bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Epigenetic Landscape Models: The Post-Genomic Era

Davila-Velderrain J[1,2,*], Martinez-Garcia JC[3,2], Alvarez-Buylla ER[1,2,*]

[1] Instituto de Ecología, Universidad Nacional Autónoma de México, Cd. Universitaria, México, D.F. 04510, México
[2] Centro de Ciencias de la Complejidad (C3), Universidad Nacional Autónoma de México, Cd. Universitaria, México, D.F. 04510, México
[3] Departamento de Control Automático, Cinvestav-IPN, A. P. 14-740, 07300 México, DF, México
[*] E-mail: Corresponding eabuylla@gmail.com

Abstract

Complex networks of regulatory interactions orchestrate developmental processes in multicellular organisms. Such a complex internal structure intrinsically constrains cellular behavior allowing only a reduced set of attainable and observable cellular states or cell types. Thus, a multicellular system undergoes cell fate decisions in a robust manner in the course of its normal development. The epigenetic landscape (EL) model originally proposed by C.H. Waddington was an early attempt to integrate these processes in a universal conceptual model of development. Since then, a wealth of experimental data has accumulated, the general mechanisms of gene regulation have been uncovered, and the placement of specific molecular components within modular gene regulatory networks (GRN) has become a common practice. This has motivated the development of mathematical and computational models inspired by the EL aiming to integrate molecular data and gain a better understanding of development, and hopefully predict cell differentiation and reprogramming events. Both deterministic and stochastic dynamical models have been used to described cell state transitions. In this review, we describe recent EL models, emphasising that the construction of an explicit landscape from a GRN is not the only way to implement theoretical models consistent with the conceptual basis of the EL. Moreover, models based on the EL have been shown to be useful in the study of morphogenic processes and not just cell differentiation. Here we describe the distinct approaches, comparing their strengths and weaknesses and the kind of biological questions that they have been able to address. We also point to challenges ahead.

Introduction

The development of multicellular organisms comprises cellular differentiation and morphogenesis (spatio-temporal arrangement of different cell types). The progressive loss

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of potency from pluripotent stem cells to mature, differentiated cells, as well as the reproducible emergence of spatiotemporal patterns through the course of development has been always perceived as strong evidence of the robustness and *deterministic* nature of development. The explanation of such a robust process has always puzzled researchers. For a long time, although not always stated explicitly, the prevailing paradigm in developmental biology was supported on two fundamental paradigms: (1) a mature cell, once established, displays an essentially irreversible phenotype; and (2) the developmental process is controlled by a "program" as a genomic blueprint following a simplistic linear scheme of causation in an essentially deterministic fashion. Discoveries in the last decade have challenged these assumptions. It has been shown that a differentiated state of a given cell is not as irreversible as previously thought, and that in fact it is possible to reprogram differentiated cells into pluripontent states with a plethora of protocols [1–3]. On the other hand, technological advances in time-lapse microscopy and live imaging allowed the *in-vivo* exploration of the role of physical constraints and non-genetic heterogeneity in tissue patterning and morphogenesis [4–6]. There is now compelling empirical evidence supporting intrinsic physical processes as a fundamental source of order (mainly ruled by stochastic self-organization processes) instead of deterministic pre-programmed rules [7]. Although these observations have just recently shift the focus of study of developmental biology and biomedical research, the new evidence supports the proposals that visionary theoretical biologists posited decades ago [8–12]. C.H. Waddington was one of the first to point out that the physical implementation of the information coded in the genes and their interactions imposes developmental constraints while forming an organism. The heuristic model of the epigenetic landscape (EL) was a visionary attempt to consolidate these ideas in a conceptual framework that allows to discuss in an intuitive manner the relationship between genetics, development, and evolution [8, 13, 14].

Now in the data-rich, post-genomic era the EL has been consolidated as the preferred conceptual model for the discussion of the mechanistic basis underlying cellular differentiation, trans-differentiation, and reprogramming events [15–17]. The holy grail of this prolific area of research, the so-called stem cells systems biology, is the possibility of discovering efficient reprogramming or therapeutic strategies by combining mathematical and computational modeling with experimental techniques [18–22]. The formalization of the EL in the context of the study of the dynamical properties of GRNs is a salient theoretical framework which is starting now to show great promise in this direction. Given the current relevance of the intended purpose of modeling the EL associated to molecular networks involved on processes such as stem cell differentiation [23], tissue morphogenesis [24], and carcinogenesis [25]; and the fact that different approaches have been proposed in order to reach similar goals [26–28], here we present an integrative and comparative review which considers the most successful approaches used up to now. Our objective is two-fold: (1) to help different research groups attempting to formalize the EL to reduce the gap existing between current different approaches and (2) to contribute to shape a common discussion ground on EL models between experimentalists and theorists. With this in mind, we present both a conceptual and rigorous exposition of otherwise rather technical and disjunct subjects.

In the sections that follow, we first explain briefly the dynamical systems view of the cell, *i.e.* the standard theoretical framework adopted for the mathematical and computational study of cell differentiation. This formal conceptualization of the cell gives rise to

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the rigorous definition of a given cell phenotype and is the starting point for the construction of EL models. We also elaborate on why there is a growing interest in extending GRN models towards the implementation of EL models. We then summarize some recent work in which derivations of the EL associated with gene regulatory circuits and networks are presented. Moreover, we discuss both the biological and mathematical difficulties associated with each of the considered approaches that have been used. Finally, we conclude with some perspectives and open problems regarding the generalization of computational models motivated by the conceptual basis of EL models.

The Dynamical-Systems View of Cell Biology

Recently, a modern picture of the EL has been emerging in the context of GRNs [27]. An assumption of the "systems dynamics" approach to biology is that cell behavior can be understood in terms of the dynamical properties of the involved molecular regulatory networks [29, 30]. This idea, particularly in the context of development, aligns very well with Waddington's intuition [15], and is now supported by a wealth of consolidated theoretical and experimental work [31, 32]. As a consequence, the orchestrating role of GRNs in developmental dynamics is now widely accepted [33–36]. Furthermore, with the use of modern experimental techniques, the number of known specific gene interactions is growing, and the mechanisms of gene regulation are fairly well understood. This knowledge enables the mathematical description of experimentally grounded GRNs in terms of dynamical systems. There are several excellent reviews on the dynamical-systems view of cell biology [24, 33, 37–39]; in this section, we briefly explain this view.

Under dynamical-systems framework, a cell is considered a dynamical system, assuming that its state at a certain time can be described by a set of time-dependent variables. As a first approximation, it is assumed that the amount of the different proteins or the levels of gene expression (*e.g.* expression profiles) are sufficient to describe the cellular phenotype [40]. Thus, the expression profile is taken as the set of variables representing the state of the cell; each gene in the cell's GRN represents one variable (see Figure 1). Mathematically, the set of variables is represented as a state vector given by $\mathbf{x}(t) = [x_1(t), x_2(t)..., x_n(t)]$ for a GRN with n genes; conceptually, this vector represents the corresponding phenotype. It is useful, both conceptually and mathematically, to imagine an abstract space termed the *state space* of the system, which in the context of GRNs corresponds to the abstract space comprising all the theoretically possible phenotypes a cell can exhibit; each point in this abstract space represents one particular phenotype of the cell that specifically corresponds to a configuration or vector with the expression states or levels in discrete and continuous models, respectively, for each molecular component (node) considered in the network (Figure 1b). Furthermore, it is assumed that the phenotype at a certain time and the phenotype at a later time are connected by a state trajectory in a causal way. Mathematically, the current phenotype is a function or a more general mapping of the initial phenotype and certain additional parameters, which may take into account the influence of the environment and possibly also actions of external control. The connection between phenotypes can be formally expressed by a dynamical equation,

$$\mathbf{x}(t + \delta t) = \mathbf{F}(\mathbf{x}(t), \mathbf{u}, \delta t), \tag{1}$$

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $\mathbf{F}$ represents the map that connects one phenotype with immediately previous phenotype ($\mathbf{F}$ is also known as the *transition map*), $\mathbf{x}(t)$ denotes the phenotype at a certain time t, and $\mathbf{u}$ stands for the vector of additional parameters. Both the time increments δt and the state variables $x_i(t)$ can be either continuous or discrete, depending on the chosen mathematical formalism used. Within the cell, the map $\mathbf{F}$ is implemented by the architecture of the GRN, which specifies both the topology of the network and the nature and form of the corresponding gene regulations [26]. Thus, through the causal connections between the phenotypes, the GRN imposes dynamical constraints and limits the permissible behavior of the cell, specifically of the transient and stable configurations that it may attain. The existence of a dynamical map, which defines the phenotype at a certain time, by means of the corresponding phenotype at an earlier time, expresses the causality of the cellular developmental process. We must point out that when modeling developmental dynamics, two different cases of causal relationships can be considered between two given phenotypes. If the relationship between the two given phenotypes is unambiguous, *e.g.* if the future phenotype is given by the map in a unique way by the initial one, the corresponding model is deterministic. In contrast, if there are several possibilities for the future phenotype depending on *random effects*, the model is stochastic. One of the most salient and impressive features of GRNs is the existence of a small number of stationary gene configurations within the state space [10]. Given a specific GRN, a specific set of phenotypes could satisfy the constraints imposed by the GRN, that is, each of these phenotypes would be connected to itself by the map $\mathbf{F}$, *e.g.* $\mathbf{x}^* = \mathbf{F}(\mathbf{x}^*, \mathbf{u})$. When these steady states ($\mathbf{x}^*$) are also resilient to perturbations, that is, if they return back to the steady state after being kicked away by state variations either of intrinsic or external origin, we refer to them as *attractors*; and the theory states that these correspond to the observable phenotypes or cell types [33]. In formal terms, an attractor state is both stationary (time-invariant) and stable [40]. All other states are either unstable or form part of transitory trajectories channeled towards one of these attractor states. From a geometric point of view, attractors represent regions in the state space toward which the system trajectories are channeled to due to dynamic constraints imposed by the GRN (see [41] for a through and intuitive explanation).

In the last two decades, researchers in the area of systems biology have witnessed the great success of GRN dynamical models as integrative and explanatory frameworks for specific developmental processes. Available experimental techniques in molecular biology enable the physical characterization of regulatory interactions, and from these it is possible to postulate experimentally grounded GRN models which translate to mapping functions $\mathbf{F}$ in a dynamical model. The study of the resulting GRNs of specific real developmental processes has shown that the emergent attractor states of the models are indeed consistent with the observable cell type phenotypes (expression configurations or profiles) in both wild-type and mutant backgrounds (see, for example: [35, 36, 42–44]). Experimental evidence supporting the hypothesis of cell fates as high-dimensional attractors has also been presented [31, 32]. The role of experimentally controllable parameters in differentiation and cell-fate decision events is also commonly studied [45–48]. Thus, it seems that the postulation of experimentally grounded GRN dynamical models, their qualitative analysis and dynamical characterization in terms of control parameters, and the validation of predicted attractors against experimental observations has become a well-established,

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

successful framework for the study of developmental dynamics [49] (Figure 1a-b).

Extending GRN to EL models

The qualitative analysis of the dynamics of GRN models is perfectly suited for the study of the specification of cell fates as a result of the constrains imposed by the associated GRN. This conventional analysis includes the identification and local characterization of attractor states and the comparison of these predicted cell types configurations with the ones that are actually observed in the corresponding biological system. If one is interested in studying the potential transition events among the already characterized stable cellular phenotypes, however, several difficulties arise. Standard analysis of dynamical systems, which focuses on the existence and local properties of a given attractor, fail to capture the main problem which is concerned with the relative properties of the different attractors [28]. In deterministic GRN models, given certain values for the related control parameters, the system under study always converges to a single attractor if initialized from the same state, and once it attains such steady-state it remains there indefinitely. In contrast, during a developmental process, cells change from one stable cell configuration to another one in specific temporal and spatial or morphogenic patterns. Thus, we need additional formalisms in order to explore questions regarding how cells in the course of differentiation decide between one of the available given attractors; or the order in which the system converges to the different attractors, given an initial condition; and to predict how these mechanisms can be altered by external control.

Modeling Goals

The need for extending GRN dynamical models beyond standard local analysis is related with the interest in addressing the following – and similar – questions. Conceptually, given an experimentally determined GRN, how can we study and predict both specific "normal" and altered cellular differentiation events or morphogenic patterns? Is it possible to control the fate of differentiation events through well-defined stimuli? Can we deliberately cause altered morphogenic patterns by means of either genetic or environmental perturbations? Or formally, given a specific dynamical mapping $\mathbf{x}(t+\delta t) = \mathbf{F}(\mathbf{x}(t), \mathbf{u}, \delta t)$, and its associated state space, how can we study the conditions under which a transition event occurs among the attractor states $\mathbf{x}^*$? Is there a reproducible pattern of transitions? Are we able to alter the observed pattern through specific external control perturbations $\mathbf{u}$? To what extent are the observed temporal or spatial morphogenetic patterns emergent consequences of the GRNs? The extension of GRN dynamical models and their analysis in order to address these and similar questions has shown to be a fruitful area of research in recent years [23, 25, 28, 50–54]. The conceptual basis for most of these efforts is the EL.

In the following sections we describe some of the different approaches that have been proposed recently in order to tackle the formulated questions. We start from the simplest deterministic approaches, which primary goal is to derive an explicit landscape (potential) function from the associated gene regulatory circuits. Then we focus on probabilistic approaches: these have shown to be the most appropriate framework to extend GRN dynamical models. Within the stochastic approaches, we distinguish those which attempt to derive an explicit probabilistic landscape from those that are concerned with the transition

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

probabilities between attractors and the temporal evolution of the probability distribution over the cellular states. We give special attention to the kind of biological questions that may be better addressed by the different approaches reviewed here.

Deterministic EL Models From Genetic Circuits

An Introductory Toy Model

A quite simple auto-activating single-gene circuit, a basic model of cell differentiation induction, is exposed in [55] as a conceptual tool to discuss some difficulties regarding Waddington's EL. In this work, the EL is mathematically described by a potential function. In dynamical systems theory, besides the state space approach explained briefly above, there is another way to visualize the dynamics of a system, but applicable only if the system is simple enough: the potential function [56, 57]. The potential is a function $V(x)$ which (in one-dimensional systems) fulfills the relation given by:

$$\frac{dx}{dt} = f(x) = -\frac{dV(x)}{dx}, \tag{2}$$

i.e. $f(x)$ is the negative derivative of the potential, which can be found by direct integration:

$$V(x) = -\int f(x)dx. \tag{3}$$

Such a function defines an attractor landscape for the given dynamical system, and its plot graphically represents the dynamics of the system (Figure 2). Specifically, *minima* of the potential correspond to fixed-point attractors (*e.g.* cell types), and *maxima* correspond to unstable fixed-points. The motion, *i.e.* the state trajectories are given by the gradient lines (the lines of steepest descent of the potential). The trajectories are attracted by the minima of the potential. This corresponds to an intuitive, direct derivation of the EL: a "height" value is associated to each of the points in the state-space in a way that those regions corresponding to attractors (cell types) will have a lower value than that of the other transitory states (Figure 2c). Conceptually, the rolling ball of Waddington's EL will represent the state of a differentiating cell moving from higher to lower regions in state space. Thus, the calculated heights of the different attractors are expected to reflect their developmental potential in a hierarchical way: the lower height the lower potential for differentiation. All one-dimensional systems have a potential function, but most two- or higher-dimensional systems do not [57]. This means that one could only apply this method if the cell is represented by a single-gene (single variable) circuit. Furthermore, note that here the EL plays the role of a "toy" model useful in conceptual discussions, a role quite relevant (see [55]) but similar to that of the original metaphorical proposal of Waddington. In this review we devote more attention to the application of EL models to real specific developmental processes with explanatory and predictive purposes. A more "realistic" sub-network model incorporating several transcription factors in a modular structure would probably be necessary in such cases. The application of the integration-based potential function approach, however, cannot be applied to cases with a higher number of genes. Also one should be cautious when postulating the existence of a potential for living systems in strict sense: a cell is an open non-equilibrium thermodynamical

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

system, and its dynamics in general does not follow a gradient (since the transition rate between two given attractor states is not path-independent) [28, 40]. For this reason authors use the term "quasi-potential" when speaking about cellular dynamics from a system-dynamics point of view (see below). In many cases, in continuous-time models of GRNs the dynamics is given by more general types of autonomous differential equations (DEs) that cannot be derived from a potential function. The time evolution of the cell state $\mathbf{x}(t)$ is commonly modeled by the system of DEs:

$$\frac{dx_i(t)}{dt} = F_i(x_1, x_2, ..., x_f, \mathbf{u}), \tag{4}$$

where $i = 1, 2, ..., n$ for a GRN of n genes. A dynamics defined by such a general DE is a special form of the map in equation (1).

A Deterministic Model of Generic Cell Fate Specification

In the past few years, it has been proposed that small gene circuits with a specific topology drive cell-fate decision at branch points of cell development [22]. One of these circuits, which has been extensively studied both experimentally and theoretically, is that in which the component genes have both cross-inhibitory and self-activating interactions [45,58,59]. Bhattacharya and collaborators recently proposed a numerical approach to derive a deterministic path-integral "quasi-potential" from this two-gene regulatory circuit as a general model of cell fate specification [60]. As mentioned in the previous section, the number of cases where it is possible to derive an analytical potential function is in fact very reduced: in general, the functions $\mathbf{F}$ in the continuous-time model for cellular dynamics (equation (4)) are nonlinear, and cannot be analytically integrated. In order to derive a function with similar analytical properties than the potential function, the authors proposed a computational-based strategy. They defined a quantity U_q that changes incrementally along a trajectory in state space. According to the definition of a potential function (equation (3)), the change of such a quantity $\triangle U_q$ for positive time increments is always negative along a trajectory. This ensures that trajectories flow downhill, and that the attractors of the system would correspond to local minima where the change in $\triangle U_q = 0$. They termed the quantity (U_q) as "quasi-potential" , and calculated its overall change by integrating numerically the change ($\triangle U_q$) along trajectories originating from different points in the phase space. The quasi-potential surface is then drawn graphically by interpolation of trajectories (see [60] for details). Among the principal contributions of this work, according to the authors, are the applicability of their approach to systems for which an analytical potential function cannot be derived as well as to high-dimensional GRNs. Although simple and intuitive, this computational approach suffers from several important limitations. The proposed quasi-potential is minimized along a trajectory *by construction*. It is not clear what information regarding the system can be obtained from this mapping which is not already attainable through the characterization of the 2-D phase space using standard analysis techniques from the qualitative theory of nonlinear dynamics [56, 57]. On the other hand, if one considers a gene regulatory module whose typical number of composing genes range from 10 to 20 – but may contain more than 50 genes (see below) – even if theoretically possible, it is impractical to apply to these higher-dimensional systems either the proposed numerical integration approach or the proposed

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

methodology for drawing the quasi-potential surface. For these and related reasons, the proposed deterministic "quasi-potential" has been criticized; for example, Zhou and collaborators noted that it does not represent either Lyapunov stability or transition rates among different attractors [28]. Consequently, even when a graphical EL is effectively drawn, the general questions raised in the previous section cannot be addressed using this approach.

Stochastic EL Models from Boolean GRNs

The first computational model envisioned for the simulation and analysis of the dynamic behavior of GRNs was the Boolean Network (BN) model [10,11]. Such approach has been extended to model various developmental processes in the context of the EL [25, 50, 51, 61, 62]. A BN models a dynamical system assuming both discrete time and discrete state. This is expressed formally with the mapping:

$$x_i(t+1) = F_i(x_1(t), x_2(t), ..., x_f(t)), \tag{5}$$

where the set of functions F_i are logical prepositions expressing the relationship between the genes that share regulatory interactions with the gene i, and where the state variables $x_i(t)$ can take the discrete values 1 or 0 indicating whether the gene i is expressed or not at a certain time t, respectively. An experimentally grounded Boolean GRN model is then completely specified by the set of genes proposed to be involved in the process of interest and the associated set of logical functions derived from experimental data [24]. A dynamics defined by such a mapping is a special form of the map in equation (1).

Attractor Transition Probability Approach to Explore the EL

In a deterministic framework once a cell state corresponds to an attractor, it will remain there indefinitely. The set of conditions that lead to each attractor comprise the attracting *basin*. Under stochastic fluctuations, the borders of attractor regions in state space may be reached and may be crossed, leading to transitions between attractors [63]. Thus, the implementation of an stochastic dynamical model opens the opportunity to study signal-independent transitions among attractors. There are several approaches to include probabilities in dynamical models. One approach is based on the idea of introducing transition probabilities. As discussed above, when studying cellular developmental dynamics, the transitions of interest are those among attractor states. Can these transitions be studied in terms of probabilities? The answer to this question is positive, since Boolean GRN can be extended in order to include stochasticity and then transition probabilities among attractors can be estimated. Several ways have been proposed to include stochasticity in a Boolean GRN model [64]. One way is the so-called stochasticity in nodes (SIN) model. Here, a constant probability of error ξ is introduced for the deterministic Boolean functions. In other words, at each time step, each gene "disobeys" its Boolean function with probability ξ. Formally:

$$\begin{aligned} P_{x_i(t+1)}[F_i(\mathbf{x}_{reg_i}(t))] &= 1-\xi, \\ P_{x_i(t+1)}[1-F_i(\mathbf{x}_{reg_i}(t))] &= \xi. \end{aligned} \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The probability that the value of the now random variable $x_i(t+1)$ is determined or not by its associated logical function $F_i(\mathbf{x}_{reg_i}(t))$ is $1-\xi$ or ξ, respectively.

Alvarez-Buylla and collaborators used this extended BN model to explore the EL associated with an experimentally grounded GRN [50]. In a BN model the set of possible states (or phenotypes) is finite. Specifically, the state space of a Boolean GRN with n genes has a size of 2^n and is composed by the set of all possible binary vectors of length n (see Figure 3a). Thus, by simulating a stochastic one-step transition starting from each of the possible states a large number of times it is possible to estimate the probability of transition from an attractor i to an attractor j as the frequency of times the states belonging to the basin of the attractor i were mapped into a state within the basin of the attractor j. In [50], the authors followed this simulation approach to estimate a transition probability matrix Π with components:

$$\pi_{ij} = P(A_{t+1} = j | A_t = i), \tag{7}$$

representing the probability that an attractor j is reached from an attractor i (Figure 3b). Once the set of attractors is known and the transition probability matrix is estimated, it is straightforward to implement a discrete time Markov chain model (DTMC) and obtain a dynamic equation for the probability distribution [65]. Indeed:

$$P_A(t+1) = \Pi P_A(t), \tag{8}$$

where $P_A(t)$ is the probability distribution over the attractors at time t, and Π is the transition probability matrix previously estimated. This equation can be iterated to simulate the temporal evolution of the probability distribution over the attractors starting from a biologically meaningful initial distribution. The extension of a Boolean GRN in order to apply this approach is quite simple and intuitive; however, there is a limitation that impedes its general applicability: as the size of the GRN grows, it is not possible to exhaustively characterize the attractor's landscape associated with the GRN in terms of the emergent attractors and its corresponding basins of attraction. If the dynamics of the Boolean GRN is not exhaustively characterized, the corresponding transition probabilities among attractors cannot be estimated using the proposed approach.

EL Exploration in Flower Morphogenesis

Alvarez-Buylla and collaborators applied the previous approach to study the role of stochastic perturbations of the GRN underlying floral organ primordial cell fates in the observed robust temporal morphogenetic pattern observed in most flowering plant species [50]. In flowering plants, a floral meristem is sequentially partitioned into four regions from which the floral organ primordia are formed and eventually give rise to sepals in the outermost whorl, then to petals in the second whorl, stamens in the third, and carpels in the fourth whorl in the central part of the flower. This spatiotemporal pattern is widely conserved among angiosperms. Can the temporal pattern of cell-fate attainment be explained by the interplay of stochastic perturbations and the constraints imposed by a nonlinear GRN? Starting from the previously characterized Boolean GRN of organ identity genes in the *A. thaliana* flower [35], and applying the stochastic approach described above, the authors showed that the most probable order in which the

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

attractors are attained, is in fact, consistent with the temporal sequence in which the specification of corresponding cellular phenotypes are observed *in-vivo*. The model provided then a novel explanation for the emergence and robustness of the ubiquitous temporal pattern of floral organ specification, and also allowed predictions on the population dynamics of cells with different genetic configurations during development [50]. Note that this approach constitutes a landscape-independent implementation of an EL model which successfully addressed an specific biological question; it also constitutes a new approach to understanding a morphogenic process. As seen in this example, it is important to stress that, through the calculation of transition probabilities among attractors, it is possible to explore the EL associated with a GRN independently of an explicit derivation of a landscape or quasi-potential. In the same paper, the authors also showed that a stochastic continuous approximation of the GRN under analysis yielded consistent results.

Probabilistic Landscape (quasi-potential) Approach

Han and Wang proposed a different approach in order to extend a BN model. Their goal was to first estimate the one-step transition probabilities among all the possible points in the state space and not just among given attractors [51]. For this, they implemented a variation of the BN that was previously proposed by Li and collaborators [66] and which has been called the threshold network formalism [67]. In this model, the structure of the network is formally represented with an adjacency matrix $\mathbf{C}$, whose components c_{ij} take a value from the set $[-1, 0, 1]$ indicating the nature of the interaction – inhibition (-1), absence (0), or activation (1) – from the gene j to gene i. The dynamic mapping for the BN takes the form:

$$x_i(t+1) = \begin{cases} 1, & \sum_j c_{ij} x_j(t) + b_i > 0, \\ 0, & \sum_j c_{ij} x_j(t) + b_i < 0, \\ x_i(t), & \sum_j c_{ij} x_j(t) + b_i = 0, \end{cases} \tag{9}$$

where b_i is a parameter representing the ground state of the gene i: its state in the absence of regulation. In words, if the total input of a gene in the network is positive (activation), negative (repression) or zero; the future state of the gene will be active, inactive or unchanged from its previous state. Here, the total input of a gene is the sum of the previous states of the genes regulating it. The characterization of the entire attractor's landscape can then be done through numerical iteration of this dynamical map as long as the network has a moderate size.Han and Wang extended this deterministic BN model into a probabilistic framework as follows. If the interest is focused on the computation of the probability of transition from one phenotype (state) to another phenotype for each of the 2^n possible phenotypes in state space, then it is necessary to introduce a transition probability matrix with the probability of all possible transitions; this calculation is practically impossible. In order to work this limitation, Han and Wang introduced a simplification: they assumed that the one-step transition probability of one phenotype to another can be expressed as the product of the probability of each gene in the network being activated or not given the phenotype of the network in the previous time. Formally:

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\pi_{kj} = P(\mathbf{x}(t+1) = k|\mathbf{x}(t) = j) = \prod_{i=1}^{n} P(x_i(t+1)|\mathbf{x}(t) = j), \tag{10}$$

where j and k represent two different phenotypes and can take values from $[1, ..., 2^n]$; n is the number of genes in the network. The factorized transition probabilities are calculated by inserting a nonzero regulatory input ($\sum_{j=1}^{n} c_{ij}x_j(t) + b_i(t) \neq 0$) as the argument of a sigmoidal function whose range spans from 0 to 1, which is to say:

$$P(x_i(t+1) = 1|\mathbf{x}(t) = j) = \frac{1}{2} \pm \frac{1}{2} \tanh \left[\mu \sum_{j=1}^{n} c_{ij}x_j + b_i \right]. \tag{11}$$

In the case of no input (*i.e.* $\sum_{j=1}^{n} c_{ij}x_j(t) + b_i = 0$) a small-valued parameter d is introduced:

$$P(x_i(t+1) = x_i(t)|\mathbf{x}(t) = j) = 1 - d.$$

In words, the probability that a gene i will be on (1) at the future time $t+1$ will be closer to one as long as its total input at the previous time t is high. Similarly, the probability of being off (0) at the future time will be closer to 1 as long as the regulatory input is low (see Figure 3c). On the other hand, if there is no input to the gene, the probability of no change from its previous state is close to 1, and the closeness depends on the parameter d, a small number representing self-degradation. These rules ensure that the state of a gene will flip only if its total input is large enough.

After having calculated these probabilities, the general idea is then to use this information to obtain an appropriate "height" measure for each of the 2^n states. With this in mind, the interest is first in calculating a steady-state probability distribution $P_{SS}(\mathbf{x})$. This stationary probability distribution is analogous to stationary configurations in the deterministic case; however, in the stochastic framework what keeps invariant in time is not the state of the system but the probability of being in any particular state. In other words, when this stationary distribution is reached, the probability of observing a cell in a particular phenotype does not change. Intuitively, one would expect that attractors would have a higher probability of being reached than transitory states. Thus, from a landscape perspective, the potency of differentiation and height should be inversely related with the probability. The approach that has been followed is to associate this $P_{SS}(\mathbf{x})$ with a height value. Wang has proposed that the probability distribution for a particular state $P(\mathbf{x}_i) = \exp[U(\mathbf{x}_i)]$, and from this expression then $U(\mathbf{x}_i) = -\ln P(\mathbf{x}_i)$, where $i = 1, ..., 2^n$. This function U has been termed the (probabilistic) quasi-potential [26, 54, 68]. The association between such a "quasi-potential" and the steady-state probability is still an open research area (see below) [28]. The key point which has been championed by Wang and coworkers is that, although there is (in general) no potential function directly obtainable from the deterministic equations for a network, a generalized potential (or "quasi-potential") function can be constructed from the probabilistic description. This generalized potential function is inversely related to the steady-state probability [51, 69, 70]. For the case of the extended BN model, once the transition matrix is calculated, the information of the steady-state probabilities can be obtained by solving a discrete set of master equations (ME) for the network [51]. The so-called ME is a dynamical equation for the temporal evolution of a probability distribution [71]. In discrete form it is written:

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\frac{\partial}{\partial t}P(\mathbf{x}_i) = \sum_j W_{ji}P(\mathbf{x}_j) - \sum_j W_{ij}P(\mathbf{x}_i), \tag{12}$$

where we used W_{ij} to denote the transition probabilities resulting from (11). The difference between this dynamical equation and the one discussed in the previous section is that here the time variable is treated as a continuous one. It is in general quite complicated to analyze MEs. In the case of this model, one ME is obtained for each of the 2^n states. Han and Wang conveniently analyzed the whole set of equations following a numerical (iterative) method starting from uniform initial conditions $P_{\mathbf{x}_i}(t_0) = 1/2^n$ and iterating until a stationary distribution is reached [51].

From Probabilistic Landscapes to Putative Cancer Therapies

The previous approach has been applied to two specific processes: cell cycle regulation [51], and DNA damage response [25]. In the former case, the focus was on the global robustness properties of the network. Here we discuss the biological implications derived from the latter case. Choi and collaborators applied this BN probabilistic landscape approach to study state transition in a simplified network of the p53 tumor suppressor protein. The analysis of this network using the probabilistic landscape approach allowed the systematic search for combinatorial therapeutic treatments in cancer [72]. Given the network, key nodes and interactions that control p53 dynamics and the cellular response to DNA damage were identified by conducting single node and link mutation simulations; as a result, one network component, the molecule Wip1, was identified as one of the critical nodes. The flexibility of the BN model also enabled the specification of a MCF7 cancer cell by fixing the state of three nodes of the "normal" network in the course of simulations (for details, see [25, 72]). Having specified two different network models, it was possible to compare the dynamics and associated quasi-potential of both normal cells and cancer cells in the absence and presence of DNA damage. Previous experimental observations indicated that prolonged p53 activity induces senescence or cell death; this behavior was shown to result from the inhibition of the interaction between the molecules Mdm2 and p53 caused by the action of the small molecule Nutlin-3 [73]. Using the model, Choi and collaborators predicted that neither Wip1 nor Mdm2-p53 interaction mutation alone were sufficient to induce cell death for MCF7 cancer cells in the presence of DNA damage; furthermore, the model provided a mechanistic explanation for this behavior: the effect of each of this perturbations alone is not enough to move the system out of an specific attractor's basin. However, the simultaneous application of the two perturbations drives cancer cells to cell death or cell senescence attractors. These theoretical predictions were then validated using single-cell imaging experiments [25, 72]. This study illustrated in an elegant way how cancer therapeutic strategies can be studied in mechanistic terms using a computational EL model. It must be pointed out that this result opened the door to the rational design of system dynamics cancer therapeutical techniques, in contrast to trial and error and reductionist approaches that have dominated the biomedical field.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stochastic EL Models from Continuous GRNs

Like the BN model, the general system of DEs can be extended in order to include stochasticity (in order to approach the modeling formalism to actual biological circunstances). The most intuitive extension considers the introduction of driving stochastic forces; in this approach, equation (4) is extended as:

$$\frac{d\,x_i(t)}{dt} = F_i(\mathbf{x}, \mathbf{u}) + \xi_i(t), \tag{13}$$

where $\xi_i(t)$ is the ith component of a driving stochastic force with zero mean value (*i.e.* $< \xi_i(t) >= 0$). This description, the so-called Langevin equation, is frequently used to model cellular dynamics under stochastic fluctuations [23, 53, 54, 74]. Although intuitively simple at first sight, the consideration of a randomly varying quantity affecting the dynamics of the system implies several conceptual issues that should be considered with care. Any single cell will follow an erratic trajectory in state space, and its developmental dynamics will make each realization different *even if it starts from exactly the same initial condition*. Under this stochastic scenario, two equivalent perspectives to study the stochastic dynamics can be considered. On the one hand, the analysis could be focused on trajectories described by Langevin type equations, which describe the developmental dynamics of a single cell (Figure 4a). On the other hand, as the stochastic forces $\xi_i(t)$ vary from cell to cell in an *ensemble* (population) of cells, the state $\mathbf{x}(t)$ will also vary from cell to cell at any given time. One therefore may ask for the probability $P(\mathbf{x}, t)$ to find the state of a cell in a given state interval of the state space or, equivalently, for the frequency of cells in the ensemble whose states are in that state interval. The focus then shifts from the dynamics of the state of one cell to the dynamics of the distribution over the states in a given ensemble of cells. Indeed, an equation for the temporal evolution of this distribution $P(\mathbf{x}, t)$ can be constructed, the so-called *Fokker-Plank equation (FPE)* [75]:

$$\frac{\partial P}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}[A_i(\mathbf{x})P] + \frac{1}{2}\sum_{i,j} Q_{i,j}(\mathbf{x})\frac{\partial^2}{\partial x_i \partial x_j}P. \tag{14}$$

In mathematical terms, the corresponding process is known as a *diffusion process*, a mathematical model for stochastic phenomena evolving in continuous time [76, 77]; the vector $\mathbf{A}(\mathbf{x})$ is known as the drift vector and the matrix $\mathbf{Q}(\mathbf{x})$ as the diffusion matrix. The FPE describes the change of the probability distribution of a cell state during the course of time (Figure 4b). Conceptually, the latter modeling perspective can be interpreted as the temporal evolution of a cloud (ensemble) of cells diffusing across the state space following both attracting and stochastic forces [78]. The stochastic nature of the trajectories also allows qualitatively richer dynamics in state space. Specifically, if one is interested in the developmental connection between one specific initial phenotype and one specific final phenotype – for example two different given attractors – there is no longer a single possible path connecting them. Instead, the same final phenotype can be reached following different paths in state space (Figure 4c). This situation raises yet another interesting problem: are all the paths equally probable? Is there a dominant path for the transition? Physicists have proposed the so-called path-integral formalism in order to tackle these and similar questions [79]. Specifically, one may want to answer what is the probability of starting from an initial phenotype at a certain time and ending in another phenotype

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

at a future time. The conceptual basis of this strategy is based on the idea of calculating an average trajectory (*e.g.* integrating over the possible paths). The calculated averaged path corresponds to the dominant path that the underlying process follows (Figure 4c).

To summarize this section: when a stochastic component with specific properties is introduced in a continuous-time dynamical model of developmental dynamics the behavior of the system can be studied from different, mathematically equivalent perspectives; one of the perspectives could be more appropriate than the others given the biological question of interest; the different perspectives complement each other, nonetheless. It is important to note that the three approaches mentioned above (e.g. Langevin, FPE and path-integral) although just recently introduced in systems biology [53, 54, 80]; are actually well established tools in nonequilibrium statistical mechanics and the stochastic approach to complex systems [71, 76, 81].

A Diffusion Approach to Study the EL

The three perspectives to study continuous-time stochastic models of developmental dynamics, briefly described in the previous subsection and represented in Figure 4, have been applied to the understanding of actual developmental cases from an EL point of view. Villarreal and collaborators recently proposed a procedure to construct a probabilistic EL by calculating the probability distribution of stable gene expression configurations arising from the topology of a general N-node GRN [53]. In this approach, the focus of study is the temporal evolution of the distribution over state space (Figure 4b) starting from a position centered on a specific attractor configuration. In other words, the proposed framework predicts how a cloud of cells distributed over a particular attractor will diffuse in time to the neighboring regions (attractors) in state space, given a specific GRN (which constraints the state trajectories). The method has been applied to the case of early flower morphogenesis (see subsection above); and its behavior, in both wild-type and mutant conditions, was in agreement with the temporal developmental pattern of floral organs attainment in *A. thaliana* and most flowering species [53]. This implementation, although consistent with the conceptual basis of the EL, does not depend on an underlying (quasi)-potential function and uncovers the interplay of nonlinear constraints and stochastic forces driving developmental dynamics.

The existence of a potential or "potential-like" function associated with diffusive systems, on the other hand, has been an intensive focus of study in theoretical physics and applied mathematics. Ao and co-workers have published important contributions on this regard [82–85]. Particularly, Ao has proposed a transformation that allows the definition of a function $U(\mathbf{x})$ which succesfully acquires the dynamical meaning of a potential function. The corresponding approach has been applied successfully to study several biological systems such as the *phage lambda life cycle* [86], and the carcinogenesis processes, [87, 88] from a landscape perspective. This transformation has also been discussed recently in the context of general methods for the decomposition of multivariate continuous mappings $F(\mathbf{x})$ and their associated quasi-potentials [28]. These, as well as some associated problems, are still an open area of theoretical research [28, 89]. From the available decomposition methods, the one that has been applied the most to specific developmental processes is the potential landscape and flux framework proposed by Wang and co-workers [90]. In this framework, the continuous dynamical mapping $F(\mathbf{x})$ is decomposed into a gradient

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

part and a flux, curl part (for details, see [91]). This approach has been applied, for example, to the study of the yeast cell cycle [69,92]; a circadian oscillator [93]; the bi-potent stem cell differentiation process [54]; and neural differentiation [52]. More recently, this method has been applied in the context of the differentiation and reprogramming of a human stem cell network [23]. Here we further discuss the latter as a diffusion landscape approach to study stem cell differentiation.

Cell Fate Decisions in the Human Stem Cell Landscape

Recently, Li and Wang adopted the diffusion approach to study a previously published [94] human stem cell developmental network composed of 52 genes [23]. In this study they showed how the three perspectives represented in Figure 4 can complement each other in the study of cellular differentiation: (1) through the numerical analysis of the Langevin-like equations for the complete network they acquired a landscape directly from the statistics of the trajectories of the system; (2) by means of approximations they studied the evolution of the probabilistic distribution and obtained an steady-state distribution; and (3) they calculated the dominant paths following a path-integral formalism [80]; the obtained paths were interpreted as the biological paths for differentiation and reprogramming [23]. As Li and Wang showed, from the results of the three perspectives it is possible to quantitatively describe the underlying landscape. One then may be interested in how the landscape changes in response to specific perturbations. A general problem of concern in stem cell research is why do the known reprogramming strategies, which commonly consist on combining perturbations to specific transcription factors, actually work. Li and Wang systematically tested which genes and regulatory interactions affect the most the quantitative properties of the landscape (*e.g.* height values and transition rates) when perturbed. Interestingly, several biological observations associated with the manipulation of the so-called Yamanaka factors (Oct3/4, Sox2, Klf4, c-Myc) - the transcription factors considered the core regulators in the induction of pluripotency – were consistent with the observed modeling results; for example, simulated knockdown perturbations to these factors consistently increased (lowered) the probability (height) of the differentiation state. On the other hand, the path-integral formalism allowed them to show how specific perturbations to these factors cause the differentiation process to be easier or harder in terms of the time spent during transitions and the characteristics of the differentiation paths. In other words, this study presented an important contribution towards the mechanistic, dynamical explanation of the characterized reprogramming strategies in terms of the properties of the underlying EL.

Concluding Remarks

For the practical implementation of EL models W. Wang recently proposed the following overall four-steps strategy [72]: (1) establishment of a GRN; (2) characterization of the attractor (and quasi-potential) landscape through dynamical modeling; (3) computational prediction of cell state responses to specific perturbations; (4) analysis of the prevailing paths of cell fate change. The first step (1) is already a well-established research problem that includes expert curation of experimental data and/or statistical inference. In this review we focused on the second step and presented examples of how steps (3) and (4)

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

can be achieved once a EL models is effectively constructed. As shown here, there are several ways to implement an EL model starting from a GRN. The specific choice should be made considering the properties of the network and the associated questions of interest. The methodologies reviewed here are mostly well-suited to approach the problem of differentiation at the cellular level in a mechanistic setting. The observed behavior results from constraints given by the joint effect of nonlinear regulatory interactions and the inherent stochasticity; however, the actual physical implementation of these generic mechanisms in a multicellular system would imply additional sources of constraint. Tissue-level patterning mechanisms such as cell-cell interactions; chemical signaling; cellular growth, proliferation, and senescence; inevitably impose physical limitations in terms of mechanical forces which in turn affect cellular behavior. This would thus imply non-homogenous GRNs with contrasting additional chemical and physical constraints. Given this fact, the next logical step to extend EL and associated dynamical models would be to account for these physical processes in an attempt to understand how cellular decisions occur during tissue patterning and not just in cell cultures. Although some progress has been presented in this direction [95, 96], the problem certainly remains open. From a theoretical perspective, a further challenge would be to carefully evaluate the assumptions implicit in the EL models. For example, the adoption of the diffusive perspective briefly explained above – which is often taken as a standard in stem cell systems biology – implicitly assumes certain properties about the forces driving the temporal evolution of the system [81]. Do these conditions are universally met by developmental systems? Recent interesting work is starting to suggest the biological relevance of additional constraints such as state-dependent fluctuations [97, 98].

Overall, the application of the methodologies discussed in this review to specific developmental processes has shown the practical relevance of dynamical models consistent with the conceptual basis of the EL and the fundamental role of the constraints imposed by the GRN interactions. Both approaches, landscape-(in)dependent, are useful to answer specific questions and can complement each other. So far, EL models have shown to be an adequate framework to understanding stem cell differentiation and reprogramming events in mechanistic terms; and are also starting to show promise as the basis for rational cancer therapeutic strategies.

Acknowledgments

This work was supported by grants from CONACYT, Mexico: 124909 to ERAB; from PAPIIT-UNAM, IN229009-3 (ERAB). JDV receives a Phd scholarship from CONACYT. The authors acknowledge technical support of Rigoberto V. PérezRuiz and logistical and administrative help of Diana Romo. The authors aknowledge the Centro de Ciencias de la Complejidad (C3), UNAM.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

References

1. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. cell 126: 663–676.

2. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. cell 131: 861–872.

3. González F, Boué S, Belmonte JCI (2011) Methods for making induced pluripotent stem cells: reprogramming a la carte. Nature Reviews Genetics 12: 231–242.

4. Trier SM, Davidson LA (2011) Quantitative microscopy and imaging tools for the mechanical analysis of morphogenesis. Current opinion in genetics & development 21: 664–670.

5. Gleghorn JP, Manivannan S, Nelson CM (2013) Quantitative approaches to uncover physical mechanisms of tissue morphogenesis. Current opinion in biotechnology .

6. Keller PJ (2013) Imaging morphogenesis: Technological advances and biological insights. Science 340.

7. Mammoto T, Ingber DE (2010) Mechanical control of tissue and organ development. Development 137: 1407–1420.

8. Waddington CH (1957) The strategy of genes. London: George Allen & Unwin, Ltd.

9. Goodwin BC, et al. (1963) Temporal organization in cells. a dynamic theory of cellular control processes. Temporal organization in cells A dynamic theory of cellular control processes .

10. Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. Journal of theoretical biology 22: 437–467.

11. Kauffman S (1993) The origins of order: Self organization and selection in evolution. Oxford University Press.

12. Goodwin BC (2001) How the leopard changed its spots: The evolution of complexity. Princeton University Press.

13. Gilbert SF (1991) Epigenetic landscaping: Waddington's use of cell fate bifurcation diagrams. Biology and Philosophy 6: 135–154.

14. Slack JM (2002) Conrad hal waddington: the last renaissance biologist? Nature Reviews Genetics 3: 889–895.

15. Fagan MB (2012) Waddington redux: models and explanation in stem cell and systems biology. Biology & Philosophy 27: 179–213.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

16. Ladewig J, Koch P, Brüstle O (2013) Leveling waddington: the emergence of direct programming and the loss of cell fate hierarchies. Nature Reviews Molecular Cell Biology .

17. Enver T, Pera M, Peterson C, Andrews PW (2009) Stem cell states, fates, and the rules of attraction. Cell Stem Cell 4: 387–397.

18. Roeder I, Radtke F (2009) Stem cell biology meets systems biology. Development 136: 3525–3530.

19. MacArthur BD, Ma'ayan A, Lemischka IR (2008) Toward stem cell systems biology: from molecules to networks and landscapes. Cold Spring Harbor symposia on quantitative biology 73: 211–5.

20. Huang S (2011) Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. Philosophical Transactions of the Royal Society B: Biological Sciences 366: 2247–2259.

21. MacArthur BD, Ma'ayan A, Lemischka IR (2009) Systems biology of stem cell fate and cellular reprogramming. Nature Reviews Molecular Cell Biology 10: 672–681.

22. Zhou JX, Huang S (2011) Understanding gene circuits at cell-fate branch points for rational cell reprogramming. Trends in Genetics 27: 55–62.

23. Li C, Wang J (2013) Quantifying Cell Fate Decisions for Differentiation and Reprogramming of a Human Stem Cell Network: Landscape and Biological Paths. PLoS Computational Biology 9: e1003165.

24. Alvarez-Buylla ER, Azpeitia E, Barrio R, Benítez M, Padilla-Longoria P (2010) From abc genes to regulatory networks, epigenetic landscapes and flower morphogenesis: making biological sense of theoretical approaches. Seminars in cell & developmental biology 21: 108–117.

25. Choi M, Shi J, Jung SH, Chen X, Cho KH (2012) Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to dna damage. Science signaling 5: ra83.

26. Huang S (2009) Reprogramming cell fates: reconciling rarity with robustness. Bioessays 31: 546–560.

27. Huang S (2012) The molecular and mathematical basis of waddington's epigenetic landscape: A framework for post-darwinian biology? Bioessays 34: 149–157.

28. Zhou JX, Aliyu M, Aurell E, Huang S (2012) Quasi-potential landscape in complex multi-stable systems. Journal of The Royal Society Interface 9: 3539–3553.

29. Huang S (2004) Back to the biology in systems biology: What can we learn from biomolecular networks? Briefings in functional genomics & proteomics 2: 279–297.

30. Han J (2008) Understanding biological functions through molecular networks. Cell research 18: 224–237.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

31. Huang S, Eichler G, Bar-Yam Y, Ingber DE (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. Physical review letters 94: 128701.

32. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 453: 544–547.

33. Huang S, Kauffman S (2009) Complex gene regulatory networks-from structure to biological observables: cell fate determination. Encyclopedia of Complexity and Systems Science Meyers RA, editors Springer : 1180–1293.

34. Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. Science 311: 796–800.

35. Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER (2004) A gene regulatory network model for cell-fate determination during arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. The Plant Cell Online 16: 2923–2939.

36. Mendoza L, Alvarez-Buylla ER (1998) Dynamics of the genetic regulatory network for arabidopsis thaliana flower morphogenesis. Journal of theoretical biology 193: 307–319.

37. Kaneko K (2006) Life: An introduction to complex systems biology, volume 171. Springer Heidelberg, Germany:.

38. Tyson J, Chen K, Novak B, et al. (2001) Network dynamics and cell physiology. Nature Reviews Molecular Cell Biology 2: 908–916.

39. Forgacs G, Newman S (2005) Biological physics of the developing embryo. Cambridge University Press.

40. Huang S (2013) Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer and Metastasis Reviews : 1–26.

41. Kaneko K (2011) Characterization of stem cells and cancer cells on the basis of gene expression profile stability, plasticity, and robustness. Bioessays 33: 403–413.

42. Von Dassow G, Meir E, Munro EM, Odell GM (2000) The segment polarity network is a robust developmental module. Nature 406: 188–192.

43. Albert R, Othmer HG (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in¡ i¿ drosophila melanogaster¡/i¿. Journal of theoretical biology 223: 1–18.

44. Jaeger J, Crombach A (2012) Lifes attractors. In: Evolutionary Systems Biology, Springer. pp. 93–119.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45. Huang S, Guo YP, May G, Enver T (2007) Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Developmental biology 305: 695–713.

46. Graham TG, Tabei SA, Dinner AR, Rebay I (2010) Modeling bistable cell-fate choices in the drosophila eye: qualitative and quantitative perspectives. Development 137: 2265–2278.

47. Sciammas R, Li Y, Warmflash A, Song Y, Dinner AR, et al. (2011) An incoherent regulatory network architecture that orchestrates b cell diversification in response to antigen signaling. Molecular systems biology 7.

48. Hong T, Xing J, Li L, Tyson J (2012) A simple theoretical framework for understanding heterogeneous differentiation of cd4+ t cells. BMC systems biology 6: 66.

49. Azpeitia E, Davila-Velderrain J, Villarreal C, Alvarez-Buylla ER (2014) Gene regulatory network models for floral organ determination. In: Flower Development: Methods and Protocols, Springer.

50. Alvarez-Buylla ER, Chaos A, Aldana M, Benítez M, Cortes-Poza Y, et al. (2008) Floral morphogenesis: stochastic explorations of a gene network epigenetic landscape. PloS one 3: e3626.

51. Han B, Wang J (2007) Quantifying robustness and dissipation cost of yeast cell cycle network: the funneled energy landscape perspectives. Biophysical journal 92: 3755–63.

52. Qiu X, Ding S, Shi T (2012) From understanding the development landscape of the canonical fate-switch pair to constructing a dynamic landscape for two-step neural differentiation. PloS one 7: e49271.

53. Villarreal C, Padilla-Longoria P, Alvarez-Buylla ER (2012) General theory of genotype to phenotype mapping: derivation of epigenetic landscapes from N-node complex gene regulatory networks. Physical review letters 109: 118102.

54. Wang J, Xu L, Wang E, Huang S (2010) The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. Biophysical journal 99: 29–39.

55. Ferrell JE (2012) Bistability, Bifurcations, and Waddington's Epigenetic Landscape. Current Biology 22: R458–R466.

56. Strogatz S (2001) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering. Perseus Books Group.

57. Fuchs A (2013) Nonlinear Dynamics in Complex Systems. Springer.

58. Chang HH, Oh PY, Ingber DE, Huang S (2006) Multistable and multistep dynamics in neutrophil differentiation. BMC cell biology 7: 11.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59. Andrecut M, Halley JD, Winkler Da, Huang S (2011) A general model for binary cell fate decision gene circuits with degeneracy: indeterminacy and switch behavior in the absence of cooperativity. PloS one 6: e19358.

60. Bhattacharya S, Zhang Q, Andersen ME (2011) A deterministic map of Waddington's epigenetic landscape for cell fate specification. BMC systems biology 5: 85.

61. Flöttmann M, Scharp T, Klipp E (2012) A stochastic model of epigenetic dynamics in somatic cell reprogramming. Frontiers in physiology 3: 216.

62. Ding S, Wang W (2011) Recipes and mechanisms of cellular reprogramming: a case study on budding yeast saccharomyces cerevisiae. BMC systems biology 5: 50.

63. Ebeling W, Feistel R (2011) Physics of Self-organization and Evolution. Wiley. com.

64. Garg A, Mohanram K, De Micheli G, Xenarios I (2012) Implicit methods for qualitative modeling of gene regulatory networks. In: Gene Regulatory Networks, Springer. pp. 397–443.

65. Allen LJ (2010) An introduction to stochastic processes with applications to biology. CRC Press.

66. Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. Proceedings of the National Academy of Sciences of the United States of America 101: 4781–4786.

67. Thompson EG, Galitski T (2012) Quantifying and analyzing the network basis of genetic complexity. PLoS computational biology 8: e1002583.

68. Huang S (2012) The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? BioEssays : news and reviews in molecular, cellular and developmental biology 34: 149–57.

69. Wang J, Huang B, Xia X, Sun Z (2006) Funneled landscape leads to robustness of cell networks: yeast cell cycle. PLoS computational biology 2: e147.

70. Lapidus S, Han B, Wang J (2008) Intrinsic noise, dissipation cost, and robustness of cellular networks: The underlying energy landscape of mapk signal transduction. Proceedings of the National Academy of Sciences 105: 6039–6044.

71. Haken H (1977) Synergetics. An Introduction. Nonequilibrium Phase Trasitions and Self-organization in Physics, Chemistry, and Biology. Berlin.

72. Wang W (2013) Therapeutic hints from analyzing the attractor landscape of the p53 regulatory circuit. Science signaling 6: pe5.

73. Purvis JE, Karhohs KW, Mock C, Batchelor E, Loewer A, et al. (2012) p53 dynamics control cell fate. Science 336: 1440–1444.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

74. Hoffmann M, Chang HH, Huang S, Ingber DE, Loeffler M, et al. (2008) Noise-driven stem cell and progenitor population dynamics. PLoS One 3: e2922.

75. Risken H (1984) Fokker-Planck Equation. Springer.

76. Gardiner CW (2009) Stochastic methods. Springer Berlin.

77. Fuchs C (2013) Inference for Diffusion Processes: With Applications in Life Sciences. Springer.

78. Huang S (2010) Cell lineage determination in state space: a systems view brings flexibility to dogmatic canonical rules. PLoS biology 8: e1000380.

79. Wio H (1999) Application of path integration to stochastic processes: an introduction. Fundamentals and Applications of Complex Systems, Nueva Ed Univ, UN San Luis : 253.

80. Wang J, Zhang K, Xu L, Wang E (2011) Quantifying the waddington landscape and biological paths for development and differentiation. Proceedings of the National Academy of Sciences 108: 8257–8262.

81. Lindenberg K, West BJ (1990) The nonequilibrium statistical mechanics of open and closed systems. VCH New York.

82. Ao P (2004) Potential in stochastic differential equations: novel construction. Journal of physics A: mathematical and general 37: L25.

83. Kwon C, Ao P, Thouless DJ (2005) Structure of stochastic dynamics near fixed points. Proceedings of the National Academy of Sciences of the United States of America 102: 13029–13033.

84. Yin L, Ao P (2006) Existence and construction of dynamical potential in nonequilibrium processes without detailed balance. Journal of Physics A: Mathematical and General 39: 8593.

85. Ao P, Kwon C, Qian H (2007) On the existence of potential landscape in the evolution of complex systems. Complexity 12: 19–27.

86. Zhu XM, Yin L, Hood L, Ao P (2004) Calculating biological behaviors of epigenetic states in the phage λ life cycle. Functional & integrative genomics 4: 188–195.

87. Ao P, Galas D, Hood L, Zhu X (2008) Cancer as robust intrinsic state of endogenous molecular-cellular network shaped by evolution. Medical hypotheses 70: 678–684.

88. Wang G, Zhu X, Hood L, Ao P (2013) From phage lambda to human cancer: endogenous molecular-cellular network hypothesis. Quantitative Biology : 1–18.

89. Ge H, Qian H (2012) Landscapes of non-gradient dynamics without detailed balance: Stable limit cycles and multiple attractors. Chaos: An Interdisciplinary Journal of Nonlinear Science 22: 023140–023140.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

90. Wang J, Xu L, Wang E (2008) Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. Proceedings of the National Academy of Sciences 105: 12271–12276.

91. Wang J (2011) Potential landscape and flux framework of nonequilibrium biological networks. Annual Reports in Computational Chemistry 7 7: 1.

92. Wang J, Li C, Wang E (2010) Potential and flux landscapes quantify the stability and robustness of budding yeast cell cycle network. Proceedings of the National Academy of Sciences 107: 8195–8200.

93. Wang J, Xu L, Wang E (2009) Robustness and coherence of a three-protein circadian oscillator: landscape and flux perspectives. Biophysical journal 97: 3038–3046.

94. Chang R, Shoemaker R, Wang W (2011) Systematic search for recipes to generate induced pluripotent stem cells. PLoS computational biology 7: e1002300.

95. Barrio RÁ, Hernandez-Machado A, Varea C, Romero-Arias JR, Alvarez-Buylla E (2010) Flower development as an interplay between dynamical physical fields and genetic networks. PloS one 5: e13523.

96. Barrio RA, Romero-Arias JR, Noguez MA, Azpeitia E, Ortiz-Gutiérrez E, et al. (2013) Cell patterns emerge from coupled chemical and physical fields with cell proliferation dynamics: The arabidopsis thaliana root as a study system. PLoS computational biology 9: e1003026.

97. Pujadas E, Feinberg AP (2012) Regulated noise in the epigenetic landscape of development and disease. Cell 148: 1123–1131.

98. Weber M, Buceta J (2013) Stochastic stabilization of phenotypic states: The genetic bistable switch as a case study. PloS one 8: e73487.

Figure Legends

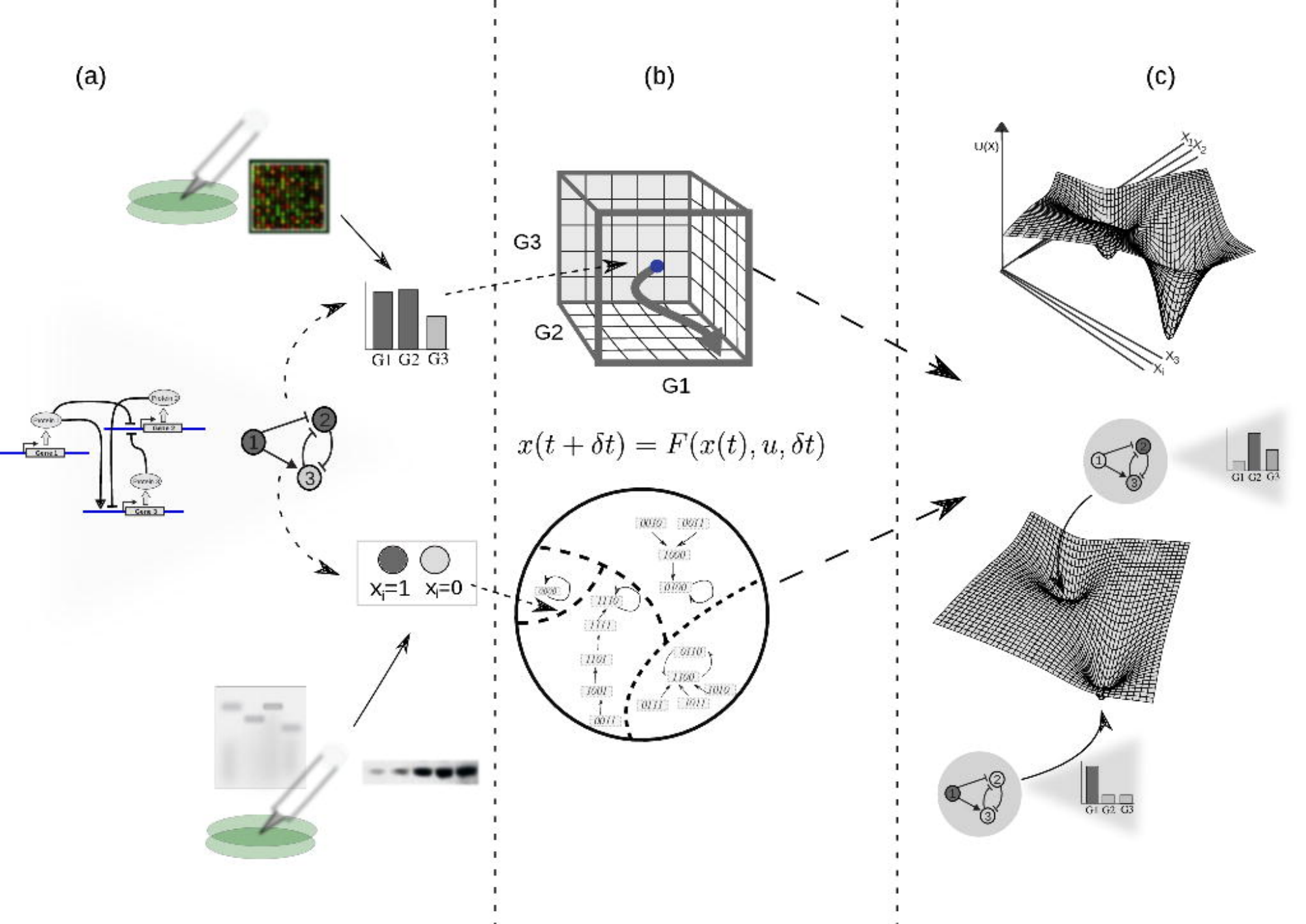
Figure 1. From experimental data on gene function and interactions to a dynamic gene regulatory network and epigenetic landscape model.
(a) The architecture of a GRN is proposed given available experimental molecular data; the state of the network is specified as a gene expression profile or gene on/off (1/0) configuration for the case of continuous or discrete state models, respectively. Boolean or differential equations are used, respectively. (b) The complete set of possible states define a continuous (above) or discrete (below) state space, where each state corresponds to a point; changes in gene expression during developmental dynamics manifest as trajectories in this abstract space (here depicted as arrows). (c) In an intuitive characterization of an EL, an "elevation" value $U(x)$ is associated to each network state x. The association of "elevation" values to network states, or more generally, the quantitative characterization of their relative stability is the ultimate goal of EL modeling efforts. For illustrative purposes, the EL is depicted here as a hypothetical low-dimensional projection.

bioRxiv preprint doi: https://doi.org/10.1101/004192; this version posted April 14, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2. The derivation of a potential function to visualize the epigenetic landscape and the dynamics of a system.
(a) The causal connection between the state of the system (an auto-activating gene circuit) at a certain time and its state at a later time is modeled by a differential equation. (b) Attractors in rate-balance analysis. The blue dotted line and the red curve represent respectively a linear degradation rate, and a nonlinear synthesis rate for the circuit's gene; the restrictions imposed by the circuit to the systems dynamics are met when both rates are balanced. The states that meet this balanced condition are stationary, and if stable (filled circles), are denoted as attractor states x^*. (c) The potential function. The attractor states x^* lie at the bottoms of valleys (minima). The trajectories starting from unstable, transitory states are attracted by the minima of the potential. The relative stability of the left (right) attractor with respect to the other is lower (higher) as quantified by the lower (higher) barrier height between them.

Figure 3. Stochastic epigenetic landscape models from Boolean dynamics.
(a) A simple mutual-inhibition circuit is modeled as a Boolean network: discrete temporal evolution and binary (0,1) state variable. The discrete state space corresponds to the set of binary vectors (here 4 possible states) and is partitioned by two basins of attraction. (b) There are $4(2^2)$ possible transitions among the two attractors. A 2×2 transition probability matrix specifies the probability of each possible transition. (c) Han and Wang proposed the use of a sigmoidal function of the total regulatory input (Ri) to calculate the probability of a one-step state transition of one gene i [51]. A specific value of the function (vertical red dotted line) gives the probability of the gene i becoming active (left) or inactive (right), given its regulatory input (Ri) in the current time. (d) There are $16(2^2 \times 2^2)$ possible transitions among the $4(2^2)$ possible states. A $2^2 \times 2^2$ transition probability matrix specifies the probability of each possible transition.

Figure 4. Differente approaches to study continuous-time stochastic models of the epigenetic landscape and developmental dynamics.
(a) A continuous-time stochastic (diffusion) model is driven by a drift (deterministic) component **F** and an stochastic (**Noise**) force. The graph shows 10 different realizations of the stochastic dynamics of the same, single cell starting form exactly the same initial condition (red dot). This realizations perspective corresponds to the Langevin equation description. The right histogram represents an approximation of the corresponding distribution over the realizations. (b) The picture represents the time evolution of a hypothetical probability distribution. A population of cells initially presents a narrow distribution centered at an intermidiate state value: most cells have an intermidiate state and no individuals show an extreme (low or high) value. As time evolves the shape of the distribution changes – gets wider – , and the population reaches lower and higher values. This perspective corresponds to the Fokker-Planck equation description. (c) A cell can follow different paths (grey dotted lines) to reach a final state x_f starting from an initial state x_0. A finer quantitative characterization of the specific transition from state x_0 to state x_f in terms of highly probable paths and difficulty of differentiation processes can be gained by means of calculating a dominant path (red line) for the transition using a path-integral formalism. For simplicity, the cell state is represented by one variable x in all three cases.


(a)
(b)
(c)
G1 G2 G3
G3
G2
G1
$x(t+\delta t) = F(x(t), u, \delta t)$
x_i=1 x_i=0
U(X)
x_1 x_2
x_3
x_i
G1 G2 G3

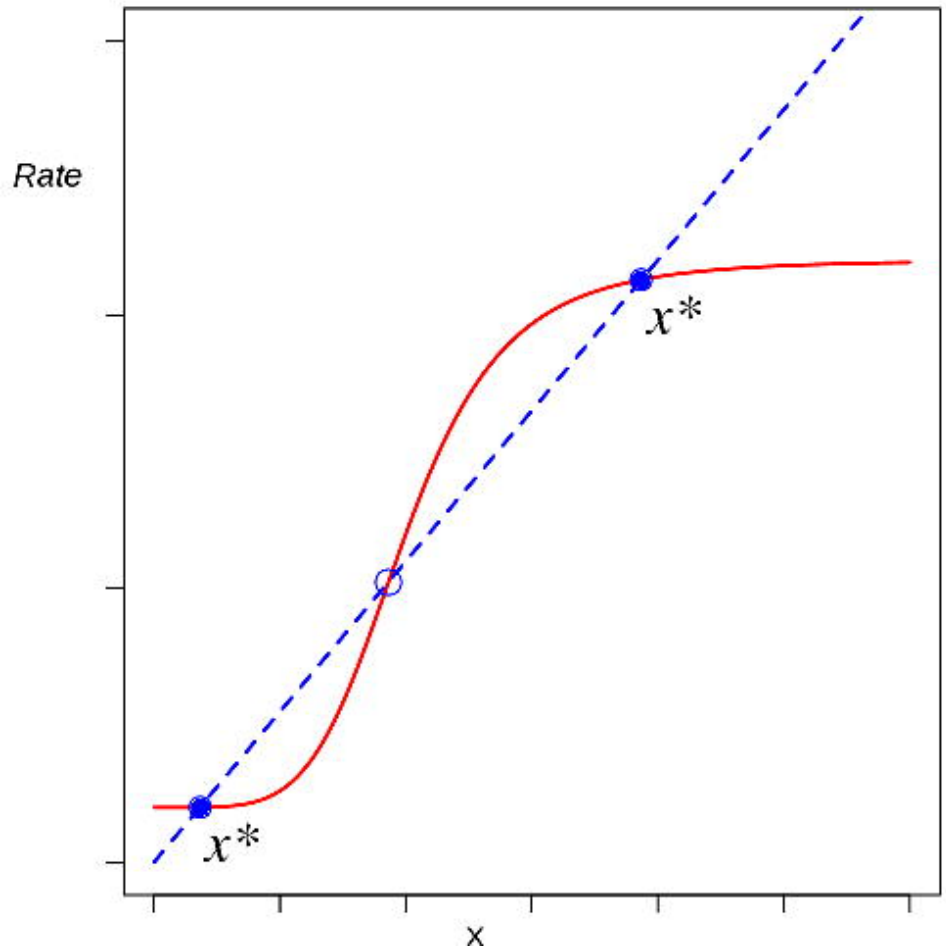
(a)

$$x(t+\delta t) = F(x(t), u, \delta t) \longrightarrow dx = F(x,t)dt$$

(b)

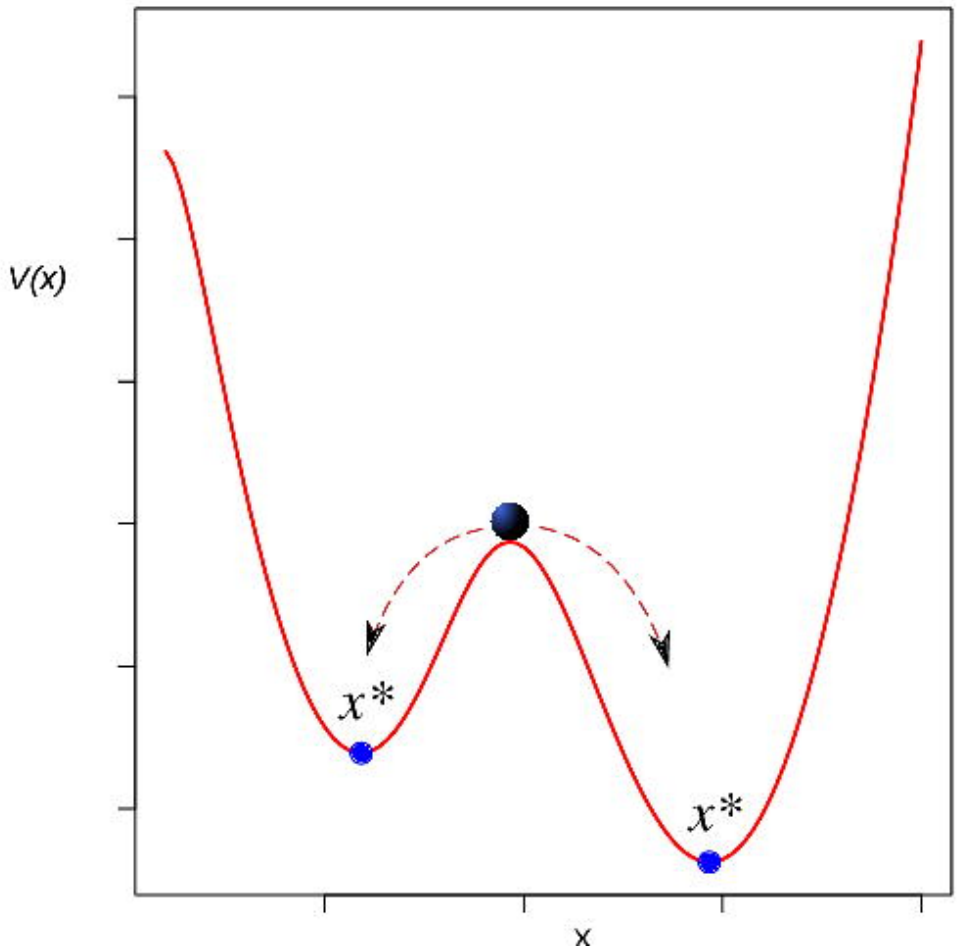
$F(x,t)$= *Synthesis rate - Degradation rate*

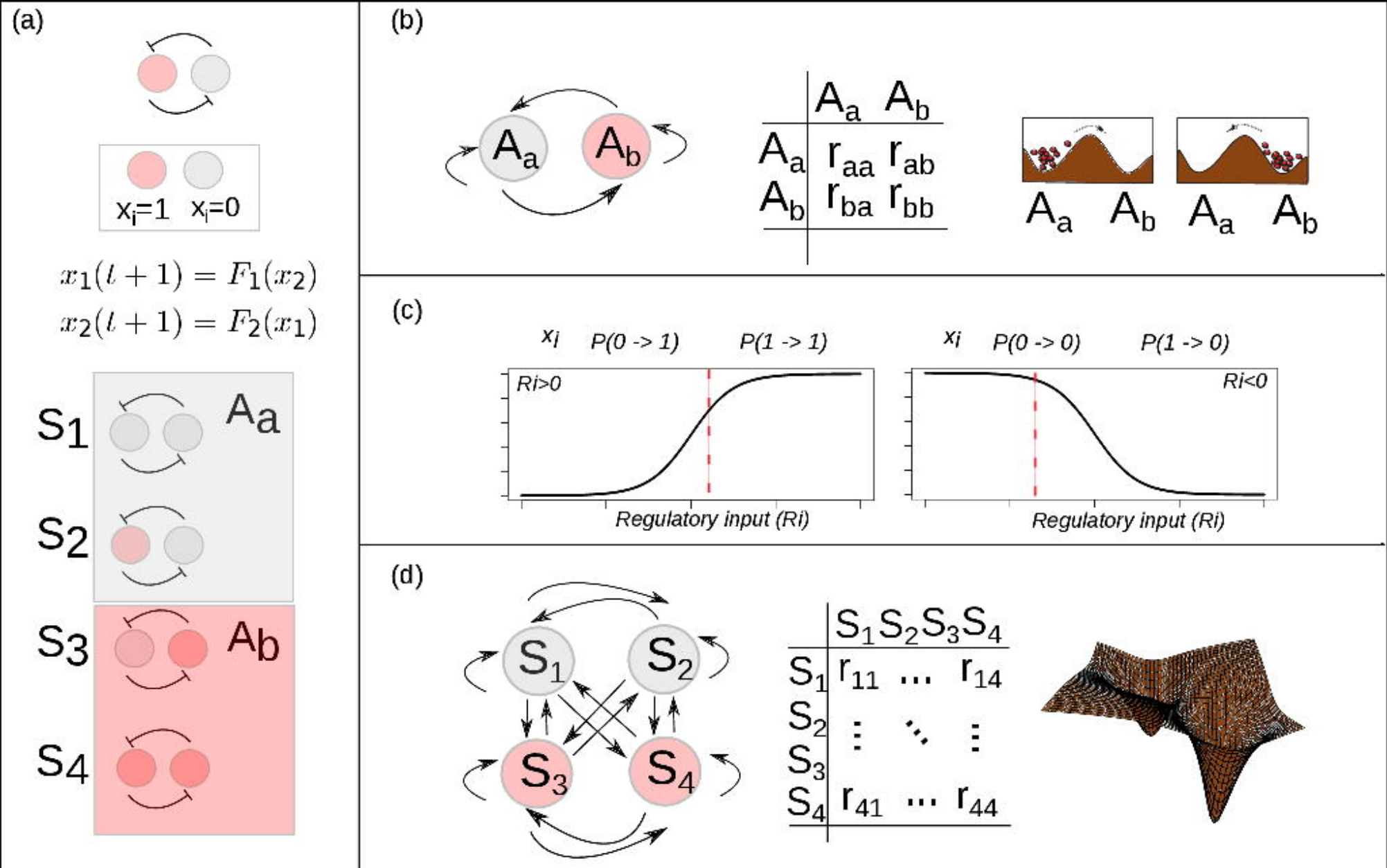
(x^*), *Synthesis rate = Degradation rate*



(c)

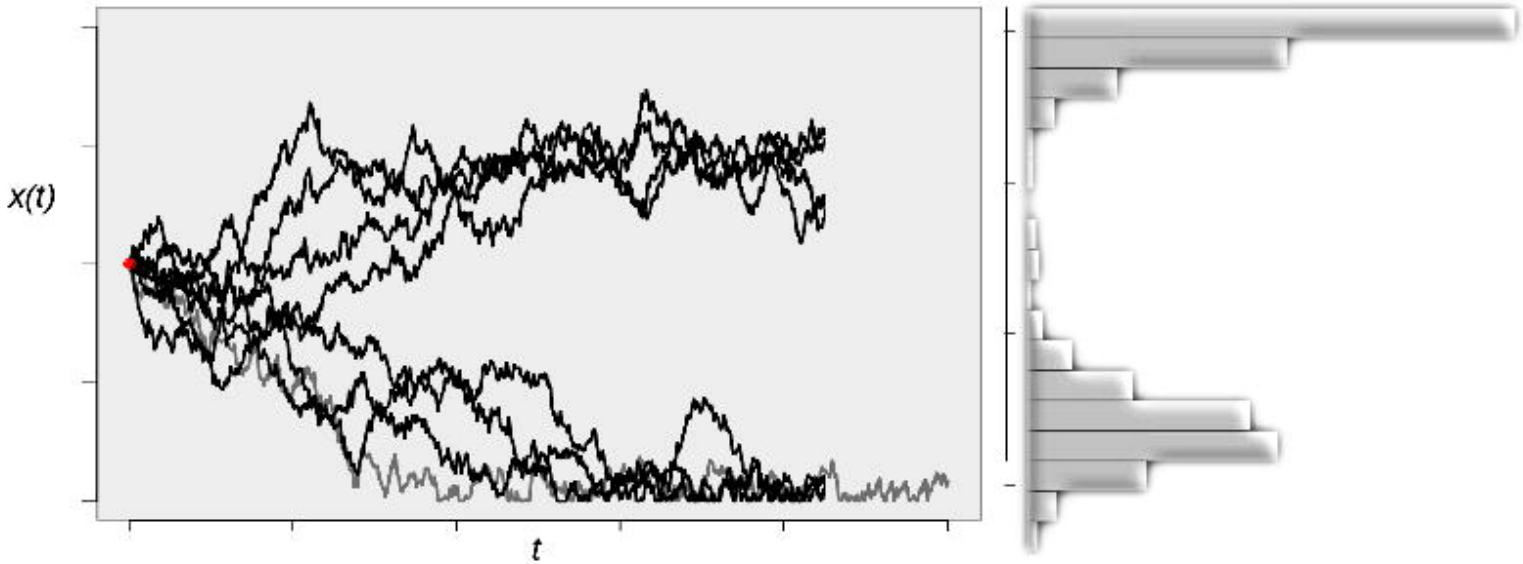
$$-\int F(x,t)dx$$




(a)
x_i=1 x_i=0
$x_1(t+1) = F_1(x_2)$
$x_2(t+1) = F_2(x_1)$
S_1
S_2
A_a
S_3
S_4
A_b
(b)
A_a A_b
A_a r_{aa} r_{ab}
A_b r_{ba} r_{bb}
A_a A_b A_a A_b
(c)
x_i P(0 -> 1) P(1 -> 1)
Ri>0
Regulatory input (Ri)
x_i P(0 -> 0) P(1 -> 0)
Ri<0
Regulatory input (Ri)
(d)
S_1 S_2 S_3 S_4
S_1 S_2 S_3 S_4
S_1 r_{11} ... r_{14}
S_2 ⋮ ⋱ ⋮
S_3
S_4 r_{41} ... r_{44}

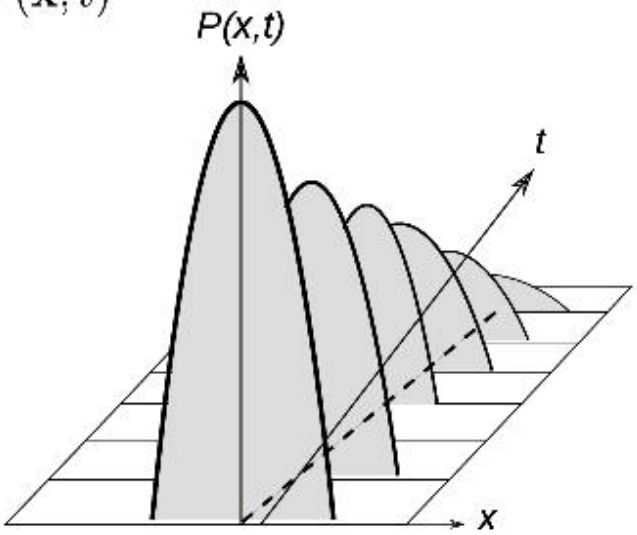
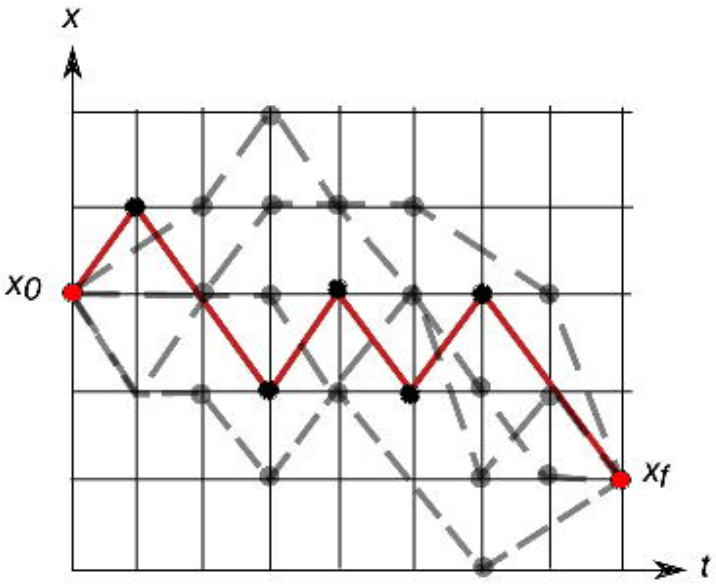
(a)

$$\frac{d}{dt}\mathbf{x} = \mathbf{F}(\mathbf{x}) + \mathbf{Noise}(\mathbf{x}, t)$$



(b)

$$\frac{d}{dt}P(\mathbf{x}, t)$$



(c)