

Incorporating stochastic gene expression, signaling-mediated intercellular interactions, and regulated cell proliferation in models of coordinated tissue development

Casey O. Barkan*

Department of Physics & Astronomy, University of California, Los Angeles, CA 90095-1766 USA

Tom Chou†

Department of Mathematics, University of California, Los Angeles, CA 90095-1555 USA

Formulating quantitative and predictive models for tissue development requires consideration of the complex, stochastic gene expression dynamics, its regulation via cell-to-cell interactions, and cell proliferation. Including all of these processes into a practical mathematical framework requires complex expressions that are difficult to interpret and apply. We construct a simple theory that incorporates intracellular stochastic gene expression dynamics, signaling chemicals that influence these dynamics and mediate cell-cell interactions, and cell proliferation and its accompanying differentiation. Cellular states (genetic and epigenetic) are described by a Waddington vector field that allows for non-gradient dynamics (cycles, entropy production, loss of detailed balance) which is precluded in Waddington potential landscape representations of gene expression dynamics. We define an epigenetic fitness landscape that describes the proliferation of different cell types, and elucidate how this fitness landscape is related to Waddington’s vector field. We illustrate the applicability of our framework by analyzing two model systems: an interacting two-gene differentiation process and a spatiotemporal organism model inspired by planaria.

I. INTRODUCTION

Cell and tissue development is a complex and multifaceted dynamical process, involving gene regulatory dynamics, cell-to-cell interactions, and differential cell proliferation and death rates. The conceptual model of Waddington’s landscape has been a guiding paradigm for understanding the mechanisms of cell development [1–3]. Waddington’s landscape suggests that the internal dynamics of a differentiating cell follow that of a particle in a rugged high-dimensional “potential” landscape. However, the internal dynamics of cell states (*e.g.*, gene expression profiles) are only one aspect of tissue development. Proliferation and death of individual cells give rise to population dynamics akin to that seen under Darwinian evolution [4, 5]. Additionally, intracellular gene dynamics and population-level birth-death processes are coupled through cell-to-cell signaling and metabolic interactions [5, 6]. Thus, a complete description of development must describe an interacting population of cells, not just the gene dynamics within a single cell.

The conceptual framework of Waddington’s landscape has been translated into a variety of mathematical models. For example, Waddington’s landscape has been interpreted as a potential landscape acting on single-cell gene expression [3, 6–11] and on population level phenotypes [12]. In the typical single-cell interpretation, a cell’s state is modeled as a particle with position \mathbf{x} in a potential energy landscape $U(\mathbf{x})$ in a noisy environment. This model implies that one can reconstruct $U(\mathbf{x})$ from data using the simple relationship $U(\mathbf{x}) = -\log p_{\text{ss}}(\mathbf{x})$, where $p_{\text{ss}}(\mathbf{x})$

is the steady-state distribution of states empirically measured. Such formulae have been applied to real datasets [13] but its use is hampered by its inherent limitation to a single cell (or to non-interacting and non-proliferating populations) and the assumption of both gradient dynamics *and* dynamics that obey detailed balance among gene regulation states.

When cell division and death are considered, the concept of a fitness landscape becomes relevant. In evolutionary population dynamics, high fitness genotypes elicit high proliferation rates; this concept must be adapted to the context of development, where all cells have the same genotype. Mathematical models of development often consider a single-cell fates and state changes [3, 8, 13–22]; those that incorporate cell population dynamics through birth and death typically neglect cell-cell interactions [23–28]. In kinetic theories that combine internal cell state dynamics with state-dependent division and death rates, tractable dynamical equations describing population-level quantities can be derived if there are no cell-cell interactions [29–32]. However, in developing tissue, cell division and death are tightly coordinated through cell-cell interactions mediated by secreted chemical signals that influence gene expression in other cells [33, 34]. Failure of this regulated tissue development can lead to pathologies such as cancer [35, 36]. Thus, effective theoretical approaches that incorporate regulated intracellular gene expression with population-level birth-death processes will provide critical tools for modeling and analyzing tissue development and evolution.

In this paper, we present a general framework for modeling a general developmental process in which cell-cell interactions regulate gene expression and cell proliferation. Our framework describes a population of individual (discrete) cells, rather than assuming an infinite continuum of cells, as has been done in other works [5, 6, 25, 26]. In

* barkanc@ucla.edu
† tomchou@ucla.edu

the tissue development context, our framework also includes spatial structure, enabling models of coordinated tissue development in which a small initial population of stem cells robustly develops into a stable and self-healing tissue. Our approach incorporates all known important processes in an interpretable way, allowing us to better define the concepts of the Waddington landscape and fitness landscape in the context of development. Rather than use a Waddington landscape that assumes gradient dynamics, we use a *Waddington vector field* to account for non-gradient dynamics of gene expression. Such detailed-balance-violating stochastic dynamics is expected in an energy-consuming process such as cell state regulation and homeostasis.

We also define an effective *epigenetic fitness landscape*. Whereas a fitness landscape describes differential proliferation of different genotypes, our epigenetic fitness landscape describes the differential proliferation of cells of the same genotype type but differing epigenetic state, capturing the fact that different attractors of the gene regulatory dynamics induce different cell division and death rates. In addition, we show how our general model can be reduced to simpler forms by applying common assumptions. Often, such assumptions are implicitly included in models of development, leading to confusion regarding the realm of applicability of such models. Our general framework provides a starting point for deriving simpler, interpretable models by applying explicitly-stated assumptions. We present two models of the robust development of a stable tissue population as proof-of-concept examples. The first is a well-mixed model of gene-regulated stem cell differentiation, and the second is a spatiotemporal model of a regenerating organism, similar to planaria.

II. MATHEMATICAL FRAMEWORK

Here, we assemble all the relevant physical processes that are understood to play roles in a developing tissue. The dependencies (and feedback) across scales and among these mechanisms are clearly delineated by proper definitions and judicious approximations/assumptions.

A. Intracellular state dynamics

Let $\mathbf{z} \in \mathbb{R}^{d_z}$ denote the internal state of a cell. \mathbf{z} specifies the state of the relevant gene regulatory network(s) in the cell, and, in principle, could also specify the methylation of DNA, chromatin structure, spatial organization of organelles, and/or any other relevant variables. While a complete specification of the gene regulatory network state requires specification of both protein and mRNA concentrations within the cell, the relatively fast timescales of mRNA transcription and degradation compared to protein translation and degradation can justify a quasi-steady state approximation where only pro-

tein concentrations need be specified. We label cells by a superscript α , *i.e.*, \mathbf{z}^α is the state of cell α , where $\alpha = 1, \dots, N(t)$, and $N(t)$ is the total number of cells which varies from cell division and death [37].

To allow for spatial resolution, let $\mathbf{r}^\alpha \in \mathbb{R}^3$ denote cell α 's position in 3-dimensional space. Interactions between cells are typically mediated by signaling molecules; suppose there are L relevant signaling molecules and let $c_\ell(\mathbf{r}, t)$ denote the concentration of signaling molecule ℓ ($\ell = 1, \dots, L$) at position \mathbf{r} at time t . For notational convenience, let $\mathbf{c}(\mathbf{r}, t) = (c_1(\mathbf{r}, t), \dots, c_L(\mathbf{r}, t))$ denote the list of signaling molecule concentrations. For cell α at position \mathbf{r}^α , the dynamics of its state $\mathbf{z}^\alpha(t)$ are modeled by a stochastic differential equation (SDE):

$$d\mathbf{z}^\alpha = \mathbf{F}(\mathbf{z}^\alpha; \mathbf{c}^\alpha)dt + \sigma(\mathbf{z}^\alpha; \mathbf{c}^\alpha)d\mathbf{B}^\alpha, \quad (1)$$

where $\mathbf{c}^\alpha \equiv \mathbf{c}(\mathbf{r}^\alpha, t)$ are the signaling concentrations experienced by cell α at time t . $\mathbf{F}(\mathbf{z}; \mathbf{c})$ defines a \mathbf{c} -dependent vector field that maps $\mathbb{R}^{d_z} \rightarrow \mathbb{R}^{d_z}$ and which describes the deterministic component of the cell state dynamics. We call $\mathbf{F}(\mathbf{z}; \mathbf{c})$ *Waddington's vector field*, and it is presumed to have stable attractors corresponding to stable cell types. Importantly, the signaling molecule concentration \mathbf{c}^α influences the dynamics of the state \mathbf{z}^α , and the dynamics of $\mathbf{c}(\mathbf{r}, t)$ are governed by the states and positions of all cells (as described below). Hence, the signaling molecules can mediate cell-to-cell interactions, allowing the tissue to develop in a coordinated way. In the special case that $\mathbf{F}(\mathbf{z}; \mathbf{c})$ is the negative gradient of a function, $\mathbf{F}(\mathbf{z}; \mathbf{c}) = -\nabla_{\mathbf{z}}U(\mathbf{z}; \mathbf{c})$, then $U(\mathbf{z}; \mathbf{c})$ would be the traditional *Waddington landscape*. However, we do not assume that $\mathbf{F}(\mathbf{z}; \mathbf{c})$ is the gradient of any function, and in section IV we review and discuss the limitations of the traditional Waddington landscape picture. Noise is characterized by the d_b -dimensional Brownian noise $d\mathbf{B}^\alpha$ and by the amplitude $\sigma(\mathbf{z}; \mathbf{c})$, a $d_z \times d_b$ matrix which may depend on the cell state and signaling molecule concentrations that the cell experiences.

B. Cell-cell interactions through signaling molecules

The dynamics of signaling molecule concentrations can be modeled by any variety of molecular transport models. For example, a deterministic PDE describing excretion, absorption, and diffusion of signaling molecules takes on the typical form [38]

$$\partial_t c_\ell(\mathbf{r}, t) = D_\ell \nabla^2 c_\ell - d_\ell c_\ell + \sum_{\text{cells } \alpha} \lambda_\ell(\mathbf{r} - \mathbf{r}^\alpha, \mathbf{z}^\alpha), \quad (2)$$

where $\lambda_\ell(\mathbf{r} - \mathbf{r}^\alpha, \mathbf{z}^\alpha)$ is the excretion rate of signaling molecule ℓ at position \mathbf{r} from a cell in state \mathbf{z}^α centered at \mathbf{r}^α , and the sum runs over all cells. The function $\lambda_\ell(\mathbf{r} - \mathbf{r}^\alpha, \mathbf{z}^\alpha)$ is localized about \mathbf{r}^α and represents the spatial extent of a cell that exports signaling molecule. Diffusion and degradation of molecule ℓ is described by

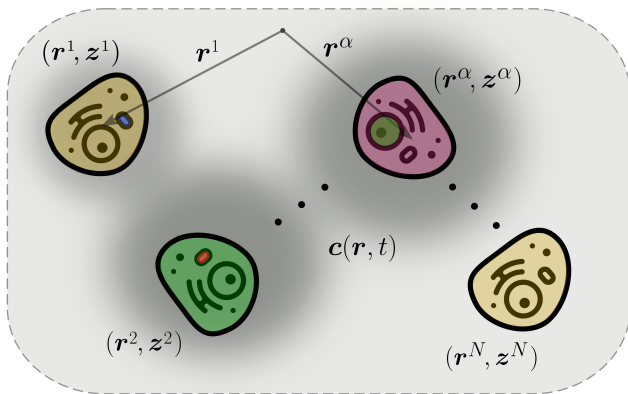


FIG. 1. Schematic of cells at various positions \mathbf{r}^α and in various gene expression states \mathbf{z}^α , $\alpha = 1, 2, \dots, N$. Cell-cell interactions mediated by position-dependent signaling molecule concentrations $\mathbf{c}(\mathbf{r}, t) \equiv (c_1(\mathbf{r}, t), \dots, c_L(\mathbf{r}, t))$ which are represented in gray-scale. Cells can produce signaling molecules at rates that depend on \mathbf{z} ; gene expression is in turn influenced by the concentrations that a cell senses, leading to potentially complex cell-cell interactions.

the coefficients D_ℓ and d_ℓ , respectively. A natural choice of boundary condition for Eq. 2 is $c_\ell(\mathbf{r}) \rightarrow 0$ as $|\mathbf{r}| \rightarrow \infty$. Alternatively, to model in subcompartments or spatially defined microenvironments, Eq. 2 can be considered in finite regions of space each with appropriate boundary conditions. For a closed environment, one would use reflecting boundary conditions for $c_\ell(\mathbf{r})$. Eq. 2 neglects reactions amongst the signaling molecules and models deterministic or locally averaged concentrations, appropriate for molecular transport at high concentrations and/or fast molecular timescales. At very low signaling molecule densities, intrinsic molecular stochasticity may be important and any number of stochastic models [39–41] and simulation approaches [42, 43] are available. Fig. 1 provides a schematic of signaling-molecule-mediated interactions among cells at positions \mathbf{r}^α , with potentially different internal states \mathbf{z}^α .

C. Cell migration

Cells may also migrate by exerting forces on neighboring cells or on their substrate; this migration may be influenced by signaling molecules (chemokines). Such coordinated movement is essential for tissue development [44, 45]. One can straightforwardly couple molecular signaling with cell motion [46], and a general stochastic model for cell motion can be expressed using an SDE such as

$$d\mathbf{r}^\alpha(t) = \mathbf{g}(\mathbf{r}^\alpha, S(t))dt + \eta(\mathbf{r}^\alpha, S(t))d\mathbf{W}^\alpha \quad (3)$$

where $S(t) = \{(\mathbf{r}^\alpha(t), \mathbf{z}^\alpha(t))\}_{\alpha=1}^{N(t)}$ is the full state of the tissue population and $d\mathbf{W}^\alpha$ is a 3-dimensional Brownian noise. The mean velocity \mathbf{g} and noise amplitude η

(a 3×3 matrix) can depend on the cell's position \mathbf{r}^α as well as on the full state of the tissue $S(t)$. Although explicit dependence of \mathbf{g} and η on the spatial profile of signaling molecule concentrations is not included in Eq. 3, one could consider a velocity \mathbf{g} which depends on $\nabla \mathbf{c}(\mathbf{r}^\alpha, t)$, or on higher derivatives of the concentration profile. Eq. 3 is a Langevin equation for fluctuating cell positions. Many related models for cell motion, including anomalous diffusion and longer-ranged hopping have been developed [47–49].

D. Cell proliferation (and death)

During cell division—a process involving cellular-scale disruption and reorganization—a mother cell divides into two daughter cells which may acquire states that are different from the state of the mother. In our model, cells have a state-dependent and concentration-dependent division rate, denoted $\beta(\mathbf{z}; \mathbf{c})$ for a cell in state \mathbf{z} and with signaling molecule concentrations \mathbf{c} at its location (to be precise, cell α has division rate $\beta(\mathbf{z}^\alpha; \mathbf{c}^\alpha)$, but we will drop the superscript α to avoid cumbersome notation). $\beta(\mathbf{z}; \mathbf{c})$ is a marginal probability over the joint distribution for the probability of a cell in state \mathbf{z} dividing into daughters with states \mathbf{z}' , \mathbf{z}'' and with positional displacements from the mother $\Delta \mathbf{r}'$, $\Delta \mathbf{r}''$. This joint distribution is denoted as

$$\tilde{\beta}(\mathbf{z}; \mathbf{c}; \mathbf{z}', \mathbf{z}'', \Delta \mathbf{r}', \Delta \mathbf{r}'') \quad (4)$$

and is invariant under exchange of \mathbf{z}' and \mathbf{z}'' and under exchange of $\Delta \mathbf{r}'$ and $\Delta \mathbf{r}''$. Marginalizing, we find the overall division rate

$$\beta(\mathbf{z}; \mathbf{c}) = \int \tilde{\beta}(\mathbf{z}; \mathbf{c}; \mathbf{z}', \mathbf{z}'', \Delta \mathbf{r}', \Delta \mathbf{r}'') d\mathbf{z}' d\mathbf{z}'' d\Delta \mathbf{r}' d\Delta \mathbf{r}'' \quad (5)$$

Finally, cell death occurs at rate $\mu(\mathbf{z}; \mathbf{c})$, which can depend on both \mathbf{z} and \mathbf{c} .

Eqs. 1, 2, and 3, along with system-specific models for birth and death rates, $\tilde{\beta}(\mathbf{z}; \mathbf{c}; \mathbf{z}', \mathbf{z}'', \Delta \mathbf{r}', \Delta \mathbf{r}'')$ and $\mu(\mathbf{z}; \mathbf{c})$, form our general description of developing cell populations. The following section illustrates how common assumptions can be incorporated to significantly simplify the model.

III. SIMPLIFYING ASSUMPTIONS

Here we describe a series of assumptions that can be applied to further simplify our mathematical model. Our examples in Sections V and VI show how various combinations of these assumption can be used to obtain models that capture different degrees of detail.

A. Fast equilibration of signaling molecule concentrations

Suppose the concentrations $c_\ell(\mathbf{r}, t)$ reach steady state on timescales shorter than those associated with gene expression dynamics, cell movement, and cell division and death. We can approximate $c_\ell(\mathbf{r}, t)$ by the quasi-steady state solution of Eq. 2 under the instantaneous state of the population $S(t) = \{\mathbf{z}^\alpha(t), \mathbf{r}^\alpha(t)\}_{\alpha=1}^{N(t)}$. We denote this quasi-steady state concentration profile by $c_\ell^{\text{ss}}(\mathbf{r}; S(t))$. For signaling molecule dynamics described by Eq. 2, $c_\ell^{\text{ss}}(\mathbf{r}; S(t))$ takes the explicit form

$$c_\ell^{\text{ss}}(\mathbf{r}; S(t)) = \sum_{\text{cells } \alpha} c_\ell^{(1)}(\mathbf{r} - \mathbf{r}^\alpha, \mathbf{z}^\alpha) \quad (6)$$

where $c_\ell^{(1)}(\mathbf{r}, \mathbf{z})$ is the concentration profile produced by a single isolated cell at position $\mathbf{r} = \mathbf{0}$, given by the solution to

$$0 = D_\ell \nabla_{\mathbf{r}}^2 c_\ell^{(1)}(\mathbf{r}, \mathbf{z}) - d_\ell c_\ell^{(1)}(\mathbf{r}, \mathbf{z}) + \lambda_\ell(\mathbf{r}, \mathbf{z}) \quad (7)$$

with appropriate boundary conditions.

B. Well-mixed population

In a further approximation, that may apply to *e.g.*, well-mixed populations of stem cells that differentiate in microenvironments, spatial distributions of cells and signaling molecules can both be treated as spatially averaged quantities. Hence, Eq. 3 can be ignored and c_ℓ is independent of \mathbf{r} . If the dynamics of c_ℓ are given by Eq. 2 with reflecting boundary conditions, then Eq. 2 can be replaced by an ODE for the mean concentration $\bar{c}_\ell(t) = (1/V) \int c_\ell(\mathbf{r}) d\mathbf{r}$ over the closed volume V

$$\frac{d\bar{c}_\ell}{dt} = \sum_{\text{cells } \alpha} \bar{\lambda}_\ell(\mathbf{z}^\alpha) - d_\ell \bar{c}_\ell \quad (8)$$

where $\bar{\lambda}_\ell(\mathbf{z}) = (1/V) \int \lambda_\ell(\mathbf{r}, \mathbf{z}) d\mathbf{r}$ is the spatially averaged excretion rate of each of the uniformly distributed cells. If signaling molecules reach steady state rapidly, then the signaling molecule concentrations are

$$c_\ell^{\text{ss}} = \frac{1}{d_\ell} \sum_{\text{cells } \alpha} \bar{\lambda}_\ell(\mathbf{z}^\alpha). \quad (9)$$

C. Fast \mathbf{z} dynamics and discrete cell types

For fast cell state dynamics, the state \mathbf{z}^α of cell α rapidly reaches an attractor of $\mathbf{F}(\mathbf{z}^\alpha; \mathbf{c}^\alpha)$ but can make

stochastic transitions to different attractors. Hence, we can approximate the continuous state space by the discrete set of attractors, labeled $q = 1, \dots, Q$. Such a discrete approximation of an otherwise continuous or highly granular set of cell states has been shown to be useful in many developmental contexts such as embryonic stem cell differentiation [21, 22, 50]. The Waddington vector field $\mathbf{F}(\mathbf{z}^\alpha; \mathbf{c}^\alpha)$ is then replaced by a matrix of jump rates, $F_{q,q'}(\mathbf{c}^\alpha)$, specifying the jump rate from q to q' . These jump rates can, in principle, be estimated numerically via simulations of Eq. 1. Alternatively, if $\mathbf{F}(\mathbf{z}^\alpha; \mathbf{c}^\alpha)$ is the gradient of a Waddington landscape, then Kramers' rate formula can be used to approximate the jump rates [51].

The division rate at which cell α in attractor q divides into daughter cells at attractor states q', q'' with positional displacements $\Delta\mathbf{r}', \Delta\mathbf{r}''$ is denoted $\tilde{\beta}_{q,q',q''}(\mathbf{c}; \Delta\mathbf{r}', \Delta\mathbf{r}'')$. Just as for Eq. 4, $\tilde{\beta}_{q,q',q''}(\mathbf{c}; \Delta\mathbf{r}', \Delta\mathbf{r}'')$ is invariant under exchange of q' and q'' , as well as under exchange of $\Delta\mathbf{r}'$ and $\Delta\mathbf{r}''$. The diagonal elements $\tilde{\beta}_{q,q,q}$ correspond to symmetric division in which both daughters have the same state as the mother, whereas elements with $q' \neq q''$ correspond to asymmetric division. Letting \mathbf{z}_q denote the attractor state q , these division rates are given by $\tilde{\beta}_{q,q',q''}(\mathbf{c}; \Delta\mathbf{r}', \Delta\mathbf{r}'') = \tilde{\beta}(\mathbf{z}_q; \mathbf{c}; \mathbf{z}_{q'}, \mathbf{z}_{q''}, \Delta\mathbf{r}', \Delta\mathbf{r}'')$. The death rate in attractor q is $\mu_q(\mathbf{c}) = \mu(\mathbf{z}_q; \mathbf{c})$.

D. Unifying the assumptions

If all three of the above simplifying approximations (fast equilibration of signaling molecules, well-mixed population, and discrete cell types) hold, our model simplifies significantly. Let the vector $\mathbf{n} \equiv (n_1, \dots, n_Q)$ specify the cell populations in the tissue, where n_q is the number of cells of type q . Let $\beta_{q,q',q''}(\mathbf{c}) \equiv \int \tilde{\beta}_{q,q',q''}(\mathbf{c}; \Delta\mathbf{r}', \Delta\mathbf{r}'') d\Delta\mathbf{r}' d\Delta\mathbf{r}''$ be the division rate of a mother of type q into daughters of type q', q'' . Signaling concentrations $\mathbf{c}(\mathbf{n})$ are functions of only \mathbf{n} , and are determined by Eq. 9.

The probability $P(\mathbf{n}, t)$ of having a tissue state \mathbf{n} at time t evolves accord to the following master equation:

$$\begin{aligned}
\partial_t P(\mathbf{n}, t) = & \sum_{q, q', q''} \left[(n_q + 1) \beta_{q, q', q''}(\mathbf{c}(\mathbf{n}_{q+, q'-, q''-})) P(\mathbf{n}_{q+, q'-, q''-}) - n_q \beta_{q, q', q''}(\mathbf{c}(\mathbf{n})) P(\mathbf{n}) \right] \\
& + \sum_{q, q'} \left[(n_q + 1) F_{q, q'}(\mathbf{c}(\mathbf{n}_{q+, q'-})) P(\mathbf{n}_{q+, q'-}) - n_q F_{q, q'}(\mathbf{c}(\mathbf{n})) P(\mathbf{n}) \right] \\
& + \sum_q \left[(n_q + 1) \mu_q(\mathbf{c}(\mathbf{n}_{q+})) P(\mathbf{n}_{q+}) - n_q \mu_q(\mathbf{c}(\mathbf{n})) P(\mathbf{n}) \right]
\end{aligned} \tag{10}$$

In analogy with the bookkeeping notation used in [29, 30], $\mathbf{n}_{q+, q'-, q''-}$ denotes a system state with, compared to system state \mathbf{n} , one more type q cell and one fewer cell of type q' and one fewer cell of type q'' . Hence, $\mathbf{n}_{q+, q'-, q''-}$ is the system state that converts into state \mathbf{n} when a mother cell of type q divides into daughter cells of type q' and q'' . Similarly, $\mathbf{n}_{q+, q'-}$ denotes a system state with one more type q cell and one fewer type q' cell relative to system state \mathbf{n} . \mathbf{n}_{q+} denotes a system state with one more type q cell than in system state \mathbf{n} . Equation 10 represents a master equation for the probability density for populations of cells of multiple types and is a simpler form of master equations studied in [29, 30, 52]. For small populations and few accessible cell types, solving this linear master equation or simulating the process is computationally feasible.

E. Developmental Fitness landscape

Proliferation of certain cell types over others due to differing division and death rates is a crucial aspect of development. Differential proliferation is also key to evolutionary dynamics, where genotypes that induce high net growth rates eventually comprise larger portions of the overall population. In evolutionary dynamics, preferential growth is described by a fitness landscape, typically defined as a function which specifies the net growth (division minus death) rate ϕ_g of a cell or organism given its genotype g . Here, we adapt this concept to the context of cell development to define an epigenetic fitness landscape which specifies which cell types proliferate in the developing tissue.

A key difference between evolutionary dynamics (or speciation) and development is that, in a developing tissue, all cells have the same genotype (aside from rare mutations) and differences in cells' division rates are due to differences in their cell type, determined by the internal state \mathbf{z} . We define the developmental fitness of cell type q , denoted $\phi_q(\mathbf{n})$, as the net proliferation rate of type q cells, absent production of type q cells through division of other cell types or stochastic jumps from other cell types. $\phi_q(\mathbf{n})$ depends on the tissue population state \mathbf{n} due to cell-cell interactions via signaling molecules. In defining $\phi_q(\mathbf{n})$, we adopt the assumptions from the previous sections, namely, fast equilibration of signaling molecules, well-mixed population, and discrete cell types. Within

our framework, we can define fitness mathematically as

$$\begin{aligned}
\phi_q(\mathbf{n}) = & \beta_{q, q, q}(\mathbf{c}(\mathbf{n})) - \beta_{q, \text{lost}}(\mathbf{c}(\mathbf{n})) \\
& - F_{q, \text{lost}}(\mathbf{c}(\mathbf{n})) - \mu_q(\mathbf{c}(\mathbf{n}))
\end{aligned} \tag{11}$$

where $\beta_{q, q, q}(\mathbf{c})$ is the rate of symmetric division in which a mother cell of type q divides into two daughter cells of type q . $\beta_{q, \text{lost}}(\mathbf{c}) \equiv \sum_{q' \neq q, q'' \neq q} \beta_{q, q', q''}(\mathbf{c})$ is the rate at which a mother cell of type q divides into daughter cells, both of whose cell types differ from that of the mother cell. Note that asymmetric division, where a cell of type q divides into daughters of type q and $q' \neq q$, leaves the population of type q cells unchanged, and therefore does not factor into the developmental fitness landscape. $F_{q, \text{lost}}(\mathbf{c}) = \sum_{q' \neq q} F_{q, q'}(\mathbf{c})$ is the jump rate of type q cells to a different type.

The following subsection shows how, in the limit of large tissue size, the developmental fitness landscape contributes to tissue dynamics through a simple equation.

F. Deterministic limit

In the limit of large tissue size, Eq. 10 predicts that cell densities in the tissue follow deterministic dynamics. This limit is analogous to how deterministic mass action equations (ODEs) emerge from a more microscopic chemical master equation description in the limit of large system size. Recall that Eq. 10 describes the population within a region of fixed volume, and denote this volume by V_0 . Then, $\rho \equiv \mathbf{n}/V_0$ are the cell type densities in the tissue described by Eq. 10. Now, consider a system with volume $V > V_0$ and take the limit that $V \rightarrow \infty$ while ρ remains fixed. The mass-action dynamics in this limit are described by

$$\frac{d\rho}{dt} = \text{diag}(\phi(\rho V_0))\rho + \mathbf{m}(\rho V_0)\rho \tag{12}$$

where $\phi(\mathbf{n}) = (\phi_1(\mathbf{n}), \dots, \phi_Q(\mathbf{n}))$ is the developmental fitness landscape defined in the previous section, and $\text{diag}(\phi(\rho V_0))$ denotes the $Q \times Q$ matrix with the elements of $\phi(\rho V_0)$ on the diagonal. $\mathbf{m}(\rho V_0)$ is a zero-diagonal matrix which contains the jump rates and asymmetric division rates through which cells of one type produce cells of another type. The matrix elements are explicitly

$$\begin{aligned}
m_{q, q'}(\mathbf{n}) = & F_{q', q}(\mathbf{c}(\mathbf{n})) + 2\beta_{q', q, q}(\mathbf{c}(\mathbf{n})) \\
& + 2 \sum_{q'' \neq q} \beta_{q', q, q''}(\mathbf{c}(\mathbf{n})).
\end{aligned} \tag{13}$$

The factor of 2 in the third term arises from the symmetry of $\beta_{q',q''}$ with respect to exchange of the final two indices.

Deterministic dynamical models of interacting population dynamics, as could be described by Eq. 12, have been used to model aspects of development [53]. Note that the stochastic fluctuations in the dynamics of ρ are proportional to $(V/V_0)^{-1/2}$, which vanishes as $V \rightarrow \infty$. For large but finite V , the stochastic component of the dynamics can be obtained with the van Kampen system size expansion [54].

IV. WADDINGTON'S LANDSCAPE VS WADDINGTON'S VECTOR FIELD

From among the various interpretations of Waddington's landscape that have been studied. The most common interpretation that has emerged is that the vector field $\mathbf{F}(\mathbf{z}; \mathbf{c})$ in Eq. 1 is a negative gradient of a function $U(\mathbf{z}; \mathbf{c})$,

$$\mathbf{F}(\mathbf{z}; \mathbf{c}) = -\nabla_{\mathbf{z}} U(\mathbf{z}; \mathbf{c}), \quad (14)$$

where $U(\mathbf{z}; \mathbf{c})$ is called a Waddington landscape [3, 9, 23, 25]. We call a stochastic dynamical system of this form, where the deterministic part is the negative gradient of a potential landscape, a *gradient system*. If the noise in Eq. 1 is additive (meaning σ is independent of \mathbf{z}), then $U(\mathbf{z}; \mathbf{c})$ can be reconstructed from experimental measurement of the steady state distribution of cell states *at fixed signaling molecule concentrations*, $p_{\text{ss}}(\mathbf{z}; \mathbf{c})$, via [3]

$$U(\mathbf{z}; \mathbf{c}) = -\log p_{\text{ss}}(\mathbf{z}; \mathbf{c}). \quad (15)$$

The validity and generality of this Waddington landscape framework has been a topic of considerable discussion and disagreement. On one hand, it has been pointed out that realistic dynamics of protein and mRNA concentrations are not gradient systems, but rather have *curl* (or the higher-dimensional generalization of curl) [14, 16, 28]. On the other hand, it has been noted that for nearly all vector fields \mathbf{F} without limit cycles (specifically, for Morse-Smale systems), the vector field \mathbf{F} can be converted to the negative gradient of a potential landscape with a coordinate transformation [8]. Hence, a very broad class of models, including most models of protein and mRNA dynamics, can be expressed as gradient systems under a suitable choice of coordinates. This suggests that the Waddington landscape picture is highly general.

However, noise has a crucial effect which significantly hinders the generality of the landscape picture. Eq. 15 is valid only for additive noise, yet the noise in the dynamics of protein and mRNA concentrations is typically non-additive [9]. It was recently demonstrated how the mis-application of Eq. 15 to systems with non-additive noise can 'add' or 'delete' attractors, meaning that the reconstructed $U(\mathbf{z}; \mathbf{c})$ from Eq. 15 may have attractors

not present in the true dynamics, or may not have attractors that are present in the true dynamics [9]. We demonstrate a similar effect for systems with curl: if Eq. 15 is mis-applied to a system with curl and with additive noise, the reconstruction of $U(\mathbf{z}; \mathbf{c})$ can also 'add' or 'delete' attractors (see appendix A). In fact, these two mis-applications of Eq. 15, to systems with non-additive noise and to systems with curl, are two sides of the same coin: if a non-gradient system with additive noise is converted via coordinate transformation to a gradient system, the noise will no longer be additive, so Eq. 15 will be invalid. Conversely, if a gradient system with non-additive noise is converted to a system with additive noise via coordinate transformation, the dynamics will no longer be gradient, so Eq. 14 will be invalid. In short, one can change coordinates to remove curl or non-additive noise, but not both. Hence, coordinate transformations cannot allow a system to simultaneously satisfy both Eqs. 14 and 15, so the Waddington landscape is not a general framework for studying gene regulatory dynamics.

We advocate for working directly with the vector field $\mathbf{F}(\mathbf{z}; \mathbf{c})$ that describes gene regulatory dynamics, which we call *Waddington's vector field*, regardless of whether it is the negative gradient of a landscape. Importantly, the key conceptual clarity offered by Waddington's landscape—that cell states are drawn into attractors governed by gene regulatory networks—remains valid in a 'Waddington vector field' picture. While some authors propose decomposing the vector field into gradient and curl parts [6, 7, 14, 28], this decomposition can be misleading because basins of attraction of the gradient part may not contain any attractor of $\mathbf{F}(\mathbf{z}; \mathbf{c})$, and we see no advantage to this decomposition. The Waddington vector field picture is, we argue, a clearer and mathematically precise framework to describe the gene regulatory dynamics in cell development.

V. EXAMPLE 1: STEM CELL DIFFERENTIATION IN A WELL-MIXED MICROENVIRONMENT

Here, we present a model of tissue development which is derived by augmenting a prototypical model of Waddington's landscape studied in [14, 16, 55]. In our model, $\mathbf{F}(\mathbf{z}; \mathbf{c})$ describes a hypothetical gene regulatory circuit involving two proteins, A and B. $\mathbf{z} = (z_A, z_B)$ specifies the concentrations of these two proteins. $\mathbf{F}(\mathbf{z}; \mathbf{c})$ has three attractors corresponding to three cell types, which, for reasons that will become clear, we call *stem cells* \mathcal{S} , *triggered stem cells* \mathcal{S}^* , and *differentiated cells* \mathcal{D} . Cells interact via two signaling molecules with concentrations $\mathbf{c} = (c_1, c_2)$. These interactions induce stem cells to differentiate in a coordinated way, so that a stable population (*i.e.* a tissue) of differentiated cells is produced and maintained. Hence, while the previous models in [14, 16, 55] describe only single-cell differentiation, our

model describes the coordinated development of tissue. After analyzing the dynamics of the full model, we show how the model can be simplified into a simple system of ODEs using the approximations developed in section III.

The gene regulatory circuit that governs the dynamics of \mathbf{z} is illustrated in Fig. 2A. Each gene (A and B) upregulates itself and downregulate the other. We make the common quasi-steady state assumption that mRNA transcription and degradation are fast relative to protein translation and degradation, so the state and dynamics are described in terms of only the protein concentrations $\mathbf{z} = (z_A, z_B)$. Signaling molecule “1” influences expression of gene A , but not gene B . Specifically, B tetramer (which downregulates gene A) can form a complex with signaling molecule 1, and this complex prevents any expression of A when bound to the silencer of A . This binding scheme gives rise to a Waddington vector field, shown in Fig. 2B, whose components are

$$\begin{aligned} F_m(\mathbf{z}; \mathbf{c}) = & r_{0,m}(1 - E_m(z_m))(1 - S_m(z_{m'})) \\ & + r_{E,m}E_m(z_m)(1 - S_m(z_{m'})) \\ & + r_{S,m}(c_1)(1 - E_m(z_m))S_m(z_{m'}) \\ & + r_{ES,m}(c_1)E_m(z_m)S_m(z_{m'}) - k_m z_m \end{aligned} \quad (16)$$

where $m, m' = A, B$ or B, A . $E_m(z_m)$ and $S_m(z_{m'})$ are, respectively, the probabilities that the enhancer and silencer of gene m have a bound tetramer, and are modeled by Hill functions:

$$E_m(z_m) = \frac{z_m^4}{K_{E,m}^4 + z_m^4}, \quad S_m(z_{m'}) = \frac{z_{m'}^4}{K_{S,m}^4 + z_{m'}^4}. \quad (17)$$

The rates $r_{0,m}$, $r_{E,m}$, $r_{S,m}(c_1)$, and $r_{ES,m}(c_1)$ are the mean expression rates of gene m with, respectively, neither enhancer nor silencer occupied, only enhancer occupied, only silencer occupied, and both enhancer and silencer occupied. The rates $r_{S,A}(c_1)$ and $r_{ES,A}(c_1)$ have a Michaelis-Menten type dependence on c_1 :

$$r_{S,A}(c_1) = \frac{r_{S,A}(0)\gamma}{\gamma + c_1}, \quad r_{ES,A}(c_1) = \frac{r_{ES,A}(0)\gamma}{\gamma + c_1}. \quad (18)$$

The factor $\gamma/(\gamma + c_1)$ is the probability that a given B tetramer is *not* bound to a signaling molecule. All other rate parameters, $r_{0,m}$, $r_{E,m}$, $r_{S,B}$, $r_{ES,B}$, $r_{S,A}(0)$, and $r_{ES,A}(0)$ are constants (listed in Table 1, along with all other model parameters). Fig. 2B shows the Waddington vector field for three values of c_1 : $c_1 = 0$ (left), $c_1 = 30\text{nM}$ (middle), and $c_1 = 50\text{nM}$ (right). For a particular choice of parameters, $\mathbf{F}(\mathbf{z}; \mathbf{c})$ reduces to the model in [14, 55].

The key role of signaling molecule 1 is to regulate how stem cells are triggered for differentiation. Fig. 2B shows the attractors (black squares) for the three cell types (\mathcal{S} , \mathcal{S}^* , and \mathcal{D}). At intermediate c_1 , stable attractors exist for both \mathcal{S} and \mathcal{S}^* , but when c_1 drops sufficiently low, the \mathcal{S} attractor disappears and stem cells move toward the triggered state \mathcal{S}^* . Hence, low c_1 triggers stem cells for

differentiation, and as we will see below, these triggered stem cells undergo asymmetric division to produce differentiated cells. On the other hand, very high c_1 reverses the triggering (the \mathcal{S}^* attractor disappears, so triggered stem cells revert to untriggered stem cells). Fig. 2C shows how the locations of the attractors (black) and unstable fixed points that separate the attractors (red) vary continuously with c_1 . The unstable fixed point that separates the \mathcal{S} and \mathcal{S}^* attractors merges with \mathcal{S} at low c_1 , but merges with \mathcal{S}^* at high c_1 . These mergers of fixed points are saddle-node bifurcations.

Cell division rates depend on both the expression level z_B of protein B and the concentration c_2 of signaling molecule 2. The dependence on z_B gives rise to the defining behavior of the three cell types: stem cells divide symmetrically (both daughters are type \mathcal{S}), triggered stem cells divide asymmetrically (one daughter is \mathcal{S} and the other \mathcal{D}), and differentiated cells do not divide. Mathematically, we model the birth rate by

$$\tilde{\beta}(\mathbf{z}; \mathbf{z}', \mathbf{z}'') = \begin{cases} 0 & z_B \leq z^{\mathcal{D}} \\ \beta_{\mathcal{S}^*} \delta(\mathbf{z} - \mathbf{z}' - \mathbf{w}(\mathbf{z})) \\ \quad \times \delta(\mathbf{z} - \mathbf{z}'' + \mathbf{w}(\mathbf{z})) & z^{\mathcal{S}} \geq z_B > z^{\mathcal{D}} \\ \frac{a_1 \delta(\mathbf{z} - \mathbf{z}') \delta(\mathbf{z} - \mathbf{z}'')}{1 + a_2 \exp(a_3 c_2)} & z_B > z^{\mathcal{S}}. \end{cases} \quad (19)$$

where $z^{\mathcal{D}} < z^{\mathcal{S}}$ are threshold values for z_B . Note that we do not include the $\Delta \mathbf{r}'$, $\Delta \mathbf{r}''$ dependencies because this model assumes a well-mixed (spatially uniform) system. The threshold concentrations $z^{\mathcal{S}}$ and $z^{\mathcal{D}}$ are indicated by dashed lines in Fig. 2B, and are determined by the (approximate) z_B values of the unstable fixed points of \mathbf{F} .

The vector $\mathbf{w}(\mathbf{z})$ induces asymmetric division: when a triggered stem cell divides, one daughter has a state shifted by $\mathbf{w}(\mathbf{z})$ relative to the mother, and the other daughter has state shifted by $-\mathbf{w}(\mathbf{z})$. As a result, one daughter returns to the \mathcal{S} attractor, while the other moves to the \mathcal{D} attractor. Note that for \mathbf{z} defined by concentrations, the concentrations \mathbf{z}' and \mathbf{z}'' in the daughter cells are related to that of the mother through $(\mathbf{z}' + \mathbf{z}'')/2 = \mathbf{z}$. The death rate μ for all cells is assumed constant.

In this model, signaling molecule 1 is excreted by differentiated cells and signaling molecule 2 is excreted by stem cells. We use Eq. 9 to model signaling molecule concentrations, which assumes a well-mixed population and that concentrations rapidly reach steady state. The excretion rates of c_1 and c_2 are, respectively,

$$\bar{\lambda}_1(\mathbf{z}) = \begin{cases} 0 & z_B > z^{\mathcal{D}} \\ L_1 & z_B \leq z^{\mathcal{D}} \end{cases} \quad (20)$$

and

$$\bar{\lambda}_2(\mathbf{z}) = \begin{cases} L_2 & z_B > z^{\mathcal{S}} \\ 0 & z_B \leq z^{\mathcal{S}} \end{cases} \quad (21)$$

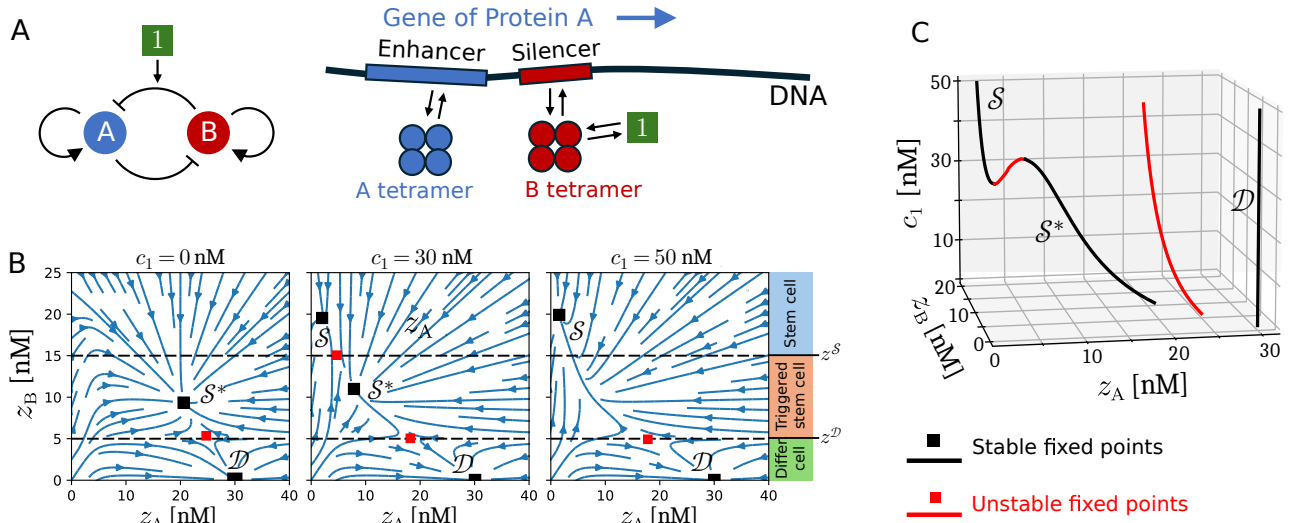


FIG. 2. Example 1: Model of regulated stem cell differentiation. (A) Gene regulatory circuit. Left: Schematic of circuit showing upregulation (arrows), down regulation (flat arrows), and the effect of signaling molecule 1 concentration c_1 on amplification of the downregulation of A by B. Right: Diagram of molecular mechanisms. Black arrows indicate molecular binding and unbinding of the A and B tetramers to the enhancer and silencer, and of signaling molecule 1 (green square) to B tetramer. Signaling molecule 2 controls cell proliferation and is not featured in this picture of individual-cell gene regulation dynamics. (B) Vector field $\mathbf{F}(\mathbf{z}; \mathbf{c})$ at three values of c_1 . Attractors (black squares) are labelled by cell type: stem cells S , triggered stem cells S^* , and differentiated cells D . Unstable fixed points indicated by red squares. Black dashed lines at $z_B = z^S$ and z^D denote the boundary between stem cells and triggered stem, and between triggered stem cells and differentiated cells, respectively. (C) Fixed points of $\mathbf{F}(\mathbf{z}; \mathbf{c})$ as a function of c_1 . Two saddle-node bifurcations result in the gain/loss of attractors S and S^* .

Thus, according to Eq. 9, in a tissue with N_S stem cells and N_D differentiated cells, the concentrations are simply proportional to the cell numbers n_S and n_D ,

$$(c_1^{ss}, c_2^{ss}) = \left(n_D \frac{L_1}{d_1}, n_S \frac{L_2}{d_2} \right). \quad (22)$$

The signaling molecules' influence on gene dynamics and birth rates gives rise to robust tissue development: from a wide range of initial conditions, the system dynamically approaches a stable tissue state with regulated relative cell type populations, as illustrated in Fig. 3A. The two signaling molecules govern the two key regulatory mechanisms: (i) the stem cell population, n_S , is self-regulated due to the dependence of stem cell division on c_2 . When n_S is low (high), c_2 is low (high), so stem cell division rate is high (low), replenishing (diminishing) n_S towards the stable, developed state; (ii) the dependence of $\mathbf{F}(\mathbf{z}; \mathbf{c})$ on c_1 leads to a thermostat-like behavior of stem cell triggering. When n_D is low (high), c_1 is low (high), so more (less) stem cells are triggered for asymmetric division. This thermostat-like behavior is apparent from Fig. 3A; whenever n_D drops below a threshold, stem cells are triggered, rapidly replenishing n_D . In Fig. 3A specifically, an initial population of only four stem cells develops into the stable tissue state.

A. Deterministic limit

The deterministic limit described in section III F can be applied to obtain a highly simplified form of the model. Denote the cell densities for the three cell types as $\boldsymbol{\rho} = (\rho_S, \rho_{S^*}, \rho_D)$. The developmental fitness landscape is

$$\phi(\mathbf{n}) = (\beta_S(n_S) - F_{D,S}(n_D) - \mu, -\beta_{S^*} - \mu, -\mu) \quad (23)$$

where $\beta_S(n_S) \equiv a_1/(1 + a_2 \exp(a_3 n_S L_2/d_2))$, as determined by Eqs. 19 and 22. $F_{D,S}(n_D)$ is the rate at which stem cells are triggered. $F_{D,S}(n_D)$ can be estimated numerically by simulating the gene dynamics, and it is well approximated by a step function, equal to 0 for $n_D > 25$ and greater than 0 for $n_D \leq 25$. The deterministic equations given by Eq. 12 written out explicitly are

$$\begin{aligned} \frac{d\rho_S}{dt} &= (\beta_S(\rho_S V_0) - F_{D,S}(\rho_D V_0) - \mu)\rho_S + \beta_{S^*}\rho_{S^*} \\ \frac{d\rho_{S^*}}{dt} &= F_{D,S}(\rho_D V_0)\rho_S - (\beta_{S^*} + \mu)\rho_{S^*} \\ \frac{d\rho_D}{dt} &= \beta_{S^*}\rho_{S^*} - \mu\rho_D. \end{aligned} \quad (24)$$

Fig. 3B plots the solutions of these equations for the same initial conditions as in Fig. 2D. The transient thermostat-like behavior is lost, but the longer timescale results are qualitatively similar.

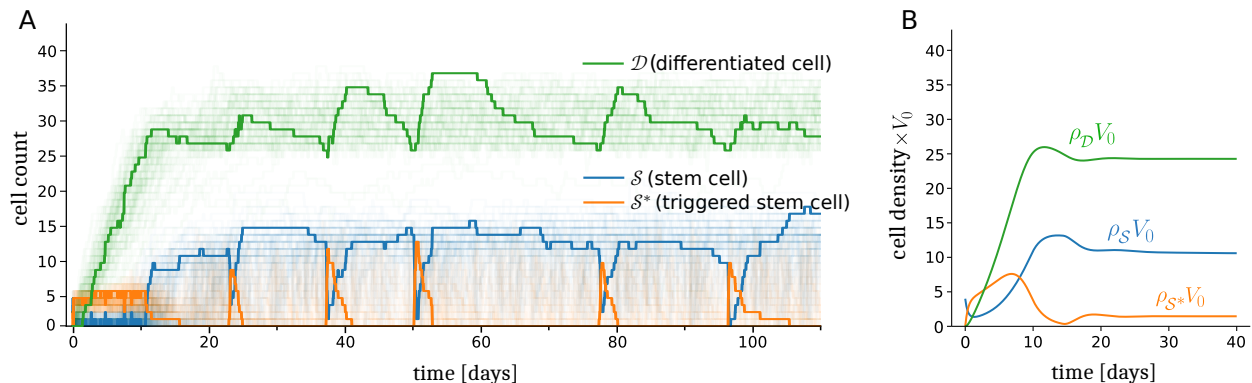


FIG. 3. Example 1: Dynamics of cell populations. (A) Dynamical simulation starting from an initial condition with $N_S = 4$ and $N_{S^*} = N_D = 0$. The population dynamics exhibit thermostat-like behavior: when N_D falls below 26, stem cells are triggered for asymmetric division, which results in N_D being replenished. One-hundred trajectories were simulated by evaluating Eq. 1, coupled with a discrete cell birth-death process, using the standard Euler-Maruyama method. One representative trajectory is highlighted, while the other ninety-nine are lightly shaded. (B) Evolution of cell populations in the deterministic limit. The densities are multiplied by the normalizing volume V_0 to express densities on the same scale as particle counts. Note that the deterministic trajectories qualitatively reflect the overall dynamics shown in (A).

VI. EXAMPLE 2: SPATIAL MODEL INSPIRED BY PLANARIA

Here, we present a model of a spatially-structured multicellular organism which develops from a single cell. Additionally, if the developed organism is cut into pieces, each piece will regenerate into a new fully-developed organism (a phenomenon called fissiparous reproduction). This model is inspired by planaria worms, which exhibit a remarkable capacity for fissiparous reproduction: a fragment of a planarian as small as 1/279th of the full worm can regenerate into a new worm [56, 57].

In this model, cell state is described by the expression of a single gene, denoted gene A, whose expression is governed by a bistable regulatory circuit. The state z is the concentration of protein A (we assume fast mRNA dynamics, as in example 1). Two stable states corresponding to two values of z define two cell types: exterior cells (which have low z) and interior cells (which have high z). Cells excrete two signaling molecules, one which governs the z dynamics and another which governs cell division rate. The cells live and migrate on a 2D substrate ($\mathbf{r} \in \mathbb{R}^2$), and they exert forces on one another which govern their movement.

The gene regulatory circuit is illustrated in Fig. 4A. Gene A upregulates itself when signaling molecule 1 is present. Specifically, protein A can bind to signaling molecule 1, and this heterodimer binds to the enhancer of gene A, increasing the expression of protein A. The Waddington vector field thus takes the form

$$F(z; c) = r_0 + \frac{r_1 c_1 z^2}{K^2 + z^2} - bz. \quad (25)$$

The values of the parameters r_0 , r_1 , K , and b , as well of the parameters introduced in the following equations, are given in Table 1.

Fig. 4B shows the dynamics (blue arrows) as well as the locations of the attractors (black) and unstable fixed points (red) of $F(z; c)$, for values of c_1 between 4 and 12 nM. The attractor at low (high) z , which corresponds to exterior (interior) cells, merges with the unstable fixed point at high (low) c_1 via a saddle-node bifurcation; when c_1 reaches this bifurcation, exterior (interior) cells convert to interior (exterior) cells. In this way, cell type is regulated by c_1 .

Cell movement is described by Eq. 3 with

$$\mathbf{g}(\mathbf{r}^\alpha, S) = \sum_{\alpha' \neq \alpha} \mathbf{v}(\mathbf{r}^\alpha - \mathbf{r}^{\alpha'}) \quad (26)$$

where $\mathbf{v}(\mathbf{r}) = v_0 \left(\frac{r_{\min}^2}{r^3} - \frac{r_{\min}}{r^2} \right) \hat{\mathbf{r}}$, where $\hat{\mathbf{r}}$ is the unit vector in the \mathbf{r} direction and v_0 is a parameter that sets the velocity scale (see Table 1). We assume the motion is deterministic, so $\eta = 0$ in Eq. 3. The velocity $\mathbf{v}(\mathbf{r})$ moves cells separated by a distance greater than r_{\min} closer together (*e.g.*, response to an attractive force), but moves cells separated by less than r_{\min} further apart, corresponding to a close-ranged repulsive force. Hence, the cells that comprise the tissue adhere together with an average spacing of approximately r_{\min} .

The spatial profile of signaling molecule concentrations is key to the regulation of cell number and cell type. We assume rapid equilibration of signaling molecule concentrations (section III A), with c_ℓ^{ss} given by Eq. 6 with z -independent $c_\ell^{(1)}(\mathbf{r}) = C_\ell \exp(-r^2/(2s_\ell^2))$ for $\ell = 1, 2$ and $r = |\mathbf{r}|$. While the concentration field $c_\ell^{(1)}(\mathbf{r})$ due to a single cell can be computed from Eq. 7, it can also be well-approximated by a Gaussian spread function with appropriately chosen variance s_ℓ^2 (see Table 1). The concentration c_1 informs cells as to whether they are surrounded by many (high c_1) or few (low c_1) neighboring cells. The gene regulatory circuit ensures that cells with

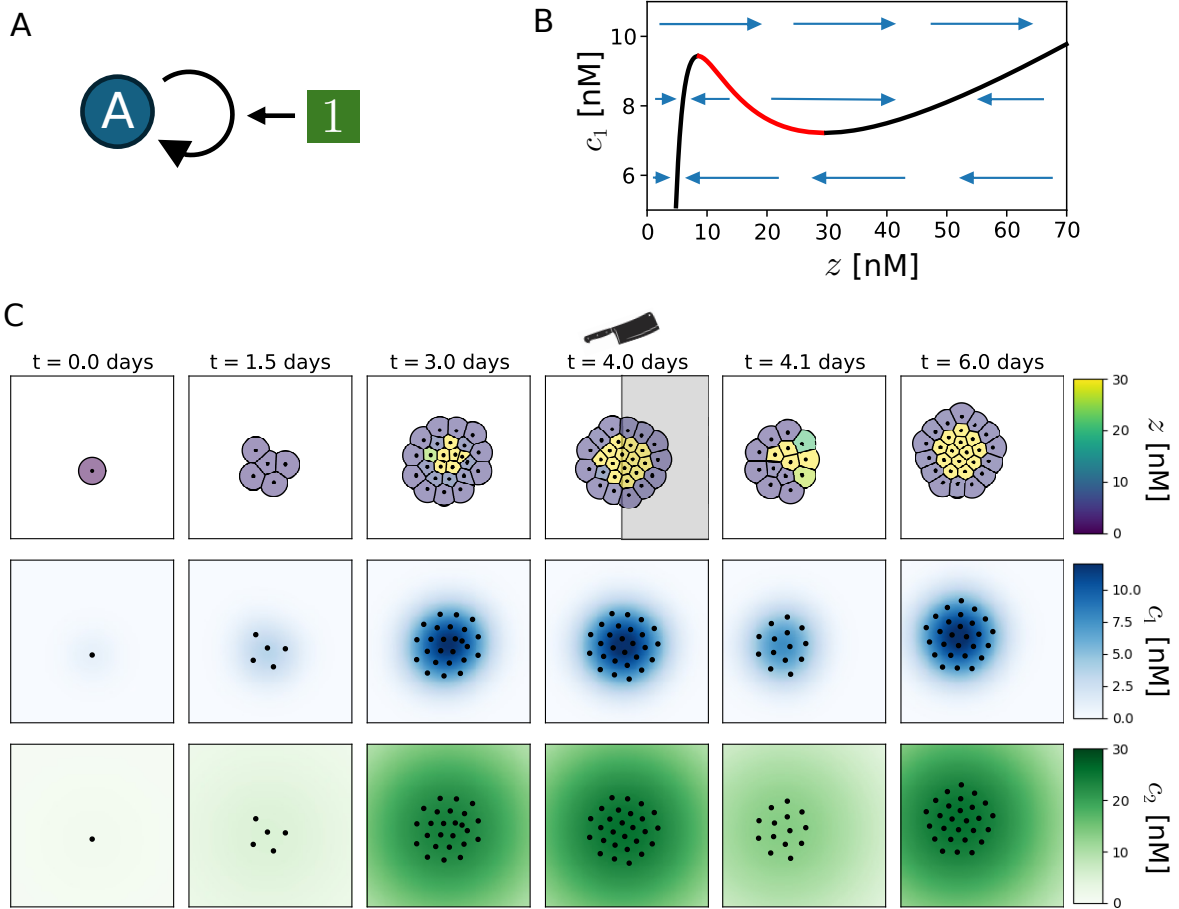


FIG. 4. Example 2: Model of a spatially structured organism with fissionous reproduction. (A) Gene regulatory circuit, where gene A upregulates itself in the presence of signaling molecule 1. (B) Waddington vector field (blue arrows), attractors (black curves) and unstable fixed points (red curve) as a function of c_1 . The lefthand attractor represents exterior cells and the righthand attractor represents interior cells. (C) Simulation of the dynamics, showing a single cell developing into a stable developed organism (day 0 through 4). At time $t = 4.0$ days, the organism is bisected and the left half is removed. The organism then regrows (4.1 to 6.0 days). Top row: cell position (black dots) and the state level z given by the color shading. Middle and bottom rows depict the corresponding concentrations c_1 and c_2 , respectively.

many neighbors become interior cells and cells with fewer neighbors become exterior cells.

As in Example 1, the concentration of molecule 2, c_2 , mediates the total cell population by influencing cell division rates, which we model by

$$\tilde{\beta}(\mathbf{c}; \Delta\mathbf{r}', \Delta\mathbf{r}'') = \begin{cases} \beta_0 \delta(\Delta\mathbf{r}') \mathcal{N}(\Delta\mathbf{r}'') & c_2 < c_2^* \\ 0 & c_2 \geq c_2^*, \end{cases} \quad (27)$$

where $\mathcal{N}(\mathbf{r}) \equiv \frac{1}{2\pi\xi^2} e^{-\mathbf{r}^2/2\xi^2}$ is the 2-dimensional Gaussian density function with variance ξ^2 . Note that the division rate is independent of z .

Fig. 4C shows a simulation of the model. Starting from a single cell at $t = 0$, the organism grows into a fully-developed organism in roughly four days. For approximately the first two days, all cells are of the exterior type (purple). After the organism grows sufficiently large that c_1 concentration reaches a sufficient level, conversion to interior-type cells (yellow) occurs. In Fig. 4C, the top

row indicates the positions and z -states of cells, the middle row shows signaling molecule 1 concentrations, and the bottom row shows signaling molecule 2 concentrations.

If the fully developed organism is bisected, each half regrows. In the simulation shown in Fig. 4C, the organism is bisected at time $t = 4.0$ days along the dotted line shown in the figure, and the righthand portion is removed. The remaining lefthand portion then regrows into a new fully-developed organism in approximately 2 days. These results illustrate an extremely robust developmental process.

VII. SUMMARY AND CONCLUSIONS

We have systematically constructed a general model for tissue development that incorporates intracellular gene expression dynamics, the effects of gene expres-

TABLE I. Model Parameters for Examples 1 and 2.

Parameter	Value [units]	Description
<i>Example 1 Model Parameters</i>		
a_1	1 [day ⁻¹]	Max. stem cell division rate (Eq. 19)
a_2	0.15 [unitless]	Parameter in Eq. 19
a_3	0.5 [unitless]	Parameter in Eq. 19
β_{S^*}	0.6 [day ⁻¹]	Triggered stem cell division rate (Eq. 19)
$\mathbf{w}(\mathbf{z})$	0.95($z_A, -z_B$) [z]	Controls asymmetric division (Eq. 19)
μ	0.02 [day ⁻¹]	Death rate
L_1, L_2	1 [nM]	Mean excretion rates of signaling molecules (Eqs. 20, 21)
z^S	15 [nM]	Threshold value of z_B between stem cells and triggered stem cells (Eqs. 19, 21)
z^D	5 [nM]	Threshold value of z_B between triggered stem and differentiated cells (Eqs. 19, 20)
$K_{E,m}$	5 [nM]	Equilibrium const. for tetramerization and binding of $m=A, B$ to Enhancer (Eq. 17)
$K_{S,m}$	5 [nM]	Equilibrium const. for tetramerization and binding of $m=A, B$ to Silencer (Eq. 17)
γ	20 [nM]	Equilibrium const. for signaling molecule 1 binding to B tetramer (Eq. 18)
$r_{0,A}$	500 [nM/day]	Expression rate of A when A enhancer and silencer of A are unoccupied (Eq. 16)
$r_{0,B}$	500 [nM/day]	Expression rate of B when B enhancer and silencer of B are unoccupied (Eq. 16)
$r(0)_{S,A}$	250 [nM/day]	Expression rate of A when A silencer is occupied and $c_1 = 0$ (Eq. 18)
$r(0)_{ES,A}$	1000 [nM/day]	Expression rate of A when A enhancer and silencer are occupied and $c_1 = 0$ (Eq. 18)
$r_{E,A}$	1500 [nM/day]	Expression rate of A when A enhancer is occupied (Eq. 16)
$r_{E,B}$	1000 [nM/day]	Expression rate of B when B enhancer is occupied (Eq. 16)
$r_{S,B}$	0 [nM/day]	Expression rate of B when B silencer is occupied (Eq. 16)
$r_{ES,B}$	500 [nM/day]	Expression rate of B when B enhancer and silencer are occupied (Eq. 16)
k_m	50 [day ⁻¹]	Degradation rate of $m=A, B$ (Eq. 16)
$\sigma(\mathbf{z}; \mathbf{c})$	0.1 $\begin{bmatrix} z_A & 0 \\ 0 & z_B \end{bmatrix}$ [day ^{-1/2}]	Noise magnitude of gene dynamics (Eq. 1)
<i>Example 2 Model Parameters</i>		
β_0	1 [day ⁻¹]	Division rate (Eq. 27)
ξ	2.5 [μm]	Std. dev. of displacement of daughter from mother (below Eq. 27)
c_2^*	23 [nM]	Threshold of c_2 above which cell division does not occur (Eq. 27)
μ	0.1 [day ⁻¹]	Death rate
C_ℓ	1 [nM]	Parameter controlling excretion rate of signaling molecule $\ell = 1, 2$
s_1	7 [μm]	Std. dev. of spatial profile of excretion of signaling molecule 1
s_2	30 [μm]	Std. dev. of spatial profile of excretion of signaling molecule 2
r_0	1000 [nM/day]	Expression rate of A when enhancer is unoccupied (Eq. 25)
r_1	2100 [nM/day]	Enhancement in expression rate when enhancer is occupied (Eq. 25)
K	35 [nM]	Equilibrium const. for dimerization and heterodimer-Enhancer binding (Eq. 25)
b	250 [day ⁻¹]	Degradation rate of protein A (Eq. 25)
σ	0.1 z [day ^{-1/2}]	Noise magnitude of gene dynamics (Eq. 1)
v_0	10 ³ [$\mu\text{m}^2/\text{day}$]	Parameter controlling velocity of cell movement (Eq. 26)
r_{\min}	8 [μm]	Spatial extent of attractive forces between cells (Eq. 26)
η	0	Noise magnitude of cell movement (Eq. 3)

sion state on proliferation and death rates, heritability of cellular states during cell division, and cell-cell interactions via chemical signaling. Dissecting our interacting-cell framework into individual components, we provide a mathematically unambiguous definition of a ‘‘Waddington landscape’’ (described as a Waddington vector field to allow for nonequilibrium, driven dynamics) and, if cell states are discretized, a developmental ‘‘fitness landscape.’’

Our general framework integrates submodels for four processes: stochastic changes in gene expression within each cell of the tissue (Eq. 1), chemical signals produced and received by cells affecting their gene expression dynamics (Eq. 2), spatial movement of cells (Eq. 3) which may affect their exposure to signaling molecules, and finally, stochastic cell proliferation (Eq. 10) that al-

lows daughter cells to acquire different states than their mother (through the differential birth rate parameter function of Eq. 4). These distinct processes are coupled together through the chemical signaling field \mathbf{c} , the drift and diffusivity of cells (which may depend on cell state \mathbf{z} and chemical field \mathbf{c}), and the differential birth rate which may also depend on mother-cell state and local chemical fields. These components represent an intermediate level of coarse-graining that allows for better interpretability than fine-grained computational models (such as CompuCell 3D [58]) or unwieldy kinetic theories [29, 30] that are difficult to marginalize to obtain equations for interpretable and measurable quantities, particularly if cells interact.

While the importance of cell-cell interactions and differential cell division and death rates for robust develop-

ment is well understood [34, 59], the complexity of incorporating these features into mathematical models has been an obstacle in modeling studies. Many theoretical works study cell fate transitions for single cells [3, 8, 13–20] or for populations of dividing but noninteracting cells without spatial structure [23–28]. For example, cell-cell interactions mediated by a chemical signaling concentration field, as in Eq. 2, typically cannot be incorporated into kinetic equations [29, 30, 60] in a straightforward way. Models that include spatial structure governed by cell-cell interactions, but which do not incorporate tissue growth via coordinated cell division or death, have also been studied [38, 61]. The models developed in [5, 6] incorporate cell-cell interactions and differential cell division rates to describe coordinated tissue development; these models also provide a more detailed description of the cell cycle than our framework. However, unlike our framework, they describe the cell population as a continuous distribution of infinitely many cells rather than describing individual cells as discrete entities, and they do not incorporate spatial structure.

We applied our intermediate-grained stochastic dynamical framework to two biological canonical processes: stem cell differentiation in a well-mixed environment and a spatially-structured growth and regeneration process. While the examples in this work are simplified models intended to provide conceptual clarity of general mechanisms of robust tissue development, one should be able to apply our framework to many other developmental processes that involve chemical signaling such as morphogenesis, embryogenesis, organogenesis, and patterning [33, 62, 63].

Appendix A: Mis-application of the Waddington Landscape Equations

The mis-application of Eqs. 14 and 15 to systems that are either non-gradient systems, or which have non-additive noise, can lead to artifacts in the reconstructed dynamics. As in Eq. 15, let $U_R(\mathbf{z}) = -\log p_{ss}(\mathbf{z})$ denote a reconstructed landscape where $p_{ss}(\mathbf{z})$ is the

steady state distribution of cell states generated by some stochastic dynamics $d\mathbf{z} = \mathbf{F}(\mathbf{z})dt + \sigma(\mathbf{z})d\mathbf{W}$. The reconstructed vector field $\mathbf{F}_R(\mathbf{z}) \equiv -\nabla U_R(\mathbf{z})$ can have ‘false’ attractors (attractors that are absent in $\mathbf{F}(\mathbf{z})$). An example of this is shown in Fig. 5A, with the false attractor indicated by a red square. Additionally, $\mathbf{F}_R(\mathbf{z})$ may fail to have attractors that do exist in $\mathbf{F}(\mathbf{z})$. An example of this is shown in Fig. 5B, with the true attractor that disappears in the reconstructed vector field indicated by an orange square.

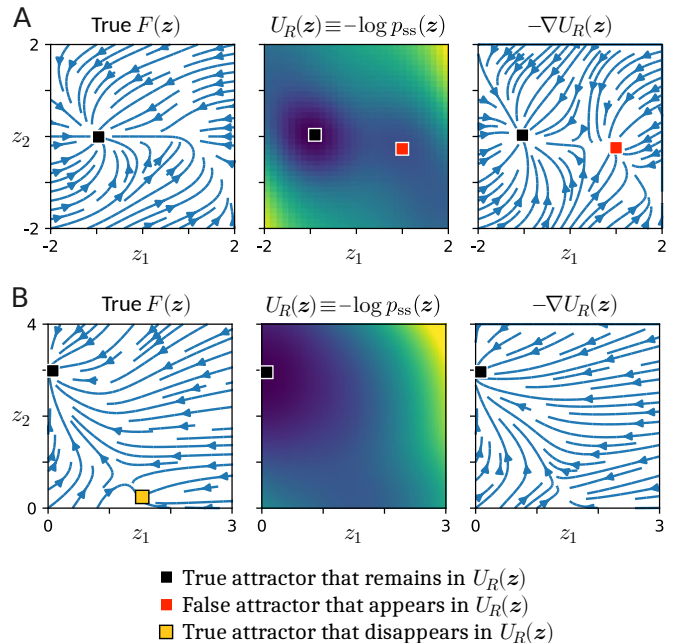


FIG. 5. Artifacts due to the mis-application of Eqs. 14 and 15. (A) An example showing a false attractor (red square) which does not exist in the true vector field, but which appears in the reconstructed vector field. (B) An example showing a true attractor which disappears in the reconstructed vector field (orange square). The reconstructed landscapes $U_R(\mathbf{z})$ are color coded where blue represents small values and yellow represents large values.

[1] Conrad H Waddington. Canalization of development and the inheritance of acquired characters. *Nature*, 150(3811):563–565, 1942.
[2] James E Ferrell. Bistability, bifurcations, and Waddington’s epigenetic landscape. *Current Biology*, 22(11):R458–R466, 2012.
[3] Geoffrey Schiebinger. Reconstructing developmental landscapes and trajectories from single-cell data. *Current Opinion in Systems Biology*, 27:100351, 2021.
[4] Leili Shahriyari and Natalia L Komarova. Symmetric vs. asymmetric stem cell divisions: an adaptation against cancer? *PLoS ONE*, 8(10):e76195, 2013.

[5] Jinzhi Lei. A general mathematical framework for understanding the behavior of heterogeneous stem cell regeneration. *Journal of Theoretical Biology*, 492:110196, 2020.
[6] Jinzhi Lei. Mathematical modeling of heterogeneous stem cell regeneration: from cell division to Waddington’s epigenetic landscape. *arXiv preprint arXiv:2309.08064*, 2023.
[7] Jin Wang, Li Xu, and Erkang Wang. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences*,

- 105(34):12271–12276, 2008.
- [8] David A Rand, Archishman Raju, Meritxell Sáez, Francis Corson, and Eric D Siggia. Geometry of gene regulatory dynamics. *Proceedings of the National Academy of Sciences*, 118(38):e2109729118, 2021.
- [9] Megan A Coomer, Lucy Ham, and Michael PH Stumpf. Noise distorts the epigenetic landscape and shapes cell fate decisions. *Cell Systems*, 13(1):83–102, 2022.
- [10] Meritxell Sáez, Robert Blassberg, Elena Camacho-Aguilar, Eric D Siggia, David A Rand, and James Briscoe. Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. *Cell Systems*, 13(1):12–28, 2022.
- [11] Peijie Zhou and Tiejun Li. Construction of the landscape for multi-stable systems: Potential landscape, quasi-potential, A-type integral and beyond. *The Journal of Chemical Physics*, 144(9), 2016.
- [12] Archishman Raju, BingKan Xue, and Stanislas Leibler. A theoretical perspective on Waddington’s genetic assimilation experiments. *Proceedings of the National Academy of Sciences*, 120(51):e2309760120, 2023.
- [13] Daniel R Sisan, Michael Halter, Joseph B Hubbard, and Anne L Plant. Predicting rates of cell state change caused by stochastic fluctuations using a data-driven landscape model. *Proceedings of the National Academy of Sciences*, 109(47):19262–19267, 2012.
- [14] Jin Wang, Kun Zhang, Li Xu, and Erkang Wang. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences*, 108(20):8257–8262, 2011.
- [15] Atefeh Taherian Fard, Sriganesh Srihari, Jessica C Mar, and Mark A Ragan. Not just a colourful metaphor: modelling the landscape of cellular development using Hopfield networks. *NPJ Systems Biology and Applications*, 2(1):1–9, 2016.
- [16] Xiaojie Qiu, Yan Zhang, Jorge D Martin-Rufino, Chen Weng, Shayan Hosseinzadeh, Dian Yang, Angela N Pogson, Marco Y Hein, Kyung Hoi Joseph Min, Li Wang, et al. Mapping transcriptomic vector fields of single cells. *Cell*, 185(4):690–711, 2022.
- [17] Anna-Simone Josefine Frank, Kamila Larripa, Hwayeon Ryu, and Susanna Röblitz. Macrophage phenotype transitions in a stochastic gene-regulatory network model. *Journal of Theoretical Biology*, 575:111634, 2023.
- [18] Yoshiyuki T Nakamura, Yusuke Himeoka, Nen Saito, and Chikara Furusawa. Evolution of hierarchy and irreversibility in theoretical cell differentiation model. *PNAS NEXUS*, 3(1):pgad454, 2024.
- [19] Gang Xue, Xiaoyi Zhang, Wanqi Li, Lu Zhang, Zongxu Zhang, Xiaolin Zhou, Di Zhang, Lei Zhang, and Zhiyuan Li. A logic-incorporated gene regulatory network deciphers principles in cell fate decisions. *Elife*, 12:RP88742, 2024.
- [20] Xinxin Chen, Ying Sheng, Liang Chen, Moxun Tang, and Feng Jiao. Quantifying cell fate change under different stochastic gene activation frameworks. *Quantitative Biology*, 13(1):e82, 2025.
- [21] Weikang Wang, Dante Poe, Yaxuan Yang, Thomas Hyatt, and Jianhua Xing. Epithelial-to-mesenchymal transition proceeds through directional destabilization of multidimensional attractor. *eLife*, 11:e74866, feb 2022.
- [22] Weikang Wang, Ke Ni, Dante Poe, and Jianhua Xing. Transiently increased coordination in gene regulation during cell phenotypic transitions. *PRX Life*, 2:043009, 2024.
- [23] Caleb Weinreb, Samuel Wolock, Betsabeh K Tusi, Merav Socolovsky, and Allon M Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, 2018.
- [24] David S Fischer, Anna K Fiedler, Eric M Kernfeld, Ryan MJ Genga, Aimée Bastidas-Ponce, Mostafa Bakhti, Heiko Lickert, Jan Hasenauer, Rene Maehr, and Fabian J Theis. Inferring population dynamics from single-cell rna-sequencing time series data. *Nature Biotechnology*, 37(4):461–468, 2019.
- [25] Jifan Shi, Tiejun Li, Luonan Chen, and Kazuyuki Aihara. Quantifying pluripotency landscape of cell differentiation from scRNA-seq data by continuous birth-death process. *PLoS Computational Biology*, 15(11):e1007488, 2019.
- [26] Stephen Zhang, Anton Afanassiev, Laura Greenstreet, Tetsuya Matsumoto, and Geoffrey Schiebinger. Optimal transport analysis reveals trajectories in steady-state systems. *PLoS Computational Biology*, 17(12):e1009466, 2021.
- [27] Aden Forrow and Geoffrey Schiebinger. LineageOT is a unified framework for lineage tracing and trajectory inference. *Nature Communications*, 12(1):4940, 2021.
- [28] Jifan Shi, Kazuyuki Aihara, Tiejun Li, and Luonan Chen. Energy landscape decomposition for cell differentiation with proliferation effect. *National Science Review*, 9(8):nwac116, 2022.
- [29] Mingtao Xia and Tom Chou. Kinetic theory for structured populations: application to stochastic sizer-timer models of cell proliferation. *Journal of Physics A: Mathematical and Theoretical*, 54(38):385601, 2021.
- [30] Mingtao Xia and Tom Chou. Kinetic theories of state- and generation-dependent cell populations. *Physical Review E*, 110:064146, 2024.
- [31] Chris D. Greenman and Tom Chou. Kinetic theory of age-structured stochastic birth-death processes. *Physical Review E*, 93:012112, 2016.
- [32] Tom Chou and Chris D. Greenman. A hierarchical kinetic theory of birth, death and fission in age-structured interacting populations. *J. Stat. Phys.*, 164:49–76, 2016.
- [33] Michael Barresi and Scott Gilbert. *Developmental Biology, 13th Ed.* Oxford University Press, 2023.
- [34] Dennis E Discher, David J Mooney, and Peter W Zandstra. Growth factors, matrices, and forces combine and control stem cells. *Science*, 324(5935):1673–1677, 2009.
- [35] Sui Huang. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer and Metastasis Reviews*, 32:423–448, 2013.
- [36] Sui Huang. Reconciling non-genetic plasticity with somatic evolution in cancer. *Trends in Cancer*, 7(4):309–322, 2021.
- [37] The labeling of cells is arbitrary. One straightforward labeling scheme is to label cells from oldest to youngest, *i.e.*, cell $\alpha = 1$ is the oldest cell. When a cell dies, all cells younger than the dying cell are then re-labeled.
- [38] Hosein Fooladi, Parsa Moradi, Ali Sharifi-Zarchi, and Babak Hosein Khalaj. Enhanced Waddington landscape model with cell–cell communication can explain molecular mechanisms of self-organization. *Bioinformatics*, 35(20):4081–4088, 2019.
- [39] Niraj Kumar, Thierry Platini, and Rahul V. Kulkarni. Exact distributions for stochastic gene expression models

- with bursting and feedback. *Phys. Rev. Lett.*, 113:268105, 2014.
- [40] Chen Jia and Ramon Grima. Small protein number effects in stochastic models of autoregulated bursty gene expression. *The Journal of Chemical Physics*, 152(8):084115, 2020.
- [41] Xinyu Wang, Youming Li, and Chen Jia. Poisson representation: a bridge between discrete and continuous models of stochastic gene regulatory networks. *Journal of The Royal Society Interface*, 20(208):20230467, 2023.
- [42] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [43] Daniel T Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.
- [44] Wang Xi, Thuan Beng Saw, Delphine Delacour, Chwee Teck Lim, and Benoit Ladoux. Material approaches to active tissue mechanics. *Nature Reviews Materials*, 4(1):23–44, 2019.
- [45] Jonas Ranft, Markus Basan, Jens Elgeti, Jean-Francois Joanny, Jacques Prost, and Frank Jülicher. Fluidization of tissues by cell division and apoptosis. *Proceedings of the National Academy of Sciences*, 107(49):20863–20868, 2010.
- [46] Sarah A Nowak, B Chakrabarti, Tom Chou, and Ajay Gopinathan. Frequency-dependent chemolocation and chemotactic target selection. *Physical Biology*, 7(2):026003, 2010.
- [47] Peter Dieterich, Rainer Klages, Roland Preuss, and Albrecht Schwab. Anomalous dynamics of cell migration. *Proceedings of the National Academy of Sciences*, 105(2):459–463, 2008.
- [48] Tajie H Harris, Edward J Banigan, David A Christian, Christoph Konradt, Elia D Tait Wojno, Kazumi Norose, Emma H Wilson, Beena John, Wolfgang Weninger, Andrew D Luster, et al. Generalized Lévy walks and the role of chemokines in migration of effector CD8+ T cells. *Nature*, 486(7404):545–548, 2012.
- [49] Felix Höfling and Thomas Franosch. Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics*, 76(4):046602, 2013.
- [50] Sumin Jang, Sandeep Choubey, Leon Furchtgott, Ling-Nan Zou, Adele Doyle, Vilas Menon, Ethan B Loew, Anne-Rachel Krostag, Refugio A Martinez, Linda Madisen, Boaz P Levi, and Sharad Ramanathan. Dynamics of embryonic stem cell differentiation inferred from single-cell transcriptomics show a series of transitions through discrete cell states. *eLife*, 6:e20487, 2017.
- [51] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Reviews of modern physics*, 62(2):251, 1990.
- [52] R. Dessalles, M. D’Orsogna, and T. Chou. Exact steady-state distributions of multispecies birth–death–immigration processes: Effects of mutations and carrying capacity on diversity. *J. Stat. Phys.*, 173:82–221, 2018.
- [53] Paras Jain, Ramanarayanan Kizhuttill, Madhav B Nair, Sugandha Bhatia, Erik W Thompson, Jason T George, and Mohit Kumar Jolly. Cell-state transitions and density-dependent interactions together explain the dynamics of spontaneous epithelial-mesenchymal heterogeneity. *iScience*, 2024.
- [54] Crispin W Gardiner. Handbook of stochastic methods for physics, chemistry and the natural sciences. *Springer Series in Synergetics*, 1985.
- [55] Sui Huang, Yan-Ping Guo, Gillian May, and Tariq Enver. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, 305(2):695–713, 2007.
- [56] Thomas Hunt Morgan. *Experimental studies of the regeneration of Planaria maculata*, volume 2. W. Engelmann, 1898.
- [57] Daniel Lobo, Wendy S Beane, and Michael Levin. Modeling planarian regeneration: a primer for reverse-engineering the worm. *PLoS Computational Biology*, 8(4):e1002481, 2012.
- [58] M. Swat, Gilberto L. Thomas, Julio M. Belmonte, A. Shirinifard, D. Hmeljak, and J. A. Glazier. Multi-scale modeling of tissues using compuCell3D. *Computational Methods in Cell Biology, Methods in Cell Biology*, 110:325–366, 2012.
- [59] Benoit G Godard and Carl-Philipp Heisenberg. Cell division and tissue mechanics. *Current Opinion in Cell Biology*, 60:114–120, 2019.
- [60] Mingtao Xia, Xiangting Li, and Tom Chou. Overcompensation of transient and permanent death rate increases in age-structured models with cannibalistic interactions. *Physica D: Nonlinear Phenomena*, 470:134339, 2024.
- [61] Matthew Smart and Anton Zilman. Emergent properties of collective gene-expression patterns in multicellular systems. *Cell Reports Physical Science*, 4(2), 2023.
- [62] Jeremy B. A. Green and James Sharpe. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development*, 142(7):1203–1211, 04 2015.
- [63] Hans Meinhardt. Hierarchical inductions of cell states: A model for segmentation in Drosophila. *Journal of Cell Science*, 1986:357–381, 1986.