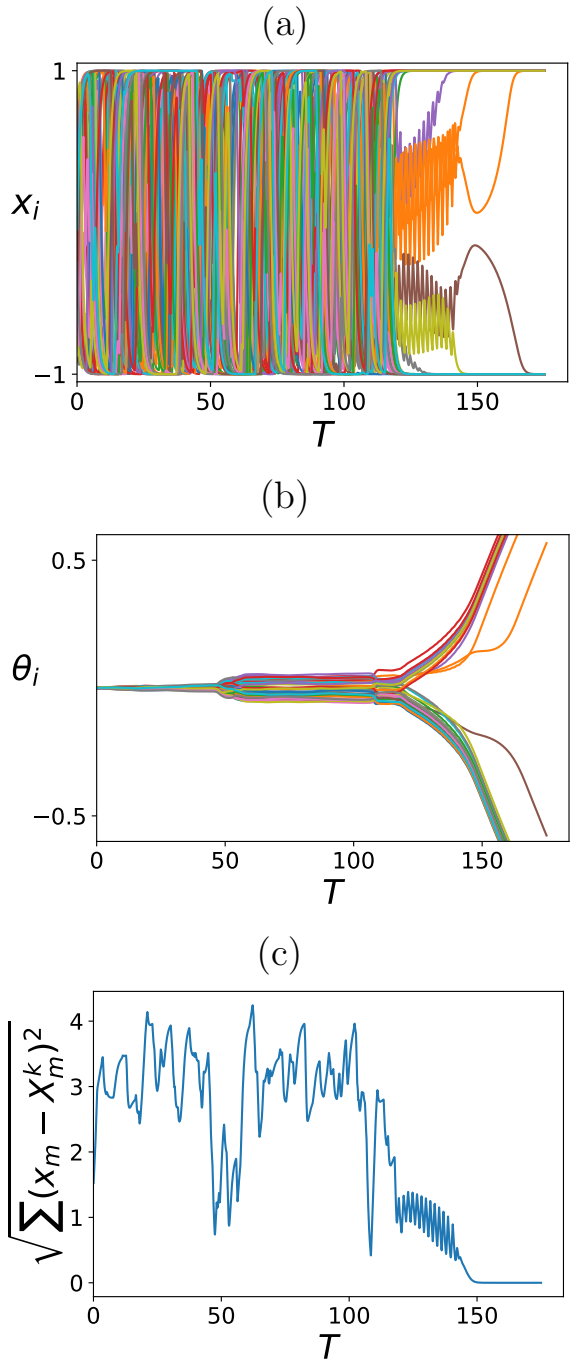


Fig. 4.4: Time series of adaptation dynamics of (a) gene expression x_i (b) epigenetic modification θ_i (c) the distance between cellular state x_m and target gene expression X_m^k.

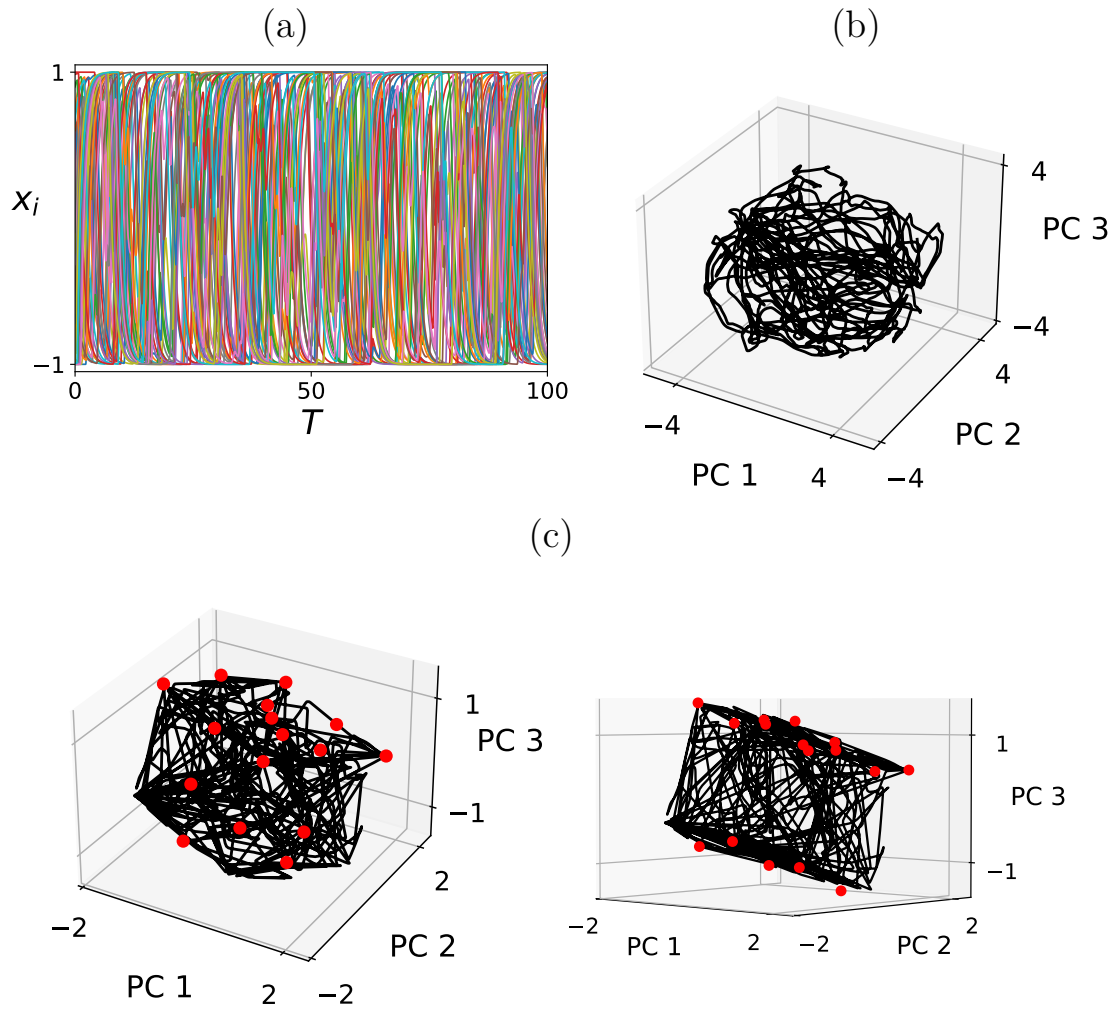


Fig. 4.5: An example of oscillatory gene expression dynamics at $\theta_i = 0$ that can adapt to various environments. Here, we choose one network which can adapt to eight environmental conditions. (a) Time series of all x_i. (b) Dynamics of x_i in PCA space. We adopt PCA obtained from the dynamics $\{x_i\}$. (c) Dynamics in the target x_m in the subspace of the PCA space. Red points represent the target gene expression patterns $\{\boldsymbol{X}^k\}$.

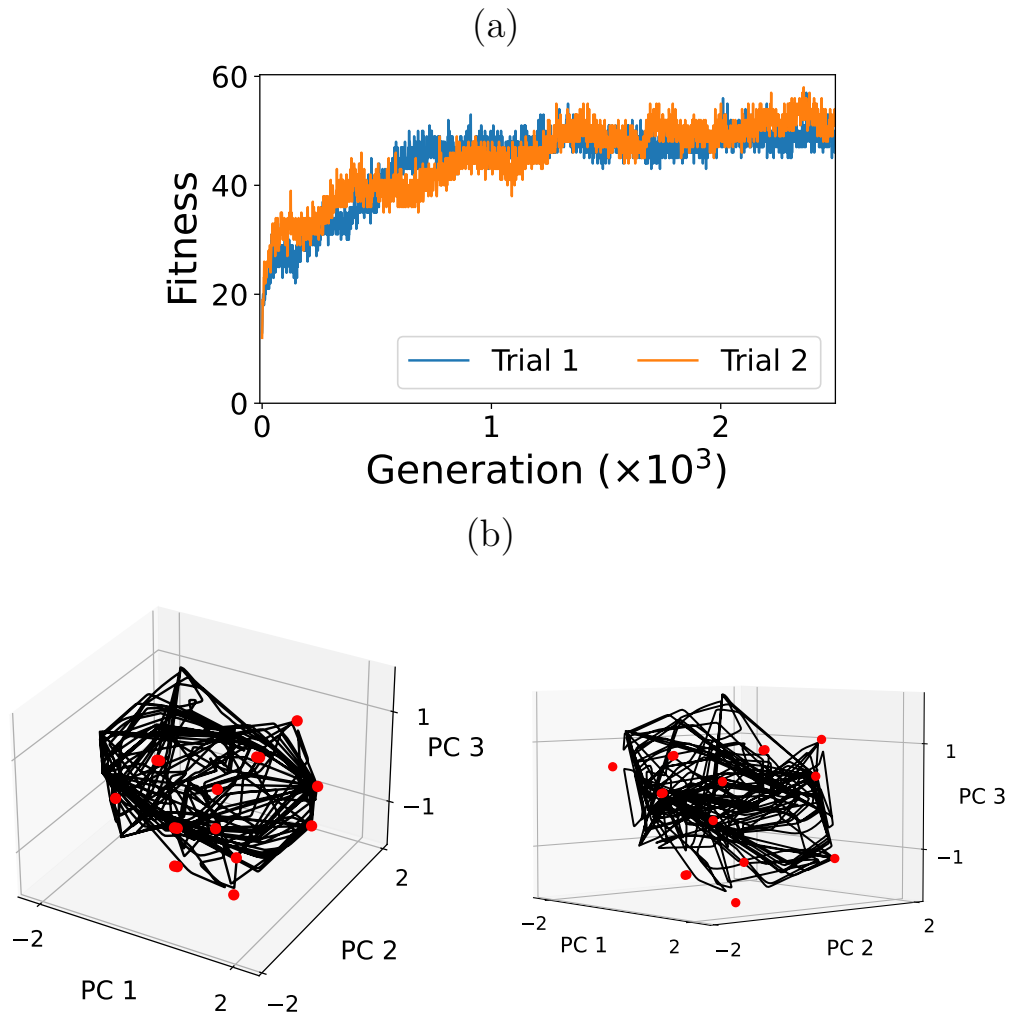


Fig. 4.6: (a) Example of time series of fitness in evolutionary simulation. We defined fitness as the number of success trials among the multiple adaptations to each environment. We adopted that the time for trials in each environment is set at four. Thus, the max fitness is $16 \times 4 = 64$. Different lines show different realizations of evolutionary simulation. As the result of evolutionary simulation, the cell is achieved to adapt to most of the environments. (b) Dynamics of evoled network at $\theta_i = 0$, in subspace x_m in PCA space. Red points represent the target gene expression patterns $\{\boldsymbol{X}^k\}$, as in Fig. 4.5c.

These results suggest that our picture of cellular differentiation and reprogramming by the interplay between oscillation in fast gene expression and slower epigenetic modification, can be applied to unicellular adaptation mechanisms. According to the results in chapter 3, we can expect that the interplay between fast oscillation and slower epigenetic modification in this adaptation model also has the global attraction to the unstable manifold of the saddle in the θ space. Thus, we can depict this adaptation mechanism on the epigenetic landscape as shown in Fig. 4.7. The adaptation in the epigenetic landscape occurs by returning to the oscillatory, "pluripotent" state. The normal adaptation process, on the other hand, is often depicted as the jumping of the cellular state from one valley to another valley. Comparing with these two pictures, we expect that adaptation on

the epigenetic landscape is superior upon the change in the environment to the jumping mechanism over the valleys. We also expect that these two pictures will be distinguishable by considering a gradient of the epigenetic landscape.

Moreover, we confirmed here that the "globalness" in oscillatory dynamics, discussed in the previous sections, is also related to the adaptation mechanism. The larger portion of phase space $\{x_i\}$ ($\{x_m\}$ in the subspace) travels over, the more environments can be adapted, whereas for cell differentiation more cell types are generated. Our result can be a candidate to bridge a gap between multicellular organisms and unicellular organisms. We need further study of this adaptation mechanism.

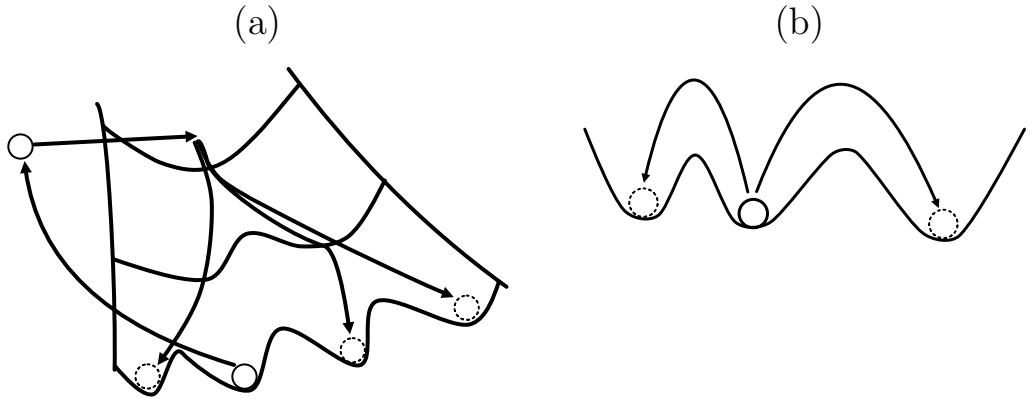


Fig. 4.7: the schematic representations of two types of adaptation dynamics. (a) Adaptation through the epigenetic landscape in our adaptation model with epigenetic change. For adaptation to different environments, the cellular state once returns to the top of the epigenetic landscape, and then a desired cellular state is selected throughout climbing down the epigenetic landscape. (b) The normal adaptation picture. Transition to the desired state is represented by jumping from one valley to another.

Bibliography

[1] Daniel E Wagner, Caleb Weinreb, Zach M Collins, James A Briggs, Sean G Megason, and Allon M Klein. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, Vol. 360, No. 6392, pp. 981–987, 2018.

[2] James A Briggs, Caleb Weinreb, Daniel E Wagner, Sean Megason, Leonid Peshkin, Marc W Kirschner, and Allon M Klein. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*, Vol. 360, No. 6392, p. eaar5780, 2018.

[3] CH Waddington. *The Strategy of the Genes*. George Allen & Unwin, 1957.

[4] Conrad H Waddington. The epigenotype. *Endeavour*, Vol. 1, pp. 18–20, 1942.

[5] John B Gurdon. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. 1962.

[6] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, Vol. 126, No. 4, pp. 663–676, 2006.

[7] Kazutoshi Takahashi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, Vol. 131, No. 5, pp. 861–872, 2007.

[8] Michael B Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, Vol. 403, No. 6767, pp. 335–338, 2000.

[9] Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, Vol. 403, No. 6767, pp. 339–342, 2000.

[10] Narito Suzuki, Chikara Furusawa, and Kunihiko Kaneko. Oscillatory protein expression dynamics endows stem cells with robust differentiation potential. *PloS one*, Vol. 6, No. 11, p. e27232, 2011.

[11] Yusuke Goto and Kunihiko Kaneko. Minimal model for stem-cell differentiation. *Physical Review E*, Vol. 88, No. 3, p. 032718, 2013.

[12] Adrian Bird. Perceptions of epigenetics. *Nature*, Vol. 447, No. 7143, p. 396, 2007.

[13] Aaron D Goldberg, C David Allis, and Emily Bernstein. Epigenetics: a landscape takes shape. *Cell*, Vol. 128, No. 4, pp. 635–638, 2007.

[14] Rudolf Jaenisch and Adrian Bird. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics*, Vol. 33, No. 3, pp. 245–254, 2003.

[15] M Azim Surani, Katsuhiko Hayashi, and Petra Hajkova. Genetic and epige-

netic regulators of pluripotency. *Cell*, Vol. 128, No. 4, pp. 747–762, 2007.

[16] Sui Huang. Non-genetic heterogeneity of cells in development: more than just noise. *Development*, Vol. 136, No. 23, pp. 3853–3862, 2009.

[17] Ruggero Cortini, Maria Barbi, Bertrand R Caré, Christophe Lavelle, Annick Lesne, Julien Mozziconacci, and Jean-Marc Victor. The physics of epigenetics. *Rev. Mod. Phys.*, Vol. 88, No. 2, p. 025002, 2016.

[18] Zakary S Singer, John Yong, Julia Tischler, Jamie A Hackett, Alphan Altinok, M Azim Surani, Long Cai, and Michael B Elowitz. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Molecular cell*, Vol. 55, No. 2, pp. 319–331, 2014.

[19] Di Zhang, Zhanyun Tang, He Huang, Guolin Zhou, Chang Cui, Yejing Weng, Wenchao Liu, Sunjoo Kim, Sangkyu Lee, Mathew Perez-Neut, et al. Metabolic regulation of gene expression by histone lactylation. *Nature*, Vol. 574, No. 7779, pp. 575–580, 2019.

[20] Federica Battistini, Pablo D Dans, Montserrat Terrazas, Chiara L Castellazzi, Guillem Portella, Mireia Labrador, Núria Villegas, Isabelle Brun-Heath, Carlos González, and Modesto Orozco. The impact of the hydroxymethylcytosine epigenetic signature on dna structure and function. *PLoS computational biology*, Vol. 17, No. 11, p. e1009547, 2021.

[21] Eran Meshorer, Dhananjay Yellajoshula, Eric George, Peter J Scambler, David T Brown, and Tom Misteli. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell*, Vol. 10, No. 1, pp. 105–116, 2006.

[22] Yaser Atlasi and Hendrik G Stunnenberg. The interplay of epigenetic marks during stem cell differentiation and development. *Nature Rev. Genet.*, Vol. 18, No. 11, pp. 643–658, 2017.

[23] Tarjei S Mikkelsen, Manching Ku, David B Jaffe, Biju Issac, Erez Lieberman, Georgia Giannoukos, Pablo Alvarez, William Brockman, Tae-Kyung Kim, Richard P Koche, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, Vol. 448, No. 7153, pp. 553–560, 2007.

[24] Kamal Tripathi and Gautam I Menon. Chromatin compaction, auxeticity, and the epigenetic landscape of stem cells. *Phys. Rev. X*, Vol. 9, No. 4, p. 041020, 2019.

[25] Oliver J Rando and Howard Y Chang. Genome-wide views of chromatin structure. *Annual review of biochemistry*, Vol. 78, pp. 245–271, 2009.

[26] Ian B Dodd, Mille A Micheelsen, Kim Sneppen, and Genevieve Thon. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell*, Vol. 129, No. 4, pp. 813–822, 2007.

[27] Kim Sneppen, Mille A Micheelsen, and Ian B Dodd. Ultrasensitive gene regulation by positive feedback loops in nucleosome modification. *Mol. Syst. Biol.*, Vol. 4, No. 1, p. 182, 2008.

[28] Masaki Sasai, Yudai Kawabata, Koh Makishi, Kazuhito Itoh, and Tomoki P Terada. Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells. *PLoS Comput. Biol.*, Vol. 9, No. 12, p. e1003380, 2013.

[29] Kazuhiro Maeshima, Kazunari Kaizu, Sachiko Tamura, Tadasu Nozaki, Tetsuro Kokubo, and Koichi Takahashi. The physical size of transcription fac-

tors is key to transcriptional regulation in chromatin domains. *J. Phys. Condens. Matter*, Vol. 27, No. 6, p. 064116, 2015.

[30] Kazutoshi Takahashi and Shinya Yamanaka. A decade of transcription factor-mediated reprogramming to pluripotency. *Nature Rev. Mol. Cell Biol.*, Vol. 17, No. 3, pp. 183–193, 2016.

[31] Sergiy Velychko, Kenjiro Adachi, Kee-Pyo Kim, Yanlin Hou, Caitlin M MacCarthy, Guangming Wu, and Hans R Schöler. Excluding oct4 from yamanaka cocktail unleashes the developmental potential of ipscs. *Cell Stem Cell*, Vol. 25, No. 6, pp. 737–753, 2019.

[32] Stuart A Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, Vol. 22, No. 3, pp. 437–467, 1969.

[33] Jin Wang, Li Xu, Erkang Wang, and Sui Huang. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys. J.*, Vol. 99, No. 1, pp. 29–39, 2010.

[34] Chikara Furusawa and Kunihiko Kaneko. A dynamical-systems view of stem cell biology. *Science*, Vol. 338, No. 6104, pp. 215–217, 2012.

[35] Gabor Forgacs and Stuart A Newman. *Biological physics of the developing embryo*. Cambridge University Press, 2005.

[36] Jin Wang, Kun Zhang, Li Xu, and Erkang Wang. Quantifying the waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences*, Vol. 108, No. 20, pp. 8257–8262, 2011.

[37] Mitra Mojtahedi, Alexander Skupin, Joseph Zhou, Ivan G Castaño, Rebecca YY Leong-Quong, Hannah Chang, Kalliopi Trachana, Alessandro Giuliani, and Sui Huang. Cell fate decision as high-dimensional critical state transition. *PLoS biology*, Vol. 14, No. 12, p. e2000640, 2016.

[38] Sui Huang, Gabriel Eichler, YaneerBar-Yam, and Donald E Ingber. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.*, Vol. 94, No. 12, p. 128701, 2005.

[39] JB Gurdon, P Lemaire, and K Kato. Community effects and related phenomena in development. *Cell*, Vol. 75, No. 5, pp. 831–834, 1993.

[40] Néstor J Oviedo, Phillip A Newmark, and Alejandro Sánchez Alvarado. Allometric scaling and proportion regulation in the freshwater planarian schmidtea mediterranea. *Developmental dynamics: an official publication of the American Association of Anatomists*, Vol. 226, No. 2, pp. 326–333, 2003.

[41] Ismael Rafols, Yasuji Sawada, Aiko Amagai, Yasuo Maeda, and Harry K MacWilliams. Cell type proportioning in dictyostelium slugs: lack of regulation within a 2.5-fold tolerance range. *Differentiation*, Vol. 67, No. 4-5, pp. 107–116, 2001.

[42] Anne Katrine Alsing and Kim Sneppen. Differentiation of developing olfactory neurons analysed in terms of coupled epigenetic landscapes. *Nucleic acids research*, Vol. 41, No. 9, pp. 4755–4764, 2013.

[43] Kunihiko Kaneko and Tetsuya Yomo. Isologous diversification for robust development of cell society. *Journal of theoretical biology*, Vol. 199, No. 3, pp. 243–256, 1999.

[44] Chikara Furusawa and Kunihiko Kaneko. Theory of robustness of irreversible

differentiation in a stem cell system: chaos hypothesis. *Journal of Theoretical Biology*, Vol. 209, No. 4, pp. 395–416, 2001.

[45] Yuri B Schwartz and Vincenzo Pirrotta. Polycomb silencing mechanisms and the management of genomic programmes. *Nature Reviews Genetics*, Vol. 8, No. 1, p. 9, 2007.

[46] Thimo Rohlf, Lydia Steiner, Jens Przybilla, Sonja Prohaska, Hans Binder, and Jörg Galle. Modeling the dynamic epigenome: from histone modifications towards self-organizing chromatin. *Epigenomics*, Vol. 4, No. 2, pp. 205–219, 2012.

[47] Rosa Karlić, Ho-Ryun Chung, Julia Lasserre, Kristian Vlahoviček, and Martin Vingron. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences*, Vol. 107, No. 7, pp. 2926–2931, 2010.

[48] Saera Hihara, Chan-Gi Pack, Kazunari Kaizu, Tomomi Tani, Tomo Hanafusa, Tadasu Nozaki, Satoko Takemoto, Tomohiko Yoshimi, Hideo Yokota, Naoko Imamoto, et al. Local nucleosome dynamics facilitate chromatin accessibility in living mammalian cells. *Cell Rep.*, Vol. 2, No. 6, pp. 1645–1656, 2012.

[49] Michael Grunstein. Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell*, Vol. 93, No. 3, pp. 325–328, 1998.

[50] Saurabh Gombar, Thomas MacCarthy, and Aviv Bergman. Epigenetics decouples mutational from environmental robustness. did it also facilitate multicellularity? *PLoS Comput. Biol.*, Vol. 10, No. 3, p. e1003450, 2014.

[51] Eric Mjolsness, David H Sharp, and John Reinitz. A connectionist model of development. *J. Theor. Biol.*, Vol. 152, No. 4, pp. 429–453, 1991.

[52] Isaac Salazar-Ciudad, Jordi Garcia-Fernández, and Ricard V Solé. Gene networks capable of pattern formation: from induction to reaction–diffusion. *J. Theor. Biol.*, Vol. 205, No. 4, pp. 587–603, 2000.

[53] I Salazar-Ciudad, SA Newman, and RV Solé. Phenotypic and dynamical transitions in model genetic networks i. emergence of patterns and genotype-phenotype relationships. *Evol. Dev.*, Vol. 3, No. 2, pp. 84–94, 2001.

[54] Kunihiko Kaneko. Evolution of robustness to noise and mutation in gene expression dynamics. *PLoS one*, Vol. 2, No. 5, p. e434, 2007.

[55] Wen Jia, Abhijeet Deshmukh, Sendurai A Mani, Mohit Kumar Jolly, and Herbert Levine. A possible role for epigenetic feedback regulation in the dynamics of the epithelial–mesenchymal transition (EMT). *Physical Biology*, Vol. 16, No. 6, p. 066004, sep 2019.

[56] Stuart L Schreiber and Bradley E Bernstein. Signaling network model of chromatin. *Cell*, Vol. 111, No. 6, pp. 771–778, 2002.

[57] John CG Spainhour, Hong Seo Lim, Soojin V Yi, and Peng Qiu. Correlation patterns between dna methylation and gene expression in the cancer genome atlas. *Cancer Inform.*, Vol. 18, p. 1176935119828776, 2019.

[58] Tadashi Miyamoto, Chikara Furusawa, and Kunihiko Kaneko. Pluripotency, differentiation, and reprogramming: a gene expression dynamics model with epigenetic feedback regulation. *PLoS computational biology*, Vol. 11, No. 8, p. e1004476, 2015.

[59] Koichi Fujimoto, Shuji Ishihara, and Kunihiko Kaneko. Network evolution

of body plans. *PLoS One*, Vol. 3, No. 7, 2008.

[60] Nirit Feldman, Ariela Gerson, Jia Fang, En Li, Yi Zhang, Yoichi Shinkai, Howard Cedar, and Yehudit Bergman. G9a-mediated irreversible epigenetic inactivation of oct-3/4 during early embryogenesis. *Nature cell biology*, Vol. 8, No. 2, p. 188, 2006.

[61] Grant A Challen, Deqiang Sun, Mira Jeong, Min Luo, Jaroslav Jelinek, Jonathan S Berg, Christoph Bock, Aparna Vasanthakumar, Hongcang Gu, Yuanxin Xi, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature genetics*, Vol. 44, No. 1, p. 23, 2012.

[62] Wolf Reik, Wendy Dean, and Jörn Walter. Epigenetic reprogramming in mammalian development. *Science*, Vol. 293, No. 5532, pp. 1089–1093, 2001.

[63] R David Hawkins, Gary C Hon, Leonard K Lee, QueMinh Ngo, Ryan Lister, Mattia Pelizzola, Lee E Edsall, Samantha Kuan, Ying Luu, Sarit Klugman, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell*, Vol. 6, No. 5, pp. 479–491, 2010.

[64] Kamal Tripathi and Gautam I. Menon. Chromatin compaction, auxeticity, and the epigenetic landscape of stem cells. *Phys. Rev. X*, Vol. 9, p. 041020, Oct 2019.

[65] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. Chaos in random neural networks. *Physical review letters*, Vol. 61, No. 3, p. 259, 1988.

[66] Hidetoshi Aoki and Kunihiko Kaneko. Slow stochastic switching by collective chaos of fast elements. *Phys. Rev. Lett.*, Vol. 111, No. 14, p. 144102, 2013.

[67] Sara Kangaspeska, Brenda Stride, Raphaël Métivier, Maria Polycarpou-Schwarz, David Ibberson, Richard Paul Carmouche, Vladimir Benes, Frank Gannon, and George Reid. Transient cyclical methylation of promoter DNA. *Nature*, Vol. 452, No. 7183, pp. 112–115, 2008.

[68] Steffen Rulands, Heather J Lee, Stephen J Clark, Christof Angermueller, Sébastien A Smallwood, Felix Krueger, Hisham Mohammed, Wendy Dean, Jennifer Nichols, Peter Rugg-Gunn, et al. Genome-scale oscillations in dna methylation during exit from pluripotency. *Cell Sys.*, Vol. 7, No. 1, pp. 63–76, 2018.

[69] Taeko Kobayashi, Hiroaki Mizuno, Itaru Imayoshi, Chikara Furusawa, Katsuhiko Shirahige, and Ryoichiro Kageyama. The cyclic gene hes1 contributes to diverse differentiation responses of embryonic stem cells. *Genes Dev.*, Vol. 23, No. 16, pp. 1870–1875, 2009.

[70] Itaru Imayoshi, Akihiro Isomura, Yukiko Harima, Kyogo Kawaguchi, Hiroshi Kori, Hitoshi Miyachi, Takahiro Fujiwara, Fumiyoshi Ishidate, and Ryoichiro Kageyama. Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. *Science*, Vol. 342, No. 6163, pp. 1203–1208, 2013.

[71] Alexander Aulehla and Olivier Pourquie. Oscillating signaling pathways during embryonic development. *Current opinion in cell biology*, Vol. 20, No. 6, pp. 632–637, 2008.

[72] Maurice A Canham, Alexei A Sharov, Minoru SH Ko, and Joshua M Brickman. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.*, Vol. 8, No. 5, p. e1000379, 2010.

[73] Hitoshi Niwa, Jun-ichi Miyazaki, and Austin G Smith. Quantitative expression of oct-3/4 defines differentiation, dedifferentiation or self-renewal of es cells. *Nature genetics*, Vol. 24, No. 4, p. 372, 2000.

[74] Laurie A Boyer, Tong Ihn Lee, Megan F Cole, Sarah E Johnstone, Stuart S Levine, Jacob P Zucker, Matthew G Guenther, Roshan M Kumar, Heather L Murray, Richard G Jenner, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *cell*, Vol. 122, No. 6, pp. 947–956, 2005.

[75] S-J Dunn, Graziano Martello, Boyan Yordanov, Stephen Emmott, and AG Smith. Defining an essential transcription factor program for naive pluripotency. *Science*, Vol. 344, No. 6188, pp. 1156–1160, 2014.

[76] Konrad Hochedlinger and Kathrin Plath. Epigenetic reprogramming and induced pluripotency. *Development*, Vol. 136, No. 4, pp. 509–523, 2009.

[77] Petra Hajkova, Katia Ancelin, Tanja Waldmann, Nicolas Lacoste, Ulrike C Lange, Francesca Cesari, Caroline Lee, Genevieve Almouzni, Robert Schneider, and M Azim Surani. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature*, Vol. 452, No. 7189, p. 877, 2008.

[78] Sui Huang. Reprogramming cell fates: reconciling rarity with robustness. *Bioessays*, Vol. 31, No. 5, pp. 546–560, 2009.

[79] Jonathan T Young, Tetsuhiro S Hatakeyama, and Kunihiko Kaneko. Dynamics robustness of cascading systems. *PLoS computational biology*, Vol. 13, No. 3, p. e1005434, 2017.

[80] John S Chuang, Zak Frentz, and Stanislas Leibler. Homeorhesis and ecological succession quantified in synthetic microbial ecosystems. *Proceedings of the National Academy of Sciences*, Vol. 116, No. 30, pp. 14852–14861, 2019.

[81] Shulamit Levenberg, Justin S Golub, Michal Amit, Joseph Itskovitz-Eldor, and Robert Langer. Endothelial cells derived from human embryonic stem cells. *Proc. Natl. Acad. Sci. USA*, Vol. 99, No. 7, pp. 4391–4396, 2002.

[82] Konrad Hochedlinger and Kathrin Plath. Epigenetic reprogramming and induced pluripotency. *Development*, Vol. 136, No. 4, pp. 509–523, 02 2009.

[83] Isabel Palmeirim, Domingos Henrique, David Ish-Horowicz, and Olivier Pourquié. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell*, Vol. 91, No. 5, pp. 639–648, 1997.

[84] Ekkehard Ullner, Aneta Koseska, Jürgen Kurths, Evgenii Volkov, Holger Kantz, and Jordi García-Ojalvo. Multistability of synthetic genetic networks with repressive cell-to-cell communication. *Phys. Rev. E*, Vol. 78, No. 3, p. 031904, 2008.

[85] Aneta Koseska, Ekkehard Ullner, Evgenii Volkov, Jürgen Kurths, and Jordi García-Ojalvo. Cooperative differentiation through clustering in multicellular populations. *J. Theor. Biol.*, Vol. 263, No. 2, pp. 189–202, 2010.

[86] Aneta Koseska, Evgenii Volkov, and Jürgen Kurths. Transition from amplitude to oscillation death via turing bifurcation. *Phys. Rev. Lett.*, Vol. 111, No. 2, p. 024103, 2013.

[87] Andrew Angel, Jie Song, Caroline Dean, and Martin Howard. A polycomb-based switch underlying quantitative epigenetic memory. *Nature*, Vol. 476, No. 7358, pp. 105–108, 2011.

[88] Chikara Furusawa and Kunihiko Kaneko. Epigenetic feedback regulation

accelerates adaptation and evolution. *PloS ONE*, Vol. 8, No. 5, p. e61251, 2013.

[89] Bin Huang, Mingyang Lu, Madeline Galbraith, Herbert Levine, Jose N Onuchic, and Dongya Jia. Decoding the mechanisms underlying cell-fate decision-making during stem cell differentiation by random circuit perturbation. *J. R. Soc. Interface*, Vol. 17, No. 169, p. 20200500, 2020.

[90] Teresa K Barth and Axel Imhof. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem. Sci.*, Vol. 35, No. 11, pp. 618–626, 2010.

[91] Christian Kuehn. *Multiple time scale dynamics*, Vol. 191. Springer, 2015.

[92] Bart Rymen, Ayako Kawamura, Alice Lambolez, Soichi Inagaki, Arika Takebayashi, Akira Iwase, Yuki Sakamoto, Kaori Sako, David S Favero, Momoko Ikeuchi, et al. Histone acetylation orchestrates wound-induced transcriptional activation and cellular reprogramming in arabidopsis. *Communications biology*, Vol. 2, No. 1, pp. 1–15, 2019.

[93] Gemma Anderson, Berta Verd, and Johannes Jaeger. Drawing to extend waddington's epigenetic landscape. *Leonardo*, Vol. 53, No. 3, pp. 256–262, 2020.

[94] Chikara Furusawa and Kunihiko Kaneko. Formation of dominant mode by evolution in biological systems. *Physical Review E*, Vol. 97, No. 4, p. 042410, 2018.

[95] Tomoya Maeda, Junichiro Iwasawa, Hazuki Kotani, Natsue Sakata, Masako Kawada, Takaaki Horinouchi, Aki Sakai, Kumi Tanabe, and Chikara Furusawa. High-throughput laboratory evolution reveals evolutionary constraints in escherichia coli. *Nature communications*, Vol. 11, No. 1, pp. 1–13, 2020.

[96] Josep Casadesús and David Low. Epigenetic gene regulation in the bacterial world. *Microbiology and molecular biology reviews*, Vol. 70, No. 3, pp. 830–856, 2006.

[97] Kunihiko Kaneko, Chikara Furusawa, and Tetsuya Yomo. Universal relationship in gene-expression changes for cells in steady-growth state. *Physical Review X*, Vol. 5, No. 1, p. 011014, 2015.

[98] Takuya U Sato and Kunihiko Kaneko. Evolutionary dimension reduction in phenotypic space. *Physical Review Research*, Vol. 2, No. 1, p. 013197, 2020.

[99] Conrad H Waddington. Canalization of development and the inheritance of acquired characters. *Nature*, Vol. 150, No. 3811, pp. 563–565, 1942.

[100] Conrad H Waddington. Genetic assimilation of an acquired character. *Evolution*, pp. 118–126, 1953.

[101] E Haeckel. *Generelle Morphologie der Organismen*. Georg Reimer, 1866.

[102] Takahiro Kohsokabe and Kunihiko Kaneko. Evolution-development congruence in pattern formation dynamics: Bifurcations in gene expression and regulation of networks structures. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, Vol. 326, No. 1, pp. 61–84, 2016.

[103] Takahiro Kohsokabe and Kunihiko Kaneko. Dynamical systems approach to evolution–development congruence: Revisiting haeckel's recapitulation theory. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 2021.

[104] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory*

and beyond: An Introduction to the Replica Method and Its Applications, Vol. 9. World Scientific Publishing Company, 1987.

[105] Leon Glass and Stuart A Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *Journal of theoretical Biology*, Vol. 39, No. 1, pp. 103–129, 1973.

[106] Chikara Furusawa and Kunihiko Kaneko. Origin of multicellular organisms as an inevitable consequence of dynamical systems. *The Anatomical Record: An Official Publication of the American Association of Anatomists*, Vol. 268, No. 3, pp. 327–342, 2002.

[107] Paul B Rainey and Katrina Rainey. Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, Vol. 425, No. 6953, pp. 72–74, 2003.

[108] Akiko Kashiwagi, Itaru Urabe, Kunihiko Kaneko, and Tetsuya Yomo. Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. *PloS one*, Vol. 1, No. 1, p. e49, 2006.

[109] Chikara Furusawa and Kunihiko Kaneko. A generic mechanism for adaptive growth rate regulation. *PLoS Computational Biology*, Vol. 4, No. 1, p. e3, 2008.

Acknowledgement

First and foremost, I would like to express my gratitude to the supervising professor Kunihiko Kaneko. I appreciate his continuous and patient support, constructive and exciting advice, thoughtful guidance, and encouragement.

I would like to appreciate professor Chikara Furusawa for discussions, opportunities, and foresight studies.

Dr. Tetsuhiro Hatakeyama, Dr. Nobuto Takeuchi, Dr. Atsushi Kamimura, Dr. Masahiko Ueda, Dr. Takahiro Kohsokabe, Dr. Qian-Yuan Tang, and Dr. Masayo Inoue gave me opportunities for discussion and fruitful comments. I would like to express my gratitude to them.

I also would like to thank all of the members, including alumni, in KKlab. They gave me enjoyable days.

Finally, I would like to thank my family for their thoughtful and financial support.