J. Indian Inst. Sci.
A Multidisciplinary Reviews Journal
ISSN: 0970-4140 Coden-JIISAD

REVIEW ARTICLE

The Mathematics of Phenotypic State Transition: Paths and Potential

*Vimalathithan Devaraj and Biplab Bose**

Abstract | Change in the phenotype of a cell is considered as a transition of a cell from one cellular state to another. Cellular state transition can be driven by an external cue or by the noise in molecular processes. Over the years, generalized physical principles, and associated mathematical models have been developed to understand phenotypic state transition. Starting with Waddington's epigenetic landscape, phenotypic state transition is seen as a movement of cells on a potential landscape. Though the landscape model is close to the thermodynamic principles of state change, it is difficult to envisage it from experimental observations. Therefore, phenotypic state transition is often considered as a discrete state jump process. This approach is particularly useful to estimate the paths of state transition from experimental observations. In this review, we discuss both of these approaches and the associated mathematical formulations. Furthermore, we explore the opportunities to connect these two approaches and the limitations of our current understanding and mathematical methods.

1 Introduction

By definition, 'phenotype' refers to the observable properties of an organism[1]. Following this definition, the phenotype of a cell refers to the specific physical and functional properties of a cell. For example, in metazoans, cells of the epithelial phenotype are polarized, non-migratory, and form tightly packed cell layers[2]. On the other hand, mesenchymal cells are migratory and have more irregular morphology. During embryonic development, epithelial cells in the primitive ectoderm undergo de-epithelization and adopt mesenchymal phenotype, leading to the formation of mesoderm[3]. This process is known as Epithelial to Mesenchymal Transition (EMT). The opposite of this process is called Mesenchymal to Epithelial Transition (MET). During embryonic development, the formation of three-dimensional structures of specialized cells, requires several rounds of EMT and MET. EMT is also involved in wound healing, tissue regeneration and cancer metastasis[3].

Many other biological phenomena, like the *differentiation of stem cells*[4], the emergence of drug-resistant cancer cells[5, 6], and *cancer stem-like cells*[7–9], also involve switching between different cellular phenotypes. Phenotypic switching does not involve any genomic alternation but is driven by the change in gene expression and possibly epigenetic alternations[10]. Switching can be driven by an external signal[4] or can happen spontaneously[8], even in the absence of an external cue.

The change in the phenotype of a cell is similar to change in the state of a physical system defined in terms of macroscopic properties, like pressure, volume. Similarly, cells of a phenotype are considered to be in that phenotypic state, and switching of the phenotype is called phenotypic state transition. Phenotypic diversification also happens through *asymmetric cell division and symmetric differentiation of stem cells*[11]. However, such phenotypic diversification through cell division is not discussed in this article.

Differentiation: It is a process by which a less specialized cell undergoes changes to become a specialized cell type.

Stem cell: Cell that can divide indefinitely, thereby renewing itself and can also give rise to specialized cells by differentiation.

Cancer stem-like cells: Cancer cells that can self-renew and gives rise to heterogenous lineages of cancer cells

Asymmetric cell division: Cell division by mitosis gives rise to two daughter cells. Usually these two daughter cells are of same type. In contrast, an asymmetric cell division gives rise to daughter cells of two different types. Asymmetric division of a stem cell gives rise to daughter cells with different fates- one cell differentiate to a particular cell lineage and the other maintains the stemness.

Symmetric differentiation: In symmetric differentiation, a stem cell divides to give rise two daughter cells that differentiate to the same cell lineage.

[1] *Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, India.*
biplabbose@iitg.ac.in

Cellular phenotypic states are often defined in terms of expression of specific molecules or markers[5, 7–9, 12]. For example, molecular markers, such as E-cadherin, N-cadherin, and Vimentin, are used to define different states during EMT/MET[13, 14]. Similarly, stem cells and cells at different stages of differentiation are also identified by molecular markers[15]. One can identify different types of cells, in a population, in terms of molecular markers using techniques like microscopy and flow cytometry. One can also segregate and isolate these cells based on the differential expression of molecular markers. High-throughput experiments, such as microarray-based gene expression analysis[16] and single-cell gene expression analysis[17], now allow biologists to categorize cell states in terms of genome-wide gene expression.

Though extensively used to define phenotypic states, gene expression is a surrogate measure of phenotype. Most cellular functions and physical properties are regulated by processes involving a large number of molecules. Therefore, it will be appropriate to define the phenotypic states directly using quantifiable functional and/or physical features. For example, changes in cell morphology have been used to understand the phenotypic state transition in EMT[18, 19]. Quantitative image analysis using machine learning tools allows the analysis of a large number of microscopy images and categorizes cells in different phenotypic states[18, 20–23]. The cellular phenotype can also be defined in terms of cell motility, which can be measured by quantitative image analysis[24]. Rimchala et al.[25] used time-lapse microscopy, followed by image analysis, to study phenotypic state transition in endothelial cells treated with angiogenic and angiostatic cytokines. They defined phenotypic states in terms of cell morphology and movement-related features.

Irrespective of our definition of a phenotypic state, a key problem in experimental biology is to estimate the paths of phenotypic state transition from experimental data. Different mathematical techniques have been developed for this purpose. There is also a growing interest in developing a generic mathematical formulation to capture the physical basis of phenotypic state transition. Concepts and tools of dynamical systems theory and statistical physics have been utilized to address this problem. In this review article, we focus on these two directions of research on the phenotypic state transition. First, we discuss the potential landscape framework for cellular state transition and its mathematical formulations. Subsequently, we delve into discrete state transition models and different approaches in the estimation of state transition paths.

Steady state: At a steady state the state variables of the system does not change with time. For a biochemical circuit, at a steady state the concentrations of the molecules involved in the circuit do not change with time. If the dynamics of the system is represented by $\dot{x} = f(x)$, then at steady state x^*, $f(x^*)=0$. A steady state can be either stable or unstable. If the system is perturbed slightly from a stable steady state, with time it will return to the stable steady state. A steady state is also called a fixed point.

2 Rolling a Ball Downhill

The emergence of a multicellular organism with an elaborate body plan from a single fertilized cell has always intrigued scientists. In 1957, Conrad Hal Waddington proposed a metaphor of landscape to explain the directionality of embryonic development and the stability of different stages of development[26]. In this model, a cell is just like a pebble that's rolling over a rugged landscape with hills, slopes, and valleys. A pluripotent cell starts rolling from a high position in this landscape and rolls down through a series of branching points where the decision of differentiation takes place. Therefore, the landscape channelizes the cell through specific paths of differentiation, leading to the formation of different differentiated somatic cells. Without going into details of the molecular processes that generate this landscape, one can argue that external interventions, like change in temperature or treatment with a drug, may change the landscape, thereby affecting the dynamics of differentiation through it.

Waddington's landscape has two key ideas—(1) time evolution of a cell from higher potential to lower and (2) alternate paths leading to different lower potential positions. Efforts from different directions have been made to provide a physical and mathematical basis to these two aspects of the landscape metaphor. One approach is to create dynamical models for molecular circuits that regulate phenotypic state transitions.

2.1 Bifurcation and State Transition

A dynamical model of a regulatory circuit is a system of ordinary differential equations (ODEs). It captures the change in concentrations of molecules in the circuit with time. The time evolution of such a dynamical system can be represented by the trajectories in the phase space. For a system with one or more stable *steady states*, the trajectories in the phase space will converge to these stable steady states[27]. In cell biology, each axis of the phase space represents the concentration of one of the molecules involved in the process. A cell will start from a particular position in the phase space and then follow a specific trajectory to reach a particular stable steady state. Each of these stable steady states represents a particular phenotypic state.

Often such a dynamical system shows bifurcation. In bifurcation, the qualitative behavior of the system changes with change in the value of a

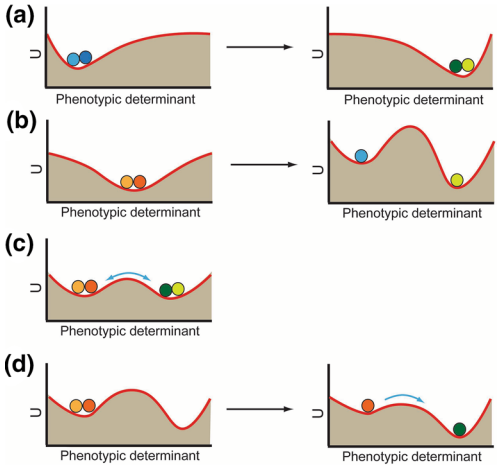

Figure 1: The landscape model for phenotypic state transition. In these examples, the phenotypic state is defined in terms of the concentration of a molecule called the phenotypic determinant. *U* is the potential. Local and global minima are different stable phenotypic states. Colored circles represent cells. In **a**, **b** an input signal causes a change in the potential landscape. In **c**, cells move from one minimum to another due to stochastic fluctuations in molecular processes. In **d**, the external signal changes the potential landscape, facilitating the transition from one minimum to another by stochastic fluctuation.

critical parameter[27]. For example, with a change in a critical parameter, the number of possible stable steady states may change. In a typical phenotypic state transition model, change in the bifurcation parameter changes the system from monostable to bi- or multi-stable. An external stimulus usually regulates the bifurcation parameter. When the value of the bifurcation parameter changes, the cell moves from one steady state to another stable one. Such a transition from one stable steady state to another one is manifested as the phenotypic state transition of the cell.

The concept of bifurcation has been successfully utilized to provide the mechanistic explanation of state transition during EMT[28–31], and differentiation[32–35]. For further details on the mathematics of dynamical modeling and bifurcation analysis in the phenotypic state transition, we refer the readers to the following articles[36–38].

2.2 The Potential Landscape of a Cell

Mechanistic dynamical models and bifurcation theory explain the existence of multiple stable phenotypic states. To get closer to Waddington's metaphor of a landscape, we need to generate a potential landscape from a dynamical model. Imagine, we have an ODE-based model for a particular phenotypic state transition, and we have derived a potential landscape from this model. In reality, the potential *U* will be a multi-dimensional function of concentrations of several molecules. However, for the ease of discussion and pictorial representation, let us consider that *U* depends upon the concentration of one molecule that we call a phenotypic determinant. The concentration of this molecule decides the phenotypic state of a cell. Therefore, the potential landscape is now just a curve in a two-dimensional space (Fig. 1).

Consider the example in Fig. 1a. In this case, *U* has only one minimum that represents a stable steady state, a stable phenotypic state. Cells starting with different levels of the phenotypic determinant will roll down spontaneously to this minimum and achieve that specific phenotype. Suppose, these cells are now treated with an external cue that changes the parameter values of ODEs. This may lead to a change in the shape of *U* and the minimum shifts to a higher value on the horizontal axes. This will force all the cells to move to this new stable steady state and they will acquire a new phenotype.

Figure 1b shows a similar phenomenon but a slightly more complicated one. In this example, *U* has one minimum, and all the cells are in that state. The external signal changes *U* such that *U* now has two new minima and the older one has disappeared. These new minima are stable steady states and represent two different phenotypic states. This is an example of bifurcation where the system is initially monostable but the external signal changes it to bistable. As the potential landscape has changed, cells will be forced to move to either of these two new minima. Though all cells were earlier present at the global minimum, they will have variation in their molecular states due to stochasticity in processes like gene expression. Such stochastic variation will decide which of the two new minima a cell will move to.

Phenotypic state transition through these types of signal-induced change in the potential landscape can explain the state transition observed in molecular signal-induced EMT[30], cell differentiation[35]. However, sometimes, cells can spontaneously jump from one phenotypic state to another, even in the absence of any external cue[8]. A potential landscape model can explain such spontaneous state transitions. Suppose *U* has two minima (Fig. 1c), and both are occupied by cells. Remember that molecular processes in every cell are noisy and that noise can change the molecular

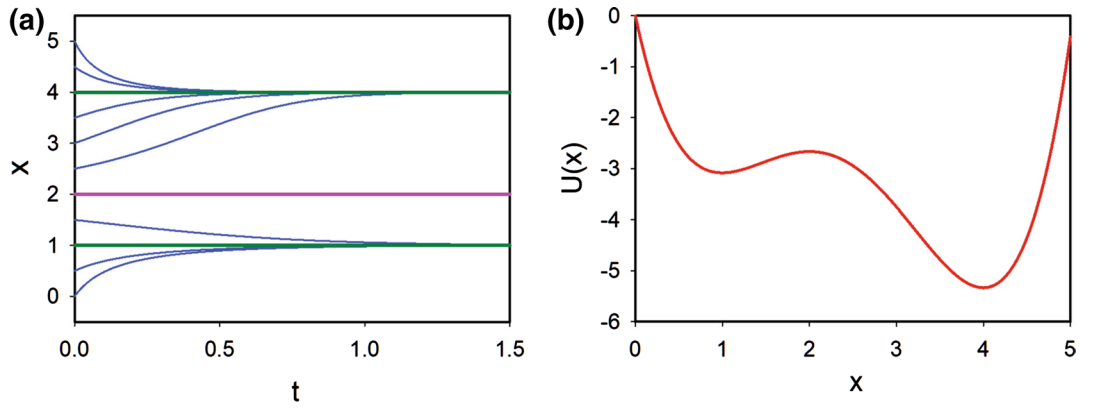

Figure 2: Dynamics and potential landscape for a one-dimensional system. The dynamics of the system follows $\frac{dx}{dt} = (1-x)(x-4)(x-2)$. **a** The trajectories of the system (blue lines) starting from different initial conditions. Green horizontal lines are stable steady states, and the pink horizontal line is an unstable steady state. **b** The potential $U(x)$ as a function of x.

state of a cell stochastically. If the energy barrier (the height of the peak between two minima) is low enough, then the noise-induced changes in the molecular state are good enough to push a cell from one minimum to another. This will manifest as a spontaneous phenotypic state change. If we observe for long, these stochastic transitions will lead to a steady-state distribution of cells in two phenotypic states. If we isolate cells in a particular state and culture those, the populations should again relax to the steady-state distribution observed earlier[8].

An external signal can work in tandem with the stochastic fluctuation. Suppose, initially, the energy barrier between two stable states was high, and the stochastic fluctuation can not force cells to cross this barrier. An external signal may reduce this barrier and facilitate the transition of cells from one state to another (Fig. 1d).

2.3 Estimating the Potential Landscape

The idea of a potential landscape is a powerful tool to understand the phenotypic state transition. However, how do we derive the potential U from our ODE-based model? In classical mechanics, the idea of potential is used to understand the motion of a particle. If free, a particle moves from a higher potential to lower potential. Let the motion of the particle is given by $\frac{dx}{dt} = f(x)$ and $U(x)$ be its potential at position x. At a mechanical equilibrium, $\frac{dU(x)}{dt} = 0$ and around this equilibrium, $\frac{dU(x)}{dt} < 0$. These conditions are met if we define the potential $U(x)$ as[27]:

$$f(x) = -\frac{dU(x)}{dx}. \tag{1}$$

Figure 2a shows the dynamics of a system for $\frac{dx}{dt} = (1-x)(x-4)(x-2)$. This system has two stable steady state ($x=1$ and 4) and one unstable steady state ($x=2$). The blue lines in Fig. 2a show the trajectories of the system, starting with different initial positions. As shown in this figure, with time, the system moves towards any of the two stable equilibriums. Figure 2b shows the potential of the system as per Eq. (1). As expected, the stable equilibriums, $x=1$ and 4 are the minima.

Differential equation-based mechanistic models for cellular state transition involve multiple dependent variables and use system of ODEs. For a system of ODEs, the relation in Eq. (1) can be written as:

$$\dot{\mathbf{x}} = -\nabla U(\mathbf{x}). \tag{2}$$

Here, $\dot{\boldsymbol{x}}$ is the derivative vector, and $U(\boldsymbol{x})$ is a continuously differentiable single valued scalar function of $\boldsymbol{x}$. If such a potential function, $U(\boldsymbol{x})$, exists, the system is called a gradient system[27].

However, most dynamical models in biology, including those for phenotypic state transition, are non-gradient systems, and we cannot derive the exact potential function from an ODE-based model[39]. In a gradient system, the energy is conserved. However, most biological systems are away from thermodynamic equilibrium, and energy is not conserved. Therefore, alternative approaches have been developed to calculate a pseudo-potential landscape to explain the transition from one phenotypic state to another.

The commonest approach for this is to use the concept of Boltzmann distribution of statistical physics. Consider that a system can be in discrete states that belong to different energy levels. If the average energy of the system remains constant, then using the principle of maximum entropy, it can be shown that the probability of the system being in the jth state is given by an exponential relation[40]:

$$p_j = \frac{e^{-U_j}}{\sum_{\text{all } j} e^{-U_j}} = \frac{e^{-U_j}}{Z}. \quad (3)$$

Here, U_j is the energy of jth state and $Z = \sum_{\text{allj}} e^{-U_j}$ is a normalization constant. Equation (3) can be re-arranged to represent U in terms of p:

$$U_j \propto -\ln(p_j). \quad (4)$$

Equation (4) re-iterates the idea that low-energy states will be more densely populated than the higher energy states.

It has been proposed that this relation between U and p, holds even for non-equilibrium systems at steady-state[41, 42]. Therefore, if we can calculate the probability distribution of different phenotypic states in the steady state, we should be able to calculate and plot U. Suppose we have a dynamical model for a molecular network. We perform stochastic simulation of this model for a very long duration and record the trajectories of the system through the phase space. Stochastic simulation can be performed using various approaches such as Gillespie algorithm and stochastic differential equations[43, 44]. Subsequently, we divide the whole phase space into small volumes, and from the simulated data, calculate the frequency of observing the system in each of these small volumes of the phase space.

This frequency is equivalent to the probability of the system being in a particular region in the phase space. Following Eq. (4), we can calculate U using these estimated probabilities.

One can also use Fokker–Planck equation to calculate the steady-state probability distribution of the system and calculate U using Eq. (4) [45, 46]. For a multi-dimensional system, the estimated U would be a continuous hypersurface. It is challenging to visualize U graphically. One can visualize it, with respect to two key molecules involved in the process, in a three-dimensional plot[47–49].

Potential U estimated by this approach is different from the potential in equilibrium thermodynamics and should be considered as a pseudo-potential. It is dimensionless and is sensitive to the noise in the system[50]. In an equilibrium system, the gradient of the potential drives the time evolution of the system. However, for a non-equilibrium cellular system, the potential estimated above is only one component of the driving force[42, 50, 51].

In several studies on phenotypic state transition, this approach has been used to estimate the potential landscape[43, 46, 48, 50, 52–54]. Others have used the concept of quasi-potential from the large deviation theory[55] and the fluctuation–dissipation theorem to generate the potential landscape[41].

3 Bouncing Over Discrete States

In dynamical systems theory and the potential landscape model, the cellular state space is considered as a continuous space, with distinct fixed points that represent different phenotypic states. The state space is defined in terms of the expression of specific molecules. Therefore, this approach is suitable for problems where we have defined phenotypic states in terms of molecular markers, and we have the knowledge of the molecular regulators.

One can argue that a discrete phenotypic state, like the ability to migrate or elongated shape, is not achieved by a unique molecular state. Cellular phenotypes are robust to usual variations in gene expression, and different molecular states can give rise to the same phenotypic state. In fact, in experimental biology, a phenotypic state is not defined by a unique value of expression of a marker, but by a range, like low, high, or medium. However, in the potential landscape model discussed above, a phenotypic state is a unique point in the state space defined by molecular expression.

These issues can be circumvented if we try to understand phenotypic state transition phenomenologically as a discrete state transition problem. This allows us to use any definition of phenotypic states as long as we define them discretely. Once defined, we can measure the distribution of cells in different states over multiple time points. Subsequently, we can use appropriate mathematical tools to study the dynamics of state transition.

3.1 State transition as a Markov process

Discrete state transition is often considered as a Markov process. A Markov process is a stochastic process where the system stochastically jumps from one discrete state to another with

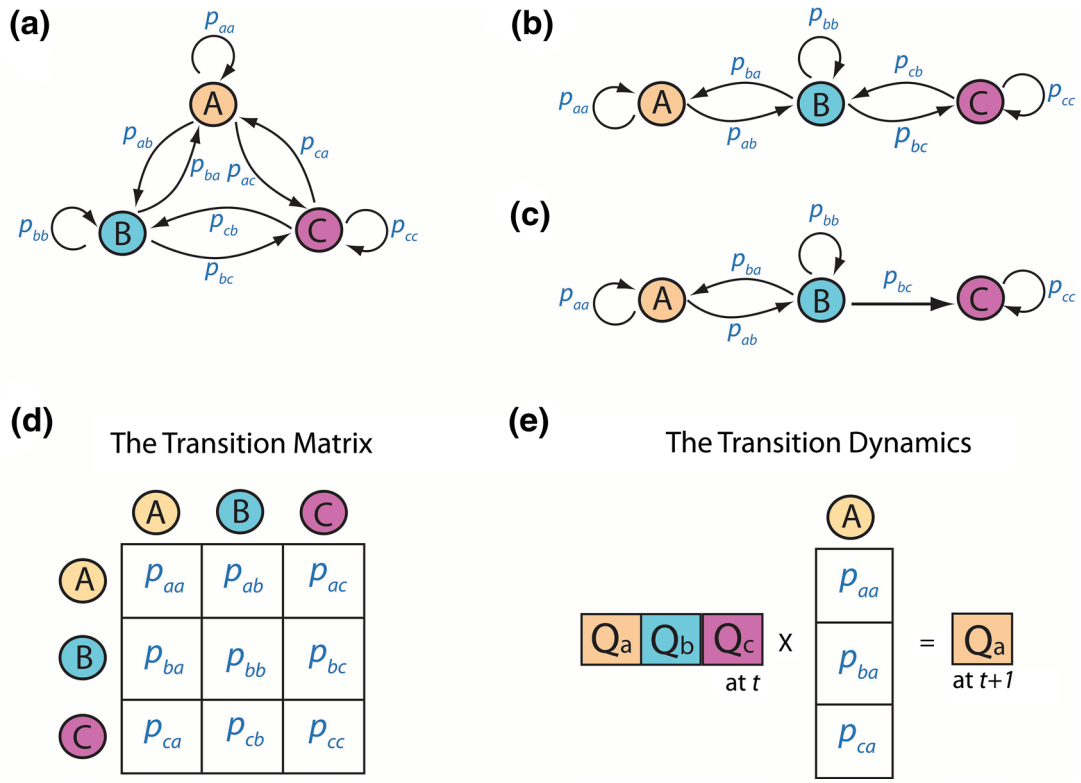

Figure 3: Discrete Markov model for cellular state transition. **a** Three state model where a cell can move from any state to any other state directly. **b** Sequential state transition. **c** State transition with an absorbing state. Here C is the absorbing state. **d** Shows the transition matrix ***P*** for the 3-state model. Each element of this matrix is the probability of transition from one state to another. For example, p_{bc} is the probability of transition from B to C. Sum of elements in a row in this matrix is equal to 1. **e** Visual representation of the mathematical method, of matrix multiplication, to estimate the probability of a cell in the state A (Q_A) at time ($t+1$), from the probability distribution of cell in three states (Q_A, Q_B, Q_C) at the previous time point (t) and a part of the transition matrix. Q_B and Q_C at the time ($t+1$) can be calculated using the same method.

a probability that depends only upon the current state of the system[56]. Figure 3a shows a three-state Markov process for a cellular system. At a particular time, a cell can be in any of these three states. In the next interval, the cell will either jump to any of the two other states or stay in the same state. The probabilities of transition over these paths are given over the arrows (Fig. 3a). The transition from one state to another can be sequential, too (Fig. 3b). Furthermore, a Markov process can have an absorbing state, such that the cells that have reached that state will not move to any other state (Fig. 3c).

Generalizing these concepts, let us consider that a cell can be in any of the k states, $S = \{S_1, S_2, \ldots, S_k\}$. Let Q_i^t be the probability of a cell being in the ith state at time t. We can represent this probability for all the states as $\boldsymbol{Q_t} = \left[\, Q_1^t \; Q_2^t \; . \; . \; Q_k^t \,\right]$. Suppose that in one time-step, a cell in state i moves to state j with a probability p_{ij}. The probabilities of all such transition can be represented by a $k \times k$ square matrix, $\boldsymbol{P} = \left[p_{ij}\right]_{i,j=1}^{k}$, such that $0 \leq p_{ij} \leq 1$ for $j = 1$ to k, and $\sum_{j=1}^{k} p_{ij} = 1$. $\boldsymbol{P}$ is called a transition matrix or a stochastic matrix. Figure 3d graphically shows a transition matrix for a three-state system.

Note that in a Markov process, the system is conserved; i.e., the total number of cells remains unchanged. Therefore, p_{ii} is the probability of a cell staying at its current state i and does not represent cell division or self-renewal of cells in the ith state. However, we can accommodate cell death, considering the dead-state as an absorbing state in the model.

The probability that a cell will be in the jth state, at $(t+1)$ can be calculated as:

$$Q_j^{t+1} = \sum_{i=1}^{k} Q_i^t . p_{ij}. \tag{5}$$

This is explained pictorially in Fig. 3e. Equation (5), can be written vectorially:

$$\mathbf{Q_{t+1}} = \mathbf{Q_t} \times \mathbf{P}. \tag{6}$$

Here, $\boldsymbol{Q}_{t+1} = \left[\, Q_1^{t+1} \;\; Q_2^{t+1} \;\; . \;\; . \;\; Q_k^{t+1} \,\right]$.

In experiments, we count the number or fraction of cells in different phenotypic states at a particular time. For a large sample size, the fraction of cells in a state (x_j^t) can be considered as equivalent to the probability of a cell to be in that state (Q_j^t). Let, the distribution of cells in different states at time t, as observed in the experiment, be $\boldsymbol{F_t} = \left[\, x_1^t \;\; x_2^t \;\; . \;\; . \;\; x_k^t \,\right]$, where x_j^t is the fraction of cells in the jth state at time t.

Assuming the equivalence of $\boldsymbol{Q}$ and $\boldsymbol{F}$, we can write Eq. (6) as:

$$\mathbf{F_{t+1}} = \mathbf{F}_t \times \mathbf{P}. \tag{7}$$

For a time-homogeneous system, $\boldsymbol{P}$ remains constant over time, and the distribution of cells after n time steps is given by:

$$\mathbf{F}_n = \mathbf{F}_0 \times \mathbf{P}^n. \tag{8}$$

Here, $\boldsymbol{F_0}$ is the distribution of cells in different states at the beginning of the experiment.

In an experiment, after an extended period of time, the distribution of cells in different states may reach a steady state. At steady state, the proportion of different cell types will not change with time. Therefore, from Eq. (8), we can calculate the steady-state distribution $\boldsymbol{F_{ss}}$ as:

$$\mathbf{F}_{SS} = \mathbf{F}_{SS} \times \mathbf{P}. \tag{9}$$

When we know $\boldsymbol{P}$, the steady-state distribution $\boldsymbol{F_{ss}}$ can be easily calculated. As $\boldsymbol{P}$ is a probability matrix, one of the eigenvalues of $\boldsymbol{P}$ is 1. The eigenvector of $\boldsymbol{P}$ for the eigenvalue 1 is $\boldsymbol{F_{ss}}$ [56].

3.2 Estimating the State Transition Probabilities

The existence of a phenotypic steady state can be checked experimentally. However, we are more interested in estimating the transition matrix $\boldsymbol{P}$ from data as it shows the paths followed by cells as they jump from one state to another.

Estimating state transition paths from an experiment where we track individual cells continuously is trivial. We can count number cells that have moved from one sate (say ith state) to another (say jth state) in a unit time. That would allow us to calculate easily $p_{ij} = \frac{n_{ij}}{\sum_{j=1}^{k} n_{ij}}$, where n_{ij} is the number of cells moved from state i to state j in an interval. Like any other problem of estimating probability from counting, we should have sufficiently large data.

However, it is usually difficult to collect such data for a large number of individual cells for several time points. Cell state analysis is often performed using flow cytometry and gene expression analysis. These are end-point assays. In such experiments, samples are collected at different time points and analyzed independently. These experiments do not generate time-series data for individual cells but provide an estimate of the proportion of cells at different states at different time points. This type of data is called aggregate data. The estimation of the transition matrix from aggregate data is not trivial.

Suppose, we have adequate aggregate time-series data for $t=0, 1, 2,\ldots, T$. Following Eq. (5), we can write:

$$x_j^t = \sum_{i=1}^{k} p_{ij} x_i^{t-1} + \varepsilon_j^t, \tag{10}$$

where x_j^t and x_i^{t-1} are observed proportions of cells in the jth state at time t, and in the ith state at $(t-1)$, respectively. The transition probability p_{ij} is unknown, and we have to estimate it from the data. ϵ_j^t is the error in our estimate, and we want to minimize the error.

Considering all the states and all the time points, we can write the Eq. (10) in a vectorial form:

$$\mathbf{X} = \mathbf{PG} + \mathbf{e}. \tag{11}$$

Here, $\boldsymbol{X}$, $\boldsymbol{G}$, and $\boldsymbol{e}$ are matrices of size $k \times (T-1)$. Both $\boldsymbol{X}$ and $\boldsymbol{G}$ are obtained from data. $\boldsymbol{P}$ is the transition matrix.

Using this formulation, one can use different regression methods to estimate $\boldsymbol{P}$ from time-series aggregate data[57–60]. In econometrics, various authors have used this approach to estimate the transition matrix of Markov models. However, like any other regression-based method, this approach is also constrained by the size of the data. The number of time points and replicate should be high enough to avoid overfitting. On the other hand, long time-series data may suffer from multi-collinearity problem[58].

As an alternative to regression, one can use a direct root-finding method to estimate the

state transition matrix for discrete-time Markov models. Suppose there are k number of cell states, and we have performed k independent experiments where distributions of cells in different states at $t=0$ and after L time steps are known. Let $\boldsymbol{D_0}$ and $\boldsymbol{D_L}$ are $k \times k$ square matrices holding these data. In these data matrices, each row represents data from one of the k experiments.

Following the state transition Eq. (8) for a Markov process, we can write:

$$\mathbf{D}_L = \mathbf{D_0} \times \mathbf{P}^L. \tag{12}$$

This relationship can be used to estimate $\boldsymbol{P}$:

$$\mathbf{P} = (\mathbf{D}_L \mathbf{D}_0^{-1})^{1/L}. \tag{13}$$

As evident from Eq. (13), estimation of $\boldsymbol{P}$ by this method requires computing the Lth root of a matrix. Such computation is possible only when $\boldsymbol{D_0}$ is invertible. The invertibility of this data matrix can be assured if we start the experiments with pure populations for each cell state. That will make $\boldsymbol{D_0}$ an identity matrix and invertible. Though a $\boldsymbol{D_0}$ with a mixed population can be invertible, achieving such a titrated sample in an experiment is not realistic.

When time-dependent data are available for several time points, $\boldsymbol{P}$ can be estimated from each time point data and then averaged. This method has been used in the R-package CellTrans[61]. Though this estimation looks simple, it is not straightforward. Note that $\boldsymbol{P}$ needs to be a stochastic matrix with $0 \leq p_{ij} \leq 1$, and $\sum_{j=1}^{k} p_{ij} = 1$. However, the Lth root calculated in Eq. (13) may not be a stochastic matrix. The CellTrans algorithm uses quasi-optimization of the root matrix to regularize the matrix and generate appropriate $\boldsymbol{P}$.

Farahat and Asada[62] proposed a Bayesian method for the estimation of the transition matrix from aggregate data. Till now, we have considered that the fraction of cells present in a state is equivalent to the probability of a cell to be in that state. Farahat and Asada's method does not require any such equivalency. They consider the count of cells in each state at two time points as the input for the estimation algorithm.

Say $n_i(t)$ and $n_i(t+1)$ are the numbers of cells in the ith state at time t and $t+1$, respectively. For all k states, we can represent this data as two vectors, $\boldsymbol{N}(t)$ and $\boldsymbol{N}(t+1)$. The number of cells in each state has changed in the interval $(t, t+1)$ as cells have moved or flown from one state to another. Consider $f_{ij}(t)$ is the number of cells that have moved from ith state to jth state in this interval. The flows between all pairs of states can be written as a square matrix, $\boldsymbol{F}(t)$, such that $\sum_{j=1}^{k} f_{ij}(t) = n_i(t)$ and $\sum_{j=1}^{k} f_{ji}(t) = n_i(t+1)$. $\boldsymbol{F}(t)$ is called a flow matrix. For a given $\boldsymbol{N}(t)$ and $\boldsymbol{N}(t+1)$, there could be a large set of possible flow matrices that satisfy all the constraints. Let us call that set $\boldsymbol{\phi}(t)$.

For a given transition matrix $\boldsymbol{P}$ and flow matrix $\boldsymbol{F(t)}$, the probability of observing $\boldsymbol{N}(t+1)$ given $\boldsymbol{N}(t)$ is a multinomial distribution problem:

$$\Pr(\mathbf{N}(t+1)|\mathbf{P}, \mathbf{N}(t)) = \prod_{i=1}^{k} n_i(t)! \prod_{j=1}^{k} \frac{p_{ij}^{f_{ij}(t)}}{f_{ij}(t)!}. \tag{14}$$

Considering all $\boldsymbol{F}(t) \in \boldsymbol{\phi}(t)$, Eq. (14) can be written as:

$$\Pr(\mathbf{N}(t+1)|\mathbf{P}, \mathbf{N}(t)) = \sum_{\mathbf{F}(t) \in \boldsymbol{\Phi}(t)} \left(\prod_{i=1}^{k} n_i(t)! \prod_{j=1}^{k} \frac{p_{ij}^{f_{ij}(t)}}{f_{ij}(t)!} \right). \tag{15}$$

Equation (15) is a likelihood function that can be used to estimate $\boldsymbol{P}$, either by the maximum-likelihood method or by the Bayesian method. However, the main obstacle for any such estimation is that $\boldsymbol{\phi}(t)$ is usually very large, and the computation for the multinomial likelihood function is intractable. Farahat and Asada[62] proposed a Gibbs sampling method for estimation of the likelihood where the multinomial distribution is approximated to a constrained multivariate Gaussian distribution.

3.3 Other Methods for Path Estimation

One can also analyze the dynamics in a discrete state model using deterministic differential equation-based approach. Consider the sequential state transition model of three states shown in Fig. 3b. Let f_A, f_B, and f_C are the fractions of cells in states A, B, and C, respectively. The change in the distribution of cells in these states can be represented by a set of ODEs:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} f_A &= -k_{ab} f_A + k_{ba} f_B, \\
\frac{\mathrm{d}}{\mathrm{d}t} f_B &= k_{ab} f_A - k_{ba} f_B - k_{bc} f_B + k_{cb} f_C, \\
\frac{\mathrm{d}}{\mathrm{d}t} f_C &= k_{bc} f_B - k_{cb} f_C.
\end{aligned} \tag{16}$$

Here, k_{ij}, for $i, j = a, b, c$, is the rate constant for the transition from ith state to jth state.

We can write a similar system of ODEs for any given model of state transition, sequential or not, with k states. One can generalize and write such a system of ODEs in vectorially:

$$\dot{\mathbf{f}}(t) = \mathbf{A}\mathbf{f}(t), \tag{17}$$

$\dot{\boldsymbol{f}}(\boldsymbol{t})$ is a vector of derivatives of $\boldsymbol{f}$. $\boldsymbol{A}$ is a matrix holding the rate constants for the transitions between states.

Note that for a time homogenous system $\boldsymbol{A}$ is constant. The coefficient matrix $\boldsymbol{A}$ is unknown, and we need to estimate it from data to understand the flow of cells in different state transition paths. The advantage of the ODE-based approach is that, with adequate experimental data, one can fit this model to data using the conventional tools of parameter estimation for ODE-based models[63,64]. For the one-directional state transition, the system of ODEs gets simpler, with a lesser number of unknown rate parameters. Such a reduced system is also more amenable to analytical methods of analysis. This approach has been used to understand the dynamics of state transition in the differentiation of stem cells[4] and EMT[65].

Till now, we have considered that the state transition system is conserved. That means either there is no birth and death, or both of these processes have the same rate. Though true in many cases, this assumption must be supported by empirical observations. In some experiments, one cannot neglect cell proliferation and cell death. As mentioned earlier cell death can be accommodated in the model considering an absorbing state for dead cells. However, one cannot represent cell division similarly.

Su et al.[12] had investigated drug-induced cellular state transition dynamics in melanoma cell lines. As expected, cells in different states had different sensitivity to the drug that affects their proliferation and survival. To include the differential effect of the drug on each cell type, they multiplied the transition matrix $\boldsymbol{P}$ in Eq. (7) with a diagonal matrix, whose diagonal elements represent the relative viability of each cell type in the presence of the drug. However, this multiplication generates a square matrix that is not a stochastic matrix. Therefore, the population distribution estimated using this matrix requires further normalization. Furthermore, the elements of the transition matrix and drug sensitivity matrix cannot be estimated by the root-finding method described earlier. Su et al.[12] have used a Monte Carlo sampling method to estimate these two matrices. However, the details of the algorithm and the quality of the estimation were not described.

The ODE-based approach can accommodate cell death and cell division, provided we have experimental measures for both of these processes. However, both stochastic and ODE-based methods have one crucial limitation—both consider the state transition process as time homogenous. In some cases, like the spontaneous state transition of cells, one can consider that the parameters for state transitions remain constant with time. A similar assumption may be appropriate for one-directional state transition induced by an external signal. However, in some cases, the state transition is reversible and the external cue changes with time. In such a situation, we can not assume time homogeneity.

For example, in EGF-induced EMT in breast cancer cell line MDA-MB-468 is reversible, and the dynamics of state transition depends upon phosphorylation status of EGFR that changes with time[18]. Furthermore, EGF also affects cell proliferation and cell death. Therefore, the conventional approaches cannot be used to estimate the state transition paths for this experimental system. Devaraj and Bose[18] formulated a difference equation-based state transition model that considers cell death and proliferation explicitly. As the parameters of this model change with time, they used a piecewise approach for parameter estimation—a set of estimated parameters for each time interval. Any piecewise parameter estimation method suffers from the problem of overfitting. Therefore, two objective functions were used for the parameter estimation—one for fitting the data to the state transition model and the other one to reduce overfitting. A genetic algorithm was used for parameter estimation using these two objective functions.

All the mathematical methods discussed above estimate either the probability of transition or rate constant for the transition between two states from experimental observations. These numerical estimates allow one to identify the predominant paths of transition between different phenotypic states. However, we need to remember that the change in the distribution of cells in different states may also arise by preferential proliferation and death, without any state transition. Therefore, for each experiment, one needs to create a null model that considers only cell death and cell proliferation[18]. This null model can be rejected if it fails to fit the experimental observation or gives rise to unreasonable predictions.

4 Transition through Microstates and Macrostates

When cell states are defined in terms of molecular markers, we can consider a continuous potential landscape. Sisan et al.[66] estimated a continuous potential landscape from experimental data, where the expression of a reporter gene was considered as the reaction coordinate. They estimated the potential following a relation, similar to Eq. (4), between steady-state distribution of the reporter expression and the potential.

For a discrete state change model, we cannot generate a continuous potential landscape from data. Nevertheless, we can use Eq. (4) to estimate the pseudo-potential of each phenotypic state from the frequency of that state in the steady state. Devaraj and Bose[18] used this approach to calculate the pseudo-potential for different morphological states in EGF-induced EMT of MDA-MB-468 cells. They observed that during EMT, these cells jump through three morphological states—cobble, elongated and circular. Cells in all three states were observed even in untreated cells, and the frequency distribution of these cell types remained constant over time, indicating a phenotypic steady state. They used this steady-state distribution to calculate the pseudo-potential for each of these states, using Eq. (4). They graphically represented those states as horizontal lines stacked vertically according to the potentials. This diagram is similar to the Jablonski diagram [67] and can be used for easy visualization of the transition of cells from one energy state to another.

However, can we connect two worlds of cell states—the molecular state of a cell and its discrete phenotypic state? Statistical physics again helps us in this. Let us consider that cells have certain discrete traits based on function or physical characters (e.g., distinct cell morphologies observed during EMT). In a way, these traits are macroscopic traits. Therefore, the cellular states defined by these macroscopic traits can be called macrostates.

These macroscopic traits emerge out of molecular processes involving a large number of molecules. Most of the time, we do not have complete information on all the molecules involved in these processes. It is also not possible to know the status of all the molecules in a cell at a particular time. Even without knowing these molecules, we can imagine that a cell can have a large number of molecular states defined by concentrations/activities of all the molecules in the cell. We can consider the molecular states of a cell as microstates. A particular macrostate can emerge out of a set of specific microstates of a cell. In a way, a microstate is an internal configuration of a cell. Even if we define a phenotypic state (or macrostate) in terms of expression of few molecular markers, different internal configurations (or microstates) can lead to a particular expression level of those markers[68].

This concept of macrostates and microstates of a cell is equivalent to the same in statistical physics. For gas in a balloon, the pressure defines a macroscopic state, whereas the position and momentum of each of the gas molecules in the balloon defines the microstates. Each of these microstates is associated with a specific macrostate. There can be a huge number of microstates, and just like cellular microstates, it may not be even possible to know all the microstates of the gas in the balloon. Even then, the tools of equilibrium statistical physics allow us to connect microstates and macrostates.

Stumpf et al.[4] introduced a similar concept of macrostate and microstate in their study on the differentiation of mouse embryonic stem cells to neuroprogenitor cells. Using gene expression analysis, they established that cells had three distinct phenotypic states. They used an ODE-based model for one-directional, sequential state transition to explain the observed state transition dynamics. However, the model fits well with the data only when they considered multiple intermediate states for each of these three states. The three phenotypic states defined by gene expression analysis were called macrostates, and these intermediate states which were not distinctly observed in experiments were called microstates.

Mathematically speaking, by breaking each macrostate into multiple microstates, they achieved time delay in the dynamics, and that helped in better data fitting. In many ODE-based biological models, one or more, unknown/undefined molecule or step is introduced to achieve time delay. Very, recently Goetz et al.[65] used the same approach and considered multiple hidden intermediate states (or microstates) to model the state transition dynamics in TGF-β-induced EMT of MCF10A cells. By estimating the rate of transition over the energy barriers between these states, they have shown that an increase in the number of intermediate states can accelerate the EMT process.

We can hope that the concept of cellular macrostates and microstates would help us further to explore phenotypic state transition using tools of statistical physics. However, we need to keep in mind the nuances of statistical physics. In general, each macrostate of a system has a unique energy state/level. Each microstate corresponding to that

macrostate also has the same energy. Therefore, the potential landscape for the system would not be smooth. The landscape will be very 'rough' with a large number of local minima having the same potential corresponding to each macrostate. In place of simplifying, such a landscape may somewhat complicate the model.

Furthermore, in an equilibrium system, microstates follow detailed balance, all the microstates for a macrostate are equally likely, and the system itself is ergodic[40]. These properties are valid for a cellular system where a steady state of phenotypic states is achieved by stochastic state transitions. We may hope to analyses such systems using concepts and tools of equilibrium statistical physics. However, most state transition problems, such as EMT and stem cell differentiation, are directional processes that violate detailed balance and sometimes even ergodicity. Therefore, the application of statistical physics in cellular state transition problems is not trivial.

5 Concluding Remarks

Being a developmental biologist, Waddington proposed the epigenetic landscape for the spontaneous and unidirectional journey of cells from pluripotency to fully differentiated states. However, now we know that a fully differentiated cell can retrace its path back to stemness, and we can reproducibly generate pluripotent stem cells from differentiated somatic cells[69, 70]. Further, differentiated cells can be coaxed to 'transdifferentiate' into another type of cells[71, 72]. In reprogramming and transdifferentiation, cells are going uphill on the Waddington's landscape. While developing a generalized potential landscape-based theory for cellular state transition, we need to consider this possible uphill journey of cells.

In equilibrium thermodynamics, a system moves from a higher potential to a lower one. Sometimes the potential landscape can have multiple local minima. The system may move through these local minima and will eventually reach the global minimum, the most stable state of the system. This is how we try to explain protein folding.

However, for cell state transition, each phenotypic state is stable and a minimum (local or global) in the landscape. The transition from one minimum to another depends upon the energy barrier between those two. Furthermore, the phenotypic potential landscape for a cell is plastic. An external cue can change this landscape, facilitating the movement of a cell from one minimum to the other. Both cellular reprogramming and transdifferentiation can be explained by changes in the energy barriers and shape of the potential landscape.

Mathematical studies in cellular state transition have focused primarily on the estimation of state transition paths and the static potential landscape, either from data or from a mechanistic model. The focus should now shift towards the kinetics of a cell over a potential landscape and effects of an external cue on the topology of the landscape itself.

In this review, we have focused only on mathematical approaches that attempt to investigate phenotypic state transition at the phenomenon level, rather than going into details of molecular processes. Understanding detailed molecular processes, from experiments and mathematical models, is important, particularly when we want to intervene in a cellular process. Recent developments in high-throughput experiments at single-cell resolution are helping to investigate phenotypic state transition with ever-increasing details. However, to make a sense out of those higher dimensional data, we need to know the generalized principles of the phenomenon. This is where the phenomenon-level models discussed in this review help. These models inform us of the physical boundaries of the problem and help us to understand the possibilities.

Publisher's Note

Acknowledgements

This work is supported by the Department of Biotechnology, Government of India (Project No. BT/PR13560/COE/34/44/2015). V. D. is supported by Department of Biotechnology, Ministry of Science and Technology, Government of India (Project No. BT/PR13560/COE/34/44/2015).

Received: 29 April 2020 Accepted: 6 May 2020
Published online: 24 May 2020

References

1. Hartl DL, Jones EW (1998) Genetics principles and analysis, 4th edn. Jones and Barlett Publishers, Sudbury
2. Tyler S (2003) Epithelium–the primary building block for metazoan complexity. Integr Comp Biol 43:55–63
3. Kim DH, Xing T, Yang Z, Dudek R, Lu Q, Chen YH (2017) Epithelial mesenchymal transition in embryonic

development, tissue repair and cancer: a comprehensive overview. J Clin Med 7:1
4. Stumpf PS, Smith RCG, Lenz M, Schuppert A, Muller FJ, Babtie A, Chan TE, Stumpf MPH, Please CP, Howison SD, Arai F, MacArthur BD (2017) Stem cell differentiation as a non-markov stochastic process. Cell Syst 5(268–282):e267
5. Pisco AO, Brock A, Zhou J, Moor A, Mojtahedi M, Jackson D, Huang S (2013) Non-Darwinian dynamics in therapy-induced cancer drug resistance. Nat Commun 4:2467
6. Kumar N, Cramer GM, Dahaj SAZ, Sundaram B, Celli JP, Kulkarni RV (2019) Stochastic modeling of phenotypic switching and chemoresistance in cancer cell populations. Sci Rep 9:10845
7. Wang W, Quan Y, Fu Q, Liu Y, Liang Y, Wu J, Yang G, Luo C, Ouyang Q, Wang Y (2014) Dynamics between cancer cell subpopulations reveals a model coordinating with both hierarchical and stochastic concepts. PLoS ONE 9:e84654
8. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 146:633–644
9. Yang G, Quan Y, Wang W, Fu Q, Wu J, Mei T, Li J, Tang Y, Luo C, Ouyang Q, Chen S, Wu L, Hei TK, Wang Y (2012) Dynamic equilibrium between cancer stem cells and non-stem cancer cells in human SW620 and MCF-7 cancer cell populations. Br J Cancer 106:1512–1519
10. Tam WL, Weinberg RA (2013) The epigenetics of epithelial-mesenchymal plasticity in cancer. Nat Med 19:1438–1449
11. Murke F, Castro SVC, Giebel B, Görgens AJS (2015) Concise review: asymmetric cell divisions in stem. Cell Biol 7:2025–2037
12. Su Y, Wei W, Robert L, Xue M, Tsoi J, Garcia-Diaz A, Homet Moreno B, Kim J, Ng RH, Lee JW, Koya RC, Comin-Anduix B, Graeber TG, Ribas A, Heath JR (2017) Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. Proc Natl Acad Sci USA 114:13679–13684
13. Steinestel K, Eder S, Schrader AJ, Steinestel J (2014) Clinical significance of epithelial-mesenchymal transition. Clin Transl Med 3:17
14. Pastushenko I, Blanpain C (2019) EMT transition states during tumor progression and metastasis. Trends Cell Biol 29:212–226
15. Calloni R, Cordero EA, Henriques JA, Bonatto D (2013) Reviewing and updating the major molecular markers for stem cells. Stem Cells Dev 22:1455–1476
16. Armond JW, Saha K, Rana AA, Oates CJ, Jaenisch R, Nicodemi M, Mukherjee S (2014) A stochastic model dissects cell states in biological transition processes. Sci Rep 4:3692
17. Trapnell C (2015) Defining cell types and states with single-cell genomics. Genome Res 25:1491–1498
18. Devaraj V, Bose B (2019) Morphological state transition dynamics in EGF-induced epithelial to mesenchymal transition. J Clin Med 8:911
19. Mandal M, Ghosh B, Anura A, Mitra P, Pathak T, Chatterjee J (2016) Modeling continuum of epithelial mesenchymal transition plasticity. Integr Biol (Camb) 8:167–176
20. Sommer C, Hoefler R, Samwer M, Gerlich DW (2017) A deep learning and novelty detection framework for rapid phenotyping in high-content screening. Mol Biol Cell 28:3428–3436
21. Buggenthin F, Buettner F, Hoppe PS, Endele M, Kroiss M, Strasser M, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Schroeder T, Theis FJ, Marr C (2017) Prospective identification of hematopoietic lineage choice by deep learning. Nat Methods 14:403–406
22. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, Golland P, Sabatini DM (2006) Cell Profiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 7:R100
23. Wang W, Douglas D, Zhang J, Chen YJ, Cheng YY, Kumari S, Enuameh MS, Dai Y, Wallace CT, Watkins SC, Shu W, Xing J (2019) M-TRACK: a platform for live cell multiplex imaging reveals cell phenotypic transition dynamics inherently missing in snapshot data. bioRxiv. https://doi.org/10.1101/2019.12.12.874248
24. Kimmel JC, Chang AY, Brack AS, Marshall WF (2018) Inferring cell state by quantitative motility analysis reveals a dynamic state system and broken detailed balance. PLoS Comput Biol 14:e1005927
25. Rimchala T, Kamm RD, Lauffenburger DA (2013) Endothelial cell phenotypic behaviors cluster into dynamic state transition programs modulated by angiogenic and angiostatic cytokines. Integr Biol (Camb) 5:510–522
26. Waddington CH (1957) The strategy ofthe genes: a discussion of some aspects of theoretical biology. George Allen & Unwin Ltd, London
27. Strogatz SH (2018) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. CRC Press, Boca Raton
28. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E (2013) MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. Proc Natl Acad Sci USA 110:18144–18149
29. Jolly MK, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, Onuchic JN, Levine H (2015) Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. Front Oncol 5:155
30. Zhang J, Tian XJ, Zhang H, Teng Y, Li R, Bai F, Elankumaran S, Xing J (2014) TGF-beta-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. Sci Signal 7:91

31. Tian XJ, Zhang H, Xing J (2013) Coupled reversible and irreversible bistable switches underlying TGFbeta-induced epithelial to mesenchymal transition. Biophys J 105:1079–1089
32. Huang S, Guo YP, May G, Enver T (2007) Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Dev Biol 305:695–713
33. Chickarmane V, Troein C, Nuber UA, Sauro HM, Peterson C (2006) Transcriptional dynamics of the embryonic stem cell switch. PLoS Comput Biol 2:e123
34. Chickarmane V, Peterson C (2008) A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. PLoS ONE 3:e3478
35. Ferrell JE Jr (2012) Bistability, bifurcations, and Waddington's epigenetic landscape. Curr Biol 22:R458–466
36. Tripathi S, Xing J, Levine H, Jolly MK (2019) Mathematical modeling of plasticity and heterogeneity in EMT
37. Bose I, Pal M (2017) Criticality in cell differentiation. J Biosci 42:683–693
38. Schiesser WE (2014) Stem cell differentiation. Differential equation analysis in biomedical science and engineering. John Wiley & Sons Inc, Hoboken, pp 217–239
39. Bhattacharya S, Zhang Q, Andersen ME (2011) A deterministic map of Waddington's epigenetic landscape for cell fate specification. BMC Syst Biol 5:85
40. Dill KA, Bromberg S, Stigter D (2003) Molecular driving forces: statistical thermodynamics in chemistry and biology. Garland Sci. https://doi.org/10.1002/macp.200390113
41. Ao P (2004) Potential in stochastic differential equations: novel construction. J Phys A Math Gen 37:L25–L30
42. Wang J (2015) Landscape and flux theory of non-equilibrium dynamical systems with application to biology. Adv Phys 64:1–137
43. Biswas K, Jolly MK, Ghosh A (2019) Stability and mean residence times for hybrid epithelial/mesenchymal phenotype. Phys Biol 16:025003
44. Li C, Wang J (2013) Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. J R Soc Interface 10:20130787
45. Li C, Hong T, Nie Q (2016) Quantifying the landscape and kinetic paths for epithelial-mesenchymal transition from a core circuit. Phys Chem Chem Phys 18:17949–17956
46. Qiu K, Gao KF, Yang LJ, Zhang ZK, Wang R, Ma HS, Jia Y (2017) A kinetic model of multiple phenotypic states for breast cancer cells. Sci Rep 7:9890
47. Zhao L, Wang J (2016) Uncovering the mechanisms of *Caenorhabditis elegans* ageing from global quantification of the underlying landscape. J R Soc Interface 13:20160421
48. Li C, Wang J (2014) Quantifying the underlying landscape and paths of cancer. J R Soc Interface 11:20140774
49. Guo J, Lin F, Zhang X, Tanavde V, Zheng J (2017) NetLand: quantitative modeling and visualization of Waddington's epigenetic landscape using probabilistic potential. Bioinformatics 33:1583–1585
50. Li C, Wang J (2013) Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. PLoS Comput Biol 9:e1003165
51. Zhou JX, Aliyu MD, Aurell E, Huang S (2012) Quasi-potential landscape in complex multi-stable systems. J R Soc Interface 9:3539–3553
52. Wang J, Zhang K, Xu L, Wang E (2011) Quantifying the Waddington landscape and biological paths for development and differentiation. Proc Natl Acad Sci USA 108:8257–8262
53. Wang J, Xu L, Wang E, Huang S (2010) The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. Biophys J 99:29–39
54. Yu P, Nie Q, Tang C, Zhang L (2018) Nanog induced intermediate state in regulating stem cell differentiation and reprogramming. BMC Syst Biol 12:22
55. Lv C, Li X, Li F, Li T (2015) Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process. PLoS Comput Biol 11:e1004156
56. Allen LJS (2010) An introduction to stochastic processes with applications to biology. Chapman and Hall/CRC, New York
57. Kalbfleisch JD (1984) Least-squares estimation of transition probabilities from aggregate data. Can J Stat 12:169–182
58. Lee TC, Judge GG, Zellner A (1970) Estimating the parameters of the Markov probability model from aggregate time series data. North-Holland Pub Co., Amsterdam
59. Dent W, Ballintine R (1971) A review of the estimation of transition probabilities in Markov chains. J Aust J Agric Econ 15:69–81
60. Kaur I, Rajarshi MJC, Computation S (2012) Ridge regression for estimation of transition probabilities from aggregate data. Commun Stat 41:524–530
61. Buder T, Deutsch A, Seifert M, Voss-Bohme A (2017) Cell trans: an R package to quantify stochastic cell state transitions. Bioinform Biol Insights 11:1177932217712241
62. Farahat WA, Asada HH (2012) Estimation of state transition probabilities in asynchronous vector markov processes. J Dyn Syst Meas Control 134:6
63. Aster RC, Borchers B, Thurber CH (2018) Parameter estimation and inverse problems. Elsevier, Amsterdam
64. Chou IC, Voit EO (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math Biosci 219:57–83
65. Goetz H, Melendez-Alvarez JR, Chen L, Tian XJ (2019) A plausible accelerating function of intermediate states in cancer metastasis. bioRxiv. https://doi.org/10.1101/828343
66. Sisan DR, Halter M, Hubbard JB, Plant AL (2012) Predicting rates of cell state change caused by stochastic fluctuations using a data-driven landscape model. Proc Natl Acad Sci USA 109:19262–19267

67. Atkins P, de Paula J (2006) Atkin's physical chemistry, 8th edn. W. H Freeman and Company, New York
68. Moris N, Arias AM (2017) The hidden memory of differentiating cells. Cell Syst 5:163–164
69. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676
70. Malik N, Rao MS (2013) A review of the methods for human iPSC derivation. Methods Mol Biol 997:23–33
71. Xie X, Fu Y, Liu J (2017) Chemical reprogramming and transdifferentiation. Curr Opin Genet Dev 46:104–113
72. Cieslar-Pobuda A, Knoflach V, Ringh MV, Stark J, Likus W, Siemianowicz K, Ghavami S, Hudecki A, Green JL, Los MJ (2017) Transdifferentiation and reprogramming: Overview of the processes, their similarities and differences. Biochim Biophys Acta Mol Cell Res 1864:1359–1369

Vimalathithan Devaraj is a Ph.D. student in the Department of Biosciences and Bioengineering at the Indian Institute of Technology Guwahati, Assam, India. He is interested in mathematical modeling of biological systems. His Ph.D. thesis work is on understanding phenotypic state transition in cells using experiments and mathematical modeling.

Biplab Bose is an Associate Professor in the Department of Biosciences and Bioengineering at the Indian Institute of Technology Guwahati, Assam, India. He received his Ph. D degree, in Biochemistry, from All India Institute of Medical Sciences, New Delhi. Prior to that, he had studied M. Pharm and B. Pharm at Jadavpur University, Kolkata. He works at the interface of experimental biology and mathematics. Using the concepts of dynamical systems theory and statistical physics, he explores emergent and dynamical phenomena in biology, particularly in cellular biology. He has developed software for data analysis. He has also contributed to the development of several recombinant therapeutic molecules.