# How Criticality of Gene Regulatory Networks Affects the Resulting Morphogenesis under Genetic Perturbations

Hyobin Kim*,**
Hiroki Sayama**
Binghamton University

**Abstract**  Whereas the relationship between criticality of gene regulatory networks (GRNs) and dynamics of GRNs at a single-cell level has been vigorously studied, the relationship between the criticality of GRNs and system properties at a higher level has not been fully explored. Here we aim at revealing a potential role of criticality of GRNs in morphogenesis, which is hard to uncover through the single-cell-level studies, especially from an evolutionary viewpoint. Our model simulated the growth of a cell population from a single seed cell. All the cells were assumed to have identical intracellular GRNs. We induced genetic perturbations to the GRN of the seed cell by adding, deleting, or switching a regulatory link between a pair of genes. From numerical simulations, we found that the criticality of GRNs facilitated the formation of nontrivial morphologies when the GRNs were critical in the presence of the evolutionary perturbations. Moreover, the criticality of GRNs produced topologically homogeneous cell clusters by adjusting the spatial arrangements of cells, which led to the formation of nontrivial morphogenetic patterns. Our findings correspond to an epigenetic viewpoint that heterogeneous and complex features emerge from homogeneous and less complex components through the interactions among them. Thus, our results imply that highly structured tissues or organs in morphogenesis of multicellular organisms might stem from the criticality of GRNs.

## 1 Introduction

Many complex biological structures, including tissues and organs, are formed through a developmental process from a single fertilized cell to multicellular embryos [5]. The morphogenesis of those complex patterns is driven by a gene regulatory network (GRN) that exists within each cell and responds to cell-cell interactions [35, 36]. Random Boolean networks (RBNs) were proposed by Kauffman as a theoretical model of such GRNs [28]. In RBNs, genes (nodes) have binary states (either ON or OFF) whose dynamics are determined by a set of Boolean functions over the states of

---

Artificial Life 24: 85–105 (2018) doi:10.1162/ARTL_a_00262

other genes. Although RBNs are a highly simplified model, they have been extensively utilized in artificial life and complex systems research [1, 3, 21, 22, 38, 45, 55, 60].

In the context of GRNs, the concept of *criticality* of RBNs has been discussed as a phase transition point between ordered and chaotic regimes for the dynamics of those networks [29, 30]. The criticality of GRNs has been recognized as a property that makes robustness and adaptability coexist in living organisms [2]. When perturbations are added to GRNs, ordered GRNs are so robust that they just sustain existing cellular functions. On the contrary, chaotic GRNs are so adaptable that they vigorously create new functions rather than conserving existing ones. In between, critical GRNs stably sustain their functions against perturbations, and at the same time flexibly generate new phenotypes, which help organisms to adapt to new environments because they have optimal balance between robustness and adaptability.

Whereas many studies have been performed to elucidate the relationship between criticality of GRNs and dynamics of GRNs at the single-cell level based on RBNs [6, 43, 47, 52, 53, 56], the relationship between their criticality and the properties of organisms at a higher level has not been fully explored. Only a few studies have determined how the properties of intracellular GRNs influence properties at a multicellular level, using RBNs as GRNs [12, 13, 18, 40, 58]. They all used multicellular models where all the cells have the same RBNs in discrete spaces like cellular automata. Such a model is not realistic as a morphogenetic model representing the development from a single cell to a multicellular embryo. Moreover, research on the relationship between the criticality of GRNs and properties of multicellular systems under genetic perturbations (e.g., mutations) has not been conducted yet, although mutations do occur in cells of living organisms by stochasticity or through environmental factors [4, 7–9, 11, 49, 54].

To investigate a potential role of criticality of GRNs in morphogenesis, which is hard to uncover through the single-cell-level studies, we have recently proposed GRN-based morphogenetic systems operating in a continuous space and found that the criticality of GRNs facilitated the formation of nontrivial morphologies [31, 32]. To make our finding more solid from an evolutionary viewpoint, in this study we include an evolutionary process in our morphogenetic model by introducing genetic perturbations into intracellular GRNs. We add perturbations to the GRNs by adding, deleting, or switching a regulatory link between a pair of genes. We look into whether the role of the criticality of GRNs can be maintained even in the presence of the evolutionary perturbations. Also, we examine what the resulting morphologies are like and what kind of biological implications this model can offer from the epigenetic viewpoint in morphology.

The rest of the article is structured as follows. In Section 2, we describe our GRN-based morphogenetic model. In Section 3, parameter setting for simulations and various measures for the analysis of results are explained. In Section 4, morphologies obtained from the simulations of the morphogenetic model are analyzed. Section 5 summarizes and concludes the article.

## 2  Model

We have developed a computational model of morphogenetic processes of cell aggregation, in which all the cells have an identical intracellular GRN. Figure 1 shows the algorithm for our model. The simulation starts with one seed cell. It imitates the process in which a single zygote divides and grows into multicellular form during embryonic development. In the cell, a RBN is produced as an intracellular GRN. Neighboring cells are detected within a fixed neighborhood radius. Through the interaction with neighbors, cells' fates are determined by the GRN. We assume that there are four fundamental cell fates in our model: proliferation, differentiation, apoptosis, and quiescence. Cells expressing proliferation, differentiation, or quiescence can switch their fates through cell-cell interaction. The cells are positioned in a two-dimensional continuous space by spring-mass-damper (SMD) kinetics. Until the termination condition of the simulation is satisfied, the initial seed cell grows into an aggregation iterating the processes of finding neighboring cells and positioning cells in the space in each time step. The simulator of our model was implemented in Java.
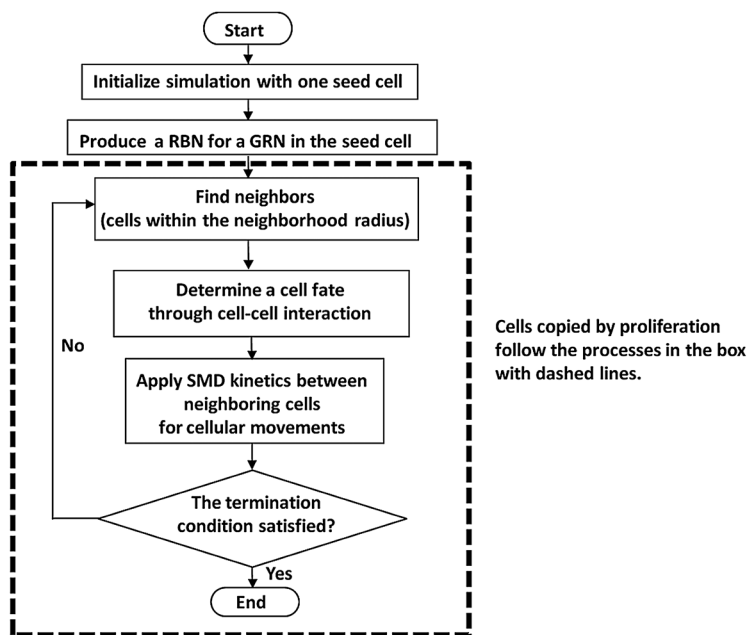
Figure 1. Algorithm for our GRN-based morphogenetic model.

## 2.1  Gene Regulatory Network (GRN)

A RBN (a.k.a. *NK* Boolean network) was suggested as a GRN model by Kauffman [28–30]. Here *N* is the number of nodes, and *K* is the number of input links per node. A node represents a gene. The state of a node can be either ON (1, activated) or OFF (0, inhibited). The node state is determined by the states of input nodes and a Boolean function assigned to each node. A state space that is constructed from the topology of a RBN and assigned Boolean functions refers to the set of all the possible configurations and all the transitions among them. Figure 2 shows schematic diagrams for an example RBN and its state space. In the state space, stationary configurations are called attractors, and the others are called basins of attraction of the attractors. The dynamics of RBNs are divided into three regimes, depending on the structure of their state space: ordered, critical, and chaotic. Using the node indegree (*K*), internal homogeneity (*p*), or canalizing functions, the dynamics of RBNs can be systematically varied. The dynamics of a RBN is known to be determined by *K*, namely, $K = 1$ is ordered, $K = 2$ is critical, and $K > 2$ is chaotic, on average [29, 30]. For our morphogenetic model, we use a RBN that consists of 16 nodes ($N = 16$) as an intracellular GRN. Adjusting *K* from 1 to 4, we vary the dynamics of RBNs: ordered ($K = 1$), critical ($K = 2$), and chaotic ($K = 3, 4$).

In view of in vitro experimental data showing that attractors of GRNs represent cell types or cell fates, Huang suggested a conceptual framework to explain stochastic and reversible transitions between cell fates, using *NK* Boolean networks [10, 25–27]. Our morphogenetic systems are based on his framework. We randomly assign the four cell fates to attractors of GRNs. Specifically, if there is only one attractor, proliferation is assigned to the attractor. If there are two attractors, proliferation and differentiation are randomly assigned to the two attractors. Likewise, if there are three attractors, proliferation, differentiation, and apoptosis are randomly assigned to those attractors. If there are four or more attractors, proliferation, differentiation, and apoptosis are randomly assigned to three attractors and quiescence is assigned to the rest of the attractors (Figure 3).

With regard to cellular behaviors, cells in proliferation are divided into two, and the daughter cells are placed within a fixed neighborhood radius (*R*) centering on the mother cells. Cells in differentiation are labeled as differentiated. Cells in apoptosis die and disappear. Cells in quiescence do not show any behaviors.

| gene1 | gene2 | gene3 | gene4 |
|---|---|---|---|
| 0 | 1 | 1 | 1 |

| Inputs | | Output |
|---|---|---|
| gene2 | gene4 | gene1 |
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

| Inputs | | Output |
|---|---|---|
| gene2 | gene3 | gene2 |
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

| Inputs | | Output |
|---|---|---|
| gene2 | gene4 | gene3 |
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

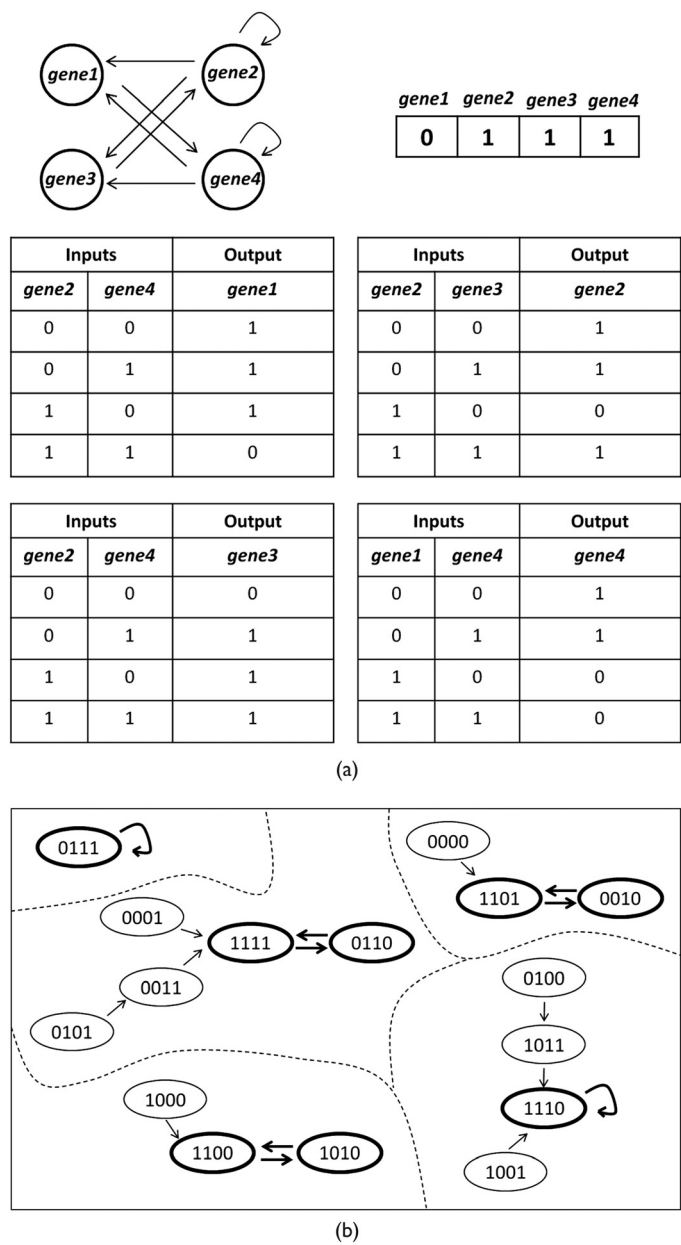| Inputs | | Output |
|---|---|---|
| gene1 | gene4 | gene4 |
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |

(a)



(b)

Figure 2. Schematic diagrams for an example GRN and its state space. (a) A RBN with $N = 4$, $K = 2$, and Boolean functions randomly assigned to each node. (b) State space of the RBN. The state space consists of $2^4 = 16$ configurations and transitions among them. The configurations with bold outlines are attractors. Dashed lines draw boundaries for each basin of attraction.

## 2.2  Cell-Cell Interaction

Switching between cell fates occurs by perturbations of internal gene expression values of an intra-cellular GRN through cell-cell interaction. Our mechanism for cell-cell interaction is based on cell signaling of Damiani et al.'s multiple random Boolean networks on 2D cellular automata [12, 13]. In our model, an intracellular GRN has $n$ genes, which consist of normal genes ($g$) and special genes ($r$) as shown in Figure 4a. The special genes exist in pairs where the genes producing signaling
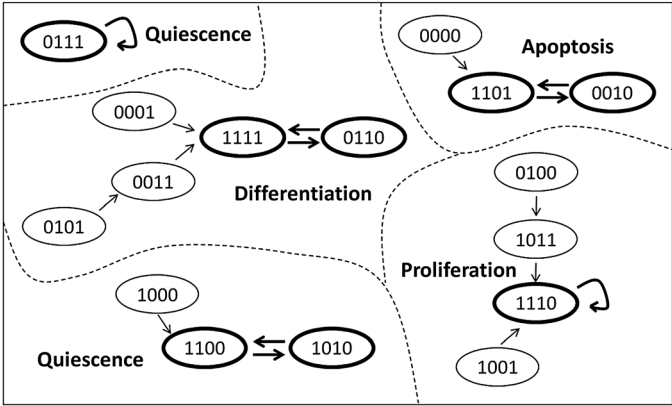
Figure 3. An example state space of a GRN where four cell fates are randomly assigned. (In actual simulations, 16 nodes were used. Thus, $2^{16}$ = 65,536 configurations exist in the state space.)
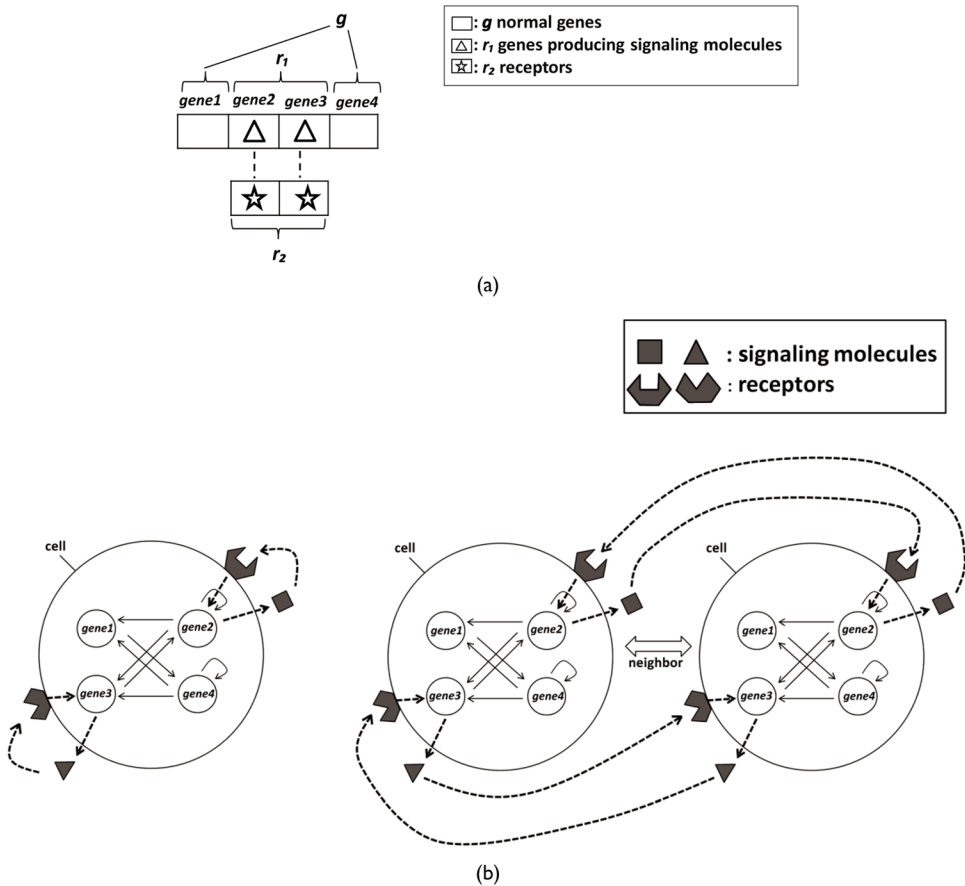


Figure 4. Cell signaling for cell-cell interaction. Schematic diagrams are examples illustrating the concept of cell signaling. (a) Assignment of genes in an intracellular GRN for cell signaling. (b) Two signal transduction mechanisms: autocrine (left) and paracrine (right).
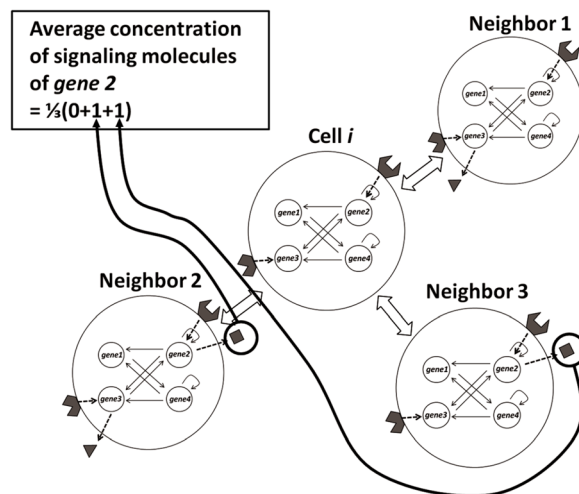
Figure 5. An example showing how to calculate the average concentration of the signaling molecules that neighboring cells produce.

molecules ($r_1$) and receptors ($r_2$) are matched one to one. This one-to-one correspondence indicates signal transduction specificity by which certain signaling molecules respond to particular receptors. The special genes $r_1$ synthesize signaling molecules and release them. Then, those molecules bind to the corresponding receptors $r_2$ on cells within the neighborhood radius $R$.

The signal transduction has two mechanisms: autocrine and paracrine. Autocrine means a cell produces signaling molecules that bind to receptors on the same cell. Paracrine means a cell produces signaling molecules that bind to receptors on its neighboring cells. Figure 4b explains the two mechanisms. In our model, when there are no neighboring cells, autocrine is used. When there are neighbors, paracrine is used.

The gene expression values of an intracellular GRN are updated as follows:

- *Normal genes g*: The states of the normal genes are updated by the states of input nodes and randomly assigned Boolean functions.

- *Genes producing signaling molecules* $r_1$: Like the normal genes *g*, the states of the genes producing signaling molecules $r_1$ are updated by the states of input nodes and randomly assigned Boolean functions. If the states of $r_1$ are 1, the genes produce signaling molecules. If the states are 0, they do not.

- *Receptors* $r_2$: The states of the receptors $r_2$ are determined by the average concentration of the signaling molecules within the neighborhood radius $R$. Figure 5 shows an example of calculating the average concentration of the signaling molecules from the neighboring cells to determine the state of the receptor *gene 2* of cell *i*. Based on a certain threshold ($\tau_{th}$), the state of the receptor *gene 2* is updated. If the average concentration value is bigger than $\tau_{th}$, the state becomes 1. Otherwise, it becomes 0.

The following steps are taken to update the gene expression values and to determine a cell fate from the change of the gene values:

(1)  Look into whether there are neighboring cells within the neighborhood radius $R$. If there are neighbors, paracrine signaling is used. Otherwise, autocrine signaling is used.

(2) Determine the states of the receptors $r_2$ through the average concentrations of signaling molecules within the neighborhood radius $R$ and the threshold value $\tau_{th}$.

(3) Activate or inhibit genes that have the receptors as input nodes in an intracellular GRN based on the states of the receptors $r_2$. If the states of the receptors are 1, the states of genes become ON (1, activated). Otherwise, the states become OFF (0, inhibited).

(4) Check the attractor that the updated gene states finally converge to.

(5) Express the cell fate that is assigned to the attractor for cellular behaviors.

(6) Assign the states of the attractor as gene expression values for the next time step.

## 2.3 Cellular Movements

Our mechanism for cellular movements is based on Doursat's approach [16]. We determine cells' positions in each time step through spring-mass-damper (SMD) kinetics. Specifically, we assume that cells within the neighborhood radius $R$ are connected by a spring with spring constant $k$ and equilibrium length $l$, and a damper with damping coefficient $c$. When cell $A$'s position is $P_A = (x_A, y_A)$ and cell $B$'s position is $P_B = (x_B, y_B)$, the equation for cellular motions is as follows:

$$m\ddot{P}_{AB} = -k\left(1 - \frac{l}{\|P_{AB}\|}\right)P_{AB} - c\dot{P}_{AB},$$

where

$$P_{AB} = P_B - P_A = (\delta \cos\theta, \delta \sin\theta),$$

$$\delta = \|P_{AB}\|, \quad \theta = \arctan\left(\frac{y_B - y_A}{x_B - x_A}\right),$$

$$\|P_{AB}\| = \|P_B - P_A\| = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}.$$

Because we neglect the effect of inertia, we replace $m\ddot{P}_{AB}$ with zero. Then, we finally obtain the following position update equation at each time step $\Delta t = 1$:

$$\Delta P_A = -\Delta P_B = \frac{\Delta P_{AB}}{2} = \frac{-k}{2c}\left(1 - \frac{l}{\|P_{AB}\|}\right)P_{AB}.$$

We can obtain different shapes of spatial patterns by the above position updating rule, allowing physical interactions such as pushing or adhesion.

To acquire much more diverse spatial patterns, we introduce the dependence of parameters $k$, $l$, and $c$ on cell fates and perturbations to the cell position $(x, y)$. In the case of $k$, $l$, and $c$, we determine the values depending on six types of cell fates possible between two cells: [*proli-proli*], [*proli-diff*], [*proli-qui*], [*diff-qui*], [*diff-diff*], and [*qui-qui*], where *proli* is proliferation, *diff* is differentiation, and *qui* is quiescence (Figure 6). Here apoptosis is excluded because cells due to apoptosis disappear from the space. In each simulation run, the values of $k$, $l$, and $c$ are randomly chosen in certain ranges given in Table 1. For the perturbations, small perturbation values are added to the updated cell positions.
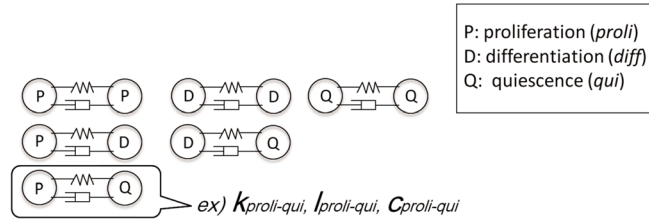
Figure 6. Six types of cell fates possible between two cells.

Table 1. Parameters of model and their values.

| Model | Parameter | Value |
|---|---|---|
| GRN | Number of nodes ($N$) | 16 |
| | Number of indegrees per node ($K$) | 1, 2, 3, 4 |
| Cell-cell interaction | Neighborhood radius ($R$) | 30 |
| | Number of special genes ($r$) | 2 |
| | Concentration threshold ($\tau_{\text{th}}$) | 0.5 |
| Cellular motions | Spring constant ($k$) | $\mathbb{R} \in$ unif (0, 1) |
| | Spring equilibrium length ($l$) | $\mathbb{R} \in$ unif (0, 100) |
| | Damper coefficient ($c$) | $\mathbb{R} \in$ unif (0, 200) |

When the dependence of $k$, $l$, and $c$ on cell fates and perturbations to the cell positions are introduced, the final position of cell $A$ whose neighbor is cell $B$ is as follows:

$$P_A(t + 1) = P_A(t) + (\Delta P_A)_{[\alpha - \beta]} + \omega_{[\alpha - \beta]},$$

where $\alpha$ is cell $A$'s fate, $\beta$ is cell $B$'s fate, and $\omega$ is the perturbation to the updated coordinate of cell $A$. This positional updating is performed for all the neighboring cells of cell $A$.

## 3 Experiments

We performed 10,000 independent simulation runs for each value of $K$, that is, $K = 1, 2, 3,$ and 4. Specifications of parameters for the simulations were the following:

- *Space:* The cells were positioned in a 2D continuous 700 × 700 (in arbitrary units) square area.

- *Limitation of cell population:* The population growth was limited to 200 cells to keep computational loads reasonable in each run.

- *Simulation termination condition:* The simulations were terminated when the time step $t$ reached 1,000 or there was no cell remaining in the space due to apoptosis.

- *Parameter values:* The values of parameters concerning GRNs, cell-cell interaction, and cellular motion are given in Table 1. The number of special genes ($r$) was determined according to Damiani et al.'s model [13] and biological evidence [44]. Damiani et al. showed that the diversity of cellular behaviors was maximized when the coupling strength (fraction of genes that are affected by neighboring cells) was around 0.1 [13]. This value is also similar to the ratio of the number of genes related to cell signaling to the number of human genes [44]. In our model, the coupling strength is set to 0.125, because the number of special genes ($r$) is 2 while the number of nodes of a GRN ($N$) is 16.

## 3.1 Perturbations to GRNs and Robust and Evolvable GRNs

As mentioned in the Introduction (Section 1), we previously presented a GRN-based morphogenetic model and revealed the role of the criticality of GRNs in morphogenesis [31, 32]. To make our finding more solid from an evolutionary viewpoint, we have included genetic perturbations in our morphogenetic model in this study. Specifically, we assumed that the genetic perturbation was due to a germinal mutation occurring in a pre-zygotic cell, which is a small-scale mutation at a genetic level. The germinal mutation is passed on to offspring, and it is present in all resulting cells during embryo development [19, 23]. We perturbed the intracellular GRN of a seed cell at the initial time step by adding, deleting, or switching one regulatory link between a pair of genes [15, 24, 37]. Because cells were duplicated through the process of cell division from the perturbed GRN in the seed cell, all the cells composing a morphogenetic pattern had the same perturbed GRNs.

Such a small regulatory link perturbation did not significantly change the average number of input links per node ($K$). For example, if $K = 2$, the total number of links of a GRN is 32, because the node size is 16. If one regulatory link is deleted as a genetic perturbation, the GRN consists of 31 links, making the value of $K$ 1.94.

Based on the studies showing that biological systems have robustness and evolvability [14, 33, 42, 46, 57, 59], we focused on analyzing robust and evolvable GRNs among the perturbed GRNs. In our model, if the GRN conserved its existing attractors and created new attractors simultaneously after perturbation, we considered the GRN a robust and evolvable GRN (Figure 7) [2].

## 3.2 Morphological Measures for Pattern Analysis

To compare how morphogenetic patterns are different between groups $K = 1, 2, 3$, and 4, we used the 12 measures (i–xii) described below [50, 51]. Among those measures, vi–xii are measures regarding network topology. To apply them to our morphologies, we constructed a network from each morphogenetic pattern by connecting each cell to other cells within the neighborhood radius $R$. Figure 8 shows an example morphology and a network constructed from it using our network
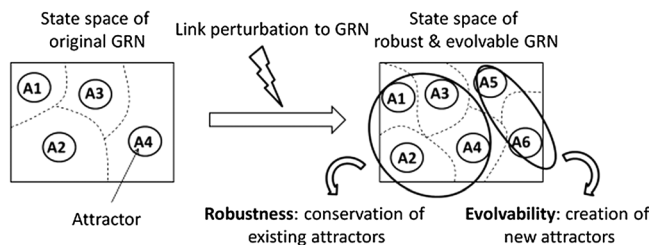


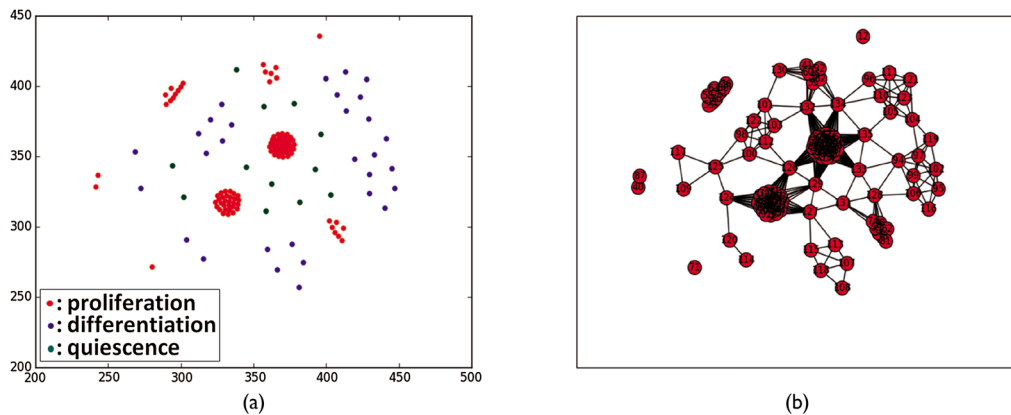Figure 7. Schematic diagrams illustrating the concept of a robust and evolvable GRN.

Figure 8. Network construction for the analysis of morphologies. (a) Snapshot of a morphogenetic pattern. (b) Network constructed using our network construction method from (a).

construction method. Such network construction allowed for detection of topological differences more effectively. All the morphogenetic measures were obtained from the final configuration of each simulation.

i. Number of cells (*numOfCells*). This is the total number of cells in a morphogenetic pattern.

ii. Average distance of cells from center of mass (*massDistance*). This is the mean of Euclidean distances between each cell position and the center of mass ($\bar{x}$, $\bar{y}$), that is, the point with the average coordinates of all the cells.

iii. Average pairwise distance (*pairDistance*). This is the mean of Euclidean distances between two randomly sampled cells' positions. The mean was calculated based on 10,000 pairs, which were sampled with replacement.

iv. Kullback-Leibler divergence between pairwise particle distance distributions of a morphogenetic pattern and a random pattern (*kld*). This quantifies how nontrivial morphogenetic patterns are, compared to randomly distributed patterns. It was calculated as the Kullback-Leibler (KL) divergence between pairwise particle distance distributions of a morphogenetic pattern (Figure 9a) and a randomly distributed pattern (Figure 9b) made of the same number of cells within the same spatial dimensions. Each pairwise particle distance distribution was obtained through 10,000 random samplings with replacement of a pair of coordinates of cells (Figure 9c). Thus, the larger *kld* is, the more structured (nonrandom) the morphogenetic pattern is.

Both *pairDistance* and *kld* used pairwise particle distances. *pairDistance* measures a rough size of a morphogenetic pattern, while *kld* quantifies the nontriviality of its morphology. Two morphogenetic patterns may have similar *pairDistance* values but very different *kld* values at the same time.

v. Mutual information of cell fates between neighboring cells (*MI*). This examines nonlinear correlation of cell fates between neighboring cells in a morphogenetic pattern. It was calculated using the frequencies of three neighboring cell fates (except for apoptosis, because cells expressing apoptosis die and disappear from the space).
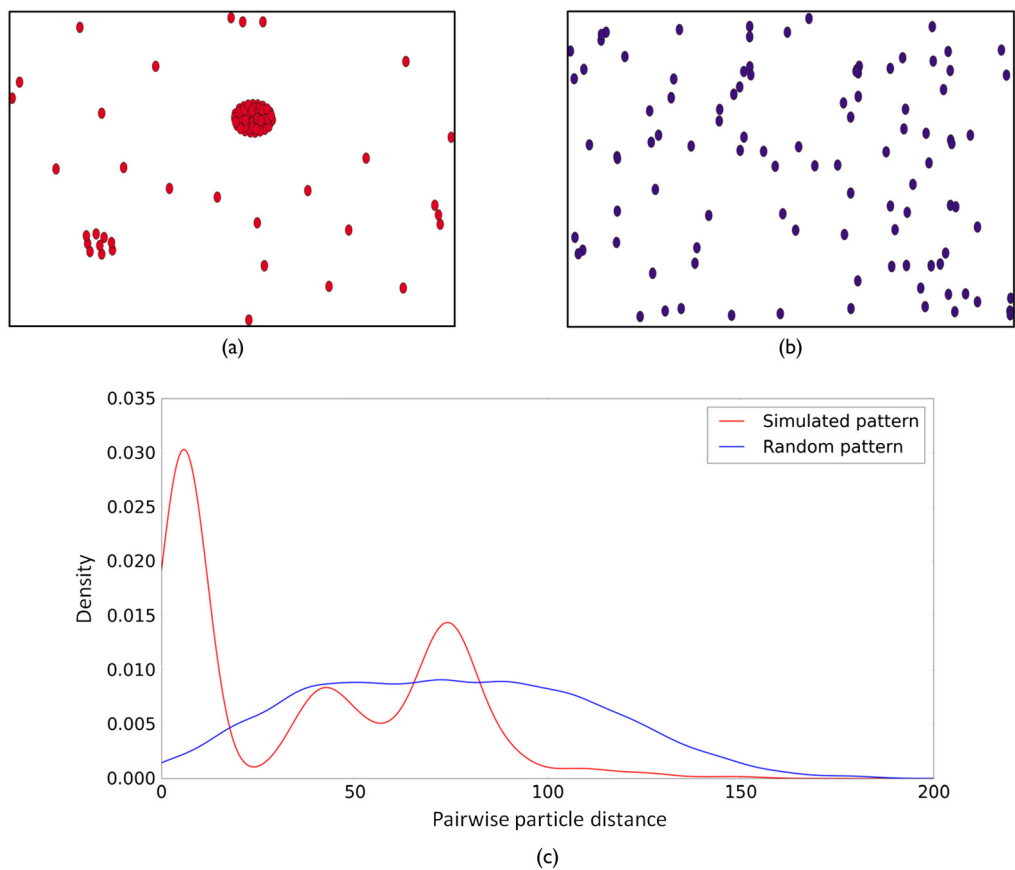
(a)



(b)



(c)

Figure 9. Nontrivial morphology detection using KL divergence. (a) A morphogenetic pattern acquired from a simulation. (b) A random pattern obtained from a uniform distribution. (c) Pairwise particle distance distributions of a simulated pattern and a random pattern. The curves are estimated by Gaussian kernel density estimation.
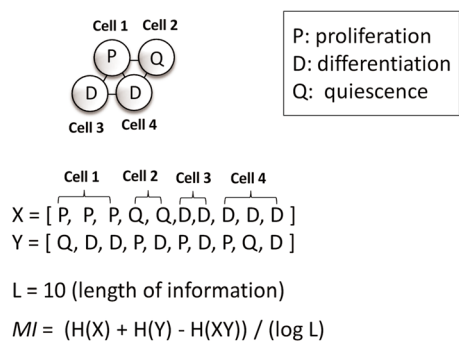


Figure 10. An example showing how to calculate the mutual information ($MI$) between fates of cells and their neighboring cells. The value of the computed mutual information was divided by log $L$ for the purpose of normalization.

Figure 10 shows an example calculating $MI$ in a morphogenetic pattern. Counting the frequencies of combinations of fates of neighboring cells, $MI$ captures how much informational correlation would exist between the fate of a cell and that of its neighbors. If there was only one cell remaining, the value of $MI$ was set to 0.

vi. Average clustering coefficient (*avgCluster*). This explains how densely connected the nodes (cells) are to each other in a network. The clustering coefficient $C(i)$ of node $v_i$ in a network is defined as follows:

$$C(i) = \frac{\left|\left\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\right\}\right|}{k_i(k_i - 1)/2}.$$

Here, $e_{jk}$ is a link that connects node $v_j$ and node $v_k$ within the set of neighboring cells $N_i$ around a node $v_i$, $E$ is a set of links in the network, and $k_i$ is the degree of $v_i$ (i.e., the size of $N_i$). The denominator is the total number of possible node pairs within node $v_i$'s neighborhood. The numerator is the number of actually connected node pairs among them. The average clustering coefficient is given by

$$\bar{C} = \frac{\sum_i c(i)}{n},$$

where $n$ is the total number of nodes.

vii. Link density (*linkDensity*). This describes the density of connections in a network. For a network $G$ composed of nodes $n$ and $m$ links, the link density $\rho(G)$ is given by

$$\rho(G) = \frac{m}{\frac{n(n-1)}{2}} = \frac{2m}{n(n-1)}$$

for an undirected network, where $m$ is the number of links.

viii. Number of connected components (*numConnComp*). This is the number of connected components in a network. A connected component is a subgraph in which there is a path between every pair of nodes. A single isolated cell was also considered one connected component by itself.

ix. Average size of connected components (*meanSizeConnComp*). This is the mean of the numbers of nodes in each connected component in a network. When there was no connected component, the value was set to 0.

x. Homogeneity of sizes of connected components (*homoSizeConnComp*). This quantifies how similar the sizes of connected components are in a network. This measure was calculated as one minus the normalized entropy in the distribution of sizes of connected components.



list of size of connected components: X = [2, 4, 2]
length of information: L = 3
normalized entropy: H(X)/log L = $\frac{-(1/3)*\log(1/3)-(2/3)*\log(2/3)}{\log(3)}$ = 0.58

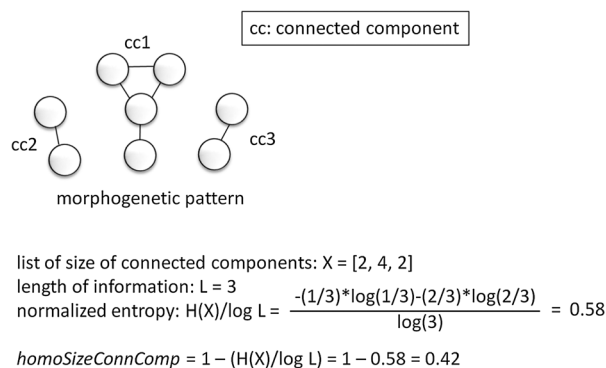*homoSizeConnComp* = 1 − (H(X)/log L) = 1 − 0.58 = 0.42

Figure 11. An example showing how to calculate homogeneity of sizes of connected components (*homoSizeConnComp*). The value of the computed entropy $H(X)$ was divided by log $L$ for the purpose of normalization.

Figure 11 shows an example calculating *homoSizeConnComp* in a morphogenetic pattern. When there was only one connected component, the value was set to 1.

xi.  Size of the largest connected components (*sizeLarConnComp*). This is the maximum size of the connected components in a network.

xii.  Average size of connected components smaller than the largest one (*meanSizeSmaller*). This is the mean of the sizes of connected components except for the largest connected component in a network. If there was only one connected component in the network, the value was set to 0.

For all of the above measures, if there were no cells remaining in the space, their values were set to 0.

## 3.3  Measures to Investigate the Relationship between GRNs and Expressed Cell Fates

To investigate the relationship between intracellular GRNs and expressed cell fates, we calculated the basin entropy and cell fate entropy from the sizes of basins of attractions and cell fates distributed in a morphogenetic pattern. We thought that the numbers of actually expressed cell fates in a morphology might be proportional to the basin sizes of attractors where each cell fate was assigned. We calculated basin and cell fate entropies to look into whether or not our expectation would be correct. As in the computation of *MI*, only three cell fates (proliferation, differentiation, and quiescence) were considered (i.e., apoptosis was ignored).

i.  Basin entropy:

$$H_{\text{basin}} = -\sum_{\rho} P_{\rho} \cdot \log_2 P_{\rho},$$

where $P_{\rho}$ is the size of the basin of the attractor $\rho$ (to which proliferation, differentiation, or quiescence was assigned) divided by the sum of sizes of all the basins except for the basin size of the attractor for apoptosis. Thus,

$$\sum_{\rho} P_{\rho} = 1.$$

Originally, basin entropy was suggested by Krawitz as a measure of the complexity of information that a system can store in *NK* Boolean networks [34]. We used it as a measure to examine the versatility of the three cell fates (proliferation, differentiation, quiescence).

ii.  Cell Fate Entropy:

$$H_{\text{cell fate}} = -\sum_{f} P_{f} \cdot \log_2 P_{f},$$

where $P_f$ is the number of cells expressing a cell fate $f$ (proliferation, differentiation, quiescence), divided by the numbers of cells (except for those in apoptosis) in a morphogenetic pattern at the final time step. Hence,

$$\sum_{f} P_{f} = 1.$$

In the case that there were no cells expressing a fate (proliferation, differentiation, quiescence), its log value was set to 0. Also, when there was no cell in the space, the value was set to 0.

## 4   Results

Figure 12a shows probabilities of producing robust and evolvable GRNs against perturbations for $K = 1$, 2, 3, and 4 in 10,000 simulation runs. We found that robust and evolvable GRNs were generated with the highest probability at $K = 2$. Figure 12b shows samples of visualized morphogenetic patterns produced by robust and evolvable GRNs for each $K$.
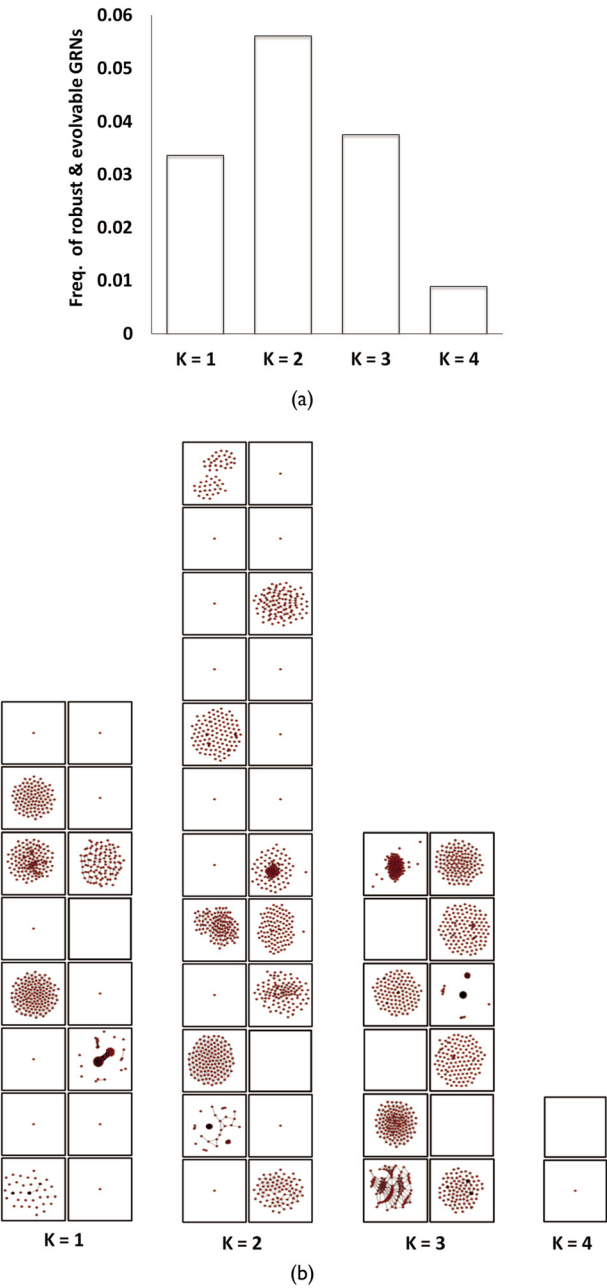


Figure 12. (a) Probabilities of generating robust and evolvable GRNs for $K = 1$, 2, 3, 4 in 10,000 simulation runs. (b) Different morphogenetic patterns obtained from robust and evolvable GRNs for $K = 1$, 2, 3, 4. The numbers of the patterns were counted from robust and evolvable GRNs produced in 500 simulation runs.
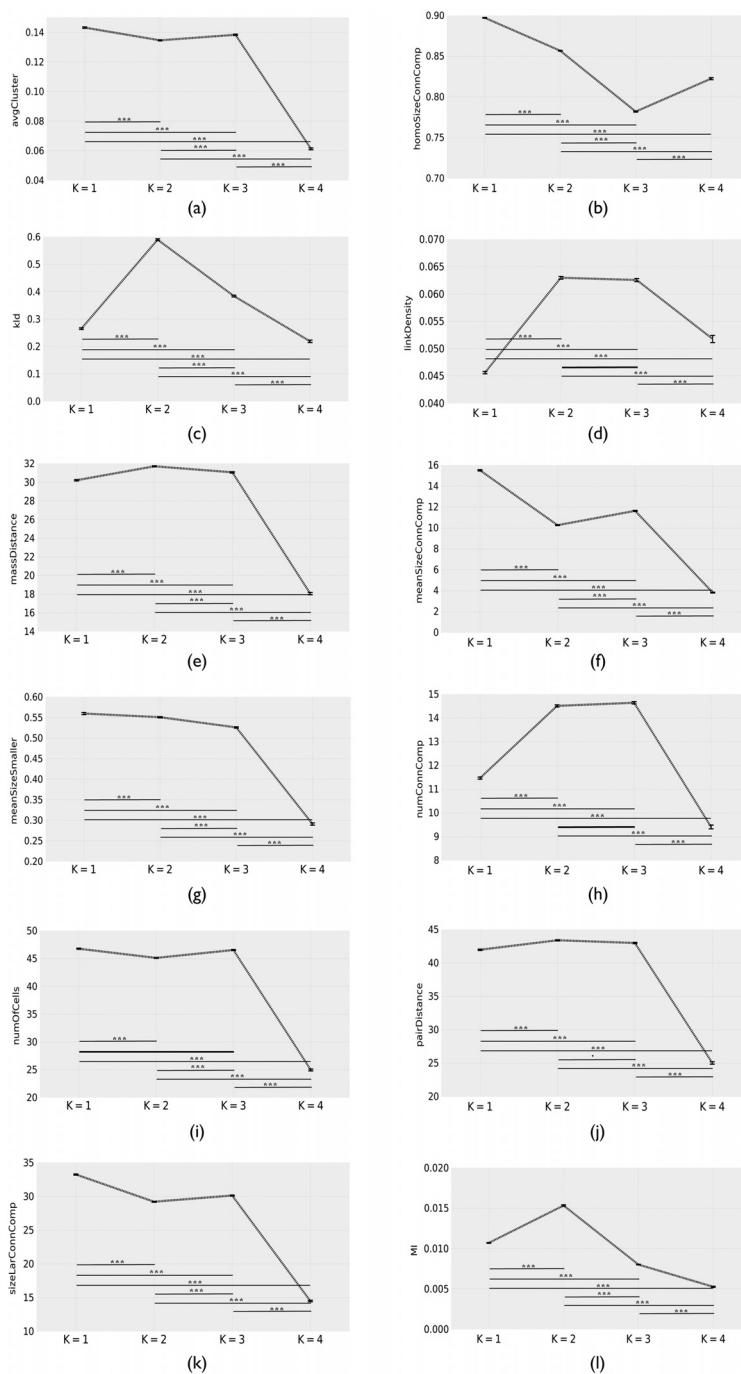
Figure 13. Comparison of means between groups ($K = 1, 2, 3, 4$) for 12 morphological measures (Kruskal-Wallis test: $p < 2.2e{-}16$, Nemenyi test (post hoc): unmarked, $p < 1.0$; `, $p < 0.1$; ***, $p < 0.001$). In the case that there is no difference between two groups, a bold line without an asterisk is drawn in the plot. (a) Average clustering coefficient (*avgCluster*). (b) Homogeneity of sizes of connected components (*homoSizeConnComp*). (c) KL divergence between pairwise particle distance distributions of morphogenetic pattern and a random pattern (*kld*). (d) Link density (*linkDensity*). (e) Average distance of cells from center of mass (*massDistance*). (f) Average size of connected components (*meanSizeConnComp*). (g) Average size of connected components smaller than the largest one (*meanSizeSmaller*). (h) Number of connected components (*numConnComp*). (i) Number of cells (*numOfCells*). (j) Average pairwise distance (*pairDistance*). (k) Size of the largest connected component (*sizeLarConnComp*). (l) Mutual information between cell fates of cells and their neighboring cells (*MI*).

We calculated the 12 morphological measures of morphogenetic patterns generated by robust and evolvable GRNs. Because the robust and evolvable GRNs were generated with different probabilities for different values of $K$, we applied bootstrap sampling 1,000 times to the values of the 12 measures for comparison between groups ($K = 1$–4) with unequal sample sizes. Figure 13 indicates the comparison of means between groups for the measures, where Kruskal-Wallis and Nemenyi (as post hoc analysis) tests were performed to show statistically significant differences among the groups.

Furthermore, we produced a correlation matrix to investigate correlations between the 12 measures (Figure 14). We found that the following six measures were highly correlated with *numOfCells*: *avgCluster*, *massDistance*, *meanSizeConnComp*, *meanSizeSmaller*, *pairDistance*, and *sizeLarConnComp*. These correlations were found in Figure 13 as well. The values of *numOfCells* at $K = 1, 2, 3$ were similar but fell sharply at $K = 4$. This trend was also shown in the six measures highly correlated with *numOfCells*. Meanwhile, *kld*, *MI*, *homoSizeConnComp*, *linkDensity*, and *numConnComp* showed different trends. In Figure 13c, *kld* was highest at $K = 2$, which means that nontrivial morphogenetic patterns were generated most frequently when the GRNs were critical under the genetic perturbations. This result demonstrates that the role of criticality of GRNs is maintained even in the presence of evolutionary perturbations.

In addition, from *MI*, *homoSizeConnComp*, *linkDensity*, and *numConnComp*, we found two interesting properties of the nontrivial morphologies at criticality. Firstly, certain combinations of cell fates between neighboring cells occurred most frequently. In Figure 13l, *MI* was highest at $K = 2$. It indicates that the fate of a cell is strongly correlated with the fate of its neighboring cells in a morphogenetic pattern generated at the criticality. To examine the relationship between intracellular GRNs and expressed cell fates, we measured basin entropy and cell fate entropy (Figure 15). Our original expectation was that if the basins of attraction for the three cell fates were most evenly distributed at $K = 2$, the expressions of different cell fates would be maximally balanced in a morphogenetic pattern. However, the cell fate entropy was highest at $K = 1$, although the basin entropy was highest at $K = 2$. This means that the distribution of cell fates in a morphology was not a simple reflection of the basin sizes of a GRN at a single-cell level, but more like an emergent property at a multicellular level obtained through the developmental process involving cell-cell interactions.
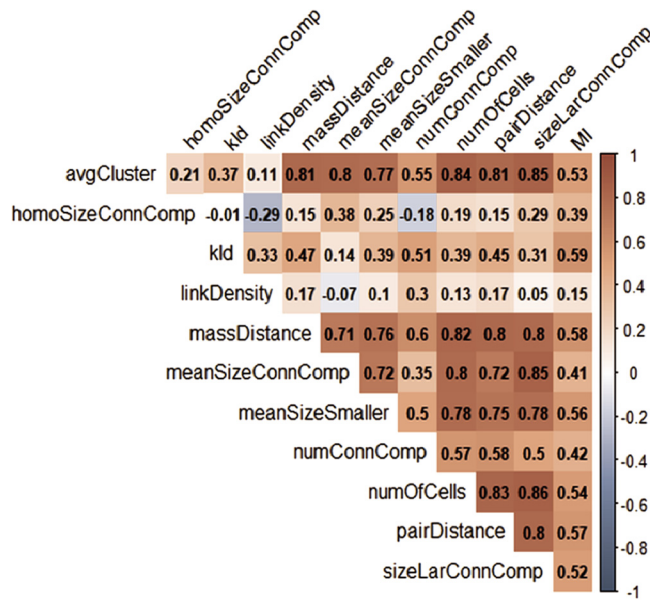


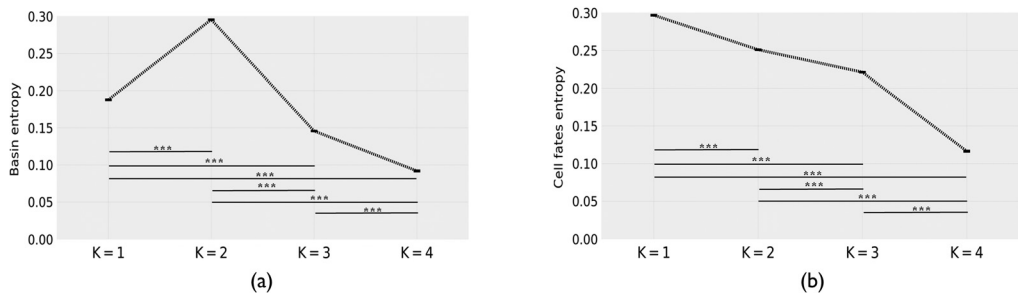Figure 14. Correlation matrix for 12 morphological measures.

Figure 15. Comparison of means between groups for basin and cell fate entropy computed from three cell fates (proliferation, differentiation, quiescence). (a) Average basin entropy for K = 1, 2, 3, 4. (b) Average state entropy of cell fates performed in a simulation at the final time step for K = 1, 2, 3, 4.

Secondly, the nontrivial morphologies emerged typically in topologically homogeneous cell clusters. In Figure 13b,d,h, $K = 1$ showed relatively high *homoSizeConnComp*, low *linkDensity*, and low *numConnComp* values, on average, compared to the corresponding measures at $K = 2$, 3, and 4. Here we simply express the observations qualitatively as (high, low, low). Similarly, in the same order, $K = 2$ showed (high, high, high), $K = 3$ showed (low, high, high), and $K = 4$ showed (low, low, low). These can be interpreted as follows: The morphologies at $K = 1$ consisted of homogeneous large-size connected components, which had a shape like a long chain. The morphologies at $K = 2$ were composed of homogeneous-size connected components where cells were interconnected. The morphologies at $K = 3$ were composed of heterogeneous-size connected components where cells were interconnected. The morphologies at $K = 4$ were composed of heterogeneous-size small connected components that had a shape like a short chain. Figure 16 summarizes the typical topological properties for $K = 1$–4 schematically.

In our morphogenetic model, there is a feedback relationship (Figure 17). Interactions with neighboring cells determine a cell fate. Depending on the cell fates, the parameters of SMD kinetics are applied. The cells are positioned by SMD kinetics. The positions of cells influence the number of neighboring cells. In the feedback relationship, we found that the nontrivial morphologies were
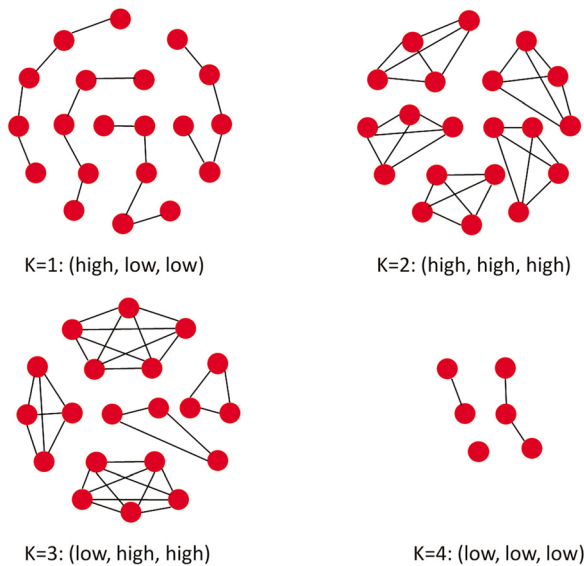


Figure 16. Topological properties of morphogenetic patterns for K = 1, 2, 3, 4. "Low" and "high" mean the relative values against K in the order of (b) *homoSizeConnComp*, (d) *linkDensity*, (h) *numConnComp* in Figure 13.
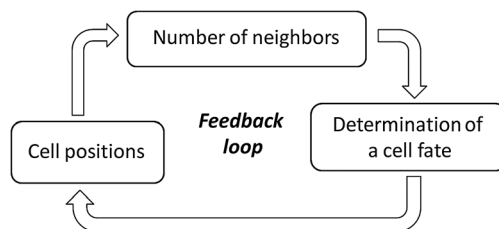
Figure 17. Feedback relationship in our morphogenetic process.

produced most frequently when the GRNs were critical under the genetic perturbations. Besides, the nontrivial morphologies at criticality had the most frequent occurrence of certain combinations of cell fates between neighbors, and were composed of topologically homogeneous cell clusters. Because the parameter values of SMD kinetics determining cells' positions depended on the cell fates, the more frequent those combinations of cell fates between neighbors were, the more likely the same SMD parameter values were to be applied among the cells. Thus, the most frequent combinations of cell fates between neighbors would naturally produce more homogeneous-size connected components where cells were interconnected. Such spatial arrangements of multicellular patterns due to the criticality of GRNs are not easily predictable from the single-cell level. This justifies the value of the present study on the role of criticality of GRNs in the context of multi-cellular settings.

## 5   Conclusions

In this study, we introduced genetic perturbations into our morphogenetic model, which uses Kauffman's RBNs as intracellular GRNs, and SMD kinetics for cellular movements. We looked into whether the previously reported role of the criticality of GRNs could be maintained even in the presence of evolutionary perturbations, and what the resulting morphogenesis would look like. We found that nontrivial morphologies were generated most frequently when the GRNs were critical under the genetic perturbations, which is consistent with the previous result obtained for morphogenetic systems without evolutionary perturbations. Moreover, we found that the nontrivial morphologies at criticality tended to be made of topologically homogeneous cell clusters due to the spatial arrangements in which certain combinations of cell fates between neighboring cells occurred most frequently. Based on these findings, we conclude that the criticality of GRNs facilitates the formation of nontrivial morphologies by adjusting the spatial arrangements of cells in GRN-based morphogenetic systems, even under the genetic perturbations.

Our findings have implications from an epigenetic viewpoint. Researchers in epigenesis have suggested that heterogeneous and complex features emerge from homogeneous and less complex components through the interactions among them [41, 48]. In our model, we showed that the non-trivial morphologies were produced most frequently at criticality, typically with topologically homo-geneous cell clusters. Thus, the result not only supports the theory of epigenesis in developmental biology, but also implies that highly structured tissues or organs in morphogenesis of multicellular organisms might stem from cell aggregations with critical GRNs.

The present study has limitations. The properties of morphogenetic patterns were explored only using our artificial model based on RBNs as GRNs. To make our findings more relevant to real biological systems, we need to develop more biologically plausible models of pattern formation in various tissues, such as the embryonic epithelium of *Xenopus* and *Drosophila* brain tumors, using empirically obtained biological Boolean networks. For example, one could investigate the epigenetic mechanisms of tumorigenesis depending on micro-environmental factors such as nutrient or oxygen gradients. One could build a genome-scale metabolic Boolean network model in a cancer cell, measure which regime the dynamics of that network would be in, and simulate cancer growth

processes in a given micro environment. Such studies would allow for the analysis of a potential relationship between the properties of metabolic Boolean networks at the single-cell level and morphologies at a collective level, using biologically more realistic model assumptions.

For future work, we plan to track the growing processes from the seed cell to the cell aggregation to fully account for the spatial and temporal distribution of cells in terms of the criticality of GRNs in morphogenesis. Also, in order to solidify our findings on the nontrivial patterns, we will measure the complexity of our morphogenetic patterns using other measures such as *information complexity*. Furthermore, we will introduce somatic mosaic mutations in post-zygotic cells that are not inherited genetic alterations in the course of cell division, which are known to be present in actual embryo development [17, 20, 39]. Adding this kind of mutation to our model may reveal different roles of the criticality of GRNs under diverse forms of mutations.

## Acknowledgments

## References

1. Aldana, M. (2003). Boolean dynamics of networks with scale-free topology. *Physica D: Nonlinear Phenomena*, *185*(1), 45–66.

2. Aldana, M., Balleza, E., Kauffman, S., & Resendiz, O. (2007). Robustness and evolvability in genetic regulatory networks. *Journal of Theoretical Biology*, *245*(3), 433–448.

3. Aldana, M., Coppersmith, S., & Kadanoff, L. P. (2003). Boolean dynamics with random couplings. In E. Kaplan, J. E. Marsden, & K. R. Sreenivasan (Eds.), *Perspectives and problems in nonlinear science* (pp. 23–89). New York: Springer.

4. Aminetzach, Y. T., Macpherson, J. M., & Petrov, D. A. (2005). Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. *Science*, *309*(5735), 764–767.

5. Ball, P. (2009). *Shapes: Nature's patterns: A tapestry in three parts* (1st ed.). New York: Oxford University Press.

6. Balleza, E., Alvarez-Buylla, E. R., Chaos, A., Kauffman, S., Shmulevich, I., & Aldana, M. (2008). Critical dynamics in genetic regulatory networks: Examples from four kingdoms. *PLOS ONE*, *3*(6), e2456.

7. Bertram, J. S. (2000). The molecular biology of cancer. *Molecular Aspects of Medicine*, *21*(6), 167–223.

8. Burrus, V., & Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology*, *155*(5), 376–386.

9. Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2013). *From DNA to diversity: Molecular genetics and the evolution of animal design* (2nd ed.). Oxford: Wiley.

10. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., & Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, *453*(7194), 544–547.

11. Chen, J., Miller, B. F., & Furano, A. V. (2014). Repair of naturally occurring mismatches can induce mutations in flanking DNA. *eLife*, *3*, e02001.

12. Damiani, C. (2013). Modelling the influence of cell signaling on the dynamics of gene regulatory networks. In P. Lecca (Ed.), *Biomechanics of cells and tissues* (pp. 103–130). Dordrecht: Springer.

13. Damiani, C., Serra, R., Villani, M., Kauffman, S. A., & Colacci, A. (2011). Cell–cell interaction and diversity of emergent behaviours. *IET Systems Biology*, *5*(2), 137–144.

14. de Visser, J. A. G. M., Hermisson, J., Wagner, G. P., Meyers, L. A., Bagheri-Chaichian, H., Blanchard, J. L., Chao, L., et al. (2003). Perspective: Evolution and detection of genetic robustness. *Evolution*, *57*(9), 1959–1972.

15. Ding, S., & Wang, W. (2011). Recipes and mechanisms of cellular reprogramming: A case study on budding yeast Saccharomyces cerevisiae. *BMC Systems Biology*, *5*(1), 50.

16. Doursat, R. (2008). Programmable architectures that are complex and self-organized: From morphogenesis to engineering. In S. Bullock, J. Noble, R. Watson, & M. A. Bedau (Eds.), *Artificial life XI: Proceedings of the*

*Eleventh International Conference on the Simulation and Synthesis of Living Systems* (pp. 181–188). Cambridge, MA: MIT Press.

17. Fitzgerald, P. H., Donald, R. A., & Kirk, R. L. (1979). A true hermaphrodite dispermic chimera with 46, XX and 46, XY karyotypes. *Clinical Genetics*, *15*(1), 89–96.

18. Flann, N. S., Mohamadlou, H., & Podgorski, G. J. (2013). Kolmogorov complexity of epithelial pattern formation: The role of regulatory network configuration. *BioSystems*, *112*(2), 131–138.

19. Foulkes, W. D., & Real, F. X. (2013). Many mosaic mutations. *Current Oncology*, *20*(2), 85.

20. Freed, D., Stevens, E. L., & Pevsner, J. (2014). Somatic mosaicism in the human genome. *Genes*, *5*(4), 1064–1094.

21. Gershenson, C. (2012). Guiding the self-organization of random Boolean networks. *Theory in Biosciences*, *131*(3), 181–191.

22. Gershenson, C., Kauffman, S. A., & Shmulevich, I. (2006). The role of redundancy in the robustness of random Boolean networks. In L. M. Rocha, L. S. Yaeger, M. A. Bedau, D. Floreano, R. L. Goldstone, & A. Vespignani (Eds.), *Artificial life X: Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems* (pp. 35–42). Cambridge, MA: MIT Press.

23. Griffiths, A. J., Wessler, S. R., Lewontin, R. C., Gelbart, W. M., Suzuki, D. T., & Miller, J. H. (2005). *An introduction to genetic analysis* (8th ed.). New York: W. H. Freeman.

24. Han, B., & Wang, J. (2007). Quantifying robustness and dissipation cost of yeast cell cycle network: The funneled energy landscape perspectives. *Biophysical Journal*, *92*(11), 3755–3763.

25. Huang, S. (1999). Gene expression profiling, genetic networks, and cellular states: An integrating concept for tumorigenesis and drug discovery. *Journal of Molecular Medicine*, *77*(6), 469–480.

26. Huang, S., & Ingber, D. E. (2000). Shape-dependent control of cell growth, differentiation, and apoptosis: Switching between attractors in cell regulatory networks. *Experimental Cell Research*, *261*(1), 91–103.

27. Huang, S., Eichler, G., Bar-Yam, Y., & Ingber, D. E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical Review Letters*, *94*(12), 128701.

28. Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, *22*(3), 437–467.

29. Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution* (1st ed.). New York: Oxford University Press.

30. Kauffman, S. A. (1996). *At home in the universe: The search for the laws of self-organization and complexity* (reprint ed.). New York: Oxford University Press.

31. Kim, H., & Sayama, H. (2017). Criticality of gene regulatory networks and the resulting morphogenesis. In C. Knibbe, G. Beslon, D. P. Parsons, D. Misevic, J. Rouzaud-Cornabas, N. Bredeche, S. Hassas, O. Simonin, & H. Soula (Eds.), *ECAL 2017: Proceedings of the 14th European Conference on Artificial Life* (pp. 245–246). Cambridge, MA: MIT Press.

32. Kim, H., & Sayama, H. (2017). *The role of criticality of gene regulatory networks in morphogenesis*. Available at https://arxiv.org/abs/1708.07882.

33. Kirschner, M., & Gerhart, J. (1998). Evolvability. *Proceedings of the National Academy of Sciences of the U.S.A.*, *95*(15), 8420–8427.

34. Krawitz, P., & Shmulevich, I. (2007). Basin entropy in Boolean network ensembles. *Physical Review Letters*, *98*(15), 158701.

35. Lander, A. D. (2007). Morpheus unbound: Reimagining the morphogen gradient. *Cell*, *128*(2), 245–256.

36. Lander, A. D. (2011). Pattern, growth, and control. *Cell*, *144*(6), 955–969.

37. Li, F., Long, T., Lu, Y., Ouyang, Q., & Tang, C. (2004). The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the U.S.A.*, *101*(14), 4781–4786.

38. Lizier, J. T., Prokopenko, M., & Zomaya, A. Y. (2008). The information dynamics of phase transitions in random Boolean networks. In S. Bullock, J. Noble, R. Watson, & M. A. Bedau (Eds.), *Artificial life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems* (pp. 374–381). Cambridge, MA: MIT Press.

39. De Marchi, M., Carbonara, A. O., Carozzi, F., Massara, F., Belforte, L., Molinatti, G. M., Bisbocci, D., Passarino, M. P., & Palestro, G. (1976). True hermaphroditism with XX/XY sex chromosome mosaicism: Report of a case. *Clinical Genetics*, *10*(5), 265–272.

40. Mohamadlou, H., Podgorski, G. J., & Flann, N. S. (2016). Modular genetic regulatory networks increase organization during pattern formation. *BioSystems*, *146*, 77–84.

41. Moss, L. (2004). *What genes can't do* (1st ed.). Cambridge, MA: MIT Press.

42. Nehaniv, C. (2003). Evolvability. *BioSystems*, *69*, 77–81.

43. Nykter, M., Price, N. D., Aldana, M., Ramsey, S. A., Kauffman, S. A., Hood, L. E., Yli-Harja, O., & Shmulevich, I. (2008). Gene expression dynamics in the macrophage exhibit criticality. *Proceedings of the National Academy of Sciences of the U.S.A.*, *105*(6), 1897–1900.

44. Papin, J. A., Hunter, T., Palsson, B. O., & Subramaniam, S. (2005). Reconstruction of cellular signalling networks and analysis of their properties. *Nature Reviews Molecular Cell Biology*, *6*(2), 99.

45. Paul, U., Kaufman, V., & Drossel, B. (2006). Properties of attractors of canalyzing random Boolean networks. *Physical Review E*, *73*(2), 026118.

46. Poole, A. M., Phillips, M. J., & Penny, D. (2003). Prokaryote and eukaryote evolvability. *BioSystems*, *69*(2), 163–185.

47. Rämö, P., Kesseli, J., & Yli-Harja, O. (2006). Perturbation avalanches and criticality in gene regulatory networks. *Journal of Theoretical Biology*, *242*(1), 164–170.

48. Robert, J. S. (2004). *Embryology, epigenesis and evolution: Taking development seriously* (1st ed.). Cambridge and New York: Cambridge University Press.

49. Rodgers, K., & McVey, M. (2016). Error-prone repair of DNA double-strand breaks. *Journal of Cellular Physiology*, *231*(1), 15–24.

50. Sayama, H. (2014). Four classes of morphogenetic collective systems. In H. Sayama, J. Rieffel, S. Risi, R. Doursat, & H. Lipson (Eds.), *Artificial life 14: Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems* (pp. 320–327). Cambridge, MA: MIT Press.

51. Sayama, H., & Wong, C. (2011). Quantifying evolutionary dynamics of swarm chemistry. In T. Lenaerts, M. Giacobini, H. Bersini, P. Bourgine, M. Dorigo, & R. Doursat (Eds.), *Advances in artificial Life, ECAL 2011: Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems* (pp. 729–730). Cambridge, MA: MIT Press.

52. Serra, R., Villani, M., & Semeria, A. (2004). Genetic network models and statistical properties of gene expression data in knock-out experiments. *Journal of Theoretical Biology*, *227*(1), 149–157.

53. Serra, R., Villani, M., Graudenzi, A., & Kauffman, S. A. (2007). Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. *Journal of Theoretical Biology*, *246*(3), 449–460.

54. Sharma, S., Javadekar, S. M., Pandey, M., Srivastava, M., Kumari, R., & Raghavan, S. C. (2016). Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death & Disease*, *6*(3), e1697.

55. Shmulevich, I., & Kauffman, S. A. (2004). Activities and sensitivities in Boolean network models. *Physical Review Letters*, *93*(4), 048701.

56. Shmulevich, I., Kauffman, S. A., & Aldana, M. (2005). Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proceedings of the National Academy of Sciences of the U.S.A.*, *102*(38), 13439–13444.

57. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J., & Doyle, J. (2004). Robustness of cellular functions. *Cell*, *118*(6), 675–685.

58. Villani, M., Serra, R., Ingrami, P., & Kauffman, S. A. (2006). Coupled random Boolean network forming an artificial tissue. In S. El Yacoubi, B. Chopard, & S. Bandini (Eds.), *Proceedings of the International Conference on Cellular Automata (ACRI 2006)* (pp. 548–556). Berlin: Springer.

59. Wagner, A. (2005). *Robustness and evolvability in living systems* (1st ed.). Princeton: Princeton University Press.

60. Willadsen, K., Triesch, J., & Wiles, J. (2008). Understanding robustness in random Boolean networks. In S. Bullock, J. Noble, R. Watson, & M. A. Bedau (Eds.), *Artificial life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems* (pp. 694–701). Cambridge, MA: MIT Press.