CSIAM Trans. Life Sci.
doi: 10.4208/csiam-ls.SO-2024-0009

Vol. **1**, No. 2, pp. 320-353
June 2025

Integrating Gene Regulatory Network Dynamics with Heterogeneous Stem Cell Regeneration

Yakun Li[1], Xiyin Liang[1,2] and Jinzhi Lei[1,2,*]

[1] *School of Mathematical Sciences, Tiangong University, Tianjin 300387, China.*
[2] *Center for Applied Mathematics, Tiangong University, Tianjin 300387, China.*

Received 1 December 2024; Accepted 24 February 2025

Abstract. Stem cell regeneration is crucial for development and maintaining tissue homeostasis in self-renewing tissues. The dynamics of gene regulatory networks (GRNs) play a vital role in regulating stem cell renewal and differentiation. However, integrating the quantitative dynamics of GRNs at the single-cell level with population-level stem cell regeneration poses significant challenges. This study presents a computational framework that links GRN dynamics to stem cell regeneration through an inheritance function. This function captures epigenetic state transitions during cell division in heterogeneous stem cell populations. Our model derives this function using a hybrid approach that integrates cross-cell-cycle GRN dynamics, effectively connecting cellular-level GRN structures with population-level regeneration processes. By incorporating GRN structure directly into stem cell regeneration dynamics, this framework simulates cross-cell-cycle gene regulation using individual-cell-based models. The scheme is adaptable to various GRNs, providing insights into the relationship between gene regulatory dynamics and stem cell regeneration. Additionally, we propose a future perspective that integrates single-cell ribonucleic acid sequencing data, GRN analysis, and cell regeneration dynamics using AI-driven tools to enhance the precision of regenerative studies.

AMS subject classifications: 92-10, 92C42

Key words: Cell plasticity, inheritance function, gene regulatory network, cell division, stem cell regeneration.

1 Introduction

Biological processes such as immune responses and cancer evolution are inherently multiscale dynamics that occur at molecular, cellular, and tissue levels [10, 12, 38]. Stem cell

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

regeneration is crucial in linking molecular and tissue-level development, thereby maintaining tissue homeostasis [8, 40, 50]. The intricate interplay between gene expression regulation and cell population dynamics is essential for sustaining dynamic equilibrium during stem cell regeneration. However, gene regulation and population evolution dynamics operate on different scales and are typically described by separate mathematical models. It is imperative to bridge the gap between these scales of dynamics.

Cell division is central to stem cell regeneration as it connects molecular and population-level dynamics. At the molecular level, gene regulation networks governing cell cycling dictate cell growth, differentiation, and division decisions. The dynamics occurring within a cell at the molecular level determine the signals that trigger cell division [9]. At the population level, these cell divisions directly influence the evolutionary dynamics of different cell types. Furthermore, variability in the epigenetic state of cells during deoxyribonucleic acid (DNA) replication and mitosis may lead to transitions between cell types post-division, which is vital for maintaining tissue homeostasis and promoting development [43, 44].

The evolution of gene expression in a single cell is often mathematically described using a set of ordinary differential equations

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{F}(\boldsymbol{X}) - \boldsymbol{K}\boldsymbol{X}, \tag{1.1}$$

where $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ represents the transcription level (messenger ribonucleic acid (mRNA) concentration) or protein concentration of multiple genes. The nonlinear function $\boldsymbol{F} = (F_1, F_2, \cdots, F_n)$ describes the rates of transcription or protein production and can be specified based on the gene regulatory network, $\boldsymbol{K} = \mathrm{diag}(k_1, k_2, \cdots, k_n)$ represents the rates of mRNA (or protein) degradation and dilution. Differential equations of the form (1.1), which can be extended to include delays or stochastic fluctuations, have been effectively used to study the dynamics of gene regulatory networks [25, 30, 52]. However, cell division is not accounted for in the differential equation model (1.1), making it unsuitable for long-term processes related to cell proliferation and differentiation.

Extending the differential equation (1.1) to include cell divisions and population dynamics is challenging. However, a top-down approach using the Euler coordinate model can be employed to describe the dynamics of stem cell regeneration while considering cell proliferation and differentiation. Recent studies [28, 29] established a mathematical model for stem cell regeneration that integrates cell division, cell heterogeneity, and the random transition of epigenetic states. This model builds upon the classical G0 cell cycle model proposed by Burns and Tannock [4], introducing a variable $\boldsymbol{x}$ – typically a high-dimensional vector – to represent the epigenetic state of a cell.

In this framework, the rates of cell proliferation, differentiation, senescence, and apoptosis are assumed to depend on the epigenetic state of the cell. Let $Q(t, \boldsymbol{x})$ denote the number of cells possessing the epigenetic state $\boldsymbol{x}$. The evolution of $Q(t, \boldsymbol{x})$ is governed by the following differential-integral equation [28, 29]:

$$\begin{cases} \dfrac{\partial Q(t,\boldsymbol{x})}{\partial t} = -Q(t,\boldsymbol{x})\big(\beta(c(t),\boldsymbol{x})+\kappa(\boldsymbol{x})\big) \\ \qquad\qquad +2\displaystyle\int_\Omega \beta\big(c\big(t-\tau(\boldsymbol{y})\big),\boldsymbol{y}\big)Q\big(t-\tau(\boldsymbol{y}),\boldsymbol{y}\big)e^{-\mu(\boldsymbol{y})\tau(\boldsymbol{y})}p(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \\ c(t) = \displaystyle\int_\Omega Q(t,\boldsymbol{x})\zeta(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \end{cases} \tag{1.2}$$

Here, β,κ, and μ represent the rates of cell proliferation, differentiation/senescence, and apoptosis, respectively. The variable τ indicates the duration of cell proliferation, while ζ signifies the rate of cytokine secretion. The concentration of growth factors secreted by all cells is denoted by $c(t)$. Additionally, $p(\boldsymbol{x},\boldsymbol{y})$ quantifies the probability of epigenetic state inheritance during cell division, as defined by the conditional probability

$$p(\boldsymbol{x},\boldsymbol{y}) = P(\text{state of daughter cell} = \boldsymbol{x} \,|\, \text{state of mother cell} = \boldsymbol{y}). \tag{1.3}$$

Eq. (1.2) provides a general mathematical model framework for analyzing the dynamics of heterogeneous stem cell regeneration, incorporating epigenetic transition.

The epigenetic state $\boldsymbol{x}$ discussed in Eq. (1.2) represents intrinsic cellular conditions that change over time during the cell cycle or cell division. Biologically, the epigenetic state of a cell refers to molecular changes that occur independently of DNA sequences. These changes can include DNA methylation, modifications to nucleosome histones, and variations in gene expression [42, 46, 47, 49, 51, 53]. As such, the epigenetic state $\boldsymbol{x}$ is not necessarily identical to the transcription level $\boldsymbol{X}$ in Eq. (1.1). The epigenetic state may influence the quantification of cellular behaviors, such as proliferation $(\beta(c,\boldsymbol{x}))$, differentiation $(\kappa(\boldsymbol{x}))$, apoptosis $(\mu(\boldsymbol{x}))$, and the cell cycle $(\tau(\boldsymbol{x}))$. Together, these behaviors are defined as the kinetotype of a cell [28]. The proliferation rate $\beta(c,\boldsymbol{x})$ depends on both cell-to-cell interactions, mediated by the cytokine concentration $c(t)$, and the epigenetic state of the cell. This concept of kinetotype highlights the importance of estimating single-cell kinetic rates based on single-cell ribonucleic acid (RNA) sequencing data [7, 13].

In the Eq. (1.2), cell division is represented as a factor 2 for the increase in cell population, alongside the inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ for the variation in the epigenetic state. During cell division, numerous biochemical reactions occur, including DNA replication and the re-establishment of DNA methylations and histone modifications, which are essential for preparing and completing the process. These molecular activities introduce variability in the daughter cells, leading to differences in their epigenetic states and cellular characteristics.

The inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ captures the relationship between the epigenetic states of daughter cells and their mother cells while abstracting the complex intermediate biochemical processes. Deriving the mathematical formulation of the inheritance function from the cell division process poses significant challenges. Numerical simulations based on epigenetic state inheritance laws [18, 19, 45] can provide phenomenological representations of this function. For instance, normalized nucleosome modifications have been demonstrated to follow a beta distribution [18], whereas transcript levels are often described by a gamma distribution [6]. Thus, if nucleosome modification is considered

as the epigenetic state, the beta distribution can serve as the inheritance function. Conversely, the gamma distribution should be employed if transcription levels are regarded as the epigenetic state. Biologically, nucleosome modification can influence gene expression by regulating the transcription process. Nonetheless, it remains unclear how the topology and dynamics of gene regulation networks may affect the inheritance function.

The mathematical framework presented in (1.2) is a foundational model that integrates essential components of stem cell regeneration, including cell cycling, cellular heterogeneity, and plasticity. This framework has been applied to various challenges, such as cell-type transitions, tumorigenesis, and tumor evolution [15,20,35,60]. However, stem cell systems can exhibit considerably greater complexity, necessitating the incorporation of additional biological processes into the framework. In particular, this framework does not include the gene regulation networks that underlie the dynamics of epigenetic states within the cell cycle. There exists a gap in how the gene network-based approach (1.1) and the stem cell regulation-based approach (1.2) can be integrated to provide a cohesive description of multiscale dynamics in tissue development.

This study aims to develop a computational scheme that links gene regulatory network dynamics with stem cell regeneration by formulating the inheritance function. We focused on a well-researched gene network that governs cell fate decisions and transitions between cell types, devising a numerical scheme to calculate the conditional probability density $p(\boldsymbol{x},\boldsymbol{y})$. This density function, validated through stochastic simulation, offers insights into how the structure of gene regulatory networks influences inheritance. Importantly, our numerical scheme paves the way for extending these findings to broader gene regulatory networks, enhancing our understanding of the connection between gene regulation dynamics and stem cell regeneration.

2 Model and method

We referred to a hybrid model of stem cell differentiation developed to study the dynamics of cell-type transitions influenced by epigenetic modifications [20]. This model combines individual-cell-based modeling of a multicellular system with gene regulation network dynamics. It also includes a $G0$ cell cycle model that governs cell regeneration and incorporates stochastic inheritance of epigenetic states during cell divisions.

2.1 Hetereogeneous $G0$ cell cycle model

In this study, we concentrated exclusively on cells capable of cycling, excluding any that have lost this ability from our simulation pool. We employed the $G0$ cell cycle model to illustrate the dynamics of stem cell regeneration [4, 36]. According to this model, cells progress through a resting phase known as $G0$. During this phase, they grow and prepare to enter the proliferative phase upon receiving signals that trigger cell cycling checkpoints (Fig. 1). Each cell in the resting phase can either enter the proliferating phase at β or exit the resting pool at a rate of κ due to processes such as terminal differentiation,

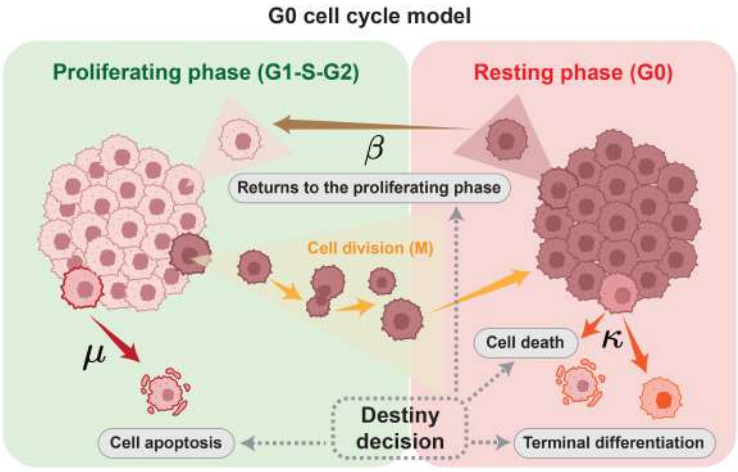

Figure 1: $G0$ cell cycle model of cell regeneration. Schematic representation of the $G0$ model depicting stem cell regeneration. Cells in the resting phase either enter the proliferative phase at β or exit the resting pool at κ due to differentiation, senescence, or death. Proliferating cells undergo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

senescence, or death. Cells in the proliferating phase may be randomly eliminated at a rate μ due to apoptosis or undergo mitosis after a fixed period, τ, following their entry into the proliferative compartment. Mitosis produces two daughter cells from each mother cell, and these newborn cells then enter the resting phase, starting the next cycle.

Let $Q(t)$ denote the number of resting-phase stem cells, and let $s(t,a)$ represent the population of proliferating stem cells. The age $a=0$ indicates the time of entry into the proliferative state. Based on these assumptions, we can derive the following $G0$ cell cycle model [33]:

$$\begin{cases} \nabla s(t,a) = -\mu s(t,a), & t>0, \quad 0<a<\tau, \\ \dfrac{\mathrm{d}Q}{\mathrm{d}t} = 2s(t,\tau) - (\beta+\kappa)Q, & t>0. \end{cases} \tag{2.1}$$

Here, $\nabla = \partial/\partial t + \partial/\partial a$. The boundary condition at $a=0$ is given by

$$s(t,0) = \beta Q. \tag{2.2}$$

Biologically, the ability of a cell to self-renewal, denoted as β, is closely linked to microenvironmental conditions such as growth factors, various cytokines, and intracellular signaling pathways [39,41,55]. Despite the complexity of these pathways, the characteristic Hill function dependence can be derived from simple assumptions about the interactions between signaling molecules and their receptors [3,28]. This can be expressed mathematically as

$$\beta = \beta_0 \frac{\theta^n}{\theta^n + c^n}, \tag{2.3}$$

where β_0 represents the maximum proliferation rate, c signifies the influence of cytokine signals released from all stem cells, θ indicates the half-effective concentration of cy-

tokines, and n is the Hill coefficient. The cytokine signal c typically depends on the collective state of all stem cells within the niche, creating a feedback loop in cell growth. Specifically, while we assume that the stem cells are homogeneous, the cytokine signals c can be considered proportional to the cell population Q.

In the Eq. (2.1), the processes of proliferation, differentiation, senescence, and cell apoptosis (or cell death during the resting phase) play a crucial role in determining the population dynamics of a multicellular system. Additionally, cells within a multicellular system are not homogeneous; they exhibit variability in their epigenetic states. Each cell's kinetic rates of proliferation, differentiation, and apoptosis can depend on its specific epigenetic state. To represent this epigenetic heterogeneity, we introduce a variable $\boldsymbol{x}$ (typically a high-dimensional vector) that denotes the epigenetic state of a cell [28,29,32]. The kinetic rates β,κ,μ, and the duration of the proliferating phase τ are all influenced by the epigenetic state $\boldsymbol{x}$. Consequently, the quadruple $(\beta(c,\boldsymbol{x}),\kappa(\boldsymbol{x}),\mu(\boldsymbol{x}),\tau(\boldsymbol{x}))$ defines the kinetotype of a cell [28].

Through the epigenetic state $\boldsymbol{x}\in\Omega$, where Ω represents the space encompassing all potential epigenetic states, let $Q(t,\boldsymbol{x})$ denote the count of cells in the resting phase at time t with epigenetic state $\boldsymbol{x}$. The total cell count can be expressed as

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

Additionally, let $\zeta(\boldsymbol{x})$ represent the rate of cytokine secretion by a cell with the state $\boldsymbol{x}$, the effective cytokine concentration c regulating cell proliferation is given by

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

Thus, the proliferation rate is expressed as

$$\beta\big(c(t),\boldsymbol{x}\big)=\beta_0(\boldsymbol{x})\frac{\theta(\boldsymbol{x})^n}{\theta(\boldsymbol{x})^n+c^n},\quad c(t)=\int_\Omega Q(t,\boldsymbol{x})\zeta(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.$$

During cell division, mother cells undergo several critical processes, including DNA replication, reconstruction of DNA methylations and histone modifications, and the separation of molecules during mitosis. Even though these processes are highly regulated, random perturbations can disrupt biochemical reactions, which may result in daughter cells that do not inherit the same epigenetic state as their mother cells. To quantify this, let $p(\boldsymbol{x},\boldsymbol{y})$ represent the probability density of a daughter cell having state $\boldsymbol{x}$, given that the mother cell was in state $\boldsymbol{y}$. This can be expressed as a conditional probability

$$p(\boldsymbol{x},\boldsymbol{y})=P(\text{state of daughter cell}=\boldsymbol{x}\mid\text{state of mother cell}=\boldsymbol{y}).\tag{2.4}$$

The inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ describes the transition of epigenetic states that occurs during cell division.

Considering cell heterogeneities following the kinetotype $(\beta(c,\boldsymbol{x}),\kappa(\boldsymbol{x}),\mu(\boldsymbol{x}),\tau(\boldsymbol{x}))$, the $G0$ cell cycle model (2.1) is modified as [28]

$$\begin{cases} \nabla s(t,a,\boldsymbol{x}) = -\mu(\boldsymbol{x})s(t,a,\boldsymbol{x}), & t>0, \quad 0<a<\tau(\boldsymbol{x}), \qquad (2.5a)\\ \dfrac{\partial Q(t,\boldsymbol{x})}{\partial t} = 2\displaystyle\int s\big(t,\tau(\boldsymbol{y}),\boldsymbol{y}\big)\,p(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\ \qquad\qquad -\big(\beta(c(t),\boldsymbol{x})+\kappa(\boldsymbol{x})\big)Q(t,\boldsymbol{x}), \quad t>0, \qquad (2.5b)\\ c(t) = \displaystyle\int_\Omega Q(t,\boldsymbol{x})\zeta(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \qquad (2.5c) \end{cases}$$

The boundary condition is given by

$$s(t,0,\boldsymbol{x}) = \beta\big(c(t),\boldsymbol{x}\big)Q(t,\boldsymbol{x}). \tag{2.6}$$

Solving the Eq. (2.5a) along with the boundary condition (2.6) by the method of the characteristic line, we have

$$s\big(t,\tau(\boldsymbol{x}),\boldsymbol{x}\big) = \beta\big(c\big(t-\tau(\boldsymbol{x})\big),\boldsymbol{x}\big)Q\big(t-\tau(\boldsymbol{x}),\boldsymbol{x}\big)e^{-\mu(\boldsymbol{x})\tau(\boldsymbol{x})}. \tag{2.7}$$

Substituting (2.7) into (2.5), we obtain the evolution equation for the cell population $Q(t,\boldsymbol{x})$ [28]

$$\begin{cases} \dfrac{\partial Q(t,\boldsymbol{x})}{\partial t} = -Q(t,\boldsymbol{x})\big(\beta\big(c(t),\boldsymbol{x}\big)+\kappa(\boldsymbol{x})\big) \\ \qquad\qquad +2\displaystyle\int_\Omega \beta\big(c\big(t-\tau(\boldsymbol{y})\big),\boldsymbol{y}\big)Q\big(t-\tau(\boldsymbol{y}),\boldsymbol{y}\big)e^{-\mu(\boldsymbol{y})\tau(\boldsymbol{y})}p(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \qquad (2.8)\\ c(t) = \displaystyle\int_\Omega Q(t,\boldsymbol{x})\zeta(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \end{cases}$$

Eqs. (2.8) offers a mathematical framework for modeling the dynamics of heterogeneous stem cell regeneration with epigenetic transitions.

2.2 Gene regulation network dynamics

2.2.1 Differential equation model

Despite the multiscale interactions involved in the Eq. (2.8), the gene regulatory networks that play a crucial role in the evolution of epigenetic states within a cell cycle are not accounted for in (2.8). In this context, we propose a general formulation for GRN dynamics based on the RACIPE modeling scheme [16]. RACIPE is a computational method for modeling gene regulatory network dynamics without requiring detailed knowledge of the exact network topology or precise parameter values.

Consider a GRN with m genes. The products of these genes can either activate or repress the expression of other genes. To model gene regulation, we adopt a function H based on the RACIPE scheme

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

where X represents the concentration of protein that regulates gene expression, λ signifies the regulatory strength, with $\lambda<1$ indicating repression and $\lambda>1$ indicating activation, h represents the half-effetive protein concentration for the corresponding regulation, and n is the Hill coefficient. The regulation from gene j to gene i is represented by a factor $H(X_j;h_{ji},n_{ji},\lambda_{ji})$, which influences the protein production rate of gene i. It is important to note that if $\lambda_{ji}=1$, there is no interaction from gene j to gene i.

Using the notations provided, the dynamics of the GRN can be expressed via the following differential equations:

$$\frac{\mathrm{d}X_i}{\mathrm{d}t}=g_i\prod_{j=1}^{m}H(X_j;h_{ji},n_{ji},\lambda_{ji})-k_iX_i,\quad i=1,2,\ldots,m. \tag{2.9}$$

Here, X_i represents the protein concentration of gene i, g_i denotes the basal protein production rate of gene i, and k_i is the degradation rate of the protein. We omitted the transcription process for simplicity and used protein concentration as a proxy for the transcription level.

When multiple regulators target a gene, the functional form of the rate equations depends on the nature of the multivalent regulation. We adopted the RACIPE approach in this study and assumed that these regulatory interactions are independent. Thus, the overall production rate is written as the product of the innate production rate of the target gene and the shifted Hill functions for all the regulatory links. Alternative formulations can be considered similarly.

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

$$a=\bar{a}e^{\sigma\xi-\sigma^2/2}. \tag{2.10}$$

Here, σ indicates the intensity of the noise perturbation, and ξ follows a colored noise pattern defined by the Ornstein-Uhlenbeck process

$$\mathrm{d}\xi=-(\xi/\vartheta)\mathrm{d}t+\sqrt{2/\vartheta}\mathrm{d}W. \tag{2.11}$$

In this equation, W represents the Wiener process, and ϑ is the relaxation coefficient.

Unlike the conventional Langevin stochastic differential equation approach, the random perturbation η in (2.10) is expressed exponentially. This approach has been utilized in previous research studies to simulate extrinsic noise perturbations in gene expression [19, 20, 31]. The exponential form in (2.10) helps to avoid potential issues with

negative parameter values that may arise in the Langevin framework, where the perturbation is typically expressed as $a \to a+\sigma\xi$.

In biochemical reactions, a chemical rate a is often related to a chemical potential μ through the equation $a=ce^{-\beta\mu}$. Therefore, a perturbation in the chemical potential μ naturally results in an exponential perturbation of the chemical reaction rate. Furthermore, gene expression rates have been observed to follow a log-normal distribution rather than a normal distribution [2]. Thus, it is appropriate to formulate the random perturbation by multiplying a random number drawn from a log-normal distribution. In this context, the random perturbation is defined using an Ornstein-Uhlenbeck process. For a more detailed discussion, please refer to Lei [27] and Huang *et al.* [20].

The Eqs. (2.10)-(2.11) present a general formulation for noise perturbation applied to a parameter. It is important to note that the perturbations, denoted as ξ, to various parameters g_i, h_{ij}, and k_i in Eq. (2.9) are independent of each other. Furthermore, the parameters σ and ϑ, which define the noise perturbations, vary for different perturbations. Consequently, the differential equation model represented by (2.9) can be expanded into the following stochastic differential equation (SDE) model:

$$\begin{cases} \dfrac{\mathrm{d}X_i}{\mathrm{d}t}=g_i\displaystyle\prod_{j=1}^{m}H(X_j;h_{ji},n_{ji},\lambda_{ji})-k_iX_i, \quad i=1,2,\ldots,m, \\ g_i=\bar{g}_ie^{\sigma_1\xi_1-\sigma_1^2/2}, \\ h_{ij}=\bar{h}_{ij}e^{\sigma_2\xi_2-\sigma_2^2/2}, \\ k_i=\bar{k}_ie^{\sigma_3\xi_3-\sigma_3^2/2}, \\ \mathrm{d}\xi_1=-(\xi_1/\vartheta_1)\mathrm{d}t+\sqrt{2/\vartheta_1}\mathrm{d}W_1, \\ \mathrm{d}\xi_2=-(\xi_2/\vartheta_2)\mathrm{d}t+\sqrt{2/\vartheta_2}\mathrm{d}W_2, \\ \mathrm{d}\xi_3=-(\xi_3/\vartheta_3)\mathrm{d}t+\sqrt{2/\vartheta_3}\mathrm{d}W_3. \end{cases} \tag{2.12}$$

2.2.2 Cross-cell-cycle dynamics

The SDE model (2.12) describes the dynamics of gene regulation within a single cell cycle. To extend these dynamics across multiple cycles and connect them with the heterogeneous $G0$ cell cycle model (2.8), we need to incorporate the cell cycle process into Eq. (2.12).

During the proliferative phase, a cell undergoes DNA replication, which alters its genetic content. As a consequence, gene expression rates fluctuate throughout this phase. For simplicity, let T represent the duration of one cycle. The cycling age, denoted as ℓ, is defined as follows: $0<\ell<T_1$ for the $G0/G1$ phase, $T_1<\ell<T_1+T_2$ for the S phase (which encompasses DNA replication), and $T_1+T_2<\ell<T$ for the G2 and M phase that occur after DNA replication (see Fig. 2). Let $\ell(t)$ represents the cycling age, defined by the equation

$$\ell(t)=t-(k-1)T, \quad (k-1)T\le t<kT, \quad k=1,2,\ldots. \tag{2.13}$$

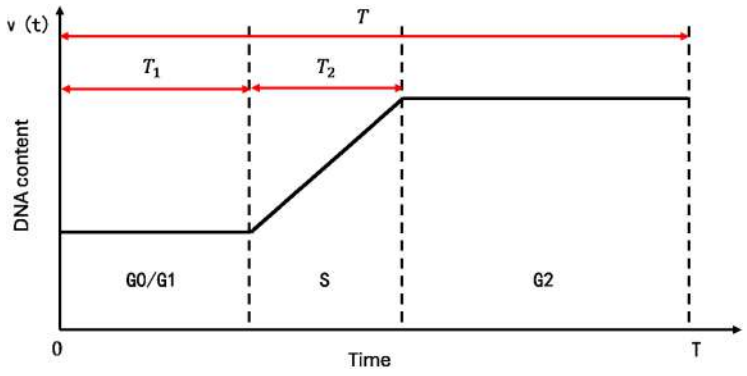

Figure 2: Illustration of a cell cycle and the cycling age. Here, T denotes the duration of one cycle, T_1 represents the duration of the $G0/G1$ phase, and T_2 indicates the duration of the S phase.

We defined the changes in DNA content with the cell's cycling age as follows:

$$\nu(t)=\begin{cases}1, & 0\leq \ell(t)<T_1,\\ 1+\dfrac{\ell-T_1}{T_2}, & T_1\leq \ell(t)<T_1+T_2,\\ 2, & T_1+T_2\leq \ell(t)<T.\end{cases} \tag{2.14}$$

This function simplifies our analysis.

By incorporate the DNA content $\nu(t)$, we replaced g_i with $g_i\nu(t)$, leading us to rewrite Eq. (2.9) as

$$\frac{\mathrm{d}X_i}{\mathrm{d}t}=g_i\nu(t)\prod_{j=1}^{m}H(X_j;h_{ji},n_{ji},\lambda_{ji})-k_iX_i,\quad i=1,2,\ldots,m \tag{2.15}$$

for the interval $(k-1)T\leq t<kT$.

The cell undergoes mitosis at time $t=kT$. During this process, molecules such as proteins and mRNAs redistribute between the two daughter cells with stochastic fluctuations. As a result, the protein concentrations in the newly formed cells need to be reset following mitosis while the molecules are allocated to the two daughters. Therefore, we reset the initial conditions at $t=kT$ as follows:

$$X_i(kT)=\chi_i\lim_{t\to kT^-}X_i(t), \tag{2.16}$$

where χ_i is a random number.

The coefficient χ_i measures the ratio of protein concentrations in the daughter cells to that in the mother cell. Biologically, the proteins in the mother cell are distributed between the two daughter cells, hence $0<\chi_i<1$. The binomial distribution is often used to describe the random assignment of proteins to the daughter cells. In contrast, the beta distribution is the conjugate prior probability distribution of the binomial distribution. Thus, we assume that the random number χ_i follows a beta distribution with a density

function that can exhibit various shapes based on the shape parameters a and b (see Fig. 9 in the Appendix C). These shapes include strictly decreasing (when $a<1$ and $b\leq 1$), strictly increasing (when $a>1$ and $b\leq 1$), U-shaped (when $a<1$ and $b<1$), or unimodal (when $a>1$ and $b>1$). We can select appropriate parameters to model symmetric or asymmetric division scenarios.

We denote the random variable following a beta distribution as $\chi_i\sim \text{Beta}(a,b)$, then the expectation and variance of χ_i are given by the shape parameters a and b as

$$\text{E}(\chi_i)=\frac{a}{a+b},\quad \text{var}(\chi_i)=\frac{ab}{(a+b)^2(a+b-1)}. \tag{2.17}$$

Additionally, we introduce two parameters: $0<\phi<1$ and $\eta>0$, which are associated with the mean and variance of χ_i. These parameters are defined as follows:

$$\text{E}(\chi_i)=\phi,\quad \text{var}(\chi_i)=\frac{1}{1+\eta}\phi(1-\phi). \tag{2.18}$$

From these definitions and (2.17), the shape parameters a and b can be expressed in terms of ϕ and η

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

The parameters ϕ and η are biologically more straightforward than the shape parameters a and b. While we set $\phi=1/2$, the value of η influences the distribution's shape: $\eta>2$ results in an unimodal distribution, reflecting symmetry division. In contrast, $\eta<2$ creates a U-shape distribution, indicating asymmetry division (see Fig. 9 in the Appendix C).

Thus, we integrated Eqs. (2.12) through (2.16) to develop a hybrid model for cross-cell-cycle dynamics as follows:

$$\begin{cases}
\dfrac{\text{d}X_i}{\text{d}t}=g_i\nu(t)\displaystyle\prod_{j=1}^{m}H(X_j;h_{ji},n_{ji},\lambda_{ji})-k_iX_i,\quad i=1,\dots,m,\\
g_i=\bar{g}_ie^{\sigma_1\xi_1-\sigma_1^2/2},\\
h_{ij}=\bar{h}_{ij}e^{\sigma_2\xi_2-\sigma_2^2/2},\\
k_i=\bar{k}_ie^{\sigma_3\xi_3-\sigma_3^2/2},\\
\text{d}\xi_1=-(\xi_1/\vartheta_1)\text{d}t+\sqrt{2/\vartheta_1}\text{d}W_1,\\
\text{d}\xi_2=-(\xi_2/\vartheta_2)\text{d}t+\sqrt{2/\vartheta_2}\text{d}W_2,\\
\text{d}\xi_3=-(\xi_3/\vartheta_3)\text{d}t+\sqrt{2/\vartheta_3}\text{d}W_3,\\
\quad (k-1)T\leq t<kT,\quad k=1,2,\dots,\\
X_i(kT)=\chi_i\lim_{t\to kT^-}X_i(t),\\
\chi_i\sim\text{Beta}(a,b).
\end{cases} \tag{2.19}$$

In the Eqs. (2.19), the term λ_{ji} in the function $H(x_j;h_{ji},n_{ji},\lambda_{ji})$ indicates the type of regulation between genes j and i. Specifically, $\lambda_{ji}>1$ denotes positive regulation, $\lambda_{ji}<1$

indicates negative regulation, and $\lambda_{ji}=1$ signifies no regulation. Consequently, the coefficients λ_{ij} represent the topology of the GRN. Eq. (2.19) outlines the dynamics of this network across multiple cell cycles, with protein levels reset following cell division.

It is important to note that our model does not account for the effects of changes in cell volume. Cell growth and volume alterations are significant biological processes that play crucial roles in cell fate decisions [11, 14]. Additionally, the biological mechanisms governing cell growth and division are complex and not completely understood for mammalian cells [5, 24, 58]. To simplify our analysis, we have assumed that cell volume remains constant during cell cycling, meaning that cell growth does not affect protein concentration.

Both Eqs. (2.8) and (2.19) describe the long-term biological process of stem cell regeneration. However, the differential-integral equation (2.8) focuses on population dynamics while accounting for cellular heterogeneity and plasticity during cell division. In contrast, the stochastic differential equation (2.19) captures molecular-level variations within a cell across multiple cell cycles. The population dynamics and molecular-level variations occur on distinct spatial and temporal scales. Changes in the population dynamics primarily involve processes such as cell regeneration, death, and differentiation. These are associated with cellular behaviors in tissues and occur on a slower time scale, ranging from minutes to hours. In contrast, molecular-level dynamics occur within individual cells and exhibit rapid variations on a time scale of milliseconds to seconds. These two equations represent dynamics at different scales: The differential-integral equation (2.8) emphasizes population-level behavior but does not incorporate gene regulatory networks, whereas the stochastic differential equation (2.19) is driven by the gene regulatory networks but does not address population-level dynamics. This raises a critical question: how can these two scales of dynamics be integrated?

2.3 Connection between two models for a simple gene regulatory network

While a uniform method to integrate Eqs. (2.8) and (2.19) is not well established, we study a simple gene network and explore a method to connect the dynamics of cross-cell-cycling gene expression with the stem cell regeneration process.

For simplicity, we consider a basic gene network consisting of two master transcription factors (TFs), specifically gene A and gene B. These TFs exhibit self-activation and mutual repression, as illustrated in Fig. 3. Let X_1 and X_2 represent the expression levels (protein concentration) of genes A and B, respectively. Based on the mathematical formulation provided earlier, Eq. (2.19) with $m=2$ describes the dynamical equation governing protein concentrations during cell cycling.

It is important to note that the gene network illustrated in Fig. 3 can be modeled using different approaches: additive or multiplicative regulations [23,54]. If multiple genes regulate the transcription of the target gene independently, the production rates from positive and negative feedbacks are summed, and therefore, additive regulation should be applied. However, multiplicative regulation should be applied if the gene regulations are

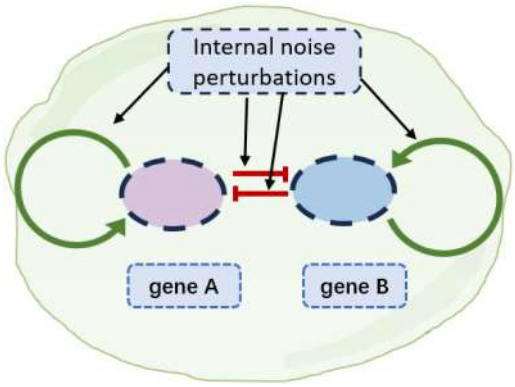

Figure 3: Schematic diagram illustrating the gene regulatory network. Green arrows represent self-activating interactions among genes, while red bars denote repressive interactions between genes.

interconnected. Moreover, more intricate mathematical formulations should be derived based on biological interactions in gene networks with more complex topological structures. In this study, we adopted the multiplicative developed by Jolly *et al.* [23] for simplicity. They applied the multiplicative regulations to model the modules of epithelial-to-mesenchymal transition (EMT). However, the computational scheme presented remains applicable even with additive feedback regulations.

To link the gene regulation dynamics with the heterogeneous stem cell regeneration described in Eq. (2.8), we need to identify the epigenetic state $\boldsymbol{x}$ and derive the inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ based on the dynamical equation (2.19).

In biology, the epigenetic state refers to a quantitative form independent of the DNA sequence, including gene expression levels, histone modifications, or DNA methylation. Among these, histone modifications and DNA methylation levels in chromosomal gene regions can influence gene expression levels. This study primarily focuses on epigenetic states associated with gene expression levels. Accordingly, we assumed a quantitative dependency between the epigenetic state $\boldsymbol{x}$ and the gene expression level $\boldsymbol{X}$. The choice of this dependency will directly impact the mathematical definition of the inheritance function $p(\boldsymbol{x},\boldsymbol{y})$.

The epigenetic state $\boldsymbol{x}$ in Eq. (2.8) represents a constant reflecting the state of a cell during its resting phase. At the same time, it changes dynamically over time in (2.19), which captures gene expression dynamics throughout a cell cycle. Thus, we must identify the epigenetic state $\boldsymbol{x}$ per the gene expression dynamics $\boldsymbol{X}(t)$ over one cell cycle. For clarity, we define the epigenetic state $\boldsymbol{x}=(x_1,x_2)$ of a cell through log normalization of gene expression $\boldsymbol{X}(t)=(X_1(t),X_2(t))$ as follows:

$$x_i=\log\big(X_i(t)+1\big)\big|_{\ell(t)=\ell^*},\quad i=1,2. \tag{2.20}$$

Here, the time t corresponds to a specific cell cycling age ℓ^*. In this study, we selected $\ell^*=T_1$, representing the time point just before DNA replication. For brevity, we can

express this as

$$\boldsymbol{x}=\log\left(\boldsymbol{X}(t)+1\right)\big|_{\ell(t)=\ell^*}. \tag{2.21}$$

It is important to note that the correspondence between the gene expression level $\boldsymbol{X}$ and the epigenetic state $\boldsymbol{x}$ remains an open question in biology. This study defined the epigenetic state $\boldsymbol{x}$ as a quantity directly related to the gene expression level within a single cell cycle. The definition in (2.20) aligns with the log normalization commonly used in analyzing single-cell RNA sequencing data. However, alternative definitions may be equally valid depending on specific requirements.

To derive the inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ from the dynamics of gene regulation, we numerically solved the cross-cell-cycle stochastic differential equation (2.19) and tracked the gene expression $\boldsymbol{X}(t)$ of individual cells during their cell divisions. Since (2.19) models the gene expression dynamics of a single cell across multiple cycles, we recorded the two daughter cells formed when a cell divided. We repeated this numerical scheme to create an ensemble of multiple cells, with each simulation starting from a randomly selected initial condition.

For each simulated cell, we recorded the state $\boldsymbol{X}=\boldsymbol{X}(t)|_{\ell=T_1}$ such that $\boldsymbol{x}=\log(\boldsymbol{X}+1)$ represents the epigenetic state of the cell. This process generated a dataset $D=\{(\boldsymbol{x}_k,\boldsymbol{y}_k)\}$, where $\boldsymbol{y}_k$ denotes the state of the k-th mother cell, and $\boldsymbol{x}_k$ denotes the state of one of its daughter cells. From this dataset D, the inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ can be derived using the conditional probability density (1.3).

For the probability density function $p(\boldsymbol{x},\boldsymbol{y})$, we assumed that the inheritances of gene A and gene B are independent of each other. Here, the inheritance of a gene refers to how the gene expression in the daughter cell depends on the state of the mother cell. This assumption implies that the mother cell's state independently influences the inheritance of each gene. However, gene inheritance may still depend on the expression levels of both genes in the mother cell due to their gene regulatory interactions. Under this assumption, the probability density function can be expressed as

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

where $p_i(x_i,\boldsymbol{y})$ represents the conditional density function of x_i (the expression level of gene i in the daughter cell) given the state of the mother cell $\boldsymbol{y}=(y_1,y_2)$.

Experimental observations showed that the steady-state probability of protein production from a single gene expression follows a gamma distribution [6]. In our study, we assumed a mixed conditional gamma distribution for the epigenetic states, represented as follows:

$$p_i(x_i,\boldsymbol{y})=\sum_{j=1}^{k}\alpha_{i,j}\mathrm{Gamma}(x_i;a_{i,j},b_{i,j}),\quad i=1,2, \tag{2.22}$$

where k denotes the number of independent distributions, $\alpha_{i,j}$ represent the combination coefficients, and $\mathrm{Gamma}(x;a,b)$ is the probability density function of the gamma distri-

bution with shape parameters a and b. It is expressed as

$$\begin{aligned} &\text{Gamma}(x;a,b)=\frac{x^{a-1}e^{-(x/b)}}{b^a\Gamma(a)},\\ &\Gamma(z)=\int_0^\infty e^{-x}x^{z-1}dx, \quad z>0. \end{aligned} \tag{2.23}$$

The combination coefficients $\alpha_{i,j}$ depend on the vector $\boldsymbol{y}$ and satisfy the following conditions:

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

Additionally, the shape parameters $a_{i,j}$ and $b_{i,j}$ are influenced by the state $\boldsymbol{y}$ of the mother cell.

We start with the probability density function given in Eq. (2.23) to determine the shape parameters. The conditional mean and variance can be expressed as follows:

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

We assume the conditional mean and variance as

$$\text{E}(x|\boldsymbol{y})=\psi(\boldsymbol{y}), \quad \text{var}(x|\boldsymbol{y})=\frac{\psi(\boldsymbol{y})^2}{\gamma(\boldsymbol{y})}. \tag{2.24}$$

From this, we can derive the following relationships:

$$a(\boldsymbol{y})=\gamma(\boldsymbol{y}), \quad b(\boldsymbol{y})=\frac{\psi(\boldsymbol{y})}{\gamma(\boldsymbol{y})}. \tag{2.25}$$

Here, we replace the shape parameters a and b with ψ and γ, which are more biologically intuitive. The function $\psi(\boldsymbol{y})$ directly measures the conditional expectation, while $\gamma(\boldsymbol{y})$ is associated with the condition variance relative to the expectation.

To obtain the inheritance function $p(\boldsymbol{x},\boldsymbol{y})$, we need to compute the functions $\psi_{i,j}(\boldsymbol{y})$ and $\gamma_{i,j}(\boldsymbol{y})$ for each component x_i using the conditional mean and variance derived from the simulation data. A detailed procedure for obtaining the inheritance function can be found in Section 3.2.

2.4 Numerical implementation

Implementing the model equation (2.19) numerically involves solving the SDE model within a cell cycle and managing the process of cell division. Cell division may result in the splitting of cells and resetting the initial conditions for newly formed cells. This numerical scheme can be realized using the object-oriented programming language `C++`. The dynamics of gene regulatory networks during a cell cycle are modeled with the SDE, which is numerically solved using the Runge-Kutta method. Additionally, we developed

an individual-cell-based numerical scheme to track the dynamics of each cell within the system. Detailed algorithm information is provided in the appendix.

In our numerical implementations, we focused on symmetric gene regulations, assuming that $k_1=k_2,\lambda_{12}=\lambda_{21},\lambda_{11}=\lambda_{22},n_{11}=n_{12}=n_{21}=n_{22}$. We normalized the protein concentrations so that $h_{11}=h_{12}=h_{21}=h_{22}$. The parameter values were estimated based on previous studies involving similar gene networks [17,20,23], with further adjustments made to ensure that cells could exhibit various phenotypes with different levels of gene expression.

3 Results

3.1 Identification of cell types through epigenetic states

To quantitatively define cell phenotypes, we first omitted the cell cycling and conducted a bifurcation analysis based on the ODE model (2.9), i.e. the following second-order differential equation:

$$\begin{cases} \dfrac{\mathrm{d}X_1}{\mathrm{d}t}=g_1\left(\dfrac{1+\lambda_{21}(X_2/h_{21})^{n_{21}}}{1+(X_2/h_{21})^{n_{21}}}\right)\left(\dfrac{1+\lambda_{11}(X_1/h_{11})^{n_{11}}}{1+(X_1/h_{11})^{n_{11}}}\right)-k_1X_1, \\ \dfrac{\mathrm{d}X_2}{\mathrm{d}t}=g_2\left(\dfrac{1+\lambda_{12}(X_1/h_{12})^{n_{12}}}{1+(X_1/h_{12})^{n_{12}}}\right)\left(\dfrac{1+\lambda_{22}(X_2/h_{22})^{n_{22}}}{1+(X_2/h_{22})^{n_{22}}}\right)-k_2X_2. \end{cases} \tag{3.1}$$

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. Fig. 10(a) in the Appendix 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 emergy (depicted by red dots in Fig. 10(a) in the Appendix C). Additionally, at $(g_1,g_2)=(0.4,0.4)$, solving the ODE (3.1) 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 Fig. 10(b) in the Appendix C. These stable steady states exhibit expression patterns of $(+,-)$ or $(-,+)$ for the maker 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 different patterns of marker gene expressions. 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 (2.19) under various noise strength σ (assuming $\sigma_1=\sigma_2=\sigma_3=\sigma$) and cell division scenarios. In our simulations, we took $T=50$, and $T_1=25,T_2=8$ in modeling cell cycling.

First, we excluded cell cycling and focused solely on gene expression dynamics with external noise perturbation. To achieve this, we solved the SDE model (2.12) with randomly selected initial conditions $0<X_i(0)<2$. Fig. 4(a) 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.

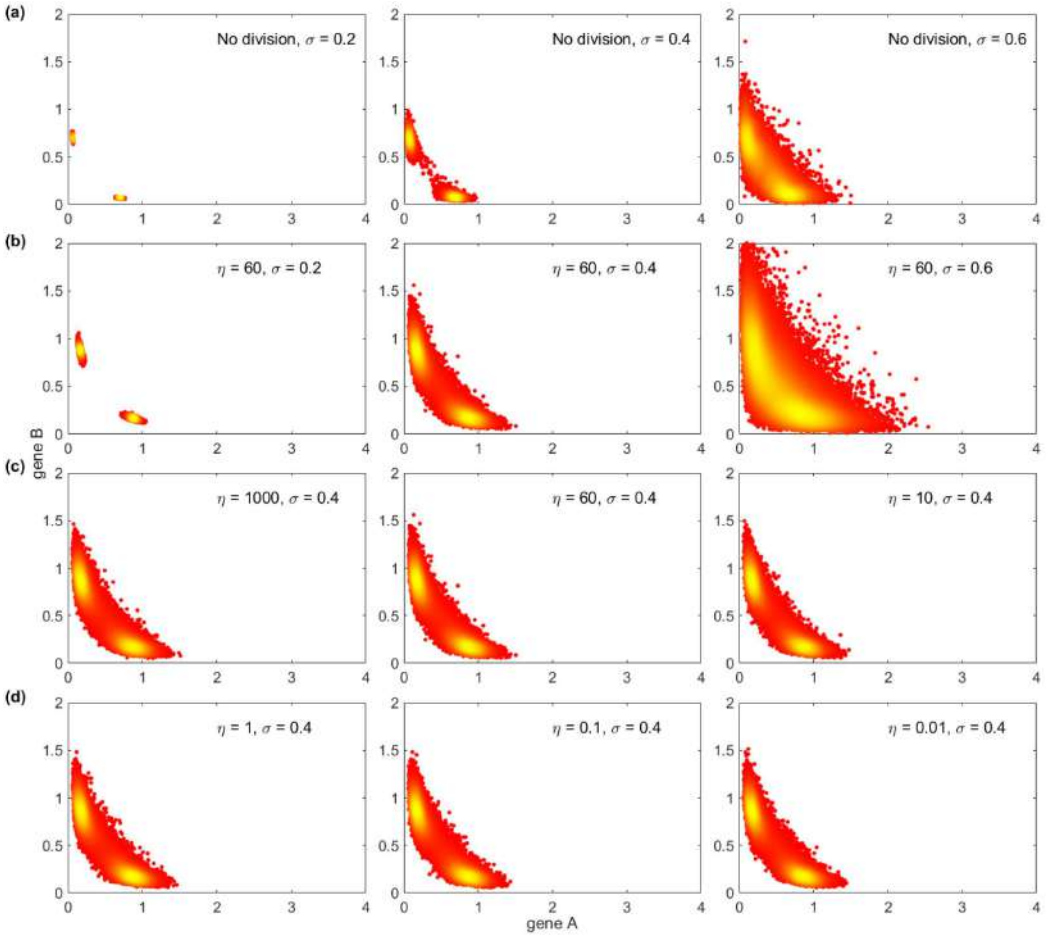

Figure 4: Distribution of the epigenetic state of cells under different model assumptions. (a). Epigenetic states of cells at $t=50$ (one cell cycle) without cell division. The cell states were obtained by solving the SDE model (2.12) with initial conditions randomly distributed over $0<X_i(0)<2$. (b). Epigenetic state of cells derived from a single cell after 15 divisions. The cell states were obtained by solving the hybrid model (2.19), 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 $\boldsymbol{x}=\log(\boldsymbol{X}(t)+1)|_{\ell(t)=T_1}$. 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 (2.18)). 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 (2.18) 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 Fig. 10 in the Appendix C.

Next, we investigated the effect of cell division using the hybrid model given by Eq. (2.19). The impact of cell division is represented by the redistribution of molecules described in Eq. (2.16), utilizing parameters ϕ and η that define the distribution of the coefficient χ_i. We fixed the values at $(\phi,\eta)=(0.5,60)$ (see Fig. 9(b) in the Appendix C) and varied the noise strength σ to solve Eq. (2.19).

We start the simulation with a single cell and ran the model dynamics for 15 cycles, resulting in 2^{15} cells from the original cell by the 15-th cycle. Fig. 4(b) illustrates the epigenetic states of all cells, displaying a pattern similar to that shown in Fig. 4(a), which was generated from multiple cells with different initial conditions. This outcome suggests that a single cell can produce a heterogeneous cell clone through cell division.

To investigate the effects of different assumptions in cell division, we varied the parameter η, which defines the distribution of the coefficient χ_i. By fixing $\sigma=0.4$, we adjusted η to reflect symmetry division, setting it between 10 and 1000, which resulted in a decreased variance of χ_i. As shown in Fig. 4(c), the distribution of epigenetic states appears to be independent of the parameter η. Similarly, when we varied η to reflect asymmetric division (ranging from 1 to 0.01), we observed consistent distributions of epigenetic states at the 15-th cycle, as depicted in Fig. 4(d). These results suggest that the distribution of epigenetic states in a cell clone remains insensitive to the variance of χ_i during cell division.

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

3.2 Data-driven inheritance function

In Fig. 4, we have demonstrated the variability of epigenetic states in individual cells originating from a single precursor cell. This variability reflects the stochastic changes in cell states during cell division. To derive the inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ based on the gene regulation dynamics outlined in Eq. (2.19), we tracked the process of cell division using numerical simulations. We recorded the states of the mother cells ($\boldsymbol{y}$) and their corresponding daughter cells ($\boldsymbol{x}$) after each cell division event.

We initiated 10^6 cells in our simulations and simulated the model over three cell cycles. In this framework, all cells at the second cycle were considered mother cells, with their epigenetic states denoted as $\boldsymbol{y}$. For each mother cell in the state $\boldsymbol{y}$, we paired it with the epigenetic state $\boldsymbol{x}$ of the corresponding daughter cell from the third cycle, forming a pair $[\boldsymbol{x},\boldsymbol{y}]=[(x_1,x_2),(y_1,y_2)]$. In this notation, x_1 and y_1 represent the state associated with gene A, while x_2 and y_2 represent the state related to gene B.

Figs. 5(a) and 5(b) present scatter plots illustrating the epigenetic states of genes A and B in mother and daughter cells. These plots demonstrate 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. Conversely, mother cells with higher expression tend to generate daughter cells with elevated expression in the same gene. However, we also observe cases of cell type switching during cell divisions, where some daughter cells exhibit gene expression patterns that differ from those of their mother cells.

To develop the mathematical formulation of the inheritance function, we used gene A as a case study to demonstrate the computational method. We identified all mother cells exhibiting a specific expression level (as indicated by the two strips in Fig. 5(a)) and ana-

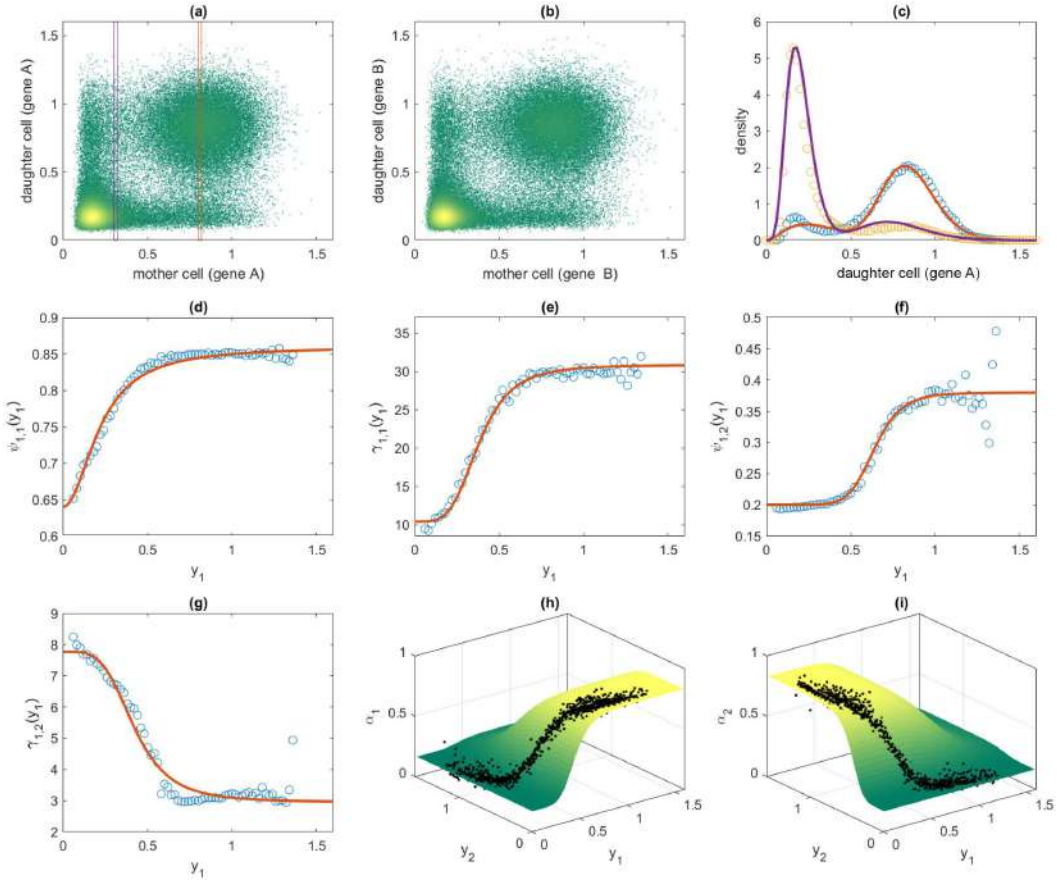

Figure 5: Data-driven inheritance function. (a). Scatter plot of daughter cell (gene A) versus mother cell (gene B) (x_1 versus y_1). Each point represents a cell, with lighter colors indicating higher density. Two vertical strips represent the sampling of data points with $0.3 \le y_1 < 0.31$ and $0.8 \le y_1 < 0.81$, 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 $\psi_{1,1}(y_1)$ to define the shape parameters in $\text{Gamma}(x_1;a_{1,1},b_{1,1})$. (e). The function $\gamma_{1,1}(y_1)$ to define the shape parameters in $\text{Gamma}(x_1;a_{1,1},b_{1,1})$. (f). The function $\psi_{1,2}(y_1)$ to define the shape parameters in $\text{Gamma}(x_1;a_{1,2},b_{1,2})$. (g). The function $\gamma_{1,2}(y_1)$ to define the shape parameters in $\text{Gamma}(x_1;a_{1,2},b_{1,2})$. (h). The coefficient $\alpha_1(y_1,y_2)$ in $p_1(x_1;\boldsymbol{y})$. (i). The coefficient $\alpha_2(y_1,y_2)$ in $p_2(x_2,\boldsymbol{y})$. In (d)-(i), dots were obtained from data analysis, while curves or surfaces were derived from Eqs. (3.3), (3.5), and (3.9), respectively.

lyzed the probability density of the expression levels in all daughter cells. Fig. 5(c) illustrates the resulting conditional probability densities corresponding to the epigenetic state of the mother cells at $y_1 = 0.3$ and $y_1 = 0.8$. These density functions display characteristics similar to a mixture of two unimodal distributions (solid lines in Fig. 5(c)).

Drawing insights from Fig. 5(c), we modeled the probability density $p_1(x_1,\boldsymbol{y})$ as a combination of two gamma distributions

$$p_1(x_1,\boldsymbol{y}) = \alpha_1 \times \text{Gamma}(x_1;a_{1,1},b_{1,1}) + (1-\alpha_1) \times \text{Gamma}(x_1;a_{1,2},b_{1,2}). \tag{3.2}$$

In this equation, the shape coefficients $a_{1,j}$ and $b_{1,j}$ depend on the state of the mother cell, $\boldsymbol{y}$, through the functions $\psi_{1,j}(\boldsymbol{y})$ and $\zeta_{1,j}(\boldsymbol{y})$. These functions are linked to the conditional mean and variance, $\text{E}(x_1|\boldsymbol{y})$ and $\text{var}(x_1|\boldsymbol{y})$, as described in Eq. (2.24). Additionally, the combination coefficient α_1 may vary according to the state of the mother cell $\boldsymbol{y}$.

To determine the coefficients α_1, $a_{1,j}$ and $b_{1,j}$, we initially assumed that their dependence was solely on the component y_1. From Fig. 5(a), we partitioned the data into multiple bins based on the values of y_1. Next, we divided the x_1 values within each bin into two subgroups and calculated the mean and variance for each subgroup. This process allowed us to derive the conditional means $E(x_1|y_1)$ and variances $\text{var}(x_1|y_1)$ for the two gamma distributions $\text{Gamma}(x_1;a_{1,j},b_{1,j})$, $j=1,2$. As a result, we obtained the functions $\psi_{1,j}(y_1)$ and $\gamma_{1,j}(y_1)$ as Hill-type functions, as described in Eq. (2.24) (Fig. 5(d)-(g))

$$\begin{cases} \psi_{1,1}(y_1) = 0.64 + 0.22 \times \dfrac{y_1^2}{0.22^2 + y_1^2}, \\ \gamma_{1,1}(y_1) = 10.38 + 20.46 \times \dfrac{y_1^4}{0.38^4 + y_1^4}, \\ \psi_{1,2}(y_1) = 0.20 + 0.18 \times \dfrac{y_1^8}{0.64^8 + y_1^8}, \\ \gamma_{1,2}(y_1) = 2.95 + 0.15 \times \dfrac{1}{0.42^4 + y_1^4}. \end{cases} \tag{3.3}$$

The shape parameters were then formulated as follows:

$$a_{1,j}(y_1) = \psi_{1,j}(y_1), \quad b_{1,j}(y_1) = \frac{\psi_{1,j}(y_1)}{\gamma_{1,j}(y_1)}. \tag{3.4}$$

We assume that the combination coefficient α_1 depends on y_1 and y_2. Therefore, we divide the (y_1,y_2) phase plane into multiple bins and collect the daughter cell epigenetic states x_1 within each bin. We then employ the expectation-maximization (EM) algorithm, as described in (3.2), to estimate the coefficients $\alpha_{1,j}$ for each bin. The results, shown in Fig. 5(h), closely resemble a Hill-type function

$$\alpha_1(\boldsymbol{y}) = 0.16 + \frac{y_1^6}{0.46^6 + y_1^6} \times \frac{1}{1.07^6 + y_2^6}. \tag{3.5}$$

Thus, the function $\alpha_1(\boldsymbol{y})$ and the shape parameters $a_{1,j}(y_1)$ and $b_{1,j}(y_1)$ together define the inheritance function $p_1(x_1;\boldsymbol{y})$ as outlined in (3.2).

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

$$p_2(x_2;\boldsymbol{y}) = \alpha_2(\boldsymbol{y}) \times \text{Gamma}(x_2;a_{2,1},b_{2,1}) + \big(1-\alpha_2(\boldsymbol{y})\big) \times \text{Gamma}(x_2;a_{2,2},b_{2,2}). \tag{3.6}$$

Interestingly, because of the symmetry between gene A and gene B, simulation data show that the forms of $\psi_{2,j}(y_2)$ and $\gamma_{2,j}(y_2)$ are the same as those of $\psi_{1,j}(y_1)$ and $\gamma_{1,j}(y_1)$, respectively. Therefore, we have

$$\begin{cases} \psi_{1,1}(y_2) = 0.64 + 0.22 \times \dfrac{y_2^2}{0.22^2 + y_2^2}, \\ \gamma_{1,1}(y_2) = 10.38 + 20.46 \times \dfrac{y_2^4}{0.38^4 + y_2^4}, \\ \psi_{1,2}(y_2) = 0.20 + 0.18 \times \dfrac{y_2^8}{0.64^8 + y_2^8}, \\ \gamma_{1,2}(y_2) = 2.95 + 0.15 \times \dfrac{1}{0.42^4 + y_2^4}. \end{cases} \tag{3.7}$$

Additionally, we define the parameters $a_{2,j}(y)$ and $b_{2,j}(y)$ as follows:

$$a_{2,j}(y_2) = \psi_{2,j}(y_2), \quad b_{2,j}(y_2) = \frac{\psi_{2,j}(y_2)}{\gamma_{2,j}(y_2)}. \tag{3.8}$$

Similarly, the function $\alpha_2(\boldsymbol{y})$ takes on the same form as $\alpha_1(\boldsymbol{y})$

$$\alpha_2(\boldsymbol{y}) = 0.16 + \frac{1}{1.07^6 + y_1^6} \times \frac{y_2^6}{0.46^6 + y_2^6}. \tag{3.9}$$

The function $\alpha_2(\boldsymbol{y})$ is illustrated in Fig. 5(i).

Finally, the inheritance function is expressed as follows:

$$p(\boldsymbol{x};\boldsymbol{y}) = p_1(x_1;\boldsymbol{y}) \times p_2(x_2;\boldsymbol{y}). \tag{3.10}$$

This procedure delineates how the inheritance function in Eq. (2.8) is derived from simulation data based on the gene regulation network model presented in Eq. (2.19). Validation of the inheritance function in Eq. (3.10) is discussed in Section 3.3.

The gamma distribution density functions, denoted as $\text{Gamma}(x_i;a_{i,j}(y_i),b_{i,j}(y_i))$, defined by Eqs. (3.2), (3.6), and (3.10), describe how daughter cells inherit characteristics from mother cells for the same gene. The combination coefficients $\alpha_1(\boldsymbol{y})$ and $\alpha_2(\boldsymbol{y})$ illustrate the properties of cell type transitions regulated by the gene regulation network. Notably, $\alpha_1(\boldsymbol{y})$ increases with y_1 and decreases with y_2, indicating that gene A activates itself while inhibiting gene B's effect on gene A. A similar trend is observed with $\alpha_2(\boldsymbol{y})$. Therefore, the structure of the gene regulation network significantly influences the inheritance function through its qualitative dependence on the state of the 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 (2.8) with that from the gene regulation network model (2.19).

To begin, we utilized the cell cycle model (2.8) to replicate generating cell clones from a single cell. We then compared our findings with those in Fig. 4. To accomplish this, we defined the kinetotype in (2.8) as follows:

$$\beta=1/T_1, \quad \kappa=\mu=0, \quad \tau=T-T_1. \tag{3.11}$$

Since directly solving the high-dimensional integral in (2.8) is computationally expensive, we adopted an individual-cell-based simulation approach, following the scheme outlined in Eq. (2.8).

To replicate the process shown in Fig. 5, we initialized the system with 10^6 cells, emulating the setup depicted in Fig. 5. We then performed individual-cell-based simulations to model the cell division process. During division, each mother cell generated two daughter cells, with the epigenetic state of each daughter cell determined by the inheritance function $p(\boldsymbol{x},\boldsymbol{y})$, as defined in (3.10) (see the Appendix B for the detailed numerical scheme). Similar to the methodology used in Fig. 5, we conducted simulations over three cycles and collected data from the second and third cycles for analysis.

Fig. 6 presents data points illustrating the epigenetic states of daughter cells with those of mother cells, obtained from both the gene regulation network model (2.19) and the $G0$ cell cycle model (2.8) with the inheritance function (3.10). A comparison of Fig. 6(a) with Fig. 6(b) shows that the data-driven inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ effectively reproduces the distribution of epigenetic states across different cell cycles. Both figures reveal epigenetic states distributed at high gene A or gene B expression, consistent with the stable steady states identified through bifurcation analysis (Fig. 4). Furthermore, both

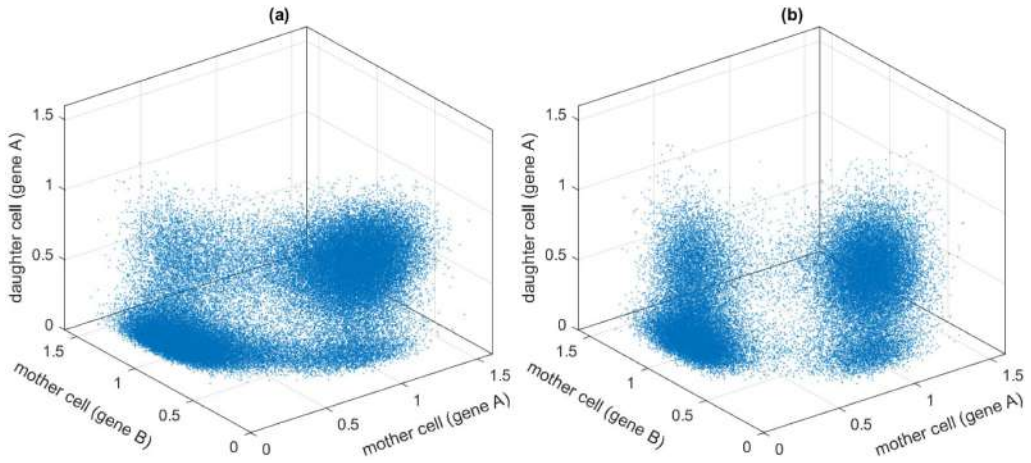

Figure 6: Comparison between the gene regulation network model and the $G0$ cell cycle model. (a). Scatter plot of daughter cell epigenetic state (gene A) versus mother cell epigenetic state (gene A and B) following the gene regulation network model (2.19). (b). Scatter plot of daughter cell epigenetic state (gene A) versus mother cell epigenetic state (gene A and B) following the $G0$ cell cycle model (2.8) with the inheritance function (3.10). Parameters were identical to those in Fig. 5, and cell cycling parameters for the $G0$ cell cycle model are given by (3.11).

simulations display scatter plots with high densities at states corresponding to either low gene A expressions in mother and daughter cells or high gene A expressions in mother and daughter cells. The epigenetic state distributions derived from both simulation approaches are closely aligned.

Next, we applied the $G0$ cell cycle model to investigate the biological process underlying heterogeneous cell growth, incorporating additional regulatory mechanisms for proliferation and differentiation. To achieve this, we defined the parameters as follows:

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

and

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

with $\beta_0 = 0.2$ and $\theta = 10^6$. Here, the regulation of proliferation depends on the total stem cell number, which is a challenging interaction to model with the stochastic dynamical equation (2.19). We simulated the cell growth process using the $G0$ cell cycle model (2.8) with the inheritance function (3.10). We initialized 10^4 cells in our simulations, randomly assigning epigenetic states to each cell. We used an individual-cell-based simulation to model cell growth up to $t = 2500$ (approximately 50 cycles).

In this scenario, we introduced nonzero rates of differentiation and apoptosis and assumed that the proliferation rate would decrease as the total cell number increased. As a result, the total cell number reached an equilibrium state after an extended period of growth (Fig. 7(a)). The system automatically reaches homeostasis with a stable cell number due to the negative feedback on cell proliferation. Fig. 7(b) shows the distribution of the epigenetic state of the cells at $t = 2000$ (indicated by the dashed line in Fig. 7(a)). Notably, two distinct cell subtypes emerge, with their epigenetic state distributions closely resembling those obtained from the gene regulation network model shown in Fig. 4. This result suggests that independent of population dynamics, the gene regulatory network determines distinct cell types.

We also examined the dynamics of transitions between cell types. Figs. 7(c) and 7(d) illustrates genes A and B's evolutionary dynamics while tracking a single cell throughout the growth process. The simulations demonstrate that cells can transition between the two subtypes as the growth continues. Both stochastic perturbation and cell plasticity during cell division may contribute to the dynamics of cell type transitions. Moreover, the cell states approach a homeostasis distribution despite these transitions, as shown in Fig. 7(b). These results highlight the $G0$ cell cycle model's ability to simulate the long-term dynamics of cell growth with heterogeneity.

4 Discussion and perspectives

Population dynamics and molecular-level variations occur on distinct spatial and temporal scales during stem cell regeneration and tissue development. Integrating the dynamics of gene regulation within individual cells with the population dynamics of stem cell

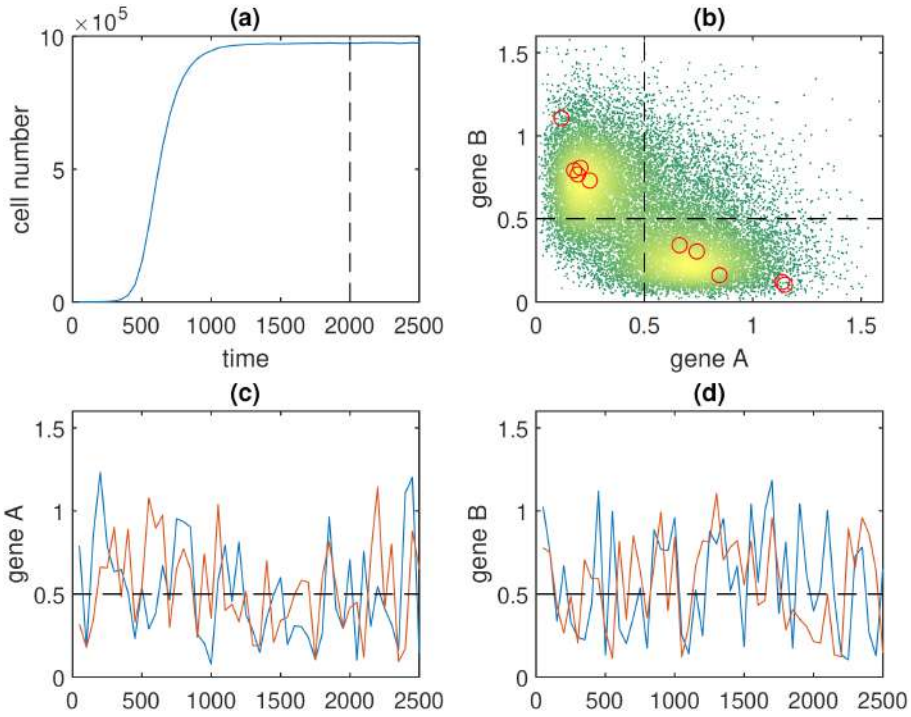

Figure 7: Heterogeneous cell growth. (a). Dynamics of cell number evolution. (b). Scatter plot of the epigenetic state of cells at $t=2000$ (dashed line in (a)). Red circles show the initial states of 10 cells. (c). Evolution of the epigenetic state of gene A in two cells during the cell growth process. (d). Evolution of the epigenetic state of gene B in two cells during the cell growth process.

regeneration remains a significant challenge. In this study, we developed a method to formulate the inheritance function for modeling stem cell regeneration dynamics based on data from cross-cell-cycle gene regulation dynamics. The regeneration dynamics of stem cells are represented using a differential-integral equation, where the inheritance function is introduced to describe cell plasticity during cell division. We demonstrated that the inheritance function can be expressed as a combined gamma distribution in Eq. (3.10), with gene circuit topology incorporated into the coefficients. Our simulations using the *G*0 cell cycle model revealed intriguing insights into cell growth dynamics with heterogeneity. The model successfully reproduced the emergence of distinct cell subtypes and their transitions over time, indicating that the heterogeneity in cell type can be attributed to underlying regulatory mechanisms within the cell cycle.

The gene circuit analyzed in this study consists of two master transcription factors (TFs). These TFs display self-activation and mutual repression, as illustrated in Fig. 3. Consequently, the inheritance function can be represented by a combined gamma distribution

$$p(\boldsymbol{x},\boldsymbol{y})=\prod_{i=1}^{m}p_i(x_i;\boldsymbol{y}), \tag{4.1}$$

where

$$p_i(x_i;\boldsymbol{y}) = \alpha_i(\boldsymbol{y})\mathrm{Gamma}(x_i;a_{i,1},b_{i,1}) + (1-\alpha_i(\boldsymbol{y}))\mathrm{Gamma}(x_i;a_{i,2},b_{i,2}). \tag{4.2}$$

The gene circuit topology influences the coefficients $\alpha_i(\boldsymbol{y})$, such that the sign of $\partial\alpha_i(\boldsymbol{y})/\partial y_j$ (positive or negative) indicates whether gene j activates or represses gene i. Furthermore, the coefficients $a_{i,j}$ and $b_{i,j}$ in (4.2) can be expressed using Hill-type functions so that the conditional expectation $E(x_i|\boldsymbol{y}) = \psi_i(\boldsymbol{y})$ increases as the component y_i rises. This observation provides insights for developing the mathematical formulation in Eqs. (4.1) and (4.2) for more complex gene regulatory networks.

The dynamics of gene expression can be modeled using ordinary differential equations as follows:

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{F}(\boldsymbol{X}) - \boldsymbol{K}\boldsymbol{X}, \tag{4.3}$$

where $\boldsymbol{K} = \mathrm{diag}(k_1,k_2,\cdots,k_n)$. The topology of the gene circuit is expressed through the nonlinear function $\boldsymbol{F}$. Following the scheme in this study, the gene expression dynamics can be extended to incorporate random perturbations to model parameters and the dynamics across different cell cycles. Additionally, our study suggests an epigenetic state defined as $\boldsymbol{x} = \log(\boldsymbol{X}(t)+1)|_{\ell(t)=\ell^*}$, which is associated with the gene expression level during a particular cell cycle. The population dynamics of stem cell regeneration are described with the differential-integral equation

$$\begin{cases} \dfrac{\partial Q(t,\boldsymbol{x})}{\partial t} = -Q(t,\boldsymbol{x})\big(\beta(c(t),\boldsymbol{x}) + \kappa(\boldsymbol{x})\big) \\ \qquad\qquad +2\displaystyle\int_\Omega \beta\big(c\big(t-\tau(\boldsymbol{y})\big),\boldsymbol{y}\big)Q\big(t-\tau(\boldsymbol{y}),\boldsymbol{y}\big)e^{-\mu(\boldsymbol{y})\tau(\boldsymbol{y})}p(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \\ c(t) = \displaystyle\int_\Omega Q(t,\boldsymbol{x})\zeta(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \end{cases} \tag{4.4}$$

The inheritance function $p(\boldsymbol{x},\boldsymbol{y})$ quantifies the variance between daughter cells and their mother cell following cell division. In our study, we propose a mathematical formulation for $p(\boldsymbol{x},\boldsymbol{y})$ as shown in (4.1)-(4.2). The coefficients $\alpha_i(\boldsymbol{y})$ in this formulation account for the gene circuit topology, satisfying the rule

$$\mathrm{sign}\left(\frac{\partial\alpha_i(\boldsymbol{y})}{\partial y_j}\right) = \mathrm{sign}\left(\frac{\partial F_i(\boldsymbol{X})}{\partial X_j}\right). \tag{4.5}$$

The rule (4.5) links gene regulation dynamics with stem cell regeneration dynamics.

Further advancements in the framework proposed in this study could facilitate the seamless integration of data-driven and model-driven approaches, effectively connecting single-cell data with population-level dynamics (Fig. 8).

Firstly, reconstructing gene regulatory networks from gene expression data remains a significant challenge in systems biology and bioinformatics. The emergence of single-cell RNA sequencing (scRNA-seq) data has spurred the development of various methods for reconstructing GRN from gene expression profiles [21, 26, 59, 61].

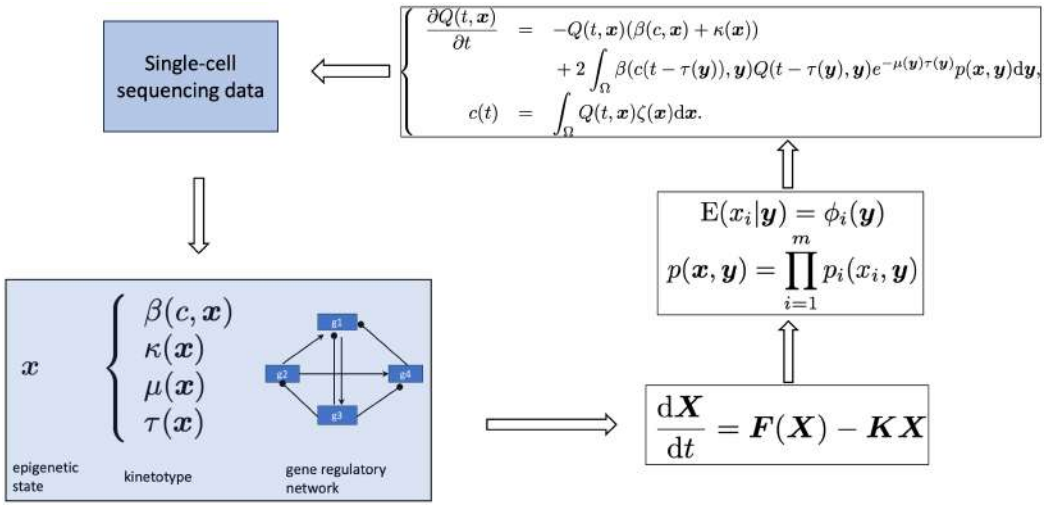

Figure 8: An integrative modeling framework connecting scRNA-seq data to heterogeneous stem cell regeneration dynamics.

Secondly, the concept of the kinetotype, represented by the quadruple

$$\big(\beta(c,\boldsymbol{x}),\kappa(\boldsymbol{x}),\mu(\boldsymbol{x}),\tau(\boldsymbol{x})\big)$$

in Eq. (4.4), is a crucial link between cell regeneration kinetics and scRNA-seq data. Several methodologies have been developed to extract cellular properties from scRNA-seq data. These include quantifying the pluripotency landscape of cell differentiation [48], identifying stemness [22,37], measuring the order of cellular transcriptome profiles using single-cell entropy [34, 57], and evaluating gene signature scores associated with signaling pathways [1, 56]. Improvements to these methods may allow for precise quantification of kinetotypes from scRNA-seq data.

Fig. 8 illustrates a vision for integrative modeling that connects scRNA-seq-data to population dynamics. Single-cell sequencing data can serve as a foundation for inferring kinetotypes and reconstructing gene regulatory networks. The structure of these networks is essential for understanding gene expression dynamics and defining the inheritance function. Together, kinetotype and inheritance function concepts allow for the modeling of evolutionary dynamics in heterogeneous stem cell regeneration. This approach integrates molecular-level processes with population-level dynamics over long-term development timescales, providing insights that connect data analysis with biological interpretation.

AI-driven tools are expected to be transformative in future studies by leveraging their capabilities for data-driven predictions and function discovery. For instance, AI-based methods can be applied to reconstruct kinetotypes and GRNs, while neural networks could represent the inheritance function. Additionally, AI methods might improve numerical techniques for solving the differential equations that govern GRNs and stem cell regeneration dynamics, further advancing the field.

Appendix A. Numerical scheme for the hybrid kinetic modeling

We applied the numerical scheme to solve the hybrid model (2.19) by combining the numerical methods for stochastic differential equations with the random process involved in cell division. During the numerical simulation, the number of cells changes over time due to cell division. Each cell corresponds to a stochastic differential equation defined by the gene regulatory network. Therefore, tracking the gene expression dynamics of each cell is necessary. The numerical scheme is provided below in Algorithm 1.

Algorithm 1. Hybrid kinetic model simulation.

- **Input:** The gene regulatory network's topology file contains gene nodes, regulatory relationships, and related parameters.
- **Initialization:** Read the topology file and model parameters. Initialize the number of simulation cycles $\texttt{cycle}=0$, set the cycle length $T=50$, and set the initial time $t=0$. Set the initial cell number N. For each cell in the multicellular system $\Sigma=\{C_j(\boldsymbol{X}_j): j=1,\ldots,N\}$, set the initial condition for the gene expression level $\boldsymbol{X}_j(0)$. Here, Σ represents the system of all cells, and C_j represents each cell, with the gene expression level of the cell given by $\boldsymbol{X}_j$.
- **Simulation process:**

for $\texttt{cycle}=0$ to $\texttt{max\_cycle}$ **do**
  Copy the system $\Sigma'=\Sigma$.
  for all cells $C_j\in\Sigma$ **do**
    Gene regulation dynamics. Solve the stochastic differential equation (2.19) with the stochastic Runge-Kutta method for $0\le t\le T$.
    Cell division. Generate two new cells $C_{j,1}(\boldsymbol{Y}_1)$ and $C_{j,2}(\boldsymbol{Y}_2)$, $\boldsymbol{Y}_l=(Y_{l,1},\cdots,Y_{l,m})$ $(l=1,2)$ 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 $Y_{l,i}(0)=\chi_i X_{j,i}(T)$.
    end for
    Replace the cell C_j in Σ' with the new cells $C_{j,1}$ and $C_{j,2}$.
  end for
  $\texttt{cycle}=\texttt{cycle}+1$.
  Copy the system $\Sigma=\Sigma'$.
end for

The two daughter cells are generated following the same rule through $\chi_i\sim\text{Beta}(a,b)$ in the above numerical scheme. This is a simplification of the actual situation. In a real biological system, the two daughter cells can be correlated, especially in the case of asymmetric division. This simplification does not affect the results if we track a single cell over a long duration. However, it may affect the results when considering the dynamics of

a cell population. In our study, we set the parameter $\eta = 60$ to represent symmetric division and reduce the correlation between two daughter cells. This simulation does not alter the primary approach of our study, which connects gene regulation dynamics with stem cell regeneration dynamics. For a more realistic scenario where the two daughter cells are correlated during cell division, further systematic discussions should be conducted in future research, which is beyond the scope of the current study.

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

The differential-integral equation (2.8) can be solved numerically using the Euler method and numerical integration. However, calculating numerical integration is computationally expensive due to the high dimensionality of the epigenetic states. Therefore, we typically do not directly solve the Eq. (2.8). Instead, we employ an individual-cell-based simulation method to model the cell cycling process.

In model simulations, a system of multiple cells is represented as a collection of epigenetic states. The individual-cell-based simulation tracks the behaviors of each cell based on its specific epigenetic state. Below is a sketch of the numerical scheme used in this approach (Algorithm 2).

To assess the epigenetic state of daughter cells based on the inheritance probability $p(\boldsymbol{x},\boldsymbol{y})$, defined by the combination gamma distributions in Eq. (3.10), we implemented the numerical scheme (see Algorithm 3).

In simulations, the number of cells can increase significantly due to cell division, which may present challenges during the simulation process. To address this issue, we employed a technique called downsampling. This process is commonly used in signal analysis to reduce data rates or the overall data size. Similarly, we utilized downsampling to decrease the number of cells being simulated.

In the current simulation, we define a maximum number of cells to be simulated and stored, denoted as $N_{\max} = 10^6$ cells. Let $N_k = N_{k,q} + N_{k,q}$ (where $N_k \le N_{\max}$) represent the number of cells under simulation at step k. Here, $N_{k,q}$ indicates the number of cells in the resting phase, and $N_{k,p}$ indicates the number of cells in the proliferative phase.

Initially, we allocate temporary storage space for $N_{k,q} + 2N_{k,p}$ cells, which accounts for the maximum possible number of cells if all cells in the proliferative phase undergo mitosis. After performing cell fate decisions for each cell, we obtain a potentially new number of cells, denoted as $N_{\text{temp}} = N'_{k,q} + N'_{k,p}$ (where $N_{\text{temp}} \le N_{k,q} + 2N_{k,p}$).

If $N_{\text{temp}} > N_{\max}$, we will select all $N'_{k,p}$ cells in the proliferative phase and at most $(N_{\max} - N'_{k,q})$ cells in the resting phase. The selection from the resting phase is made by choosing each cell with a probability $p = \max\{1, (N_{\max} - N'_{k,q})/N'_{k,q}\}$ to yield N_{next} cells for the next step simulation. Otherwise, if $N_{\text{temp}} \le N_{\max}$, we select all N_{temp} cells, setting $N_{\text{next}} = N_{\text{temp}}$.

Algorithm 2. Individual-cell based simulation.

- **Input:** The parameters, kinetotype functions β,κ,μ, and τ, inheritance function $p(\boldsymbol{x},\boldsymbol{y})$, and the time step Δt.
- **Initialization:** Set the time $t=0$, the initial cell number N_0, and all cells $\Sigma=\{[C_i(\boldsymbol{x})]_{i=1}^{Q}\}$. 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 $\mathtt{T}_{\mathtt{end}}$ 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 resting, 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 $p(\boldsymbol{x},\boldsymbol{y})$. 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

Algorithm 3. Epigenetic state inheritance.

Calculates the values of $\alpha_i, \psi_{i,j}, \gamma_{i,j}$ 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 $q=\texttt{rand}()$.
- If $q<\alpha_i$, produce the epigenetic state of gene i for the 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 $\text{Gamma}(x_i;a_{i,2},b_{i,2})$.

end for

Update the epigenetic state of the daughter cells.

Finally, we stored the state of all selected N_{next} cells, calculate the proliferation rate as $f_{\text{pro},k}=N_{\text{temp}}/N_k$, and set $N_{k+1}=N_{\text{next}}$ for the number of cells to be simulated at step $k+1$.

In the downsampling simulation approach, at step k, there are $N_k \le N_{\max}$ cells being simulated, and the states of these cells are recorded. The actual number of cells at this step is given by the formula

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

This equation indicates calculating the number of cells at step k.

Appendix C. Supplementary figures

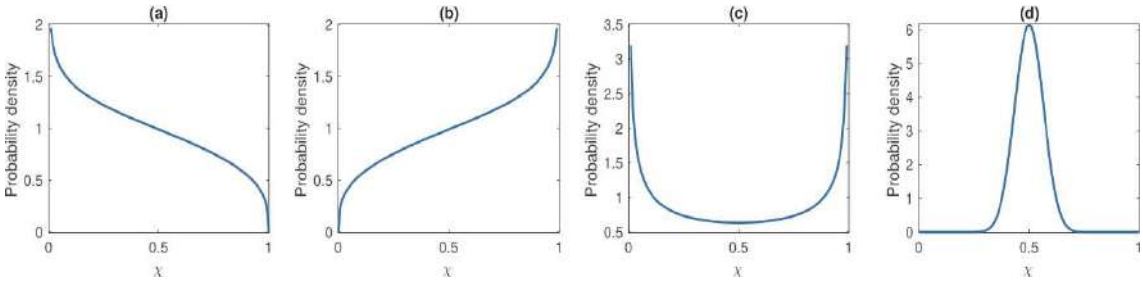

Figure 9: Probability density function of the beta distribution. (a). $\phi=0.4, \eta=2.2$. (b). $\phi=0.6, \eta=2.2$. (c). $\phi=0.5, \eta=1.0$. (d). $\phi=0.5, \eta=60$.

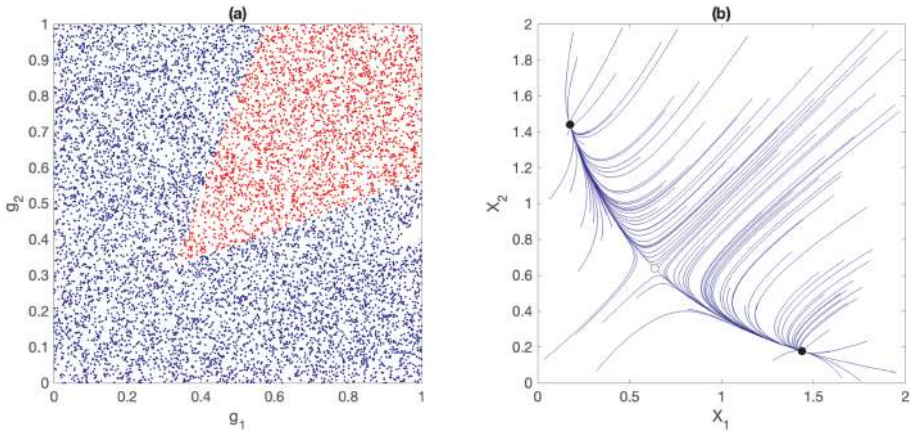

Figure 10: Bifurcation analysis of gene expression kinetics. (a). Dependence of the number of steady states of the ODE model (3.1) with randomly selected parameters (g_1,g_2). Blue dots represent one steady state, while red dots represent three steady states. (b). Trajectories for the ODE model (3.1) 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.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (Grant No. 12331018).

References

[1] M. Andreatta and S. J. Carmona, *UCell: Robust and scalable single-cell gene signature scoring*, Comput. Struct. Biotechnol. J., 19:3796–3798, 2021.
[2] D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, and M. L. Simpson, *Gene network shaping of inherent noise spectra*, Nature, 439:608–611, 2006.
[3] S. Bernard, J. Bélair, and M. C. Mackey, *Oscillations in cyclical neutropenia: New evidence based on mathematical modeling*, J. Theor. Biol., 223:283–298, 2003.
[4] F. J. Burns and I. F. Tannock, *On the existence of a G0-phase in the cell cycle*, Cell Prolif., 3:321–334, 1970.
[5] C. Cadart, S. Monnier, J. Grilli, P. J. Sáez, N. Srivastava, R. Attia, E. Terriac, B. Baum, M. Cosentino-Lagomarsino, and M. Piel, *Size control in mammalian cells involves modulation of both growth rate and cell cycle duration*, Nat. Commun., 9(1):3275, 2018.
[6] L. Cai, N. Friedman, and X. S. Xie, *Stochastic protein expression in individual cells at the single molecule level*, Nature, 440:358–362, 2006.
[7] W. Chen and A. E. Teschendorff, *Estimating dfferentiation potency of single cells using single-cell entropy (SCENT)*, in: Computational Methods for Single-Cell Data Analysis, Methods in Molecular Biology, Vol. 1935, Humana Press, 125–139, 2019.
[8] H. Clevers, *What is an adult stem cell?*, Science, 350:1319–1320, 2015.
[9] S. Dalton, *Linking the cell cycle to cell fate decisions*, Trends Cell Biol., 25:592–600, 2015.

[10] T. S. Deisboeck, Z. Wang, P. Macklin, and V. Cristini, *Multiscale cancer modeling*, Annu. Rev. Biomed. Eng., 13:127–155, 2011.
[11] A. Doncic, M. Falleur-Fettig, and J. M. Skotheim, *Distinct interactions select and maintain a specific cell fate*, Mol. Cell, 43:528–539, 2011.
[12] R. Eftimie, J. J. Gillard, and D. A. Cantrell, *Mathematical models for immunology: Current state of the art and future research directions*, Bull. Math. Biol., 78:2091–2134, 2016.
[13] D. S. Fischer, A. K. Fiedler, E. M. Kernfeld, R. M. J. Genga, A. Bastidas-Ponce, M. Bakhti, H. Lickert, J. Hasenauer, R. Maehr, and F. J. Theis, *Inferring population dynamics from single-cell RNA-sequencing time series data*, Nat. Biotechnol., 37:461–468, 2019.
[14] M. B. Ginzberg, R. Kafri, and M. Kirschner, *Cell biology. On being the right (cell) size*, Science, 348:1245075, 2015.
[15] Y. Guo, Q. Nie, A. L. MacLean, Y. Li, J. Lei, and S. Li, *Multiscale modeling of inflammation-induced tumorigenesis reveals competing oncogenic and oncoprotective roles for inflammation*, Cancer Res., 77:6429–6441, 2017.
[16] B. Huang, D. Jia, J. Feng, H. Levine, J. N. Onuchic, and M. Lu, *RACIPE: A computational tool for modeling gene regulatory circuits using randomization*, BMC Syst. Biol., 12:74, 2018.
[17] B. Huang, M. Lu, D. Jia, E. Ben-Jacob, H. Levine, and J. N. Onuchic, *Interrogating the topological robustness of gene regulatory circuits by randomization*, PLoS Comput. Biol., 13:e1005456, 2017.
[18] R. Huang and J. Lei, *Dynamics of gene expression with positive feedback to histone modifications at bivalent domains*, Int. J. Mod. Phys. B, 32:1850075, 2018.
[19] R. Huang and J. Lei, *Cell-type switches induced by stochastic histone modification inheritance*, Discrete Contin. Dyn. Syst. Ser. B, 24:5601–5619, 2019.
[20] R. Huang, Q. Situ, and J. Lei, *Dynamics of cell-type transition mediated by epigenetic modifications*, J. Theor. Biol., 577:111664, 2024.
[21] X. Huang, C. Song, G. Zhang, Y. Li, Y. Zhao, Q. Zhang, Y. Zhang, S. Fan, J. Zhao, L. Xie, and C. Li, *scGRN: A comprehensive single-cell gene regulatory network platform of human and mouse*, Nucleic Acids Res., 52:D293–D303, 2024.
[22] H. Jiang, J. Liu, Y. Song, and J. Lei, *Quantitative modeling of stemness in single-cell RNA sequencing data: A nonlinear one-class support vector machine method*, J. Comput. Biol., 31:41–57, 2024.
[23] M. K. Jolly, D. Jia, M. Boareto, S. A. Mani, K. J. Pienta, E. Ben-Jacob, and H. Levine, *Coupling the modules of EMT and stemness: A tunable 'stemness window' model*, Oncotarget, 6:25161–25174, 2015.
[24] R. Kafri, J. Levy, M. B. Ginzberg, S. Oh, G. Lahav, and M. W. Kirschner, *Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle*, Nature, 494:480–483, 2013.
[25] G. Karlebach and R. Shamir, *Modelling and analysis of gene regulatory networks.*, Nat. Rev. Mol. Cell Biol., 9:770–780, 2008.
[26] H. Kim, H. Choi, D. Lee, and J. Kim, *A review on gene regulatory network reconstruction algorithms based on single cell RNA sequencing*, Genes Genom., 46:1–11, 2024.
[27] J. Lei, *Stochastic modeling in systems biology*, arXiv:1104.4524v1, 2011.
[28] J. Lei, *A general mathematical framework for understanding the behavior of heterogeneous stem cell regeneration*, J. Theor. Biol., 492:110196, 2020.
[29] J. Lei, *Evolutionary dynamics of cancer: From epigenetic regulation to cell population dynamics – mathematical model framework, applications, and open problems*, Sci. China Math., 63:411–424, 2020.
[30] J. Lei, *Systems Biology: Modeling, Analysis, and Simulation*, Springer, 2021.

[31] J. Lei, G. He, H. Liu, and Q. Nie, *A delay model for noise-induced bi-directional switching*, Nonlinearity, 22:2845–2859, 2009.
[32] J. Lei, S. A. Levin, and Q. Nie, *Mathematical model of adult stem cell regeneration with cross-talk between genetic and epigenetic regulation*, Proc. Natl. Acad. Sci. USA, 111:E880–E887, 2014.
[33] J. Lei and M. C. Mackey, *Multistability in an age-structured model of hematopoiesis: Cyclical neutropenia*, J. Theor. Biol., 270:143–153, 2011.
[34] J. Liu, Y. Song, and J. Lei, *Single-cell entropy to quantify the cellular order parameter from single-cell RNA-seq data*, Biophys. Rev. Lett., 15:35–49, 2020.
[35] S. Ma, J. Lei, and X. Lai, *Modeling tumour heterogeneity of PD-L1 expression in tumour progression and adaptive therapy*, J. Math. Biol., 86(3):38, 2023.
[36] M. C. Mackey, *Unified hypothesis for the origin of aplastic anemia and periodic hematopoiesis*, Blood, 51:941–956, 1978.
[37] T. M. Malta et al., *Machine learning identifies stemness features associated with oncogenic dedifferentiation*, Cell, 173:338–354, 2018.
[38] A. Masoudi-Nejad, G. Bidkhori, S. Hosseini Ashtiani, A. Najafi, J. H. Bozorgmehr, and E. Wang, *Cancer systems biology and modeling: Microscopic scale and multiscale approaches*, Semin. Cancer Biol., 30:60–69, 2015.
[39] A. Moustakas, K. Pardali, A. Gaal, and C. H. Heldin, *Mechanisms of TGF-β signaling in regulation of cell growth and differentiation*, Immunol. Lett., 82:85–91, 2002.
[40] L. E. O'Brien, *Tissue homeostasis and non-homeostasis: From cell life cycles to organ states*, Annu. Rev. Cell Dev. Biol., 38:395–418, 2022.
[41] D. M. Ornitz and N. Itoh, *Fibroblast growth factors*, Genome Biol., 2:3005.1–3005.12, 2001.
[42] A. V. Probst, E. Dunleavy, and G. Almouzni, *Epigenetic inheritance during the cell cycle*, Nat. Rev. Mol. Cell Biol., 10:192–206, 2009.
[43] K. Ragkousi and M. C. Gibson, *Cell division and the maintenance of epithelial order*, J. Cell Biol., 207:181–188, 2014.
[44] P. Rué and A. Martinez Arias, *Cell dynamics and gene expression control in tissue homeostasis and development*, Mol. Syst. Biol., 11(1):792, 2015.
[45] S. Sahoo, A. Mishra, H. Kaur, K. Hari, S. Muralidharan, S. Mandal, and 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):zcab027, 2021.
[46] T. Schepeler, M. E. Page, and K. B. Jensen, *Heterogeneity and plasticity of epidermal stem cells*, Development, 141:2559–2567, 2014.
[47] A. Serra-Cardona and Z. Zhang, *Replication-coupled nucleosome assembly in the passage of epigenetic information and cell identity*, Trends Biochem. Sci., 43:136–148, 2018.
[48] J. Shi, T. Li, L. Chen, and K. Aihara, *Quantifying pluripotency landscape of cell differentiation from scRNA-seq data by continuous birth-death process*, PLoS Comput. Biol., 15:e1007488, 2019.
[49] Z. S. Singer, J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz, *Dynamic heterogeneity and DNA methylation in embryonic stem cells*, Mol. Cell, 55:319–331, 2014.
[50] A. Soufi and S. Dalton, *Cycling through developmental decisions: How cell cycle dynamics control pluripotency, differentiation and reprogramming*, Development, 143:4301–4311, 2016.
[51] K. Takaoka and H. Hamada, *Origin of cellular asymmetries in the pre-implantation mouse embryo: A hypothesis*, Philos. Trans. R. Soc. Lond. B, 369:20130536, 2014.
[52] J. J. Tyson, T. Laomettachit, and P. Kraikivski, *Modeling the dynamic behavior of biochemical regulatory networks*, J. Theor. Biol., 462:514–527, 2019.
[53] H. Wu and Y. Zhang, *Reversing DNA methylation: Mechanisms, genomics, and biological func-*

tions, Cell, 156:45–68, 2014.

[54] L. Xu, K. Zhang, and J. Wang, *Exploring the mechanisms of differentiation, dedifferentiation, reprogramming and transdifferentiation*, PLoS One, 9:e105216, 2014.

[55] L. Yang, Y. L. Pang, and H. L. Moses, *TGF-β and immune cells: An important regulatory axis in the tumor microenvironment and progression*, Trends Immunol., 31:220–227, 2010.

[56] Z. Yao, S. Jin, F. Zhou, J. Wang, K. Wang, and X. Zou, *A novel multiscale framework for delineating cancer evolution from subclonal compositions*, J. Theor. Biol., 582:111743, 2024.

[57] Y. Ye, Z. Yang, M. Zhu, and J. Lei, *Using single-cell entropy to describe the dynamics of reprogramming and differentiation of induced pluripotent stem cells*, Int. J. Mod. Phys. B, 34:2050288, 2020.

[58] E. Zatulovskiy and J. M. Skotheim, *On the molecular mechanisms regulating animal cell size homeostasis*, Trends Genet., 36:360–372, 2020.

[59] Y. Zeng, Y. He, R. Zheng, and M. Li, *Inferring single-cell gene regulatory network by non-redundant mutual information*, Brief. Bioinform., 24(5):bbad326, 2023.

[60] C. Zhang, C. Shao, X. Jiao, Y. Bai, M. Li, H. Shi, J. Lei, and X. Zhong, *Individual cell-based modeling of tumor cell plasticity-induced immune escape after CAR-T therapy*, Comput. Syst. Oncol., 1:e21029, 2021.

[61] J. Zhao, C.-W. Wong, W.-K. Ching, and X. Cheng, *NG-SEM: An effective non-Gaussian structural equation modeling framework for gene regulatory network inference from single-cell RNA-seq data*, Brief. Bioinform., 24(6):bbad369, 2023.