

A computational scheme connecting gene regulatory network dynamics with heterogeneous stem cell regeneration

Yakun Li^a, Xiyin Liang^a, Jinzhi Lei^{a,*}

^a School of Mathematical Sciences, Center for Applied Mathematics, Tiangong University, Tianjin, 300387, China

Abstract

Stem cell regeneration is a vital biological process in self-renewing tissues, governing development and tissue homeostasis. Gene regulatory network dynamics are pivotal in controlling stem cell regeneration and cell type transitions. However, integrating the quantitative dynamics of gene regulatory networks at the single-cell level with stem cell regeneration at the population level poses significant challenges. This study presents a computational framework connecting gene regulatory network dynamics with stem cell regeneration through a data-driven formulation of the inheritance function. The inheritance function captures epigenetic state transitions during cell division in heterogeneous stem cell populations. Our scheme allows the derivation of the inheritance function based on a hybrid model of cross-cell-cycle gene regulation network dynamics. The proposed scheme enables us to derive the inheritance function based on the hybrid model of cross-cell-cycle gene regulation network dynamics. By explicitly incorporating gene regulatory network structure, it replicates cross-cell-cycling gene regulation dynamics through individual-cell-based modeling. The numerical scheme holds the potential for extension to diverse gene regulatory networks, facilitating a deeper understanding of the connection between gene regulation dynamics and stem cell regeneration.

Keywords: cell plasticity, inheritance function, gene regulatory network, cell division

1. Introduction

Stem cell regeneration is a critical biological process in the development and maintenance of tissues, ensuring their self-renewal and homeostasis[1]. The intricate interplay between gene expression regulation and cell population dynamics is central to maintaining dynamic equilibrium during stem cell regeneration. However, as these processes are typically described by distinct mathematical models, bridging the gap between their scales of dynamics is imperative.

In the context of heterogeneous stem cell regenerations, a comprehensive mathematical framework has emerged, integrating cell division behavior with cell heterogeneity and random transition of epigenetic

*Corresponding author: jzlei@tiangong.edu.cn (J. Lei)

states[2, 3]. This framework introduces a variable $\mathbf{x}$, often a high-dimensional vector, to represent the epigenetic state of a cell. The evolution of the population of cells with epigenetic state $\mathbf{x}$, denoted as $Q(t, \mathbf{x})$, is governed by a differential-integral equation[2, 3]

$$\left\{ \begin{array}{l} \frac{\partial Q(t, \mathbf{x})}{\partial t} = -Q(t, \mathbf{x})(\beta(c(t), \mathbf{x}) + \kappa(\mathbf{x})) \\ \quad + 2 \int_{\Omega} \beta(c(t - \tau(\mathbf{y})), \mathbf{y}) Q(t - \tau(\mathbf{y}), \mathbf{y}) e^{-\mu(\mathbf{y})\tau(\mathbf{y})} p(\mathbf{x}, \mathbf{y}) d\mathbf{y}, \\ c(t) = \int_{\Omega} Q(t, \mathbf{x}) \zeta(\mathbf{x}) d\mathbf{x}. \end{array} \right. \quad (1.1)$$

Here, β , κ , and μ represent the rates of cell proliferation, differentiation/senescence, and apoptosis, respectively; τ indicates the duration of cell proliferation; ζ signifies the rate of cytokine secretion. The concentration of growth factors secreted by all cells is denoted by c , and $p(\mathbf{x}, \mathbf{y})$ quantifies the probability of transferring epigenetic states during cell division. The equation (1.1) was proposed by generalizing the classical G0 cell cycle model and incorporating the epigenetic state of individual cells represented by a continuous multidimensional variable[2].

The epigenetic state $\mathbf{x}$ in (1.1) represents intrinsic cellular states that may dynamically change over time within a cell cycle or during cell division. Biologically, the epigenetic state of a cell can refer to molecular level changes that are independent of the DNA sequences, including DNA methylation, nucleosome histone modifications, and gene expression[4, 5, 6, 7, 8, 9]. The epigenetic state may influence the quantification of cellular behaviors, such as proliferation (β), differentiation (κ), apoptosis (μ), and cell cycle (τ), which collectively define the *kinetotype* of a cell[2].

Gene regulatory dynamics within a cell cycle are often described by a set of ordinary or stochastic differential equations

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}) + \boldsymbol{\eta}(t), \quad (1.2)$$

where $\mathbf{X}$ represents the vector of gene expression. The function $\mathbf{F}$ describes the regulatory relationship defined by gene regulatory networks, and $\boldsymbol{\eta}(t)$ denotes the stochastic fluctuations in molecular concentration changes. When the epigenetic state $\mathbf{x}$ is associated with gene expression levels $\mathbf{X}$, e.g., $\mathbf{x} = \log(1 + \mathbf{X})$, equation (1.2) describes the dynamics of the epigenetic state inside a single cell.

In addition to gene regulatory dynamics within a cell cycle, epigenetic states undergo random changes during cell division, leading to cell plasticity in regeneration. This randomness is encapsulated by the inheritance function $p(\mathbf{x}, \mathbf{y})$ in (1.1), which plays a crucial role in characterizing cell type changes. However, precisely formulating this function poses challenges due to the complexity of biochemical reactions during cell divisions.

Biologically, $p(\mathbf{x}, \mathbf{y})$ represents the probability that a daughter cell with state $\mathbf{x}$ originated from a mother

cell with state $\mathbf{y}$ after division. Thus, treating $p(\mathbf{x}, \mathbf{y})$ as a conditional probability density

$$p(\mathbf{x}, \mathbf{y}) = P(\text{state of daughter cell} = \mathbf{x} \mid \text{state of mother cell} = \mathbf{y}), \quad (1.3)$$

we can focus on the epigenetic state before and after cell division while simplifying the analysis of the intermediate processes. The mathematical formulation of the inheritance function is often difficult to derive following the complex biological process. Numerical simulations based on epigenetic state inheritance laws[10, 11, 12] offer phenomenological representations of the inheritance function. For instance, normalized nucleosome modifications follow a beta distribution[10], while transcript levels follow a gamma distribution[13]. Nevertheless, how the structure of the gene regulation network may be involved in the inheritance function remains not known.

This study aims to develop a computational scheme linking gene regulatory network dynamics and stem cell regeneration through a data-driven formulation of the inheritance function. We focused on a well-studied gene network governing cell fate decisions and cell type switches, devising a numerical scheme to obtain the conditional probability density $p(\mathbf{x}, \mathbf{y})$. The obtained density function, verified through stochastic simulation, provides insights into how gene regulatory network structure influences inheritance. Importantly, our numerical scheme offers a pathway for extending these findings to broader gene regulatory networks, facilitating deeper insights into the nexus between gene regulation dynamics and stem cell regeneration.

2. Model and Method

We referred to the hybrid model of stem cell differentiation developed to study the dynamics of cell-type transitions mediated by epigenetic modifications[14]. This model integrates individual-cell-based modeling of a multicellular system, dynamics of a gene regulation network (GRN) involving two genes, a G0 cell cycle model governing cell regeneration, and stochastic inheritance of epigenetic states during cell divisions.

2.1. Heterogeneous G0 cell cycle model

In our model, we focused solely on cells capable of cycling, excluding those that had lost this capacity from the simulation pool. We utilized the G0 cell cycle model to describe the dynamics of stem cell regeneration[15, 16]. According to this model, cells progress through a resting phase known as G0, during which they undergo growth and prepare to enter the proliferative phase upon receiving cell cycling checkpoint signals (Fig. 1). Each cell in the resting phase has the potential to either enter the proliferating phase at a rate β or exit the resting pool at a rate κ due to processes such as terminal differentiation, senescence, or death. Cells in the proliferating phase are randomly eliminated at a rate μ due to apoptosis or undergo mitosis at a fixed time τ following entry into the proliferative compartment. Mitosis results in the generation of two daughter cells from each mother cell. The newborn cells then enter the resting phase, initiating the next cycle.

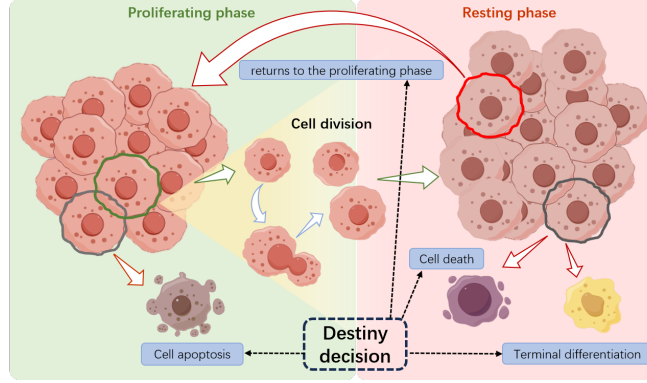


Figure 1: **G0 cell cycle model of cell regeneration.** Schematic representation of the G0 model depicting stem cell regeneration. Cells in the resting phase either enter the proliferative phase at a rate β or exit the resting pool at a rate κ due to differentiation, senescence, or death. Proliferating cells undergo apoptosis at a rate μ .

Biologically, the self-renewal capability of a cell is intricately linked to microenvironmental conditions, including growth factors, various cytokines, and intracellular signaling pathways[17, 18, 19]. Despite the complexity of these pathways, the phenomenological formation of Hill function dependence can be derived from simple assumptions regarding interactions between signaling molecules and receptors[2, 20], expressed as:

$$\beta = \beta_0 \frac{\theta^n}{\theta^n + c^n}, \quad (2.1)$$

where β_0 denotes the maximum proliferation rate, c represents the impact of cytokine signals released from all stem cells, θ represents the half-effective concentration of cytokines, and n stands for the Hill coefficient. This cytokine signal often depends on the collective state of all stem cells within the niche under consideration, forming a feedback loop in cell growth.

Proliferation, differentiation, senescence, and cell apoptosis (or cell death at the resting phase) are pivotal in determining the population dynamics of multicellular systems. Furthermore, cellular heterogeneity, characterized by variance in the epigenetic state of each cell, significantly influences these processes. The kinetic rates of proliferation, differentiation, and apoptosis of each cell are dependent on its epigenetic states. To represent this heterogeneity, we introduced a variable $\mathbf{x}$ (typically a high-dimensional vector) for the epigenetic state of a cell and Ω as the space encompassing all potential epigenetic states of the resting-phase stem cells[2, 3, 21]. Through the epigenetic state $\mathbf{x} \in \Omega$, let $Q(t, \mathbf{x})$ denote the count of cells in the resting phase at time t with epigenetic state $\mathbf{x}$, and the total cell count is given by

$$Q(t) = \int_{\Omega} Q(t, \mathbf{x}) d\mathbf{x}.$$

If $\xi(\mathbf{x})$ represents the rate of cytokine secretion by a cell with the state $\mathbf{x}$, the effective cytokine concentration

regulating cell proliferation is expressed as

$$c(t) = \int_{\Omega} Q(t, \mathbf{x}) \zeta(\mathbf{x}) d\mathbf{x}.$$

Moreover, the kinetic rates β , κ , μ , and the duration of the proliferating phase τ are dependent on the epigenetic state $\mathbf{x}$. The quadruple $(\beta(c, \mathbf{x}), \kappa(\mathbf{x}), \mu(\mathbf{x}), \tau(\mathbf{x}))$ constitutes the *kinetotype* of a cell[2].

During cell division, a single mother cell divides into two daughter cells. However, these daughter cells may not share the same epigenetic state as the mother cell, leading to cell plasticity in cellular regulation. By introducing an inheritance function (also known as the transition function) $p(\mathbf{x}, \mathbf{y})$ defined by the conditional probability density (3.9), the evolution equation (1.1) for the cell population $Q(t, \mathbf{x})$ can be derived from the G0 cell cycle model as follows[2]:

$$\left\{ \begin{array}{l} \frac{\partial Q(t, \mathbf{x})}{\partial t} = -Q(t, \mathbf{x})(\beta(c(t), \mathbf{x}) + \kappa(\mathbf{x})) \\ \quad + 2 \int_{\Omega} \beta(c(t - \tau(\mathbf{y})), \mathbf{y}) Q(t - \tau(\mathbf{y}), \mathbf{y}) e^{-\mu(\mathbf{y})\tau(\mathbf{y})} p(\mathbf{x}, \mathbf{y}) d\mathbf{y}, \\ c(t) = \int_{\Omega} Q(t, \mathbf{x}) \zeta(\mathbf{x}) d\mathbf{x}. \end{array} \right. \quad (2.2)$$

Equation (2.2) provides a comprehensive mathematical framework for modeling the dynamics of heterogeneous stem cell regeneration with epigenetic transitions.

2.2. Gene regulation network

2.2.1. Ordinary differential equation model

We introduced gene regulation networks to establish connections between the epigenetic state and gene expressions within individual cells. For simplicity, we considered a basic gene network comprising two master transcription factors (TFs), namely gene A and gene B. These factors exhibit self-activation and mutual repression, as illustrated in Figure 2. Let X_1 and X_2 denote the expression levels (protein concentration) of genes A and B, respectively. While we focus on the master genes A and B, the gene expression levels $\mathbf{X} = (X_1, X_2)$ define the epigenetic state of a cell.

The kinetics of gene expression within one cell cycle is modeled using the following ordinary differential equations (ODEs):

$$\left\{ \begin{array}{l} \frac{dX_1}{dt} = g_1 \left(\frac{1 + \lambda_{21} \left(\frac{X_2}{h_{21}} \right)^{n_{21}}}{1 + \left(\frac{X_2}{h_{21}} \right)^{n_{21}}} \right) \left(\frac{1 + \lambda_{11} \left(\frac{X_1}{h_{11}} \right)^{n_{11}}}{1 + \left(\frac{X_1}{h_{11}} \right)^{n_{11}}} \right) - k_1 X_1, \\ \frac{dX_2}{dt} = g_2 \left(\frac{1 + \lambda_{12} \left(\frac{X_1}{h_{12}} \right)^{n_{12}}}{1 + \left(\frac{X_1}{h_{12}} \right)^{n_{12}}} \right) \left(\frac{1 + \lambda_{22} \left(\frac{X_2}{h_{22}} \right)^{n_{22}}}{1 + \left(\frac{X_2}{h_{22}} \right)^{n_{22}}} \right) - k_2 X_2. \end{array} \right. \quad (2.3)$$

Here, the parameters g_i , λ_{ij} , n_{ij} , h_{ij} , and k_i ($i, j = 1, 2$) are non-negative parameters. The parameters g_i represent the basal protein production rates of the two genes; λ_{ij} denote the regulatory strength from

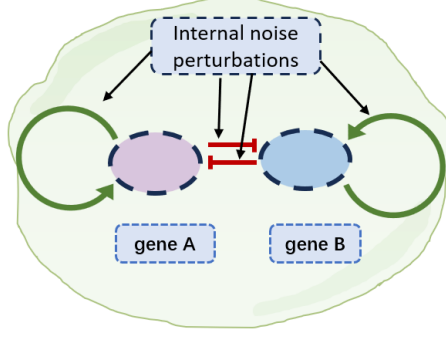


Figure 2: **Schematic diagram depicting the gene regulatory network.** Green arrows indicate self-activating interactions of genes, and red bars indicate repressive interactions between genes.

gene j to gene i , with $\lambda_{ij} < 1$ indicating repression and $\lambda_{ij} > 1$ indicating activation; h_{ij} represent the half-effective protein concentrations for the corresponding regulation; n_{ij} are the Hill coefficients; k_i denote the degradation rates of the two proteins.

Follworing the setup in RACIPE[22], we define a function H as

$$H(X; h, n, \lambda) = \frac{1 + \lambda(\frac{X}{h})^n}{1 + (\frac{X}{h})^n},$$

and equation (2.3) can be expressed as

$$\frac{dX_i}{dt} = g_i \prod_{j=1}^2 H(X_j; h_{ji}, n_{ji}, \lambda_{ji}) - k_i X_i, \quad (i = 1, 2) \quad (2.4)$$

for simplicity.

It is worth noting that the gene network in Figure 2 can be modeled using two different approaches: additive or multiplicative regulations, where the production rates from positive and negative feedbacks are either added or multiplied, respectively[23, 24]. This study adopts the multiplicative modeling approach, consistent with the model described by Jolly et al. (2015)[23]. However, the computational scheme presented in this study remains consistent with additive feedback regulations.

2.2.2. Random perturbation to model parameters

In equation (2.4), the production rates g_i , effective concentrations h_{ij} , and degradation rates k_i may experience random fluctuations due to the extrinsic noise in the cellular environment. To incorporate the effects of these random perturbations, we introduced stochastic fluctuations to each parameter a ($a = g_i, h_{ij}$, or k_i) as follows:

$$a = \bar{a} e^{\sigma \eta - \sigma^2/2}, \quad (2.5)$$

where σ represents the intensity of the noise perturbation, and η follows a colored noise pattern defined by the Ornstein-Uhlenbeck process:

$$d\eta = -(\eta/\vartheta)dt + \sqrt{2/\vartheta}dW, \quad (2.6)$$

where W denoting the Wiener process and ϑ represents the relaxation coefficient.

Unlike the conventional Langevin stochastic differential equation approach, the random perturbation η in (2.5) is expressed exponentially. This form has been previously utilized in research studies to simulate extrinsic noise perturbations in gene expression[11, 14, 25]. The exponential form (2.5) avoids potential issues with negative parameter values that may arise with the Langevin approach $a \rightarrow a + \sigma\eta$. In biochemical reactions, a chemical rate a is often associated with a chemical potential μ through $a = ce^{-\beta\mu}$. Consequently, a perturbation to the chemical potential μ naturally translates into an exponential form perturbation in the chemical reaction rate. Moreover, gene expression rates have been observed to follow a log-normal distribution rather than a normal distribution[26]. Thus, formulating the random perturbation by multiplying the log-normal distribution random number is appropriate. Here, the random perturbation is defined by an Ornstein-Uhlenbeck process. For a more detailed discussion, please refer to Lei (2011)[27] and Huang et al. (2024)[14].

It is important to note that (2.5)-(2.6) provide a general formulation for noise perturbation to a parameter. However, the perturbations η to different parameters g_i , h_{ij} and k_i in equation (2.4) are independent of each other. Additionally, parameters σ and ϑ defining the noise perturbations differ for perturbations to different parameters. Thus, the ODE model (2.3) can be extended as the following stochastic differential equation (SDE) model:

$$\left\{ \begin{array}{l} \frac{dX_i}{dt} = g_i \prod_{j=1}^2 H(X_j; h_{ji}, n_{ji}, \lambda_{ji}) - k_i X_i, \quad (i = 1, 2) \\ g_i = \bar{g}_i e^{\sigma_1 \eta_1 - \sigma_1^2/2} \\ h_{ij} = \bar{h}_{ij} e^{\sigma_2 \eta_2 - \sigma_2^2/2} \\ k_i = \bar{k}_i e^{\sigma_3 \eta_3 - \sigma_3^2/2} \\ d\eta_1 = -(\eta_1/\vartheta_1)dt + \sqrt{2/\vartheta_1}dW_1 \\ d\eta_2 = -(\eta_2/\vartheta_2)dt + \sqrt{2/\vartheta_2}dW_2 \\ d\eta_3 = -(\eta_3/\vartheta_3)dt + \sqrt{2/\vartheta_3}dW_3 \end{array} \right. \quad (2.7)$$

2.2.3. Cross-cell-cycling dynamics

The SDE model (2.7) describes the gene regulation dynamics within a single cell cycle. To extend these dynamics across multiple cycles and link them with the heterogeneous G0 cell cycle model (2.2), we must integrate the cell cycling process into the equation (2.7).

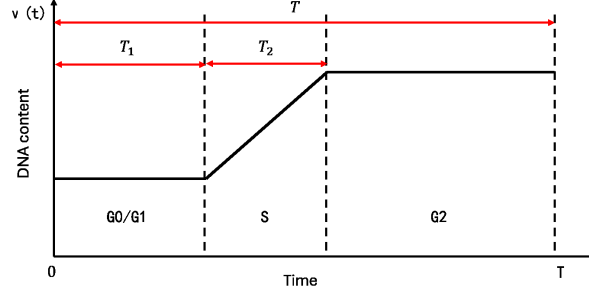


Figure 3: **Illustration of a cell cycle and the cycling age.** Here T denotes the duration of one cycle, T_1 means the duration of G0/G1 phase, and T_2 represents the duration of S phase.

During the proliferative phase, a cell undergoes DNA replication, resulting in a change in its genetic content. Consequently, the gene expression rates vary throughout this phase. For simplicity, let T represent the duration of one cycle, with the cycling age ℓ denoted as $0 < \ell < T_1$ for the G0/G1 phase, $T_1 < \ell < T_1 + T_2$ for the S phase (DNA replication), and $T_1 + T_2 < \ell < T$ for the G2 and M phase post-DNA replication (Figure 3). Let $\ell(t)$ represents the cycling age defining as

$$\ell(t) = t - kT, \quad kT \leq t < (k+1)T, \quad (k = 0, 1, 2, \dots). \quad (2.8)$$

We defined the changes in DNA content with the cell cycling age as

$$\nu(t) = \begin{cases} 1, & 0 \leq \ell(t) < T_1 \\ (1 + \frac{\ell - T_1}{T_2}), & T_1 \leq \ell(t) < (T_1 + T_2) \\ 2, & (T_1 + T_2) \leq \ell(t) < T \end{cases} \quad (2.9)$$

for simplicity. Through the DNA content $\nu(t)$, we replaced g_i with $g_i \nu(t)$ and rewrote (2.4) as:

$$\frac{dX_i}{dt} = g_i \nu(t) \prod_{j=1}^2 H(X_j; h_{ji}, n_{ji}, \lambda_{ji}) - k_i X_i, \quad (i = 1, 2) \quad (2.10)$$

when $kT \leq t < (k+1)T$.

The cell undergoes mitosis when $t = kT$. Molecules such as proteins and mRNAs redistribute stochastically to two daughter cells. Therefore, the protein concentrations X_1 and X_2 in the newborn cell should be reset post-mitosis while molecules are allocated to the two daughters. Thus, we reset the initial condition at $t = kT$ as

$$X_i(kT) = \chi_i \lim_{t \rightarrow kT^-} X_i(t) \quad (2.11)$$

where χ_i is a beta distribution random number.

We assumed a beta distribution random number with a density function exhibiting various shapes based on shape parameters a and b , including strictly decreasing ($a < 1, b \leq 1$), strictly increasing ($a > 1, b \leq 1$), U-shaped ($a < 1, b < 1$), or unimodal ($a > 1, b > 1$). We denote the beta distribution random number as $\chi_i \sim \text{Beta}(a, b)$, where shape parameters a, b satisfy

$$\mathbb{E}(\chi_i) = \frac{a}{a+b} = \frac{1}{2}.$$

We can select proper parameters to describe the situations of either symmetry or asymmetry division. In this study, we introduced two parameters $0 < \phi < 1$ and $\eta > 0$ so that

$$\mathbb{E}(\chi_i) = \phi, \text{var}(\chi_i) = \frac{1}{1+\eta}\phi(1-\phi), \quad (2.12)$$

then the shape parameters a and b are given by

$$a = \eta\phi, b = \eta(1-\phi).$$

In particular, we took $\phi = \frac{1}{2}$, then $\eta > 2$ yields a unimodal distribution, corresponding to symmetry division, while $\eta < 2$ gives a U-shape distribution, corresponding to asymmetry division (Figure 4).

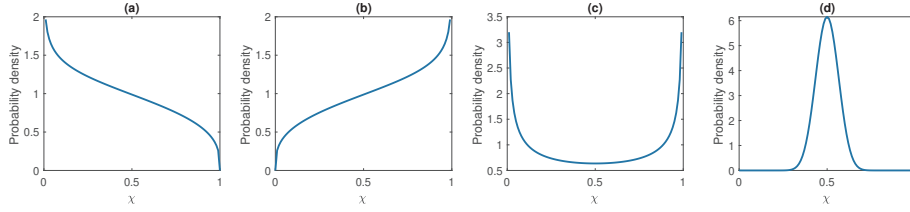


Figure 4: Probability density function of the beta distribution. **(a).** $\phi = 0.4, \eta = 2.2$. **(b).** $\phi = 0.6, \eta = 2.2$. **(c).** $\phi = 0.5, \eta = 1.0$. **(d).** $\phi = 0.5, \eta = 60$.

Therefore, considering cell cycling, we integrated equations (2.7)-(2.11) to formulate a hybrid model as

follows (where $m = 2$ represents the number of genes):

$$\left\{ \begin{array}{l} \frac{dX_i}{dt} = g_i \nu(t) \prod_{j=1}^m H(X_j; h_{ji}, n_{ji}, \lambda_{ji}) - k_i X_i, \quad (i = 1, \dots, m) \\ g_i = \bar{g}_i e^{\sigma_1 \eta_1 - \sigma_1^2/2} \\ h_{ij} = \bar{h}_{ij} e^{\sigma_2 \eta_2 - \sigma_2^2/2} \\ k_i = \bar{k}_i e^{\sigma_3 \eta_3 - \sigma_3^2/2} \\ d\eta_1 = -(\eta_1/\vartheta_1)dt + \sqrt{2/\vartheta_1}dW_1 \\ d\eta_2 = -(\eta_2/\vartheta_2)dt + \sqrt{2/\vartheta_2}dW_2 \\ d\eta_3 = -(\eta_3/\vartheta_3)dt + \sqrt{2/\vartheta_3}dW_3 \\ (kT \leq t < (k+1)T) \\ X_i(kT) = \chi_i \lim_{t \rightarrow kT^-} X_i(t) \\ \chi_i \sim \text{Beta}(a, b). \end{array} \right. \quad (2.13)$$

In the equation (2.13), the coefficient λ_{ji} in the function $H(x_j; h_{ji}, n_{ji}, \lambda_{ji})$ represents positive regulation ($\lambda_{ji} > 1$), negative regulation ($\lambda_{ji} < 1$), or no regulation ($\lambda_{ji} = 0$) between genes j and i . Thus, it is easy to extend the model equation to the dynamics of a general gene regulation network with m genes.

The equation (2.13) describes the dynamics of the gene regulation network over multiple cell cycles, with the protein levels reset after cell division.

This study does not consider the effect of cell volume change. Cell growth and volume changes are significant biological processes that play essential roles in cell-fate decisions[28, 29]. Moreover, biological mechanisms controlling cell growth and division are complex and not fully understood for mammalian cells[30, 31, 32]. In this study, we focused on the computational scheme connecting gene regulation network dynamics with stem cell regeneration to avoid complexity. We assumed that the volume remains unchanged during cell cycling, so cell growth does not affect the protein concentration.

2.2.4. Numerical implementation

Implementing the model equation (2.13) numerically includes solving the SDE model within a cell cycle and managing the process of cell division. Cell division may lead to cell splitting and the resetting of initial conditions for newborn cells. This numerical scheme can be realized using the object-oriented programming language C++. The dynamics of gene regulatory networks within a cell cycle are modeled with the SDE model, which is numerically solved through the Runge-Kutta method. Additionally, we designed an individual-cell-based numerical scheme to monitor the dynamics of each cell within the system. Specific algorithm details are provided in the Appendix.

In our numerical implementations, we focused on symmetric gene regulations, assuming $k_1 = k_2$, $\lambda_{12} = \lambda_{21}$, $\lambda_{11} = \lambda_{22}$, $n_{11} = n_{12} = n_{21} = n_{22}$, and normalized the protein concentrations so that $h_{11} = h_{12} = h_{21} = h_{22}$. The parameter values were estimated based on previous studies for the similar gene network [14, 23, 33], and further adjustments were made to ensure that cells could exhibit various phenotypes with different gene expression levels.

2.3. Inheritance function

Connecting the dynamics of the gene regulation network with the heterogeneous cell cycle model (2.2) requires identifying the epigenetic state $\mathbf{x}$ and deriving the inheritance function $p(\mathbf{x}, \mathbf{y})$ based on the gene regulation dynamics (2.13) across a cell cycle.

It is important to note that while the epigenetic state $\mathbf{x}$ in (2.2) represents a constant reflecting the state of a cell during the resting phase, it dynamically changes over time in (2.13), representing gene expression dynamics during a cell cycle. We need to identify the epigenetic state $\mathbf{x}$ in accordance with the gene expression dynamics $\mathbf{X}(t)$ in one cell cycle. For consistency, we define the epigenetic state $\mathbf{x}$ of a cell through log normalization of gene expression $\mathbf{X}(t)$ as

$$\mathbf{x} = \log(\mathbf{X}(t) + 1)|_{\ell(t)=\ell^*} \quad (2.14)$$

with the time t corresponding to a specific cell cycling age ℓ^* . In our simulations, we took $T = 50$, and $T_1 = 25, T_2 = 8$. We considered the state at $\ell^* = T_1$, which is the time point before DNA replication, as the cycling age to define the epigenetic state of a cell during a cell cycle.

To obtain the inheritance function from simulation results, we numerically solved the cross-cell-cycle stochastic differential equation (2.13) and tracked the gene expression $\mathbf{X}(t)$ of individual cells over cell divisions. Since (2.13) describes the gene expression dynamics of a single cell over multiple cycles, two daughter cells were recorded when a cell divided. Thus, we repeated the numerical scheme to obtain an ensemble of multiple cells, with each simulation starting with a randomly selected initial condition. For each cell under simulation, record the state $\mathbf{X} = \mathbf{X}(t)|_{\ell=T_1}$ so that $\mathbf{x} = \log(\mathbf{X} + 1)$ represents the epigenetic state of the cell. This process yields a dataset $D = \{\mathbf{x}_k, \mathbf{y}_k\}$, where $\mathbf{y}_k$ represents the state of the k 'th mother cell and $\mathbf{x}_k$ the state of one daughter cell of the k 'th mother cell. Through dataset D , we obtain the inheritance function $p(\mathbf{x}, \mathbf{y})$ via the conditional probability density defined by (3.9).

For the probability density function $p(\mathbf{x}, \mathbf{y})$, we assumed that the inheritances of gene A and gene B were independent of each other, yielding

$$p(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^2 p_i(x_i, \mathbf{y}),$$

where $p_i(x_i, \mathbf{y})$ denotes the conditional density function of x_i given the state of the mother cell $\mathbf{y} = (y_1, y_2)$.

In our study, we presumed the mixed conditional gamma distribution for epigenetic states so that

$$p_i(x_i, \mathbf{y}) = \sum_{j=1}^k \alpha_{i,j} \text{Gamma}(x_i; a_{i,j}, b_{i,j}), \quad (i = 1, 2), \quad (2.15)$$

where k is the number of independent distributions, $\alpha_{i,j}$ are combination coefficients, and $\text{Gamma}(x; a, b)$ represents the probability function of gamma distribution with shape parameters a and b , expressed as

$$\text{Gamma}(x; a, b) = \frac{x^{a-1} e^{-(x/b)}}{b^a \Gamma(a)}, \quad \Gamma(z) = \int_0^\infty e^{-x} x^{z-1} dx, \quad z > 0. \quad (2.16)$$

The combination coefficients $\alpha_{i,j}$ satisfy

$$\sum_{j=1}^k \alpha_{i,j} = 1, \quad \alpha_{i,j} \geq 0.$$

In (2.15), the shape parameters $a_{i,j}$ and $b_{i,j}$ depend on the state $\mathbf{y}$ of the mother cell.

To determine the shape parameters, we consider that, given the probability density function (2.16), the conditional mean and variance are

$$E(x|\mathbf{y}) = a(\mathbf{y})b(\mathbf{y}), \quad \text{var}(x|\mathbf{y}) = a(\mathbf{y})b(\mathbf{y})^2.$$

Thus, while we assume the conditional mean and variance as

$$E(x|\mathbf{y}) = \psi(\mathbf{y}), \quad \text{var}(x|\mathbf{y}) = \frac{\psi(\mathbf{y})^2}{\zeta(\mathbf{y})}, \quad (2.17)$$

we have

$$a(\mathbf{y}) = \zeta(\mathbf{y}), \quad b(\mathbf{y}) = \psi(\mathbf{y})/\zeta(\mathbf{y}). \quad (2.18)$$

Thus, to obtain the inheritance function $p(\mathbf{x}, \mathbf{y})$, we only need to derive, for each component x_i , the functions $\psi_{i,j}(\mathbf{y})$ and $\zeta_{i,j}(\mathbf{y})$ through the conditional mean and variance from the simulation data. The detailed procedure for obtaining the inheritance function is provided in Section 3.2.

3. Results

3.1. Identification of cell types through epigenetic states

To quantitatively define cell phenotypes, we conducted a bifurcation analysis based on the ODE model (2.3). We varied the production rates g_1 and g_2 while keeping other parameters fixed and examined the dependence of steady states on these parameter values. Figure 5a illustrates the number of steady states with randomly varying parameters (g_1, g_2) . When g_1 and g_2 are large, three steady states emerge (depicted

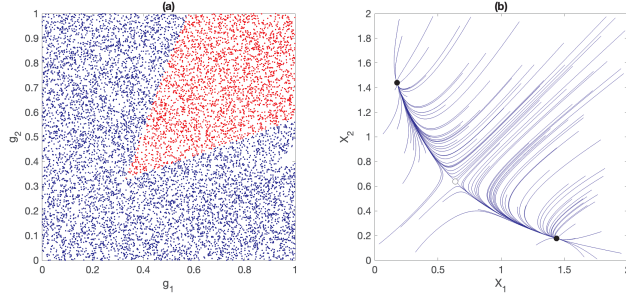


Figure 5: **Bifurcation analysis of gene expression kinetics.** (a). Dependence of the number of steady states of the ODE model (2.3) with randomly selected parameters (g_1, g_2) . Blue dots represent 1 steady state, while red dots represent 3 steady states. (b). Trajectories for the ODE model (2.3) with randomly selected initial conditions. Solid dots marked the two stable steady states, and the hollow dot represents the unstable steady state. Parameters are $\lambda_{21} = \lambda_{12} = 0.1$, $\lambda_{11} = \lambda_{22} = 5.0$, $k_1 = k_2 = 1.0$, $h_{ij} = 1$, $n_{ij} = 2$, and $(g_1, g_2) = (0.4, 0.4)$ in (b). The parameters for time units have been normalized in our study.

by red dots in Figure 5a). Additionally, at $(g_1, g_2) = (0.4, 0.4)$, solving the ODE (2.3) with varying initial conditions results in the convergence of solutions $(x_1(t), x_2(t))$ to one of the two stable steady states, as indicated by the black dots in Figure 5b. These stable steady states exhibit expression patterns of either $(+, -)$ or $(-, +)$ for the marker genes A and B, defining two distinct cell phenotypes.

The bifurcation analysis reveals bistable steady states in gene network dynamics, corresponding to different cell types defined by marker gene expression patterns. To further explore how external noise perturbation and cell divisions may shape the heterogeneous within a cell clone, we examined the dynamics of the hybrid model (2.13) under various noise strength σ (assuming $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$) and cell division scenarios.

First, we excluded cell cycling and focused solely on gene expression dynamics with external noise perturbation. To achieve this, we solved the SDE model (2.7) with randomly selected initial conditions $0 < X_i(0) < 2$. Figure 6a illustrates the distributions of the epigenetic state of cells at one cycle ($T = 50$) with varying noise strength σ . Results indicate distinct cell phenotypes when the external noise is small ($\sigma = 0.2$ or 0.4); however, the two cell types merge as the noise strength σ increases.

Next, we explored the influence of cell division using the hybrid model (2.13). The effect of cell division is depicted by the redistribution of molecules through (2.11), with parameters ϕ and η defining the distribution of the coefficient χ_i . We fixed $(\phi, \eta) = (0.5, 60)$ (Figure 4d) and varied the noise strength σ to solve (2.13). Here, we initiated the system with one cell and simulated model dynamics for 15 cycles, resulting in 2^{15} cells at the 15th cycle, all originating from a single cell. Figure 6b displays the epigenetic states of all cells, showing a similar pattern to Figure 6a, generated from multiple cells with different initial conditions. This suggests that a single cell can give rise to a heterogeneous cell clone through cell division.

Furthermore, to investigate the effects of different assumptions in cell division, we varied the parameter η

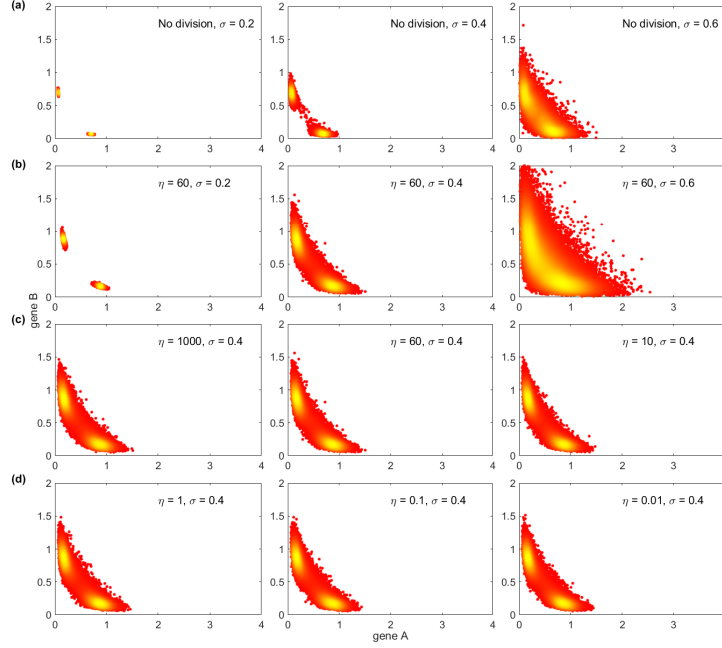


Figure 6: **Distribution of the epigenetic state of cells under different model assumptions.** (a). Epigenetic states of cells at $t = 50$ (one cell cycle) without cell division. The cell states were obtained by solving the SDE model (2.7) with initial conditions randomly distributed over $0 < X_i(0) < 2$. (b). Epigenetic state of cells derived from a single cell after 15 divisions. The cell states were obtained by solving the hybrid model (2.13), with initially a single cell having random gene expression states $0 < X_i(0) < 2$. (c). Same as (b) with symmetric division ($\eta > 2$) and different distributions for the parameter χ_i . (d). Same as (b) with asymmetric division ($\eta < 2$) and different distributions for the parameter χ_i . Here, the epigenetic states are represented by $\mathbf{x} = \log(\mathbf{X}(t) + 1)|_{\ell(t)=T_1}$. In (a), no cell division was considered, and different external perturbation strengths σ were applied. In (b)-(d), cell division was considered, with $\phi = 0.5$ and different parameters η for the distribution of χ_i (refer to (2.12)). The values of σ and η are shown in the figure; the parameters $\vartheta_1 = \vartheta_2 = \vartheta_3 = 0.3$; the conditional perspective ϕ in (2.12) was set as $\phi = 0.5$; the cell cycling parameters were $T = 50, T_1 = 25, T_2 = 8$; other parameters were the same as in Figure 5.

in defining the distribution of the coefficient χ_i . Fixing $\sigma = 0.4$, we adjusted η to reflect symmetry division (10 to 1000), resulting in a decreased variance of χ_i . The distribution of epigenetic states in Figure 6 appears independent of the parameter η . Similarly, varying η to reflect asymmetric division (ranging from 1 to 0.01) yields consistent distributions of epigenetic states at the 15th cycle, as shown in Figure 6d. These results suggest that the distribution of the epigenetic states of a cell clone remains insensitive to the variance of χ_i in cell division.

For subsequent discussions below, we focused on the method of obtaining the inheritance function through simulation data, with fixed values of $\phi = 0.5$, $\eta = 60$, and $\sigma = 0.4$.

3.2. Data-driven inheritance function

In Figure 6, we have demonstrated the variability of epigenetic states in individual cells originating from a single cell. The variability reflects the stochastic changes in cell states during cell division. To derive the inheritance function $p(\mathbf{x}, \mathbf{y})$ based on the gene regulation dynamics (2.13), we tracked the process of cell division through numerical simulations. We recorded the state of the mother cells ($\mathbf{y}$) and the daughter cells ($\mathbf{x}$) following each cell division event. In our simulations, we initiated 10^6 cells. For each cell, we simulated the model for 3 cell cycles, considering all cells at the second cycle as mother cells and denoting their epigenetic states as $\mathbf{y}$. For each mother cell with state $\mathbf{y}$, we paired it with the epigenetic state $\mathbf{x}$ of the corresponding daughter cell at the third cycle, forming a pair $[\mathbf{x}, \mathbf{y}] = [(x_1, x_2), (y_1, y_2)]$. Here, x_1 and y_1 represent the state associated with gene A, while x_2 and y_2 represent the state associated with gene B.

Figure 7a and 7b present scatter plots illustrating the epigenetic states of genes A and B in mother cells and their daughter cells. These plots reveal distinct cell types: mother cells with lower expression in gene A (or gene B) typically produce daughter cells with similarly lower expression levels in the same gene, while those with higher expression tend to generate daughter cells with elevated expression levels in the same gene. However, we also observe instances of cell type switches during cell divisions, where some daughter cells exhibit gene expression patterns differing from those of their mother cells.

To establish the mathematical formulation of the inheritance function, we utilized gene A as an example to illustrate the computational method. We identified all mother cells with a specific expression level (as delineated by the two strips in Figure 7a) and examined the probability density of the expression levels of all daughter cells. Figure 7c illustrates the resulting conditional probability densities with the epigenetic state of mother cells $y_1 = 0.3$ and $y_1 = 0.8$, respectively. These density functions exhibit characteristics akin to a combination of two unimodal distributions (solid lines in Figure 7c).

Drawing insights from Figure 7c, we modeled the probability density $p_1(x_1, \mathbf{y})$ with a combination of two gamma distributions:

$$p_1(x_1, \mathbf{y}) = \alpha_1 \times \text{Gamma}(x_1; a_{1,1}, b_{1,1}) + (1 - \alpha_1) \times \text{Gamma}(x_1; a_{1,2}, b_{1,2}). \quad (3.1)$$

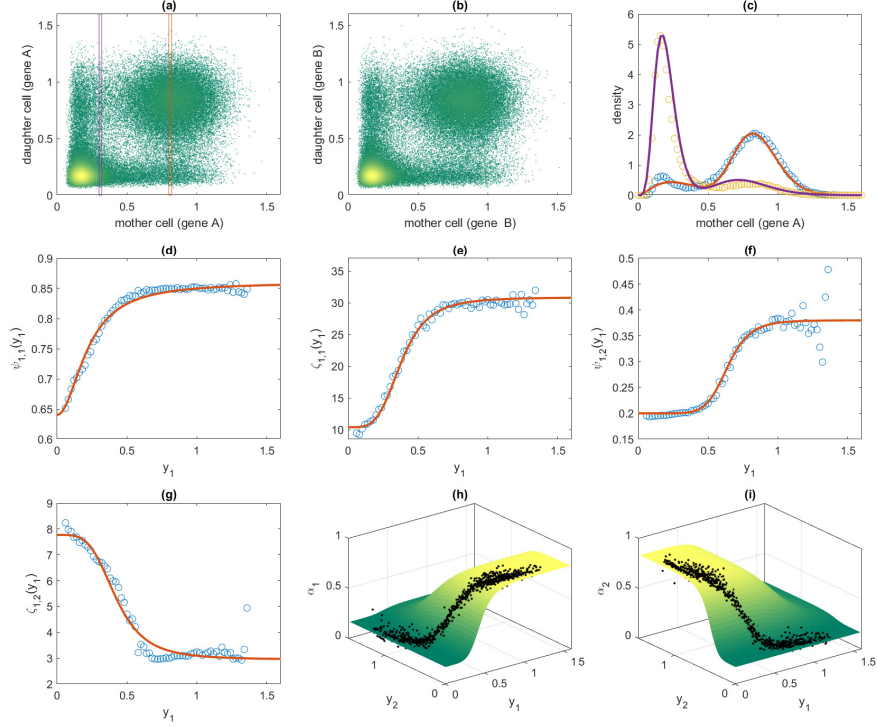


Figure 7: **Data-driven inheritance function.** (a). Scatter plot of daughter cell (gene A) versus mother cell (gene B) (x_1 versus y_1). Each point represents a cell, with lighter colors indicating higher density. Two vertical strips represent the sampling of data points with $0.3 \leq y_1 < 0.31$ and $0.8 \leq y_1 < 0.81$, respectively. (b). Scatter plot of daughter cell (gene B) versus mother cell (gene B) (x_2 versus y_2). (c). Probability density of daughter cells (gene A) given the state of the mother cell (gene A). Dots were obtained from data analysis, and solid lines represent the functions of a combined gamma distribution. The densities were obtained from data in the two strips in (a). (d). The function $\psi_{1,1}(y_1)$. (e). The function $\zeta_{1,1}(y_1)$. (f). The function $\psi_{1,2}(y_1)$. (g). The function $\zeta_{1,2}(y_1)$. (h). The function $\alpha_1(y_1, y_2)$. (i). The function $\alpha_2(y_1, y_2)$. In (d)-(i), dots were obtained from data analysis, while curves or surfaces were derived from mathematical formulations.

Here, the shape coefficients $a_{1,j}$ and $b_{1,j}$ depend on the state of the mother cell $\mathbf{y}$ through the functions $\psi_{1,j}(\mathbf{y})$ and $\zeta_{1,j}(\mathbf{y})$, which are associated with the conditional mean and variance $E(x_1|\mathbf{y})$ and $\text{var}(x_1|\mathbf{y})$ in accordance with (2.17). Additionally, the combination coefficients α_1 may vary based on the state of the mother cell $\mathbf{y}$.

To determine the coefficients α_1 , $a_{1,j}$ and $b_{1,j}$, we initially presumed their dependence solely on the component y_1 . From Figure 7a, we partitioned the data into multiple bins based on the y_1 value. Subsequently, we segregated the x_1 values in each bin into two subgroups and calculated the mean and variance in each subgroup, thus deriving the conditional means $E(x_1|y_1)$ and variances $\text{var}(x_1|y_1)$ for the two gamma distributions $\text{Gamma}(x_1; a_{1,j}, b_{1,j})$, $j = 1, 2$. This process yielded the functions $\psi_{1,j}(y_1)$ and $\zeta_{1,j}(y_1)$ as Hill-type functions following (2.17) (Figure 7d-g):

$$\begin{cases} \psi_{1,1}(y_1) = 0.64 + 0.22 \times \frac{y_1^2}{0.22^2 + y_1^2}, \\ \zeta_{1,1}(y_1) = 10.38 + 20.46 \times \frac{y_1^4}{0.38^4 + y_1^4}, \\ \psi_{1,2}(y_1) = 0.20 + 0.18 \times \frac{y_1^8}{0.64^8 + y_1^8}, \\ \zeta_{1,2}(y_1) = 2.95 + 0.15 \times \frac{1}{0.42^4 + y_1^4}. \end{cases} \quad (3.2)$$

The shape parameters were formulated as

$$a_{1,j}(y_1) = \psi_{1,j}(y_1), b_{1,j}(y_1) = \psi_{1,j}(y_1)/\zeta_{1,j}(y_1). \quad (3.3)$$

Subsequently, we assumed that the combination coefficient α_1 depends on both y_1 and y_2 . Consequently, we partitioned the (y_1, y_2) phase plane into multiple bins and gathered the daughter cell epigenetic states x_1 in each bin. We then employed the expectation-maximization (EM) algorithm, as per (3.1), to determine the coefficients $\alpha_{1,j}$ for each bin. The results, illustrated in Figure 7h, closely align with a Hill-type function:

$$\alpha_1(\mathbf{y}) = 0.16 + \frac{y_1^6}{0.46^6 + y_1^6} \times \frac{1}{1.07^6 + y_2^6}. \quad (3.4)$$

Thus, the function $\alpha_1(\mathbf{y})$ and the shape parameters $a_{1,j}(y_1), b_{1,j}(y_1)$ collectively define the inheritance function $p_1(x_1; \mathbf{y})$ per (3.1).

Similarly, we derive the inheritance function $p_2(x_2; \mathbf{y})$ for gene B:

$$p_2(x_2; \mathbf{y}) = \alpha_2(\mathbf{y}) \times \text{Gamma}(x_2; a_{2,1}, b_{2,1}) + (1 - \alpha_2(\mathbf{y})) \times \text{Gamma}(x_2; a_{2,2}, b_{2,2}). \quad (3.5)$$

Remarkably, owing to the symmetry between gene A and gene B, simulation data exhibit the same forms of

$\psi_{2,j}(y_2)$ and $\zeta_{2,j}(y_2)$ as the functions $\psi_{1,j}(y_1)$ and $\zeta_{1,j}(y_1)$, respectively. Consequently:

$$\begin{cases} \psi_{1,1}(y_2) = 0.64 + 0.22 \times \frac{y_2^2}{0.22^2 + y_2^2}, \\ \zeta_{1,1}(y_2) = 10.38 + 20.46 \times \frac{y_2^4}{0.38^4 + y_2^4}, \\ \psi_{1,2}(y_2) = 0.20 + 0.18 \times \frac{y_2^8}{0.64^8 + y_2^8}, \\ \zeta_{1,2}(y_2) = 2.95 + 0.15 \times \frac{1}{0.42^4 + y_2^4}, \end{cases} \quad (3.6)$$

and

$$a_{2,j}(y_2) = \psi_{2,j}(y_2), b_{2,j}(y_2) = \psi_{2,j}(y_2)/\zeta_{2,j}(y_2). \quad (3.7)$$

Likewise, $\alpha_2(\mathbf{y})$ assumes the same form as $\alpha_1(\mathbf{y})$, thus:

$$\alpha_2(\mathbf{y}) = 0.16 + \frac{1}{1.07^6 + y_1^6} \times \frac{y_2^6}{0.46^6 + y_2^6}. \quad (3.8)$$

The function $\alpha_2(\mathbf{y})$ is shown in Figure 7i.

Finally, the inheritance function is expressed as:

$$p(\mathbf{x}; \mathbf{y}) = p_1(x_1; \mathbf{y}) \times p_2(x_2; \mathbf{y}). \quad (3.9)$$

The above procedure delineates the processes of acquiring the inheritance function in (2.2) from simulation data based on the gene regulation network model in (2.13). Validation of the inheritance function (3.9) is presented in section 3.3.

The gamma distribution density functions $\text{Gamma}(x_i; a_{i,j}(y_i), b_{i,j}(y_i))$ defined by (3.1), (3.5), and (3.9) describe the inheritance of daughter cells from the mother cells for the same gene, while the combination coefficients $\alpha_1(\mathbf{y})$ and $\alpha_2(\mathbf{y})$ elucidate the property of cell type transition regulated by the gene regulation network. Notably, $\alpha_1(\mathbf{y})$ increases with y_1 and decreases with y_2 , signifying the activation of gene A by itself and the inhibition of gene B to gene A. A similar pattern is observed in $\alpha_2(\mathbf{y})$. Thus, the gene regulation network structure influences the inheritance function through its qualitative dependence on the state of mother cells.

3.3. Validation of the inheritance function

To verify the data-driven inheritance function, we compared simulation results obtained from the heterogeneous G0 cell cycle model (2.2) and the gene regulation network model (2.13).

Initially, we utilized the cell cycle model (2.2) to reproduce the process of cell clone generation from a single cell and compared our findings with those in Figure 6. To achieve this, we defined the kinetotype in

(2.2) as follows:

$$\beta = 1/T_1, \kappa = \mu = 0, \tau = T - T_1. \quad (3.10)$$

Given that solving the high-dimensional integral in (2.2) directly is computationally expensive, we applied an individual-cell-based simulation approach, adhering to the scheme outlined in equation (2.2).

To replicate the process illustrated in Figure 7, we initialized the system with 10^6 cells, emulating the setup in Figure 7, and performed individual-cell-based simulations to model the cell division process. During cell division, each mother cell generated two daughter cells, with the epigenetic state of each daughter cell determined by the inheritance function $p(\mathbf{x}, \mathbf{y})$ defined in (3.9) (see the Appendix for the detailed numerical scheme). Similar to the methodology in Figure 7, we conducted simulations over 3 cycles and collected the cells from the 2nd and the 3rd cycles for data analysis.

Figure 8 presents the data points of daughter cell epigenetic states against mother cell epigenetic states, obtained from both the gene regulation network model (2.13) and the G0 cell cycle model (2.2) with the inheritance function (3.9). A comparison of Figure 8a with Figure 8b demonstrates that the data-driven inheritance function (3.9) effectively reproduces the distribution of the cross-cell cycling epigenetic states.

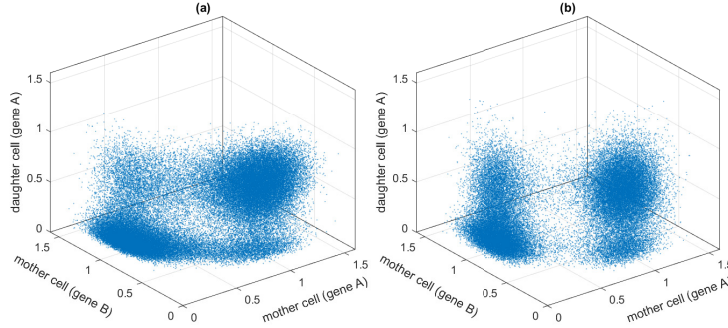


Figure 8: **Comparison between the gene regulation network model and the G0 cell cycle model.** (a). Scatter plot of daughter cell epigenetic state (gene A) versus mother cell epigenetic state (gene A and B) following the gene regulation network model (2.13). (b). Scatter plot of daughter cell epigenetic state (gene A) versus mother cell epigenetic state (gene A and B) following the G0 cell cycle model (2.2) with inheritance function given by (3.9). Parameters were identical to those in Figure 7, with cell cycling parameters given by (3.10) in G0 cell cycle model.

Next, we applied the G0 cell cycle model to investigate the biological process of heterogeneous cell growth with additional regulations in proliferation and differentiation. For this purpose, we defined the kinetotype as:

$$\kappa = 0.009, \mu = 0.0007, \tau = T - T_1,$$

and

$$\beta(c, \mathbf{x}) = \beta_0 \frac{\theta}{\theta + c}, c = \int_{\Omega} Q(t, \mathbf{x}) d\mathbf{x},$$

with $\beta_0 = 0.2$ and $\theta = 10^6$. Employing identical assumptions, we simulated the cell growth process following the G0 cell cycle model (2.2) with the inheritance function (3.9). In our simulations, we initialized 10^4 cells, randomly assigning epigenetic states to each cell. We employed individual-cell-based simulation to model cell growth up to $t = 2500$ (about 50 cycles).

In this scenario, we introduced nonzero differentiation and apoptosis rates, assuming that the proliferation rate decreases with the total cell number. Consequently, the cell number would eventually reach an equilibrium state following an extended period of cell growth (Figure 9a). Figure 9b displays the distribution of the epigenetic state of cells at $t = 2000$ (marked by the dashed line in Figure 9a). Notably, two subtypes of cells are evident, with their epigenetic state distributions closely resembling those obtained from the gene regulation network model in Figure 6. We further examined the dynamics of individual cell-type transitions. Figure 9c-d illustrates the evolution dynamics of gene A and gene B while tracking a single cell during the cell growth process. Our simulations demonstrate that cells can transition between the two subtypes as the cell growth process unfolds. These results underscore the capability of the G0 cell cycle model (2.2) to simulate the long-term dynamics of cell growth with heterogeneity.

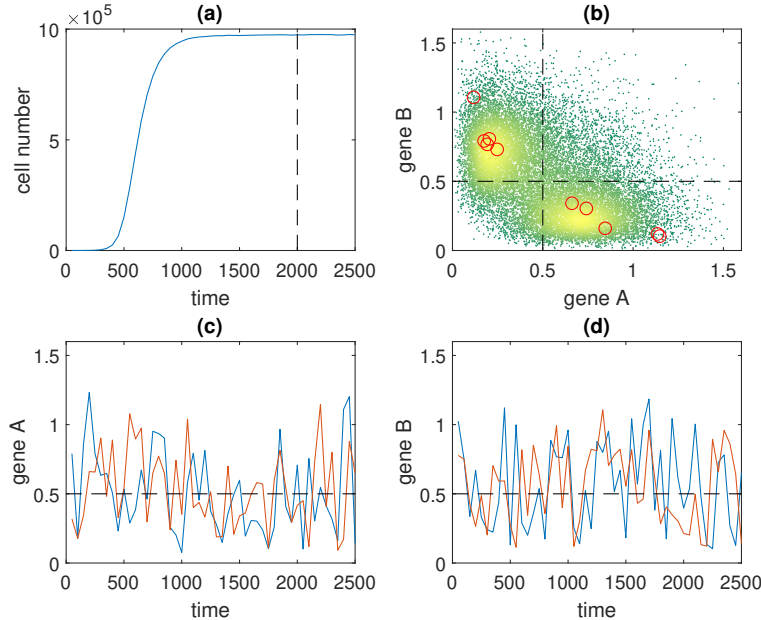


Figure 9: **Heterogeneous cell growth.** (a). Dynamics of cell number evolution. (b). Scatter plot of the epigenetic state of cells at $t = 2000$ (dashed line in (a)). (c). Evolution of gene A epigenetic state of two cells during the cell growth process. (d). Evolution of gene B epigenetic state of two cells during the cell growth process.

4. Discussions

Our study investigated the inheritance of epigenetic states in cells using a combination of computational modeling and data analysis. We developed a data-driven inheritance function based on the gene regulation dynamics model and validated it using simulations from both a gene regulation network model and a heterogeneous G0 cell cycle model. Our results demonstrate that the data-driven inheritance function effectively captures the distribution of epigenetic states across cell generations. Furthermore, we applied the G0 cell cycle model to simulate the long-term dynamics of cell growth with heterogeneity, showcasing its capability to replicate the behavior observed in the gene regulation network model.

The findings of our study shed light on the mechanisms underlying the inheritance of epigenetic states in cells. By combining computational modeling with data analysis, we were able to derive a robust inheritance function that accurately predicts the distribution of epigenetic states in daughter cells based on their mother cell's states. This suggests that epigenetic inheritance is governed by complex regulatory processes embedded within the gene regulation network.

Moreover, our simulations using the G0 cell cycle model revealed intriguing insights into the dynamics of cell growth with heterogeneity. The model successfully reproduced the emergence of distinct cell subtypes and their transitions over time, indicating that the heterogeneity observed in experimental data can be attributed to underlying regulatory mechanisms within the cell cycle.

In conclusion, our study provides valuable insights into the inheritance of epigenetic states and the dynamics of cell growth with heterogeneity. By developing and validating a data-driven inheritance function, we have contributed to a deeper understanding of the regulatory processes governing cell behavior. Moving forward, our findings could inform further research into the molecular mechanisms underlying epigenetic inheritance and their implications for cellular function and disease.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (NSFC 12331018).

Appendix

A. Numerical scheme for the hybrid kinetic modeling

We applied the numerical scheme to solve the hybrid model (2.13) by combining the numerical method of stochastic differential equation and the random process of cell division. During numerical simulation, the cell number is changed over time due to cell division. Each cell corresponds to a stochastic differential equation defined by the gene regulatory network. Therefore, we need to track the gene expression dynamics of each individual cell. The number scheme is given below:

- **input** The topology file of the gene regulatory network, containing gene nodes, regulatory relationships, and related parameters.
- **initialization** Read the topology file and model parameters. Initialize the number of simulation cycles `cycle` = 0, set the cycle length $T = 50$, and set the initial time $t = 0$. Set the initial cell number N . For each cell $\Sigma = \{C_j(\mathbf{X}_j) : j = 1, \dots, N\}$, set the initial condition for the gene expression level $\mathbf{X}_j(0)$.
- **simulation process**

for `cycle` = 0 to `max_cycle` **do**

Copy the system $\Sigma' = \Sigma$.

for all cells $C_j \in \Sigma$ **do**

- **Gene regulation dynamics** Solve the stochastic differential equation (2.13) with the stochastic Runge-Kutta method for $0 \leq t \leq T$.

- **Cell division**

Generate two new cells $C_{j,1}(\mathbf{Y}_1)$ and $C_{j,2}(\mathbf{Y}_2)$, $\mathbf{Y}_l = (Y_{l,1}, \dots, Y_{l,m})$ ($l = 1, 2$) store the initial gene expression state of the cell l .

for $l = 1$ to 2 and $i = 1$ to m **do**

Generate a parameter $\chi_i \sim \text{Beta}(a, b)$;

Reset the expression state $Y_{1,i}(0) = \chi_i X_{j,i}(T)$.

end for

Replace the cell C_j in Σ' with the new cells $C_{j,1}$ and $C_{j,2}$.

end for

`cycle` = `cycle` + 1.

Copy the system $\Sigma = \Sigma'$.

end for

B. Numerical scheme of the individual-cell-based simulation method

The differential-integral equation (2.2) can be solved numerically through the Euler method and numerical integration. However, it is expensive to calculate the numerical integration due to the high-dimensional epigenetic states. Hence, we usually do not solve the equation (2.2) directly. We apply an individual-cell-based simulation method to simulate the cell cycling process.

In the model simulation, a multiple-cell system is represented as a collection of epigenetic states, and the individual-cell-based simulation tracks the behaviors of each cell according to its own epigenetic states. The sketch of the numerical scheme is given below.

- **input** The parameters, kinetotype functions β , κ , μ , and τ , inheritance function $p(\mathbf{x}, \mathbf{y})$, and the time step Δt .
- **initialize** Set the time $t = 0$, the initial cell number N_0 , and all cells $\Sigma = \{[C_i(\mathbf{x})]_{i=1}^Q\}$. At the initial state, all cells are at the resting phase ($S = 0$), and the corresponding age at the proliferating phase is $a = 0$. Set the cell number at the resting phase as $N_q = N_0$ and the cell number at the proliferative phase as $N_p = 0$.
- **simulation process**

for t from 0 to T_{end} with step Δt **do**

for each cell in Σ **do**

Kinetotype Calculate the proliferation rate β , the apoptosis rate μ , and the terminate differentiation rate κ .

Cell fate decision Determine the cell fate during the time interval $(t, t + \Delta t)$:

- When the cell is at the resting phase, undergo terminal differentiation with a probability $\kappa\Delta t$, or enter the proliferation phase with a probability $\beta\Delta t$, or, if otherwise, remain unchanged.
- When a cell is at the proliferative phase, if the age $a < \tau$, the cell is either removed (through apoptosis) with a probability $\mu\Delta t$ or remains unchanged; if the age $a \geq \tau$, the cell undergoes mitosis and divides into two cells.

end for

system update: **for** each cell in Σ **do**

- If the cell fate is differentiation or apoptosis, remove the cell, and reduce the total cell number by 1, $N = N - 1$. Accordingly, reduce the number of resting or proliferative cells: $N_q = N_q - 1$ or $N_p = N_p - 1$.
- If the cell fate is entering the proliferative state, set the cell state at proliferative ($S = 1$) and the proliferating age $a = 0$, and let $N_q = N_q - 1$, $N_p = N_p + 1$, the total cell number N is unchanged.
- If the cell is at the proliferative phase and remains unchanged, update the proliferative age $a = a + \Delta t$, the cell numbers N_q , N_p , and N are unchanged.
- If the cell is under mitosis, it produces two daughter cells, and the epigenetic state of each daughter cell is determined according to the inheritance probability $p(\mathbf{x}, \mathbf{y})$. The proliferating age of each daughter cell is set as $a = 0$, and update the cell number: $N_q = N_q + 2$, $N_p = N_p - 1$, and $N = N + 1$.

end for
end for

To determine the epigenetic state of daughter cells according to the inheritance probability $p(\mathbf{x}, \mathbf{y})$ defined by the combination gamma distribution (3.9), we applied the following numerical scheme:

- calculates the $\alpha_i, \psi_{i,j}, \eta_{i,j}$ according to the state of the mother cell, and the coefficients $a_{i,j}$ and $b_{i,j}$;
- **for** $i = 1$ to 2 **do**
 - generates a random number $q = \mathbf{rand}()$;
 - if $q < \alpha_i$, produce the epigenetic state of gene i for daughter cell according to the first gamma distribution function $\text{Gamma}(x_i; a_{i,1}, b_{i,1})$, otherwise, produce the epigenetic state according to the second gamma distribution function $\text{Gamma}(x_i; a_{i,2}, b_{i,2})$.
- set the epigenetic state of the daughter cells.

In simulations, the cell number may increase to a very high number through cell division, which may cause a challenge issue in simulations. To overcome this issue, we applied a technique of *downsampling*. The process of downsampling is often used in signaling analysis to reduce the data rate or the size of data. Here, we applied a similar technique to reduce the number of cells under simulation.

We predefine a maximum number of cells to be simulated and stored ($N_{\max} = 10^6$ cells in the current simulation). Let $N_k = N_{k,q} + N_{k,p}$ ($N_k \leq N_{\max}$) be the number of cells under simulation at step k , where $N_{k,q}$ represents the number of cells at the resting phase, and $N_{k,p}$ represent the number of cells at the proliferative phase. We first opened a temporary storage space for $N_{k,q} + 2N_{k,p}$ cells (the maximum number of cells if all proliferative phase cells undergo mitosis). After performing cell fate decisions for each cell, we have potentially $N_{\text{temp}} = N'_{k,q} + N'_{k,p}$ ($N_{\text{temp}} \leq N_{k,q} + 2N_{k,p}$) cells. If $N_{\text{temp}} > N_{\max}$, we selected all cells ($N'_{k,p}$ cells) in the proliferative phase and at most $(N_{\max} - N'_{k,q})$ cells in the resting phase (select each cell with a probability $p = \max\{1, (N_{\max} - N'_{k,q})/N'_{k,q}\}$) to obtain N_{next} cells for the next step simulation; otherwise, we select all N_{temp} cells, and set $N_{\text{next}} = N_{\text{temp}}$. Finally, we stored the state of all selected N_{next} cells, let $f_{\text{pro},k} = N_{\text{temp}}/N_k$ for the proliferation rate, and $N_{k+1} = N_{\text{next}}$ for the number of cells simulated at step $k + 1$.

According to the approach of downsampling simulation, at step k , there are $N_k \leq N_{\max}$ cells under simulation, and the states of these cells are stored; the real cell number is given by

$$N_{\text{real number at step } k} = N_0 \prod_{i=0}^{k-1} f_{\text{pro},i}.$$

This gives the cell number at step k .

References

- [1] H. Clevers, What is an adult stem cell?, *Science* 350 (6266) (2015) 1319–1320.
- [2] J. Lei, A general mathematical framework for understanding the behavior of heterogeneous stem cell regeneration, *J Theor Biol* 492 (2020) 110196.
- [3] J. Lei, Evolutionary dynamics of cancer: from epigenetic regulation to cell population dynamics—mathematical model framework, applications, and open problems, *Sci China Math* 63 (3) (2020) 411–424. doi:10.1007/s11425-019-1629-7.
- [4] A. V. Probst, E. Dunleavy, G. Almouzni, Epigenetic inheritance during the cell cycle, *Nat Rev Mol Cell Biol* 10 (3) (2009) 192–206. doi:10.1038/nrm2640.
- [5] T. Schepeler, M. E. Page, K. B. Jensen, Heterogeneity and plasticity of epidermal stem cells., *Development* 141 (13) (2014) 2559–2567.
- [6] A. Serra-Cardona, Z. Zhang, Replication-coupled nucleosome assembly in the passage of epigenetic information and cell identity., *Trends Biochem Sci* 43 (2) (2018) 136–148.
- [7] Z. S. Singer, J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, M. B. Elowitz, Dynamic heterogeneity and DNA methylation in embryonic stem cells., *Mol Cell* 55 (2) (2014) 319–331.
- [8] K. Takaoka, H. Hamada, Origin of cellular asymmetries in the pre-implantation mouse embryo: a hypothesis., *Philos Trans R Soc Lond B Biol Sci* 369 (2014) 20130536.
- [9] H. Wu, Y. Zhang, Reversing DNA methylation: mechanisms, genomics, and biological functions., *Cell* 156 (1-2) (2014) 45–68.
- [10] R. Huang, J. Lei, Dynamics of gene expression with positive feedback to histone modifications at bivalent domains, *Int J Mod Phys B* 32 (07) (2018) 1850075.
- [11] R. Huang, J. Lei, Cell-type switches induced by stochastic histone modification inheritance, *Discrete and Continuous Dynamical Systems-B* 24 (10) (2019) 5601–5619.
- [12] S. Sahoo, A. Mishra, H. Kaur, K. Hari, S. Muralidharan, S. Mandal, M. K. Jolly, A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells, *NAR Cancer* 3 (3) (2021) zcab027.
- [13] L. Cai, N. Friedman, X. S. Xie, Stochastic protein expression in individual cells at the single molecule level, *Nature* 440 (7082) (2006) 358–362.

- [14] R. Huang, Q. Situ, J. Lei, Dynamics of cell-type transition mediated by epigenetic modifications, *J Theor Biol* 577 (2024) 111664.
- [15] F. J. Burns, I. F. Tannock, On the existence of a G0-phase in the cell cycle, *Cell Proliferation* 3 (4) (1970) 321–334.
- [16] M. C. Mackey, Unified hypothesis for the origin of aplastic anemia and periodic hematopoiesis, *Blood* 51 (5) (1978) 941–956.
- [17] A. Moustakas, K. Pardali, A. Gaal, C. H. Heldin, Mechanisms of TGF- β signaling in regulation of cell growth and differentiation, *Immunol Lett* 82 (1-2) (2002) 85–91.
- [18] L. Yang, Y. L. Pang, H. L. Moses, TGF- β and immune cells: an important regulatory axis in the tumor microenvironment and progression, *Trends Immunol* 31 (6) (2010) 220–227.
- [19] D. M. Ornitz, N. Itoh, Fibroblast growth factors, *Genome Biol.* 2 (3) (2001) 3005.1–3005.12.
- [20] S. Bernard, J. Bélair, M. C. Mackey, Oscillations in cyclical neutropenia: new evidence based on mathematical modeling, *J Theor Biol* 223 (3) (2003) 283–298.
- [21] J. Lei, S. A. Levin, Q. Nie, Mathematical model of adult stem cell regeneration with cross-talk between genetic and epigenetic regulation., *Proc Natl Acad Sci USA* 111 (10) (2014) E880–E887.
- [22] B. Huang, D. Jia, J. Feng, H. Levine, J. N. Onuchic, M. Lu, RACIPE: a computational tool for modeling gene regulatory circuits using randomization, *BMC Syst Biol* 12 (2018) 74.
- [23] M. K. Jolly, D. Jia, M. Boareto, S. A. Mani, K. J. Pienta, E. Ben-Jacob, H. Levine, Coupling the modules of EMT and stemness: a tunable ‘stemness window’ model, *Oncotarget* 6 (28) (2015) 25161–25174. doi:10.18632/oncotarget.4629.
- [24] L. Xu, K. Zhang, J. Wang, Exploring the mechanisms of differentiation, dedifferentiation, reprogramming and transdifferentiation, *PLoS One* 9 (8) (2014) e105216. doi:10.1371/journal.pone.0105216.
- [25] J. Lei, G. He, H. Liu, Q. Nie, A delay model for noise-induced bi-directional switching, *Nonlinearity* 22 (12) (2009) 2845–2859. doi:10.1088/0951-7715/22/12/003.
- [26] D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, M. L. Simpson, Gene network shaping of inherent noise spectra, *Nature* 439 (7076) (2006) 608–611. doi:10.1038/nature04194.
- [27] J. Lei, Stochastic modeling in systems biology, arXiv (2011) arXiv:1104.4524v1.

- [28] A. Doncic, M. Falleur-Fettig, J. M. Skotheim, Distinct interactions select and maintain a specific cell fate, *Mol Cell* 43 (4) (2011) 528–539. doi:10.1016/j.molcel.2011.06.025.
- [29] M. B. Ginzberg, R. Kafri, M. Kirschner, On being the right (cell) size, *Science* 348 (6236) (2015) 1245075. doi:10.1126/science.1245075.
- [30] R. Kafri, J. Levy, M. B. Ginzberg, S. Oh, G. Lahav, M. W. Kirschner, Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle, *Nature* 494 (7438) (2013) 480–483. doi:10.1038/nature11897.
- [31] C. Cadart, S. Monnier, J. Grilli, P. J. Sáez, N. Srivastava, R. Attia, E. Terriac, B. Baum, M. Cosentino-Lagomarsino, M. Piel, Size control in mammalian cells involves modulation of both growth rate and cell cycle duration, *Nat Commun* 9 (1) (2018) 3275. doi:10.1038/s41467-018-05393-0.
- [32] E. Zatulovskiy, J. M. Skotheim, On the molecular mechanisms regulating animal cell size homeostasis, *Trends Genet* 36 (5) (2020) 360–372. doi:10.1016/j.tig.2020.01.011.
- [33] B. Huang, M. Lu, D. Jia, E. Ben-Jacob, H. Levine, J. N. Onuchic, Interrogating the topological robustness of gene regulatory circuits by randomization, *PLoS Comput Biol* 13 (3) (2017) e1005456.