



Review

Criticality in cell differentiation

INDRANI BOSE* and MAINAK PAL

Department of Physics, Bose Institute, Kolkata 700 009, India

**Corresponding author (Email, indrani@jcbose.ac.in)*

MS received 12 May 2017; accepted 17 October 2017; published online 9 November 2017

Cell differentiation is an important process in living organisms. Differentiation is mostly based on binary decisions with the progenitor cells choosing between two specific lineages. The differentiation dynamics have both deterministic and stochastic components. Several theoretical studies suggest that cell differentiation is a bifurcation phenomenon, well-known in dynamical systems theory. The bifurcation point has the character of a critical point with the system dynamics exhibiting specific features in its vicinity. These include the critical slowing down, rising variance and lag-1 autocorrelation function, strong correlations between the fluctuations of key variables and non-Gaussianity in the distribution of fluctuations. Recent experimental studies provide considerable support to the idea of criticality in cell differentiation and in other biological processes like the development of the fruit fly embryo. In this review, an elementary introduction is given to the concept of criticality in cell differentiation. The correspondence between the signatures of criticality and experimental observations on blood cell differentiation in mice is further highlighted.

Keywords. Bifurcation; cell differentiation; criticality; signatures of criticality; stochasticity

1. Introduction

Cell differentiation is the process through which stem or progenitor cells diversify into different cell types such as kidney, liver and skin cells. Cells, in general, have identical sets of genes. The different cell types are distinguished by distinct gene expression profiles. A gene active in one cell type may be either silent or expressed at a lower level in another cell type. Three representations which capture well the broad aspects of cell differentiation are Waddington's epigenetic landscape (Moris *et al.* 2016), the mammalian cell-fate tree (Zhou and Huang 2011) and the potential landscape (MacArthur *et al.* 2009). Waddington's epigenetic landscape is a pictorial depiction of the transformation of an undifferentiated embryonic stem cell state, during the development of an organism, through a sequence of transient states to the final fully differentiated cell state which is stable. The landscape consists of a cascade of binary branching valleys, i.e. a single valley at the origin branches into two new valleys each of which in turn branches into two more new valleys and so on. The two branching valleys at a branch point are separated by a ridge or hill conferring stability on the valley states. A marble representing a cell rolls down the landscape and at a branch point slides into one of the two descendant valleys. The cell state at any time is defined by the gene expression profile of the cell at that time.

The valley at the origin represents the embryonic stem cell state, the intermediate valleys describe transient cell states and the terminal valleys correspond to fully differentiated cell states. In the mammalian cell-fate tree, the tree starts from a single root depicting the embryonic stem cell and has a binary branching structure. At each branch point, the cell has a choice between two different lineages with the terminal branches of the tree associated with stable, distinct cell types. In the potential landscape, the potential function (to be defined) is plotted in the space of all gene expression states, termed the state space. The landscape consists of hills and valleys with the valleys and hilltops describing stable differentiated and metastable progenitor cell states respectively. A vast body of experimental knowledge on cell differentiation provides the basis for the construction of quantitative theoretical models to gain insight on the physical principles underlying the differentiation process. The mathematical and computational formalisms used to elucidate the principles utilize the concepts and techniques of nonlinear dynamics and the theory of stochastic processes. We describe briefly a few essential concepts of dynamical systems theory which provide the basis for developing quantitative models of cell differentiation. The definitions of some basic terminologies appear in a collective form in the Glossary at the end of the paper. Cell differentiation is a dynamical process with the cell state, defined in terms of the gene expression profile,

changing as a function of time. The dynamics involve the synthesis and degradation of some key proteins participating in the differentiation process. The rate of change in the concentration of a protein is given by the difference in the rates of protein synthesis and degradation. In the steady state of the dynamics, the rates of change are zero and the proteins acquire fixed concentrations. The unchanging values of the protein concentrations define the steady state of the cell. A stable steady state, when weakly perturbed, say, by changing the steady state protein concentrations by small amounts, regains the steady state in the course of time. The stable steady states constitute the attractors, or the final destinations of the dynamics. Under certain circumstances the dynamics may result in multistability, i.e. the coexistence of multiple stable steady states for the same set of parameter values. A special case of multistability is bistability which implies the coexistence of two stable steady states. The important parameters include the synthesis and degradation rate constants, the binding and unbinding rate constants associated with the binding and unbinding of regulatory molecules at the target sites of the promoter regions of the genes of interest etc. Bifurcations in the dynamics occur at critical parameter values at which the stability property of a steady state changes (a stable steady state becomes unstable or ceases to exist) and/or the appearance of new stable steady states.

Max Delbrück in 1948 first suggested that the bistability of biochemical networks forms the basis of differentiation into distinct cell lineages (Huang 2009a). Monod and Jacob in a seldom quoted paper published in 1961 (Monod and Jacob 1961), described several theoretical models to explain epigenetic differences including those associated with differentiation as a manifestation of multistationarity. One of the models considered by them is similar to the two-gene motif shown in figure 1(i) in which the proteins synthesized by the two genes are mutual repressors of each other's expression. The specific motif defines a positive feedback loop since an increase in the concentration of one protein gives rise to a further increase in the protein concentration. The model stressed on the roles of positive feedback and autocatalysis in the process of cell differentiation. In 1978, R Thomas (Thomas 1978) pointed out that a positive feedback loop is essential for differentiation into multiple steady states. Kauffman (1993) carried out an exhaustive simulation of the dynamics of a genetic network and put forward the hypothesis that the different cell types correspond to the attractors of the dynamics of a network of thousands of interacting genes. The result is best interpreted in the landscape picture with the valleys in the landscape defining the attractors of the dynamics. The experimental evidence that a high-dimensional stable attractor describes a distinct cellular phenotype was first provided by Huang et al (Huang et al. 2005). A series of theoretical studies have viewed cell differentiation as a bifurcation phenomenon (Enver et al. 2009;

Wang et al. 2010; Ferrell 2012; Moris et al. 2016). While bifurcation can be of different types, the supercritical pitchfork bifurcation provides a simple metaphor for understanding cell differentiation as a binary decision-making process. In this type of bifurcation, the stable undifferentiated cell state (monostability) differentiates into one of two distinct stable cell types (bistability) at a critical parameter value known as the bifurcation point. In a population of differentiating cells, variations in the levels of the regulatory proteins give rise to two distinct subpopulations of differentiated cells corresponding to two lineages. The bifurcation point is characterised by the loss of stability of the monostable state and the appearance of a new pair of stable steady states. A sequence of successive pitchfork bifurcations is representative of the mammalian cell fate tree consisting of binary branches (Foster et al. 2009). A two-gene motif (figure 1(i)) provides a simple physical picture of the cell differentiation process. The evidence for the existence of the motif, as a component of the gene regulatory networks governing cell differentiation in different cell lines,

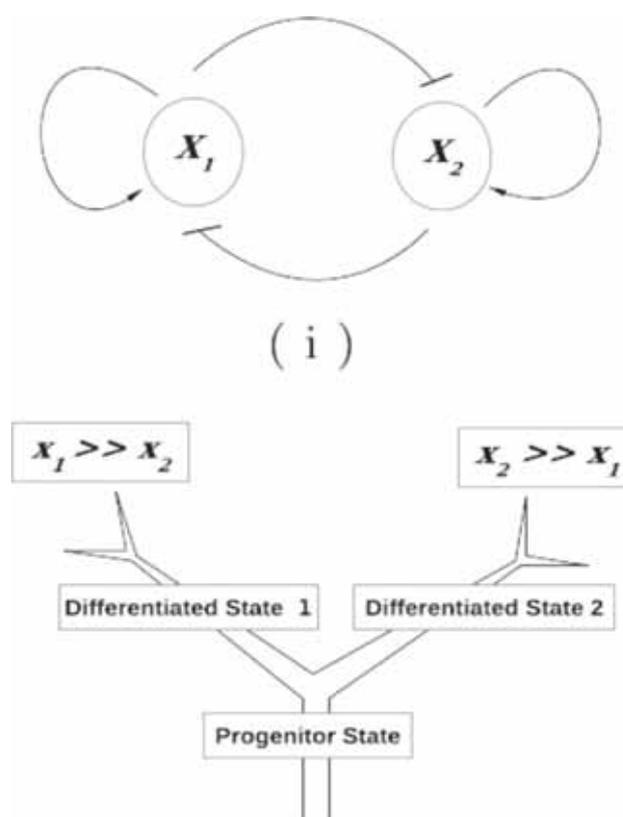


Figure 1. (i) The central motif of gene regulatory networks controlling cell differentiation. The protein products of two genes X_1 and X_2 repress each other's synthesis, thus constituting a positive feedback loop. The proteins further autoactivate their own synthesis. (ii) Illustration of a progenitor state differentiating into two states 1 and 2 distinguished by the protein concentrations $x_1 \gg x_2$ and $x_2 \gg x_1$.

has been obtained through an exhaustive analysis of the human expression data involving 2602 transcriptionally-regulated genes and 166 distinct cell types (Heinäniemi *et al.* 2013). The motif consists of the genes X_1 and X_2 the protein products of which repress each other's expression. Jacob and Monod thus correctly anticipated the dynamical relationship between genes which results in cell differentiation. The feedback loop, based on mutual antagonism, is equivalent to a positive feedback loop. The proteins also autoactivate their own production through individual positive feedback loops. Positive feedback loops coupled with sufficiently nonlinear dynamics promote bistability in specific parameter regimes. The simple genetic circuit captures the binary decision process at a branch point of the cell-fate tree (figure 1(ii)). In the undifferentiated progenitor state, the protein concentrations x_1 and x_2 are nearly equal whereas the stable steady states, $x_1 \gg x_2$ and $x_1 \ll x_2$, represent the differentiated cell states.

2. Dynamics of the two-gene motif

The rate equations describing the dynamics of the two-gene motif can be written as (Pal *et al.* 2015):

$$\frac{dx_1}{dt} = \frac{a_1 x_1}{S + x_1} + \frac{b_1 S}{S + x_2} - k_1 x_1 - g x_1 x_2 \quad (1)$$

$$\frac{dx_2}{dt} = \frac{a_2 x_2}{S + x_2} + \frac{b_2 S}{S + x_1} - k_2 x_2 - g x_1 x_2 \quad (2)$$

The first terms on the right hand sides of the two equations correspond to autoactivation, the second terms represent the cross-repression of the expression of the two genes, the third terms denote the individual protein degradation rates and the fourth terms represent the indirect repression of gene expression via heterodimer formation. In the first terms, $a_i (i = 1, 2)$ represents the maximal synthesis rate of the protein with concentration $x_i (i = 1, 2)$. Positive autoregulation requires that the protein binds to a single promoter site and activates the initiation of transcription. The protein after binding can also unbind. The binding and unbinding processes are significantly faster than the transcription and translation rates. One may then assume that a binding equilibrium is achieved at a much earlier time point than the time required for the protein concentration to reach its steady state value. The activation function has the typical Michaelis-Menten form with the parameter S being the equilibrium dissociation constant (Alon 2007). In the second terms of equations (1) and (2), the parameter $b_i (i = 1, 2)$ denotes the basal rate of expression. A reduction in the basal rate occurs due to cross-repression, an increase in the concentration $x_2 (x_1)$ reduces the protein concentration $x_1 (x_2)$. The form of the repression function is again standard (Alon 2007) with the parameter S denoting the equilibrium dissociation constant associated with the binding of a protein to a promoter

site of the gene expressing its repressor. Following Wang *et al.* (2010), the dissociation constant S is taken to be the same for both autoactivation and cross-repression. This assumption does not detract from the major conclusions of the model study. Considering the symmetric situation $a_1 = a_2 = a$, $b_1 = b_2 = b$ and $k_1 = k_2 = k$, one can compute the bifurcation diagram for the steady state values of $x_1 (x_2)$ for any one of the bifurcation parameters a , b , k and g . Figure 2 shows the bifurcation diagram, a supercritical pitchfork bifurcation, with a serving as the bifurcation parameter. The other parameter values are $b = 2.0$, $k = 1.0$, $S = 2.0$ and $g = 0.5$. The solid lines represent stable steady states whereas the dotted line describes a branch of unstable steady states. There are no further bifurcations as the parameter a is increased. The model described by equations (1) and (2) is appropriate for the case of terminal differentiation when the differentiated cell states are stable steady states. The modeling of the binary cell-fate tree (Zhou and Huang 2011) in which the intermediate states are transient states requires a more complex set of regulatory interactions and an extended mathematical framework for probing the differentiation dynamics. A sequence of successive pitchfork bifurcations (Foster *et al.* 2009) captures to a limited extent the journey from the embryonic stem cell state to the fully differentiated cell states. Each pitchfork bifurcation corresponds to a distinct pair of mutually repressing transcription factors.

If cell differentiation were totally controlled by deterministic dynamics, the stable steady states of the cells in an undifferentiated cell population, prepared in the same initial state, would be identical and described by a single gene expression level. Recent experimental observations, however, offer evidence for considerable non-genetic heterogeneity in the gene expression profiles of both embryonic and adult stem cells (Huang 2009b; Martinez Arias and Brickman 2011; Cahan and Daley 2013; Torres-Padilla and Chambers 2014). The heterogeneity implies that instead of a single gene expression, say, protein level, the undifferentiated cell population is characterised by a distribution of protein levels. The heterogeneity arises due to the randomness associated with key biological processes like gene expression (Kærn *et al.* 2005; Raj and van Oudenaarden 2008). In the presence of stochasticity, the certainty of a deterministic dynamical description is lost. Stochasticity introduces noise (random fluctuations) in gene expression levels which can be quantified using single cell and single molecule methodologies (Raj and van Oudenaarden 2009). In the case of a limited amount of noise, the broad features of deterministic dynamics are still valid. The state of a cell population at time t is no longer represented by a single point in state space but by a cloud of points. The attractor of the dynamics, instead of being a single fixed point, has a fuzzy character in the state space. Stochastic dynamics require a statistical mechanical perspective for the appropriate description of the process of cell differentiation (Garcia-Ojalvo and Arias 2012; MacArthur

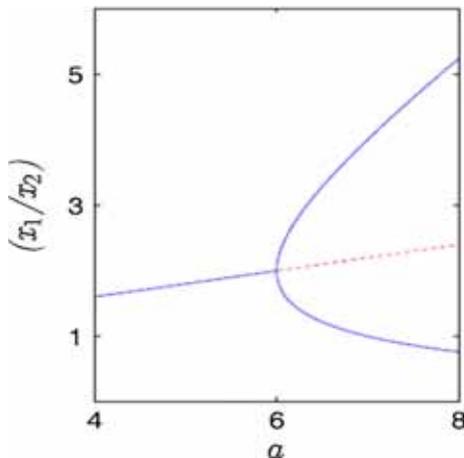


Figure 2. Bifurcation diagram depicting the steady state protein concentrations x_1 (x_2) versus the bifurcation parameter a .

and Lemischka 2013). The microstate at time t of an ideal gas in a closed container is described in terms of the instantaneous positions and momenta of the individual gas molecules. The macrostate of the system is characterized by average properties like pressure and temperature with numerous interchangeable microstates realizing the same macrostate. The pluripotent state of a progenitor cell population, which is yet to differentiate into daughter lineages, is not a single state but a macrostate arising out of a large number of microstates. The individual cells in the population are the analogs of gas molecules and the protein concentrations the analogs of positions and momenta. The microstate at time t is defined by the sets of instantaneous protein concentrations in individual cells. The macrostate is specified by the steady state probability distribution of the protein levels in the cell population. In the case of the two-gene motif, the steady state probability distribution is given by $P_{st}(x_1, x_2)$ and a potential function may be computed as $U(x_1, x_2) = -\ln P_{st}(x_1, x_2)$. The potential landscape, mentioned in the Introduction, is obtained by plotting $U(x_1, x_2)$ in the two-dimensional (2d) state space with its two axes denoting the possible values of x_1 and x_2 . In the case of a one-variable system described by the deterministic rate equation, $\frac{dx}{dt} = f(x)$, the potential function $U(x)$ is given by $f(x) = -\frac{dU}{dx}$, i.e., $U(x)$ can be computed by integrating $f(x)$ over x in the appropriate range of x values. This method of computing the potential function cannot, however, be generalized to the case of a multi-variable system (Zhou *et al.* 2012) for which the stochastic prescription for the construction of the potential is more appropriate. In the stochastic description, the valleys of the potential landscape are the regions in state space in which the steady state probability distribution is maximal. In the case of deterministic dynamics, each stable steady state has its own basin of attraction and the trajectories initiated in one basin of attraction stay confined within the boundary to ultimately end up at the stable steady state. With the

introduction of stochasticity in the dynamics, fluctuation-driven transitions of the trajectories between different basins are possible so that a cell may acquire a fate different from the one expected on the basis of deterministic dynamics.

One simple method of investigating stochastic time evolution is to make use of the formalism based on the Langevin equations (Fox *et al.* 1988; van Kampen 1992). The Langevin equation is a rate equation with a noise term added to the usual deterministic part so that it reduces to the deterministic rate equation in the absence of noise. The steady state probability distributions of the two-gene motif, with the deterministic components of the dynamics given by equations (1) and (2), have been computed for different values of the bifurcation parameter a using the formalism of the Langevin equations (Pal *et al.* 2015). Figure 3(i) depicts the steady state probability distribution, $P_{st}(x_1, x_2)$, which corresponds to the macrostate of a pluripotent cell population. The broad heterogeneity in the distribution of protein levels is consistent with the experimental observations of Chang *et al.* (Chang *et al.* 2008). In the experiment, a clonal population of mouse hematopoietic (blood-cell forming) progenitor cells was found to exhibit a nearly 1000-fold range in the levels of a stem-cell-surface marker protein Sca-1. A two-gene motif plays a key role in the differentiation process with the two antagonistic genes synthesizing the proteins GATA1 (concentration x_1) and PU.1 (concentration x_2). The cells with low (high) Sca-1 levels were found to have high (low) GATA1 and low (high) PU.1 levels. Figures 3(ii-iv) display the steady state probability distributions after bifurcation has occurred. The initial states of the cell population have been chosen from three distinct parts of the distribution in figure 3(i). The initial states chosen are $x_1 = x_2$ (figure 3(ii)), $x_2 > x_1$ (figure 3(iii)) and $x_1 > x_2$ (figure 3(iv)). The smaller subpopulations of differentiated cells in figures 3(iii) and 3(iv) are associated with cell fates distinct from those in the case of deterministic dynamics. The origin of these subpopulations lies in noise-induced transitions across the boundaries of the basins of attraction. The computational results are in conformity with the experimentally observed feature of multilineage priming (Chang *et al.* 2008) of the pluripotent cell population, i.e., the different parts of the cell population having distinct propensities for lineage choice. A recent experimental study (Mojtahedi *et al.* 2016) has detected the presence of “rebellious cells” which acquire the fate opposite to the one expected. A possible origin of the existence of such cells lies in noise-induced inter-basin transitions.

3. Criticality signatures

Bifurcation can be described as a dynamical phase transition separating two distinct dynamical regimes with the bifurcation point serving as a critical point. More familiar examples of phase transitions include thermodynamic transitions like

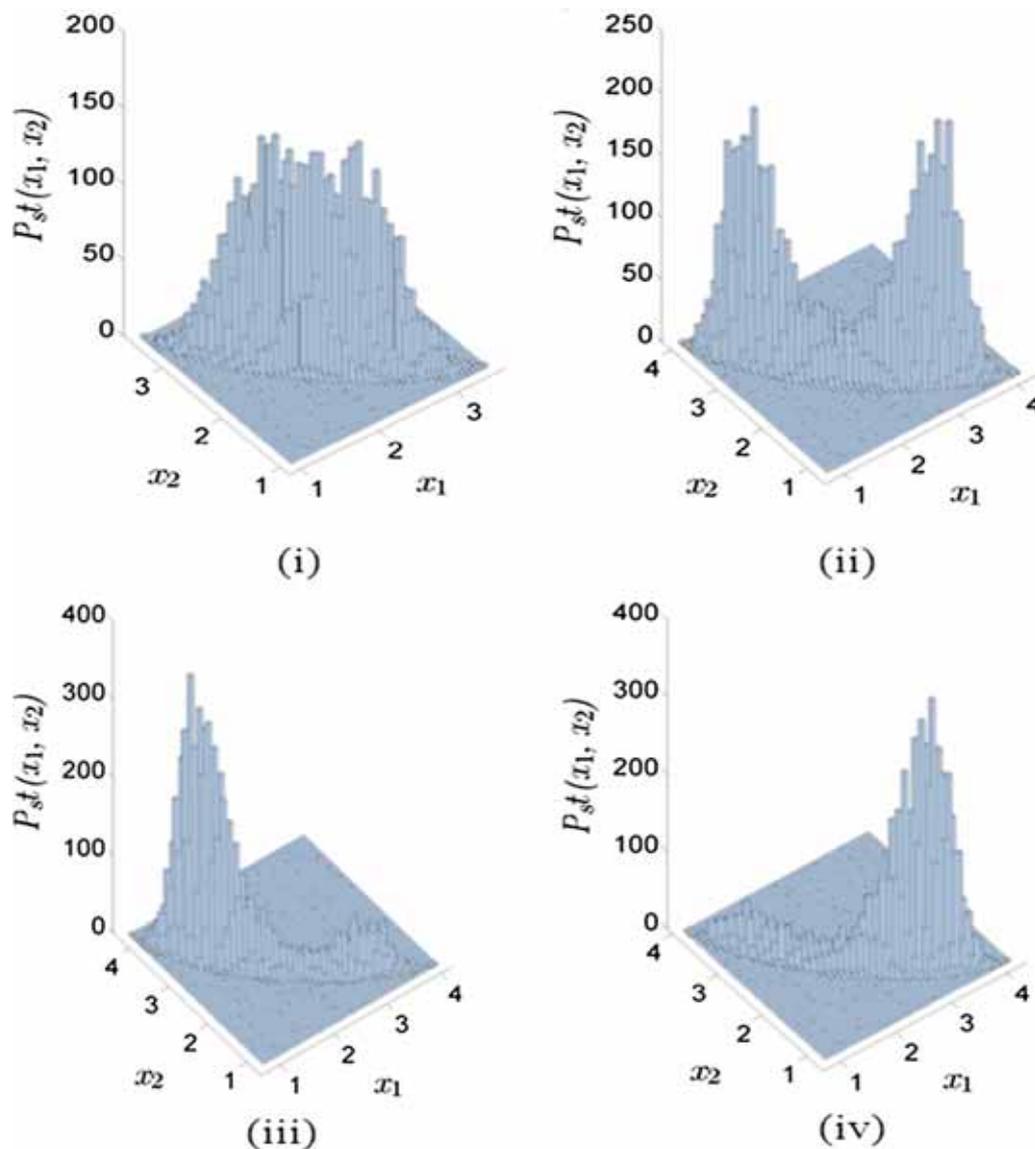


Figure 3. Steady state probability distributions $P_{st}(x_1, x_2)$ obtained by simulating the Langevin equations (12) containing only additive noise. (i) The parental histogram of the protein levels x_1 and x_2 in the case of undifferentiated cells. (ii–iv) Steady state probability distribution for differentiated cells. The forms of the distributions are dictated by the locations of the initial states in the parental distribution. The parameter a serves as the bifurcation parameter.

liquid–gas, paramagnet–ferromagnet and metal–superconductor transitions. For temperatures below a critical temperature, a magnetic system is in the ordered ferromagnetic phase whereas above the critical temperature the system is in the magnetically disordered paramagnetic phase. Thermodynamic phase transitions involve equilibrium phases of the system whereas bifurcation describes transitions between regimes defined in terms of nonequilibrium steady states. From a statistical mechanical perspective, the cell-fate decisions of embryonic stem cells have been likened to phase transitions involving the appearance of an appropriately defined order parameter below the transition point

(Garcia-Ojalvo and Arias 2012). Examples of order parameter in thermodynamic critical point transitions are: spontaneous magnetization in the case of paramagnet-ferromagnet transition and density difference between the two phases in the liquid-gas transition. The order parameter has a non-zero expectation value for temperature T less than the critical temperature T_C and has a value zero for $T > T_C$. The temperature plays the role of a control parameter in thermodynamic phase transitions. Cell differentiation involves a transition in the nonequilibrium but one can draw analogies between equilibrium and nonequilibrium transitions. The extracellular signal which induces differentiation could play

the role of a control parameter. The choice of the order parameter is more problematic but could be defined as any quantity which grows sharply when the magnitude of the control parameter exceeds a critical threshold value. A feature common to both thermodynamic critical point transitions and bifurcations is that of critical slowing down.

Consider a thermodynamic (dynamic) system to be weakly perturbed so that the system moves away from the thermodynamic equilibrium state (stable steady state). The return time T_R , defined to be the average time taken by the system to return to the equilibrium (stable steady) state, diverges as the critical point is approached, i.e. the relaxation kinetics become extremely sluggish. We illustrate the origin of critical slowing down close to a bifurcation point in the case of the two-gene motif (figure 1(i)), a key component in the gene regulatory network controlling differentiation in various cell lines. The dynamics of the two-gene motif are described by the differential rate equations

$$\frac{dx_1}{dt} = f(x_1, x_2), \frac{dx_2}{dt} = g(x_1, x_2) \quad (3)$$

where the functions $f(x_1, x_2)$ and $g(x_1, x_2)$ are given by the expressions on the right hand sides of equations (1) and (2) respectively. Let a steady state solution of the rate equations be denoted by the fixed points (x_1^S, x_2^S) . If the system is perturbed away from the steady state, the deviations of the variables from their fixed point values are represented as

$$\delta x_1 = x_1 - x_1^S, \delta x_2 = x_2 - x_2^S \quad (4)$$

Define the column vector

$$\delta x = \begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix} \quad (5)$$

The functions $f(x_1, x_2)$ and $g(x_1, x_2)$ are now Taylor expanded around the steady state values of the variables x_1 and x_2 . In the case of weak perturbation, one needs to retain terms up to only the first order and the rate equations become

$$\frac{d(\delta x)}{dt} = J \delta x \quad (6)$$

where J is the Jacobian matrix given by

$$\left(\begin{array}{cc} \frac{\partial f(x_1, x_2)}{\partial x_1} & \frac{\partial f(x_1, x_2)}{\partial x_2} \\ \frac{\partial g(x_1, x_2)}{\partial x_1} & \frac{\partial g(x_1, x_2)}{\partial x_2} \end{array} \right)_{SS} = \begin{bmatrix} -\beta_1 & r_{12} \\ r_{21} & -\beta_2 \end{bmatrix} \quad (7)$$

The diagonal elements of the Jacobian matrix represent the effective decay rates of the two proteins and the off-diagonal terms describe the repressive effects of the proteins on each other's synthesis. The suffix SS in the expression for J denotes that the different derivatives are

to be computed in the steady state. Let the eigenvalues of J be λ_1 and λ_2 with the corresponding eigenvectors V_1 and V_2 . The general solution of equation (6) is of the form

$$\delta x(t) = C_1 V_1 e^{\lambda_1 t} + C_2 V_2 e^{\lambda_2 t} \quad (8)$$

where the coefficients C_1 and C_2 are determined by the initial conditions. The stability of the steady states ($\delta x(t)$ becomes zero in the course of time) requires that the real parts of λ_1 and λ_2 are negative. Let λ_1 be the largest (dominant) eigenvalue of J and let λ_{max} be its real part. The return time $T_R = 1 / |\lambda_{max}|$. For a steady state to be stable, λ_{max} is < 0 . The steady state progressively loses stability as the bifurcation point is approached with $\lambda_{max} \rightarrow 0$. The divergence of the return time at the bifurcation point indicates critical slowing down in the vicinity of the critical point. The dynamics of the two-gene motif are characterised by two modes: slow and fast (Pal *et al.* 2015; Krotov *et al.* 2014; Strogatz 1994). The eigenvector V_i ($i = 1, 2$) sets the direction in the two-dimensional state space along which

$$\delta x(t) = e^{\lambda_i t} V_i \quad (9)$$

Substitution of equation (9) into equation (6) recovers the eigenvalue equation $J V_i = \lambda_i V_i$, $i = 1, 2$. The eigenvectors V_1 and V_2 set the directions of slow and fast dynamics respectively in the state space since λ_1 is $> \lambda_2$ (less negative) so that the $\exp(\lambda_2 t)$ term in equation (8) decays faster. In the limit of large times, the trajectories in the state space approach the stable steady state tangential to the slow eigendirection. Figure 4(i) shows some of these trajectories computed by solving equations (1) and (2). The parameter values used are $a = 5.0$, $b = 2.0$, $k = 1.0$, $g = 0.5$ and $S = 2.0$. Figure 4(ii) shows that the direction of slow relaxation is also the direction of large fluctuations (variance). The distribution of the (x_1, x_2) values has been obtained by solving the Langevin equations (Pal *et al.* 2015). The variance increases as the bifurcation point is approached. The covariance matrix CVM in the case of the two-gene motif has the form

$$CVM = \begin{bmatrix} \langle (\delta x_1)^2 \rangle & \langle \delta x_1 \delta x_2 \rangle \\ \langle \delta x_1 \delta x_2 \rangle & \langle (\delta x_2)^2 \rangle \end{bmatrix} \quad (10)$$

with the diagonal elements representing the variances. The notation $\langle \dots \rangle$ denotes the average. Near criticality, the CVM is given by

$$CVM = \sigma^2 \begin{bmatrix} 1 & \frac{\beta_1}{r_{12}} \\ \frac{\beta_1}{r_{12}} & \left(\frac{\beta_1}{r_{12}} \right)^2 \end{bmatrix} \quad (11)$$

with σ^2 denoting the variance in the protein level x_1 .

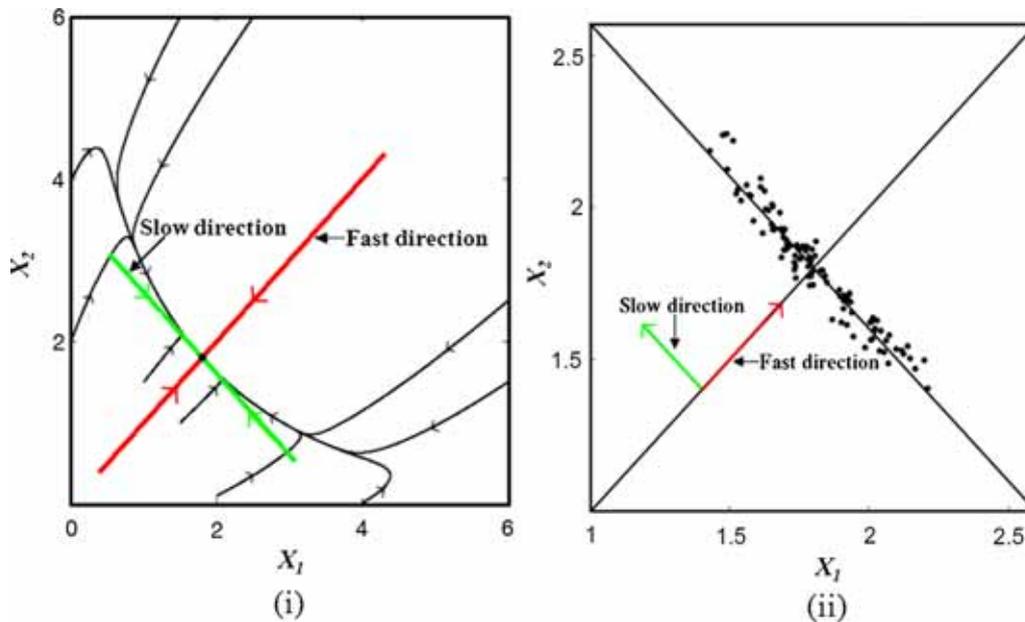


Figure 4. (i) Trajectories in state space become tangential to the slow eigendirection in the limit of large time. (ii) Variance is large (small) along the slow (fast) eigendirection.

The normalized eigenvector of the *CVM* with the larger eigenvalue, termed the first principal component in the standard principal component analysis (PCA) (Ringnér 2008), sets the direction in state space along which the variation in the data is maximal. Close to criticality, i.e., when $\lambda_1 \rightarrow 0$, the direction becomes the same as that corresponding to the slow mode of the dynamics. In the case of several genes, the state space is high-dimensional and the PCA provides an efficient tool for analyzing the data. The PCA determines new variables, termed the principal components (PCs), which are linear combinations of the original state space variables. The PCs can be shown to be the normalized eigenvectors of the covariance matrix with the eigenvectors being ordered by the magnitudes of the associated eigenvalues. The eigenvector corresponding to the largest eigenvalue defines the first PC which sets the direction of maximum variation in the state space. The second PC accounts for the dominant part of the remaining variation and so on. Thus, one needs to consider only the first few PCs to retain most of the variation present in the original data set. If the original data constitute an m -dimensional state space and one retains the first q PCs ($q < m$), there is an effective dimensional reduction in the analysis of the data without significant loss in information.

The experiment by Chang *et al.* (2008) on blood cell differentiation in mice provides evidence of both slow kinetics and broad heterogeneity (large variance). The distribution of the Sca-1 marker protein levels in the undifferentiated cell population exhibits considerable heterogeneity. In the experiment, the clonal cell population

was sorted into three subpopulations with low, medium and high Sca-1 levels respectively. The distributions of the Sca-1 levels in the subpopulations were found to revert back to the distribution in the parental cell population through slow reaction kinetics. In the study by Pal *et al.* (Pal *et al.* 2015), the conjecture was put forward, based on theoretical analysis, that the experimentally observed features are a manifestation of criticality, i.e., the proximity of the differentiating cell population to a bifurcation (critical) point. In a subsequent study, Ridden *et al.* (Ridden *et al.* 2015) analyzed the expression fluctuations of the marker protein Sca-1 in terms of a simple stochastic model (model fit to experimental data was very good) and showed that the computed distribution in the Sca-1 levels lies close to the critical lines separating bistable and monostable regimes and also the maximum entropy distribution. The latter result is consistent with Jaynes' Maximum Entropy Principle (Jaynes 1957) according to which the equilibrium distribution maximizes entropy subject to the constraints imposed on the system. The removal of a constraint increases the entropy of a system. The entropy of a gas confined by a piston in a cylinder is increased when the constraint (piston) is removed and the gas is able to expand freely. In terms of the regulatory constraints of a genetic network, the undifferentiated pluripotent cell population has entropy higher than that of the more tightly regulated differentiated cell populations. Since entropy is a measure of variability, the pluripotent cell population is expected to be more heterogeneous. This is borne out by the steady state distributions exhibited in figure 3.

Recently, a large number of studies have been carried out on the early signatures of sudden regime shifts in systems as diverse as ecosystems, financial markets, population biology and complex diseases (Scheffer *et al.* 2009, 2012). Similar studies on the signatures of regime shifts in gene expression dynamics have been carried out in Refs. (Pal *et al.* 2013, 2015). Regime shifts mostly occur at bifurcation points or may be noise-induced. The early signatures of approaching regime shifts include the critical slowing down and its associated effects, namely, a rising variance and lag-1 autocorrelation function, and an increasingly skewed steady state probability distribution. These quantities are experimentally measurable and provide advance knowledge of the regime shifts occurring at bifurcation (critical) points. In the case of a random variable x , the variance of x is expressed as $\text{Var}(x) = \langle x^2 \rangle - \langle x \rangle^2$ where $\langle \dots \rangle$ represents the mean value over an ensemble of cells. The lag-1 autocorrelation function, ACF , is given by

$$ACF = \frac{\langle x(t+1)x(t) \rangle - \langle x(t+1) \rangle \langle x(t) \rangle}{\sqrt{\text{Var}(x(t+1))\text{Var}(x(t))}} \quad (12)$$

The critical slowing down close to a bifurcation point implies that the system's intrinsic rates of change are lowered. The fluctuations in a state variable thus have a greater propensity for accumulation (rising variance) as the bifurcation point is approached. Similarly, the state of the system at time $t+1$ closely resembles the state at time t giving rise to increased lag-1 autocorrelation which reaches its maximum value at the bifurcation point. The quantities providing early signatures have been computed in the case of the two-gene motif and their behaviour is along expected lines (Pal *et al.* 2015). Another signature of criticality was obtained in terms of a perfect anticorrelation between the two variables x_1 and x_2 . The correlation coefficient, CCF (also known as Pearson's correlation coefficient), is given by the normalized covariance,

$$CCF = \frac{\langle \delta x_1 \delta x_2 \rangle}{\sqrt{\langle (\delta x_1)^2 \rangle \langle (\delta x_2)^2 \rangle}} \quad (13)$$

From equations (10) and (11), the $CCF = -1$ since the coefficient $r_{12} = -1$ (repressive influence). The first experimental evidence of critical slowing down was obtained close to the bifurcation point separating the G2 growth phase from the mitotic phase of the cell division cycle in frog egg extracts (Sha *et al.* 2003). The major predictions of a mathematical model were verified experimentally in the study. Experiments detecting the critical slowing down are difficult to carry out but some recent experiments demonstrate the feature in specific biological systems (Dai *et al.* 2012; Veraart *et al.* 2012). Quantities like the variance and the lag-1 autocorrelation function are measurable using single cell methodologies like flow-cytometry and time-lapse

fluorescence spectroscopy (Weinberger *et al.* 2008; Kaufmann and van Oudenaarden 2007). The proposal made in (Pal *et al.* 2015) that the proximity of a bifurcation point confers the experimentally observed features of broad heterogeneity and slow relaxation kinetics on the cell population can be tested via the early signatures in the experimental measurements mentioned. The signatures would be missing in the case of pure noise-driven transitions. A very recent experimental study by Mojtahedi *et al.* (2016) provides strong support to the idea of critical state transitions. In the experiment it was shown that the destabilization of a high-dimensional attractor state brings about the exit of cells from the pluripotent progenitor state paving the way for cell differentiation into specific lineages. A low-dimensional representation of the data in high-dimensional state space was obtained using the PCA. Single-cell monitoring of the gene expression profiles of cell populations in high-dimensional state space was further carried out and a new index I_C computed which provides the signature for an approaching critical point transition. I_C is defined to be

$$I_C(t) = \frac{\langle |R(g_i, g_j)| \rangle}{\langle R(S^k, S^l) \rangle} \quad (14)$$

where R is the Pearson's correlation coefficient between the cell-state vectors S^k and S^l or between the gene vectors g_i and g_j . The brackets $\langle \dots \rangle$ denote the average of R over all pairs of state or gene vectors associated with a cell population. Let n be the number of cells in the population and m be the number of genes the expressions of which are monitored in each individual cell. The state vector S^k of the k -th cell at time t is defined by the protein concentrations, x_i 's, $i = 1, \dots, m$. The gene vector g_i is defined by the expression levels (protein concentrations) of the i -th gene in each of the n cells. The magnitude of I_C increases as the bifurcation point is approached and attains its maximum value at the bifurcation point itself.

4. Concluding remarks

Cell-fate choice in the form of differentiation into lineages plays an important role in normal embryonic development and in the maintenance of homeostasis in adult tissues (Semrau and van Oudenaarden 2015). Cell differentiation occurs when the undifferentiated (pluripotent) state becomes unstable due to a number of regulatory inputs from signaling pathways. Specifically during development, the drivers of differentiation could include morphogen gradients and cyclic oscillations (Furusawa and Kaneko 2012). Cell-cell interactions are in some cases important in bringing about different fate choices for neighbouring cells (Rouault and Hakim 2012). The focus of the present review is on two fundamental processes underlying cell differentiation in a

clonal population of cells, namely, the regulation of gene expression via positive feedback and single-cell heterogeneities due to the stochastic nature of gene expression. The evidence for the presence of the two-gene motif (figure 1(i)) in gene regulatory networks controlling cell differentiation in different cell lines (Heinäniemi *et al.* 2013) motivates the choice of the motif as the core circuit studied by mathematical modelling. Bistability, an outcome of positive feedback, and bifurcations constitute a universal theme for a number of cell biological phenomena (Pomeroy 2008; Bose 2012). The analogy between a bifurcation point and a critical point in equilibrium phase transitions is not new (van Kampen 1992; Mora and Bialek 2011). The literature on critical point transitions and associated critical phenomena is vast (Goldenfeld 1992) but the studies have so far been confined to non-living matter. The new direction set by current biological research is to identify the signatures of criticality in nonequilibrium biological systems using the concepts and tools of statistical physics. The goal is to compute quantities providing such signatures as well as to obtain experimental validation. There have been suggestions that biological networks like neuronal and genetic networks operate at criticality with the dual functional advantages of robustness and flexibility (Kauffman 1993; Gros 2008; Mora and Bialek 2011). The critical state separates ordered from disordered phases, the ordered phase is resistant to perturbations, i.e., robust whereas the disordered phase favours variable (flexible) responses. A recent study has analysed the experimental gene expression data of a gap-gene network in a developing fruit fly embryo and has identified signatures that the genetic network is tuned to criticality (Krotov *et al.* 2014). The gap-gene network has four main gap genes. There are, however, regions in the expression profile with simpler dynamics involving a pair of mutually repressive gap genes. There is experimental evidence of strong anticorrelation of magnitude -1 between the fluctuations of the two protein levels. The other signatures of criticality include slow dynamics along some directions in state space with large variances of the expression data in these directions, long-range spatial correlations of expression fluctuations in the embryo and a non-Gaussian distribution of these fluctuations. In both the cases of the fruit fly embryo at the developmental stage and blood cell differentiation, the two-gene motif of mutual repression of gene expression is a key component of the regulatory genetic network. The mathematical analysis of criticality is similar in the two cases and forms the basis for identifying criticality in more general gene regulatory networks (Podolskiy *et al.* 2015; Valverde *et al.* 2015). In the cases of gene regulatory networks, dynamical regime changes mostly occur through bifurcations. There is a large body of literature on generic signatures of approaching bifurcation points in diverse systems (Scheffer *et al.* 2009, 2012) conferring a universal character on criticality associated with such systems. The search for

such signatures in the dynamical behavior of gene networks is in its early stages with only a few theoretical and experimental studies carried out to date (Garcia-Ojalvo and Arias 2012; Krotov *et al.* 2014; Pal *et al.* 2015; Mojtahedi *et al.* 2016). In this context, we mention a recent experimental result (Peláez *et al.* 2015) which may be interpreted in terms of an approaching critical point. The transcription factor Yan plays a key role in maintaining the *Drosophila* (fruit fly) eye cells in a multipotent state. Yan is a core component of the gene regulatory network which control the time and place of the multipotent cells into several differentiated lineages. A sharp spike in the Yan expression noise was observed in the experiment coinciding with the early stages of the transition of progenitor cells to differentiated states. If the transition occurs via a bifurcation, the increasing noise may be attributed to a rising variance associated with criticality. Richard *et al.* (Richard *et al.* 2016) analysed at the single cell level the progression of primary chicken erythroid progenitors through the differentiation process. The expression levels of several functionally related genes were measured in individual cells at different time points during differentiation. A significant result emerging out of the study relates to the cell-to-cell variability in gene expression peaking transiently before an irreversible commitment to cell differentiation. Based on the results obtained, the Authors concluded ‘our results fully supported a dynamical view where differentiation can be seen as a critical phase transition driven by stochasticity’. The study of Mojtahedi *et al.* (Mojtahedi *et al.* 2016) and Richard *et al.* (2016) acquire significance for providing experimental support to the criticality conjecture (Pal *et al.* 2015) in cell differentiation. The generality of the idea needs to be tested in multiple cell lines using the known measures associated with the approach to criticality. A recent study (Erez *et al.* 2017) has shown that the nonequilibrium system of biochemical feedback near a bifurcation point exhibits critical properties characteristic of equilibrium systems belonging to the mean-field universality class (Goldenfeld 1992). The quantitative theoretical predictions were tested in single-cell experiments on T cell signaling. The agreement between theory and experiments suggests the validity of the notion of criticality in immune response and more generally in the functioning of cellular networks. On a more fundamental level, one needs to understand the advantages arising from the dynamical tuning of living systems to be near criticality (Hidalgo *et al.* 2014).

Acknowledgements

IB acknowledges the support by CSIR, India, vide sanction Lett. No. 21(0956)/13-EMR-II dated 28.04.2014. MP acknowledges the support from Bose Institute, India, for carrying out the study. The Authors thank Achintya Singha for his help in preparing the manuscript. Figures 1, 2 and 3

are reprinted with permission from the paper titled “Non-genetic heterogeneity, criticality and cell differentiation” by Pal M, Ghosh S and Bose I 2015 *Phys. Biol.* **12** 016001 (IOP, UK).

Glossary: Basic Terminologies of Nonlinear Dynamics

Dynamical system: a system the state of which evolves as a function of time.

State at time t : defined by the magnitudes of the key variables at time t . The key variables could be protein concentrations.

State space: the space of all states. For an N -variable system the state space is N -dimensional with one coordinate axis for each variable.

Time evolution: described by differential rate equations, one equation for each variable. For two variables p and q , the rate equations are

$$\frac{dp}{dt} = f(p, q), \frac{dq}{dt} = g(p, q)$$

where the functions f and g have specific mathematical forms.

Trajectory in state space: each state is defined by a single point in state space. A trajectory is obtained by joining the points representing the states of the system at successive time intervals. A knowledge of the state at time t is obtained by solving the differential rate equations.

Steady state: a state in which all rates of change are zero (e.g. $\frac{dp}{dt} = 0$, $\frac{dq}{dt} = 0$), defines the fixed point of the dynamics.

Stability of steady state: a steady state is stable (unstable) if the system comes back to (moves away from) the steady state after being weakly perturbed from it.

Bistability: two stable steady states are possible for the same parameter values.

Basin of attraction of a stable steady state: defined by a region in state space. All trajectories in this region end up at the fixed point representing the stable steady state.

Attractors of dynamics: trajectories in state space end up in the attractors which include both fixed points and limit cycles. In the latter case, the attractor is a cycle of states repeatedly traversed by the system indicative of periodic motion.

Bifurcation: occurs at specific parameter values at which there is a change in the dynamical regime, involves changes in the number and/or the stability properties of steady state solutions of the differential rate equations.

References

- Alon U 2007 *An introduction to systems biology: design principles of biological circuits* (New York: Chapman and Hall/CRC)
- Bose I 2012 Routes to binary gene expression. *Sci. Culture* **78** 113–121
- Cahan P and Daley GQ 2013 Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* **14** 357–368
- Chang HH, Hemberg M, Barahona M, Ingber DE and Huang S 2008 Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453** 544–547
- Dai L, Vorselen D, Korolev KS and Gore J 2012 Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science* **336** 1175–1177
- Enver T, Pera M, Peterson C and Andrews PW 2009 Stem cell states, fates and the rules of attraction. *Cell Stem Cell* **4** 387–397
- Erez A, Byrd TA, Vogel RM Altan-Bonnet G and Mugler A 2017 Criticality of biochemical feedback. arXiv: 1703.04194 v1 [physics. bio-ph]
- Ferrell JE 2012 Bistability, bifurcations and Waddington’s epigenetic landscape. *Curr. Biol.* **22** R458–R466
- Foster DV, Foster JG, Huang S and Kauffman SA 2009 A model of sequential branching in hierarchical cell fate determination. *J. Theoret. Biol.* **260** 589–597
- Fox RF, Gatland IR, Roy R and Vemuri G 1988 Fast, accurate algorithm for numerical simulation of exponentially correlated coloured noise. *Phys. Rev. A* **38** 5938–5940
- Furusawa C and Kaneko K 2012 A dynamical-systems view of stem cell biology. *Science* **338** 215–217
- Garcia-Ojalvo J and Arias AM 2012 Towards a statistical mechanics of cell-fate decisions. *Curr. Opin. Genet. Dev.* **22** 619–626
- Gros C 2008 *Complex adaptive dynamical systems; a primer* (New York: Springer)
- Goldenfeld N 1992 *Lectures on phase transitions and the renormalization group* (Reading, MA: Addison-Wesley)
- Heinäniemi M *et al.* 2013 Gene-pair expression signatures reveal lineage control. *Nat. Method.* **10** 577
- Hidalgo J, Grilli J, Suweis S, Muñoz MA, Banavar JR and Maritan A 2014 Information-based fitness and the emergence of criticality in living systems. *Proc. Natl. Acad. Sci. USA* **111** 10095–10100
- Huang S, Eichler G, Bar-Yam Y and Ingber DE 2005 Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94** 128701
- Huang S 2009a Reprogramming cell fates: reconciling rarity with robustness. *BioEssays* **31** 546–560
- Huang S 2009b Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136** 3853–3862
- Jaynes ET 1957 Information theory and statistical mechanics. *Phys. Rev.* **106** 620–630
- Kærn M, Elston TC, Blake WJ and Collins JJ 2005 Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6** 451–464
- Kauffman SA 1993 *The origins of order: Self-organization and selection in evolution* (New York: Oxford University Press)

- Kaufmann BB and van Oudenaarden A 2007 Stochastic gene expression: from single molecules to the proteome. *Curr. Opin. Genet. Dev.* **17** 107–112
- Krotov D, Dubuis JO, Gregor T and Bialek W 2014 Morphogenesis at criticality. *Proc. Natl. Acad. Sci. USA* **111** 3683–3688
- MacArthur BD, Ma'ayan A and Lemischka IR 2009 Systems biology of stem cell fate and cell reprogramming. *Nat. Rev. Mol. Cell Biol.* **10** 672–681
- MacArthur BD and Lemischka IR 2013 Statistical mechanics of pluripotency. *Cell* **154** 484–489
- Martinez Arias A and Brickman JM 2011 Gene expression heterogeneities in embryonic stem cell populations: origin and function. *Curr. Opin. Cell Biol.* **23** 1–7
- Mojtahedi M *et al.* 2016 Cell fate decision as high-dimensional critical state transition. *PLoS Biol.* **14** e2000640
- Monod J and Jacob F 1961 Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26** 389–401
- Mora T and Bialek W 2011 Are biological systems poised at criticality? *J. Stat. Phys.* **144** 268–302
- Moris N, Pina C and Arias AM 2016 Transition states and cell fate transitions in epigenetic landscapes. *Nat. Rev. Genet.* **17** 693–703
- Pal M, Pal AK, Ghosh S and Bose I 2013 Early signatures of regime shifts in gene expression dynamics. *Phys. Biol.* **10** 036010
- Pal M, Ghosh S and Bose I 2015 Non-genetic heterogeneity, criticality and cell differentiation. *Phys. Biol.* **12** 016001
- Peláez N *et al.* 2015 Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *eLife* **4** e08924
- Podolskiy D *et al.* 2015 Critical dynamics of gene networks is a mechanism behind ageing and Gompertz law arXiv: 1502.04307 v1 [q-bio.MN]
- Pomerening JR 2008 Uncovering mechanisms of bistability in biological systems. *Curr. Opin. Biotechnol.* **19** 381–388
- Raj A and van Oudenaarden A 2008 Nature, nurture or chance: stochastic gene expression and its consequences. *Cell* **135** 216–226
- Raj A and van Oudenaarden A 2009 Single-molecule approaches to stochastic gene expression. *Ann. Rev. Biophys.* **38** 255–270
- Richard A *et al.* 2016 Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS Biol.* **14** e1002585
- Ridden SJ, Chang HH, Zygalakis KC and MacArthur BD 2015 Entropy, ergodicity and stem cell multipotency. *Phys. Rev. Lett.* **115** 208103
- Ringnér M 2008 What is principal component analysis? *Nat. Biotech.* **26** 303–304
- Rouault H and Hakim V 2012 Different cell fates from cell-cell interactions: core architectures of two-cell bistable networks. *Biophys. J.* **102** 417–426
- Scheffer M *et al.* 2009 Early-warning signals for critical transitions. *Nature* **461** 53–59
- Scheffer M *et al.* 2012 Anticipating critical transitions. *Science* **338** 344–348
- Semrau S and van Oudenaarden A 2015 Studying lineage decision-making in vitro: emerging concepts and novel tools. *Annu. Rev. Cell Dev. Biol.* **31** 317–345
- Sha W *et al.* 2003 Hysteresis drives cell-cycle transitions in *Xenopus laevis* egg extracts. *Proc. Natl. Acad. Sci.* **100** 975–980
- Strogatz SH 1994 *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry and engineering* (Reading, MA: Addison-Wesley)
- Thomas R 1978 Logical analysis of systems comprising feedback loops. *J. Theor. Biol.* **73** 631–656
- Torres-Padilla M-E and Chambers I 2014 Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development* **141** 2173–2181
- Valverde S, Ohse S, Turalska M, West BJ and Garcia-Ojalvo J 2015 Structural determinants of criticality in biological networks. *Front. Physiol.* **6** 127
- van Kampen NG 1992 *Stochastic processes in physics and chemistry* (Amsterdam: Elsevier)
- Veraart AJ, Faassen EJ, Dakos V, van Nes EH, Lürling M and Scheffer M 2012 Recovery rates reflect distance to a tipping point in a living system. *Nature* **481** 357–359
- Wang J, Xu L, Wang E and Huang S 2010 The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys. J.* **99** 29–39
- Weinberger LS, Dar RD and Simpsn ML 2008 Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat. Genet.* **40** 466–470
- Zhou JX and Huang S 2011 Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends Genet.* **27** 55–62
- Zhou JX, Aliyu MDS, Aurell E and Huang SJ 2012 Quasi-potential landscape in complex multi-stable systems. *R. Soc. Interface* **9** 3539–3553

Corresponding editor: BASUTHKAR J RAO