 DOI 10.1002/bies.200800189

Reprogramming cell fates: reconciling rarity with robustness

Sui Huang[1,2]*

[1]Institute for Biocomplexity and Informatics, University of Calgary, Calgary, AB, Canada
[2]Vascular Biology Program, Children's Hospital, Harvard Medical School, Boston, MA, USA

The stunning possibility of ''reprogramming'' differentiated somatic cells to express a pluripotent stem cell phenotype (iPS, induced pluripotent stem cell) and the ''ground state'' character of pluripotency reveal fundamental features of cell fate regulation that lie beyond existing paradigms. The rarity of reprogramming events appears to contradict the robustness with which the unfathomably complex phenotype of stem cells can reliably be generated. This apparent paradox, however, is naturally explained by the rugged ''epigenetic landscape'' with valleys representing ''preprogrammed'' attractor states that emerge from the dynamical constraints of the gene regulatory network. This article provides a pedagogical primer to the fundamental principles of gene regulatory networks as integrated dynamic systems and reviews recent insights in gene expression noise and fate determination, thereby offering a formal framework that may help us to understand why cell fate reprogramming events are inherently rare and yet so robust.

Keywords: attractor; dynamics; gene regulatory networks; pluripotency; stem cell

Introduction–the subjectivity of surprise

The recent reprogramming of murine adult dermal fibroblasts into ''induced pluripotent stem cells'' (iPS) that resemble embryonic stem (ES) cells[1,2] by overexpressing just four key transcription factors has surprised many stem cell biologists.[2,3] A surprise in research can be defined as the unexpected departure of a finding from the outcome anticipated by a prevailing paradigm. So what is this paradigm? The rarely articulated but deeply rooted and broadly accepted thinking in developmental biology is that the lineage of a mature cell, once established, is essentially irreversible.[4] Accordingly, reprogramming the ES-cell phenotype in mature cells should not have worked–except by nuclear transfer into zygotes that will provide the complex physiological context for ''rebooting'' the genome by erasing covalent chromatin marks.[5] Hence, the occasionally observed plasticity of differentiated cells (beyond simple transdifferentiation[6] between related cell lineages[7–10]) is often dismissed as idiosyncratic or is explained by experimental artifacts, such as cell fusion or impurity of the starting cell culture.[4] Another, surprise was the recent demonstration that the pluripotent and self-renewing state of ES cells is a ''ground state'',[11] that is, a natural default state that needs not to be actively maintained. As we will see, these two unexpected findings are related.

Since a surprise shines light on the hidden cracks of existing thinking, let us, in lieu of an attempt at explaining the unexpected by stretching the current paradigm, question the paradigm itself. In fact, theoretical biologists who have studied cell type diversification within the framework of the first principles of regulatory dynamics and experimental biologists who have reasoned about cell plasticity[12–17] were hardly caught off-guard by the revelation of reprogrammability or the ''ground state'' of pluripotency. Much to the contrary, the new findings have long been anticipated[18] and they corroborate the theoretical concepts that will be explained here.

In the brief period since the first report by Yamanaka and coworkers,[1] reprogramming the pluripotent state in differentiated cells has been replicated with an exploding number of protocol variations, including starting from a variety of mature cell types or relying on transient instead of stable overexpression of transgenes,[2,19–23] thus revealing the inherent robustness of this process. Yet, the actual event of reprogramming at the level of individual cells is in general rare, occurring at a frequency of $<1\%$.[24,25] Thus, one challenge for a theory will be to explain how a complicated, intuitively unlikely and rare biological event is inherently robust, that is, given appropriate conditions, almost inevitable.

Therefore, we would like to present here a conceptual framework that has long evaded the attention of experimental biologists, but will offer a different perspective in which both reprogramming and the ''ground state'' character of pluripotency will appear natural and expected. More broadly speaking, perhaps we have always been asking the wrong question: instead of taking for granted that a cell lineage is cast in stone once committed, and then wondering how transdifferentiation or retro-differentiation can occur, let us

***Correspondence to:** S. Huang, Institute for Biocomplexity and Informatics, University of Calgary, Calgary, AB, Canada.
E-mail: Sui.huang@ucalgary.ca

 BioEssays 31:546–560,

reverse the question: given that all cells in the metazoan body possess the same set of genes [with the exception of post-meiotic germ line cells, mature lymphocytes, and cells in species that exhibit chromosome diminution[26]] and that the cell's phenotype is essentially determined by its genome-wide gene expression pattern, why then do we see discretely distinct, stable, almost immutable cell types in the first place? And how do the vastly complex cell type-specific gene expression profiles, encompassing tens of thousands of genes, so reliably establish themselves in response to comparably simple triggers, as if orchestrated by an invisible hand? Considering that chromatin-modification, once thought to permanently repress unused genes, is actually dynamic so that each gene can be reversibly switched on and off, we may then raise these key hypothetical questions: Why can cells not simply alter their expression profile to morph from one cell type to another? What is the very nature of the "barriers" that prevent such *ad libitum* inter-conversion between the cell type-specific expression profiles?

The goal of this article is to explain how "epigenetic barriers," which restrict *ad libitum* inter-conversion between cell type-specific expression profiles, channel development and maintain the "ground state" characteristic of pluripotency, can be understood as a naturally emerging property of a gene regulatory network (GRN). This will require the formal treatment of a network as an integrated dynamic system. We will attempt to achieve our goal without taking refuge in mathematical equations, but instead, by using formally rigorous, yet qualitative, and sometimes graphical explanations. Using permissively simplifying and pedagogically intuitive pictures, we hope to open the eyes of experimental biologists not familiar with dynamical systems to a set of principles that will benefit our thinking about the sources of stability and flexibility of cellular states, and in particular, help reconcile the rarity and robustness of reprogramming events.

Premise: from pathways to profiles

In the current paradigm of biological understanding, the default explanation of an observed cellular behavior, such as cell fate control, lies in molecular pathways in which individual molecular regulatory events are linked by an "arrow-to-arrow" scheme (Fig. 1A). In this epistemology, a molecular 'upstream–downstream' pathway implies a linear *chain of causation*. Thus, cell behavior is explained by a system of deterministic, causal relationships between symbols. This qualitative description is disconnected from the elementary principles of a dynamical system, which the cell's regulatory network of genes has to obey.

To find this connection, let us start with the obvious molecular description of a cell phenotype: our premise is that the pattern of expression of active proteins encoded by the genes throughout the genome, or shortly, the "*gene expression profile*", is the essential determinant of a cell's phenotype. Then, each cell fate, cell lineage, and cell type in a higher metazoan constitutes the "biological observable" that is uniquely associated with a gene expression profile of the roughly $N = 25{,}000$ genes of a mammalian genome. The gene expression profile is a *configuration*, i.e., one among many combinatorial possibilities of the set of N numbers, each indicating the expression status of the N genes which can be active (expressed), to varying degrees, or inactive (silenced). This configuration can be formalized as a ***state vector***,

$$\mathbf{S} = [x_1, x_2, \ldots, x_N],$$

where x_i is the level of expression of gene i (the level of active protein encoded by gene i) in the expression profile $\boldsymbol{S}$. Then, each cell type m, which can represent a stem cell, a progenitor cell of a particular lineage, a particular cell fate or one of the hundreds of terminally differentiated cell types, is uniquely associated with a particular state $\boldsymbol{S}^m$, that is, a combination of expression values, $[x_1{}^m, x_2{}^m, x_3{}^m, \ldots, x_N{}^m]$. Thus, for instance, $x_3{}^m$ has a high value if gene 3 is highly expressed in cell type m.

The advent of genome-wide transcript profiling techniques that allows for the efficient measurement of transcriptomes offers a glimpse of the state vector $\boldsymbol{S}$—although a series of post-transcriptional cellular processes have to be taken into account when mapping a transcriptome into a cell phenotype. Nevertheless, transcriptomes are good approximations of the state $\boldsymbol{S}$ of a cell, at least in the time scale of embryonic development since the presence of mRNA in the embryo tissue, as typically detected by in situ hybridization, correlates well with the activity of that gene. Thus, it is not surprising that cluster analyses of transcript profiles so reliably separate the various cell lineages.[27–29] However, there is more we can learn from gene expression profiling than the routine classification of tissues or the identification of individual tissue-specific genes, if we examine expression profiles through the lens of dynamical systems theory and treat them as an integrated entity.

To demonstrate an elementary meaning of an 'integrative' reading of gene expression patterns, let us, for didactic purposes, assume a toy-genome of four genes. Then, a particular state $\boldsymbol{S}^m$ in cell type $m = A$ may be written as follows: $\boldsymbol{S}^A = [1, 4, 0, 2]$, meaning that in this profile, *gene 1* is expressed at the level of 1 unit, $x_1{}^A = 1$, and accordingly, $x_2{}^A = 4$, $x_3{}^A = 0$, and $x_4{}^A = 2$. In any process in which one cell phenotype changes to another, as most prosaically epitomized by cellular differentiation, e.g., from cell type A (a progenitor cell) to cell type B (a differentiated cell), the associated gene expression profile will transition from a state $\boldsymbol{S}^A = [1, 4, 0, 2]$ to a state $\boldsymbol{S}^B$—for instance, $\boldsymbol{S}^B = [1, 0, 4, 4]$. Thinking in terms of such profiles as "balance sheets" of

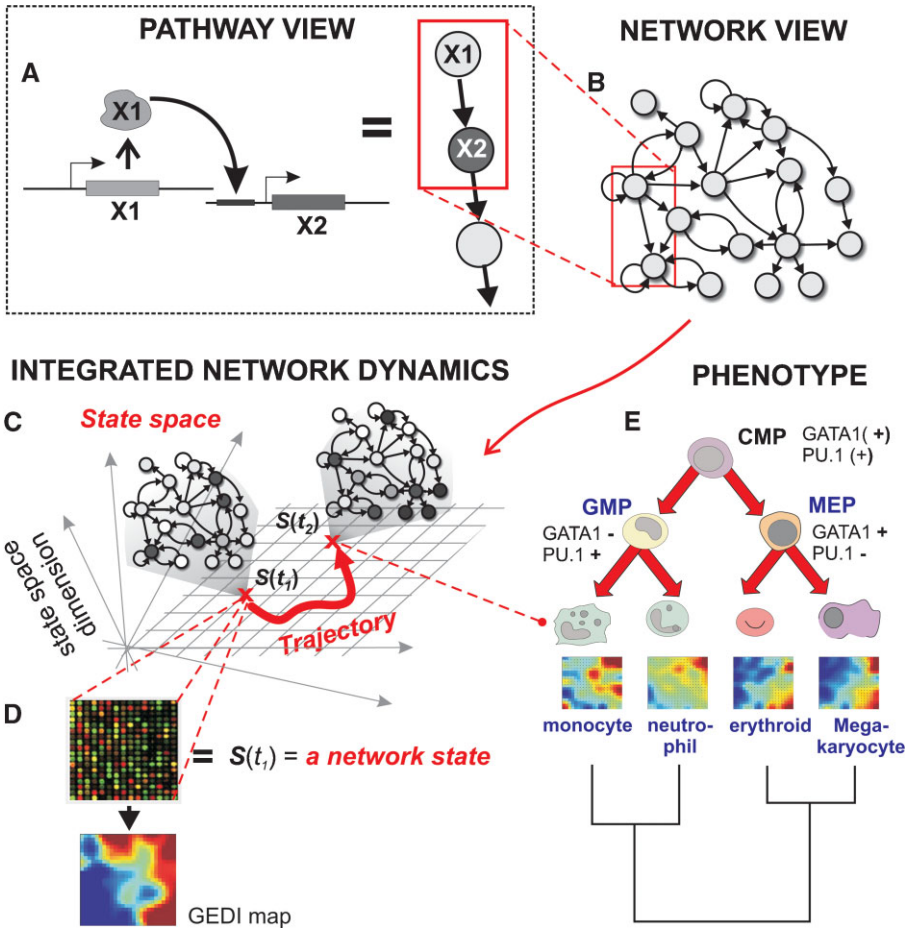

Figure 1. Overview of basic terms: Pathways, Networks, and Profiles. From traditional upstream–downstream genetic pathways (**A**) to the formalization of a regulatory network (**B**), the dynamics of which can be depicted as trajectory in a high-dimensional state space (**C**) where each point is a network state, manifest as gene expression profile (**D**) which maps into a cell state (**E**). Depicted underneath each cell type is the gene expression profile as GEDI map.[95] Note the similarity relationships, reflected in the hierarchical clustering dendrogram which distantly captures the tree-structure of the developmental paths, which corresponds to the trajectories (solid red arrows). Example shows the branching differentiation of the hematopoietic multipotent cell CMP = common myeloid progenitor. GMP = granulocyte-monocyte progenitor (myeloid lineage), MEP = megakaryocte-erythrocyte progenitor. PU.1 and GATA1 are the respective fate-determining transcription factors (see Fig. 2D); their expression in GMP and MEP reflect mutual exclusiveness. GEDI map[95] is a method for visual representation of a microarray that reorganizes the genes so as to create characteristic patterns for the human eye that reflect similarity between patterns. For details see ref.[95].

numbers without bothering with the molecular pathways involved in the transition already opens a new vista: the net amount of "work" that needs to be performed is conceivably related to how different these two states are from each other. This can be measured as a *distance* between the two profiles. Each gene has to be altered from its expression level in state $\boldsymbol{S^A}$ to meet that which helps define state $\boldsymbol{S^B}$. Thus, the "high-dimensional distance" between $\boldsymbol{S^A}$ and $\boldsymbol{S^B}$ is related to the sum of all the differences in expression levels for each of the corresponding genes in $\boldsymbol{S^A}$ and $\boldsymbol{S^B}$. Both the number of genes that are differentially expressed as well as the extent by which these genes have different levels in S^A and S^B will contribute to the distance. A simple distance measure, the squared Euclidian distance D_{AB}, sums up the square of these gene-level differences:

$$\begin{array}{rl} S_A = & [1,\ 4,\ 0,\ 2] \\ \underline{-S_B =} & \underline{[1,\ 0,\ 4,\ 4]} \\ & 0+4-4-2 \\ \rightarrow & 0+4^2+(-4)^2+(-2)^2=36=D_{AB} \end{array}$$

or in brief, $D_{AB}=(1-1)^2+(4-0)^2+(0-4)^2+(2-4)^2=36=D_{AB}$.

"Distances" between gene expression patterns, such as the (squared) Euclidian distance, are commonplace in expression profile analysis, as they form the first step used by cluster analysis to identify groups of similar expression

profiles. But do such distances between profiles have a biological meaning? The answer is yes, since hierarchical clustering based on such distance values roughly produces dendrograms that distantly replicate the familiar branching cell fate maps of development,[29,30] as exemplified for hematological development in Fig. 1E. The distance D_{AB} between cell types *A* and *B* may thus provide a first crude measure for how hard it is to "go from type *A* to *B*". Obviously, the further the distance the "harder" the transition. Transdifferentiation between neighboring branches in the developmental cell fate tree is indeed readily achieved by overexpression of one single fate-determining transcription factor.[7–9,31] Interestingly, a recent report suggests that reprogramming neural stem cells into pluripotent stem cells could be achieved with higher efficiency than starting with mature fibroblasts, and even by overexpression of only two instead of four transcription factors.[21]

We can summarize this first step of integration and formalization as follows: a cell-type-associated gene expression profile ***S*** is the molecular correlate of the loosely used metaphor of a (genetic) "program". The term "reprogramming" would imply the deliberate, targeted alteration of one profile ***S*** to another by some manipulation of individual genes. But something is missing. The above rather primitive quantitation of "distance" between lineages does not explain the inherent stability and the discrete nature of cell lineages: one could reprogram any state ***S*** to any other one if one were able to freely reconfigure the gene expression patterns. Thus, a biological version of *Maxwell's demon*, who would know every specific gene expression profile of each cell type in the body, could reprogram any cell type into any other one by scanning through the genomic DNA and switching on or off each gene until the desired pattern, ***S***, is achieved. What makes it hard, however, is the presence of "barriers" between the profiles that arise from dynamical constraints. But first, let us briefly discuss the conventional view's explanation of the stability of cell type-specific gene expression profiles ***S***.

The conventional molecular biology view and its limitations–"epigenetic marks"

In the prevailing epistemologic habit of reducing a cell phenotype to the molecular embodiments of its causation, the cell type-specific expression of a gene is explained by the action of the type-specific transcription factor (TFs) "upstream" of it.[32] This "proximate explanation"[33] is of course not satisfactory. Conversely, the absence of gene activity, for instance, lack of expression of liver-specific genes in erythrocytes, is now typically explained by the silencing of the unneeded genes through DNA methylation and covalent histone modifications. These covalent marks recruit protein complexes that further modify chromatin structure, and thereby control the access of the transcriptional machinery to the regulatory sites on the DNA.[34–38] Because of the perceived chemical stability of covalent modifications and the inertia of higher chromatin structure compared to non-covalent TF-DNA binding, and because of the maintenance of such modifications through DNA replication,[39] these mechanisms were readily adopted as the guardians of cell type-specific gene activation patterns. Such DNA and histone modifications were given the attribute "epigenetic"—an onomasiologically unfortunate choice[40]—to distinguish them from genetic, DNA sequence-based mechanism of inheritance[41] or from the type of somatic differentiation mediated by loss of genomic DNA in some species.[26]

However, there are two major conceptual flaws in the notion of a regulatory role for "epigenetic" DNA modification:

(a) Lack of stability. With the accumulating characterization of chromatin-modifying enzymes, notably those controlling histone lysine (de)methylation,[38,42–44] it has become clear that covalent "epigenetic" modifications are bidirectional (reversible) and highly dynamic.[42,43] Moreover, in assuring transmission of methylation patterns during DNA replication, the much-celebrated maintenance DNA methylase is actually quite error prone, so that DNA methylation in individual cells rapidly diversifies after a short clonal expansion.[45] So, if the marking of genes thought to provide a permanent molecular memory is actually dynamical and reversible—what maintains lineage-specific gene expression patterns in an inheritable fashion across cell divisions?
(b) Lack of locus-specificity. The enzymes whose catalytic activity is responsible for covalent "epigenetic" modification are not gene-locus specific. They can add or remove covalent modifications on virtually any gene in the genome (see below).

If DNA methylation and nucleosome modification *can* operate the switch of gene activity for any gene but cannot control *which* specific gene to actually switch on or off, then what system orchestrates the covalent modification machinery at tens of thousands of gene loci in the genome, such that the appropriate set of genes is (in)activated across the genome to generate the cell type-specific patterns of gene expression? Who writes the "histone-code"?

A systems view: the orchestrating role of the gene regulatory network

Obviously, a specific, biologically meaningful pattern of gene expression requires that the expression of an individual gene "considers" the expression status of other genes. Such coordination arises from the GRN (defined in Box 1) formed by

Gene network architecture: Basic concepts - and misconceptions

The architecture of a GRN is specified by the following attributes: (1) the gene-gene relationships and (2) their directionality (''who controls whom'' as denoted by the direction of the arrow connecting the genes), (3) the interaction modalities (inhibition vs. stimulation) as well as (4) the integrating ''transfer functions'' (mathematical functions defining the mode of cooperation between multiple upstream regulators of one gene and joint effect on the output expression kinetics, etc).[98] All these characteristics are determined by the physicality of specific protein-protein and protein-target DNA interactions, which is obviously encoded in the sequences of the proteins and the DNA response elements. These interaction determinants collectively establish what operationally can be defined as the GRN architecture. Currently, only attributes (1) and (2), often referred to as ''network topology'' have been systematically determined in simple organisms.[98,99] The GRN architecture contains the essential information for predicting the behavior or the *dynamics* of the network, which describes the coordinated change of expression levels of the individual genes of a network. The conceptual distinction between *architecture* and *dynamics* is, although rarely articulated, of fundamental significance in reaching beyond viewing the arrows in network schemata as simple symbols of causation (Fig. 1A, B).

Since the interaction specifications of the GRN architecture are determined by the structure of proteins and target DNA sequence, the GRN architecture is ''hard-wired'' in the genome: There is only one network per genome. In a commonly used abstraction, a genome of *N* genes can then be depicted as a network (mathematically, a ''graph'')[98] consisting of *N nodes* (=genes) connected by arrows (=regulatory interactions). [Attributes (1) (2) and (3) can be represented as a '*signed directed graph*'; and (3) and (4) can be captured by the same mathematical ''transfer'' function associated with each node] (Fig. 1B).

While the notion that each organism has *one* genomic network of defined architecture comes naturally to theorists, experimental biologists often use expressions, such as ''the regulatory network of the liver cell'' or ''the stem cell network,'' as if the network's wiring structure changes with each cell type or during development. What distinguishes cell lineages is not that they have different network architectures, but the gene expression profile $\boldsymbol{S}$, i.e., the activation status of each of the *N* nodes in the very same network, as can be approximately assessed by measuring the transcriptome (Fig. 1C, D). Thus, the genome-wide gene expression pattern $\boldsymbol{S}$ represents a *state of THE network* in condition (tissue) *m*, $\boldsymbol{S}^m$ or at time *t*, or mathematically, $\boldsymbol{S}(t)$. The fact that a hepatocyte does not express neuron-specific genes does not imply that it has a different network, but rather, that the very same network is in a different network state $\boldsymbol{S}$. The genes not expressed in the liver cells must somehow be suppressed in a sustainable way, directly or indirectly, by genes that establish the liver's gene expression profile – this itself is a manifestation of coordination by the entire genomic network.

the regulatory interactions exerted by genes, such as signaling and transcription factors, that can modulate the expression of other genes. The regulatory network imposes *constraints* on the collective behavior of genes, that is, on the changes of the expression profiles $\boldsymbol{S}$**.** It is obvious that as a consequence of gene regulation, individual genes cannot alter their expression independently. Thus, because of ''regulatory conflicts'' the vast majority of the combinatorially possible gene expression configurations $\boldsymbol{S}$ are ''logically forbidden,'' while few are favored. If, for instance, *gene 1* (unconditionally) inhibits *gene 2*, then all patterns $\boldsymbol{S}$ in which *gene 1* AND *gene 2* are both highly expressed will be ''forbidden.'' Such forbidden expression patterns are unstable. They embody the ''barriers'' that prevent free inter-conversion between stable gene expression profiles. We refer to these barriers between stable gene expression patterns as ''*epigenetic barriers*'' since they emerge from the interaction between the genes—a notion closer to Waddington's original meaning of ''epigenetic'' when he coined the term in 1940.[46]

In this network view, the concept of chromatin modification as the *prima causa* of lineage-specific gene expression patterns, operating ''upstream'' of the TFs by controlling their access to DNA target sites, must be revised. Indeed, evidence is accumulating for a rather reversed role: the TFs themselves, endowed with sequence-recognition capability, must be in charge of initiating the opening of chromatin at specific sites, probably followed by a bidirectional cooperation[37] with the chromatin-modifying enzymes that they recruit to their target loci.[44,47–52] Since the expression pattern of TFs is a consequence of the control of the GRN, the sites of chromatin opening and closing must, in principle, mirror the dynamics of the gene expression profile.[53] In fact, the role of ''epigenetic'' modifications as *prima causa* has recently been questioned.[40,54] While not the chief coordinator, chromatin modification could serve as an additional, important layer of stabilization of expression patterns established by the network of transcriptional regulation.

If the constraints on gene expression pattern change are dictated by the GRN, then obviously the epigenetic barriers must be encoded in the particular structure or ''architecture'' of the GRN (Box 1). But how do the specific constraints arise from the network interactions? The answer lies in what is called the *dynamics* of the network, which refers to the

temporal behavior of the state ***S***. Studying the dynamics of a networked system is a standard task in the field of 'systems dynamics' in mathematics and physics.

Network dynamics: the state space, trajectories, and attractors

To explain the network dynamics, let us further reduce our toy genome from $N=4$ to $N=2$ (Fig. 2A) and discuss a minimal gene regulatory circuit consisting of two genes, *X1* and *X2*, which inhibit each other's expression. Such an architecture of mutual regulation (inhibition or activation) is, in fact, widely found among transcription factors that have key roles in cell fate control in multipotent cells. In such a minimal system, the gene expression profile represents a circuit state $\boldsymbol{S}(t)$ at time t, defined by just two variables, the expression level value $x_1(t)$ for gene *X1* and $x_2(t)$ for *X2*:

$$S(t) = [x_1(t),\ x_2(t)].$$

A basic, convenient formalism for studying the dynamics of a network is the concept of a *state space*, in which every point corresponds to a circuit state $\boldsymbol{S}(t)$ at time t defined by the expression levels $x_1(t)$ or $x_2(t)$. For an $N=2$ gene circuit, the state space is a two dimensional plane, a coordinate system in which the variables x_1 and x_2 are represented by its two axes (Fig. 2B), so that the *position* of $\boldsymbol{S}(t)$ identifies the gene expression pattern $[x_1(t), x_2(t)]$. As explained in more detail in Box 2, the constraints dictate that a given state $\boldsymbol{S}(t)$ will move ("flow") in state space in a particular direction along a specified *trajectory* ("if x_1 changes by this much, then x_2 changes by that much"). As noted above, most states ***S*** in the state space are not stable due to the regulatory interactions which urge ***S*** to move until a stable state that satisfies the network interactions is found. Such stable equilibrium states are the ***attractor states*** and appear as "sinks" ($\boldsymbol{S^A}$ and $\boldsymbol{S^B}$) in the "flow field" of Fig. 2B (Box 2).

This particular 2-gene *X1-X2* mutual inhibition circuit generates two distinct attractor states: one ($\boldsymbol{S^A}$) has the expression pattern, $x_1 \gg x_2$, and the other ($\boldsymbol{S^B}$) has the complementary pattern, $x_1 \ll x_2$. This reversed configuration is intuitively plausible given their opposing relationship. The two attractors can coexist within the same environmental conditions, and thus, the circuit is said to be "*bistable*." Which state a given cell or circuit "occupies" in equilibrium depends solely on the history: the position of the state the cell started from, i.e., the *initial state* $\boldsymbol{S}(t=0)$. The set of initial states that

The dynamics of gene expression patterns ***S:*** attractor states

For a two-gene circuit at any given time t every circuit state $\boldsymbol{S}(t) = [x_1(t), x_2(t)]$ is a position in the *xy* state space, defined by the values, $x_1(t)$ and $x_2(t)$ which act as the coordinates of the position of ***S***. This conceptualization gives ***S*** position-like properties, allowing us to intuitively describe abstract behaviors of the circuit as the movement of $\boldsymbol{S}(t)$. The mathematical description with so-called *ordinary differential equations* (ODE) prescribes how the constraints imposed by the regulatory interactions translate into the trajectory of $\boldsymbol{S}(t)$ at each position in state space. Concretely, they compute the "arrows" (*vectors*) emanating from selected grid points shown in Fig. 2B, which indicate how a particular state ***S*** has to move ("flow") in the next time unit and collectively form the "*vector*" or "*flow field*." The vectors indicate rate and direction by which ***S*** at a certain state space position moves (Fig. 2B), describing the trajectory (for more details see also refs.[13,74,81]). Under normal circumstances a simple mathematical analysis of the equations readily identifies the points ***S*** that do not experience any driving force exerted by the regulatory interactions because they satisfy the circuit constraints. Accordingly, these points do not have outward-pointing arrows and are the *steady-states* (or *stationary states* or *fixed-points*) of the circuit, labeled here as $\boldsymbol{S^A}$, $\boldsymbol{S^B}$ and $\boldsymbol{S^C}$ in our example (Fig. 2B). All other (non-stationary) points in the state space experience a force that is represented by the vector. Looking at the flow field as a whole, a global picture of apparently coherent flow emerges, revealing two types of steady-states (Fig. 2B):

The flows *converge* in $\boldsymbol{S^A}$ and $\boldsymbol{S^B}$ which hence appear as "sinks." These two points, $\boldsymbol{S^A} = [x_1{}^A, x_2{}^A]$ and $\boldsymbol{S^B} = [x_1{}^B, x_2{}^B]$, are the "*attractors*" of the circuit. Upon perturbation of the state $\boldsymbol{S^A}$ (or $\boldsymbol{S^B}$), introduced by externally imposed changes of x_1 or x_2 and displacing it to any point $\boldsymbol{S}'$ in its close neighborhood, the circuit will return from $\boldsymbol{S}'$ back to the attractor state $\boldsymbol{S^A}$ (or $\boldsymbol{S^B}$) as long as $\boldsymbol{S}'$ remains in the *basin of attraction*. The line that separates the two attractors and their basins is the so called *separatrix* (boundary between basins).

Importantly, not every state that experiences no forces from the network interactions and hence, is a stationary (steady) state, is stable. The state $\boldsymbol{S^C}$ which sits on the *separatrix* is also a stationary state, but is not stable: the circuit stays at that position only when all regulatory influences are exactly balanced, so as not to exert any net force. In this symmetric example this is the case when $x_1 = x_2$. Any slight perturbation away from the condition $x_1 = x_2$ will trigger the circuit to move to either attractor, $\boldsymbol{S^A}$ or $\boldsymbol{S^B}$ (depending on the direction of the perturbation). Thus, $\boldsymbol{S}^C$ is called an *unstable ("repelling") steady state*.

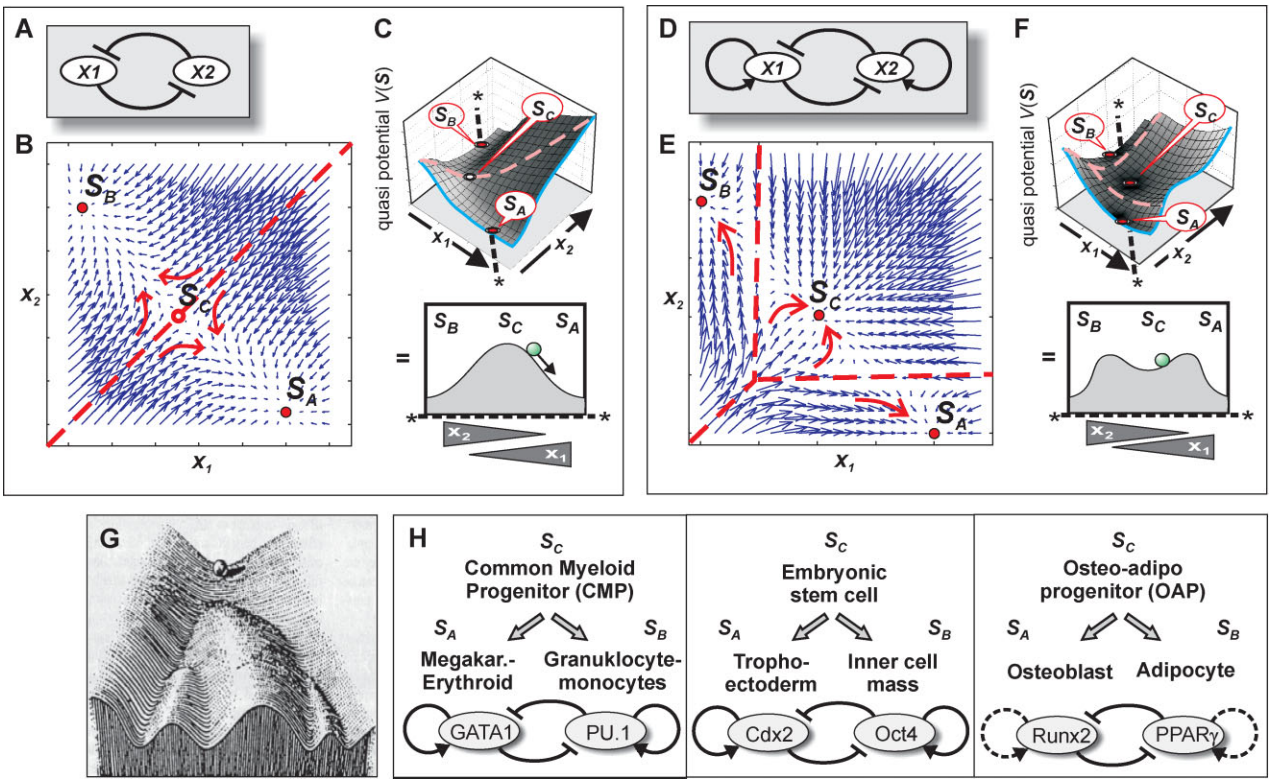

Figure 2. Dynamics of the 2-gene regulatory circuit. Qualitative explanation of basic principles of dynamical systems for the bistable (**A–C**) and tristable (**D–F**) circuits. **A**, **D**. Circuit architecture for two mutually inhibitory genes. **B**, **E**. State space (x_1–x_2)-plane with typical vector field (flow field). Each point in the plane is a gene expression configuration of the circuit (x_1, x_2). The arrows (vectors) indicate how the states $\boldsymbol{S} = (x_1, x_2)$, arbitrarily positioned on a grid to cover the state space, move within a tiny time unit. In ***B***, the states $\boldsymbol{S_A}$ and $\boldsymbol{S_B}$ (red dots) are stable steady states (attractors); $\boldsymbol{S_C}$ (empty circle) is an unstable steady state. In ***E***, the central state $\boldsymbol{S_C}$ is also an attractor. Dashed red line represents the separatrix, dividing the state space into the basins of attraction. **C**, **F**. Associated (quasi) potential landscape.[96] (For the computation of the landscape in F a circuit with different parameters than that underlying the vector field in *E*. was used to optimize 3D visualisation). Bottom: simplified schematic representations, obtained from cross section along the dashed line * ------ *. **G**. Waddington's "epigenetic landscape" (from ref.[70]) a qualitative metaphor that predates the formal quasi-potential landscape of gene networks. **H**. Examples of gene regulatory circuits with the same architecture which control binary decisions at branch points of cell differentiation in multi-potent cells, including CMP,[74] embryonic stem cells,[75] and OAP.[97] The dashed arrows indicate that the positive feedback loops are indirect. Note that these circuits are embedded in larger regulatory networks.

end up in the same attractor forms the *basin of attraction* of that attractor. Attractor states are robust, "self-stabilizing," distinct states. Once an attractor is reached, the associated expression pattern is maintained even after the original stimulus that placed it in the corresponding basin has disappeared. Thus, *bistability* is the most elementary mechanism for *memory* in nature. The basins are separated by regions of unstable states, which constitute the epigenetic barriers. (See Box 2 for more details.) Small perturbations of attractor states (imposed temporary changes in x_1 or x_2) are "buffered" by the basins of attraction, that is, the perturbed circuit state $\boldsymbol{S}'$ returns to the attractor state. In contrast, larger perturbations above a distinct threshold will "kick" a state out of the basin and into the other attractor. (Fig. 3B illustrates the idea of attractor transitions).

Cell types as high-dimensional attractors

Given the dynamical properties of attractors, it is now natural to equate each of the two stable attractor states $\boldsymbol{S^A}$ and $\boldsymbol{S^B}$ with an observable, stable gene expression pattern, and hence, with a cell fate, a lineage or a cell type. That cell types correspond to attractor states was historically a central idea, unfortunately forgotten as molecular biology turned its attention to explaining cell fate regulation in terms of molecular markers and linear pathways (Fig. 1A). It is important to note the fundamental ontological difference between arrows in the typical pathway charts that symbolize a molecular causation and the arrows in state space that represent a movement of $\boldsymbol{S}$ (a *trajectory* or *path*) driven by network dynamics (Fig. 1C).

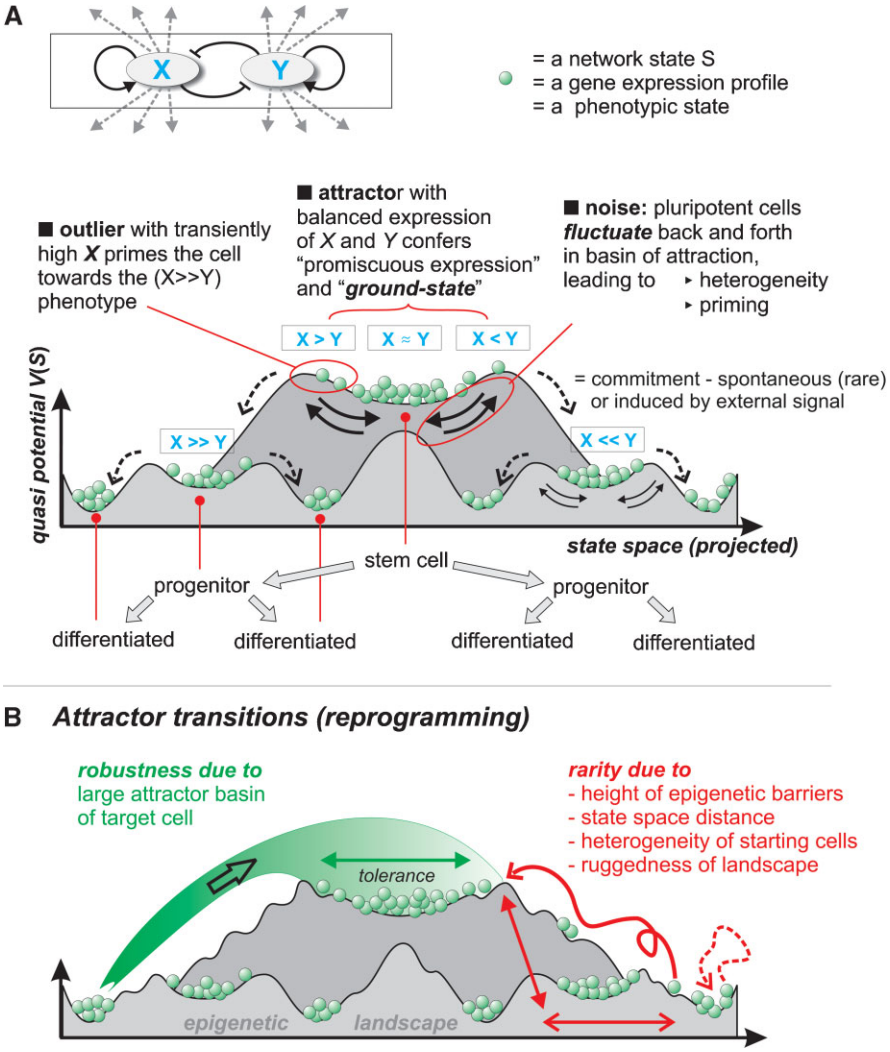

Figure 3. Schematic overview integrating the key concepts discussed in the text. **A**. A hypothetical epigenetic landscape to illustrate the concepts behind the hierarchy of cell type diversification during development. As in Fig. 2C, 2F, the horizontal axis is a schematic, projected state space-coordinate; the elevation (quasi-potential) represents the relative instability of individual states at each state space location. Each cell, defined by a position in state space is represented by a green ball. The landscape integrates two basic ideas: (1) metastable local attractors with the associated promiscuous gene expression pattern of the multipotent progenitor state, and (2) local fluctuations within attractors as the basis for cellular heterogeneity, which is manifest as a 'cloud' of states (balls). The "outlier cells" near the rim of the attractor basin are primed to differentiate. Attractor transitions to lower valleys correspond to fate commitment to more differentiated and lineage-restricted cells (dashed arrows). **B**. Illustration of the co-existence of rarity and robustness in reprogramming the pluripotent state ("jumping back"). Note the "subattractors" ("wash-board potentials") as manifestation of the ruggedness of the epigenetic landscape which imposes intermediate states that slow down the reprogramming events.

Max Delbrück first proposed in 1948 that bistability of biochemical networks may explain differentiation into discretely distinct cell states.[55] Soon after Monod and Jacob discovered gene regulation in the 1960s, they proposed that a circuit with the same architecture as in Fig. 2A[56] explains differentiation. Thomas suggested that at least a positive feedback loop (or, equivalently, a composite regulatory feedback loop with an even number of inhibitory interactions) is necessary for differentiation into multiple stationary states.[57] Bistability is a special (the simplest) case of *multistability*, an elementary concept in the theory of dynamical systems: In a network with a larger number of genes *N*, not two but multiple attractors can arise given the appropriate architecture of the gene regulatory network. In the late 1960s, Stuart Kauffman proposed that each one of the hundreds of cell types in the body corresponds to one of the

high-dimensional attractors that a large network of hundreds to thousands of genes can produce (if they are wired so as to avoid 'chaotic dynamics').[12,18]

This picture is closer to reality, since a network state is defined by an entire genome-wide profile $\boldsymbol{S} = [x_1, x_2, \ldots, x_N]$, where N can be in the thousands, so that the state space has N dimensions. After all, the small "local circuits" studied in various mathematical models[58–60] must be embedded in a genome-spanning gene network (Box 1). Nevertheless, the success of simple models in predicting cell fate behavior suggests that the core regulatory element controlling the decision between two lineages may consist of rather simple interlinked circuits, similar to the one discussed here[58–61] and that individual circuits may act quite independently in a particular cell type as the master control system that determines much of the dynamics of the genome-scale network. For instance, the regulatory core circuit that controls pluripotency in embryonic stem cells comprises the TFs Oct4, Sox2, Nanog, Klf4, etc.[62–65] and circuit models comprising few of these TFs qualitatively predict cell fate behavior well.[59,61,66]

The idea that a cell type-specific gene expression profile may be a *high-dimensional* attractor state, as first proposed by Kauffman,[18] has now found experimental support. Measurements of temporal evolution of transcriptomes during neutrophil differentiation and in a hematopoietic cell line have shown the convergence of high-dimensional trajectories (across >3000 genes)[67] or the return of noise-induced deviations of the transcriptome from the border of the basin of attraction back to the attractor state.[68] Both are dynamical hallmarks of attractor states.

In summary, attractors of regulatory networks exhibit the natural properties of cell types:[69] they are discretely distinct entities (a network can produce a "countable" number of attractors) and they are self-stabilizing, that is, *robust* to small perturbations; yet they allow "all-or-nothing" transitions to other attractors given sufficiently high perturbations. This mental picture of high-dimensional attractors explains the stunning reliability with which cells establish a particular, genome-wide gene expression profile of thousands of genes—without the help of our biological Maxwell's demon.

The ensemble of trajectories in the state space as depicted in Fig. 2B, E intuitively suggests a flow-like behavior. Indeed, the circuit dynamics can be further formalized as a kind of *quasi-potential landscape* that captures the global dynamics, as explained in Box 2. In this landscape picture the attractors behave like the lowest points in "potential wells" or valleys. They are separated by hills that correspond to unstable states. The latter then represent the epigenetic barriers. The state space and its landscape structure may be what Conrad Waddington, unaware of "gene networks" in the modern sense, had in mind when he proposed his "*epigenetic landscape*" in the 1940s to explain cell fate determination (Fig. 2G). Unfortunately, since the term "epigenetic" is used in molecular biology to describe covalent modifications—which does not do justice to Waddington's original ideas[70]—his landscape is continuously being (mis)interpreted in a loosely metaphoric manner.[71,72]

Reprogramming as transitions between attractors

It now appears that the simple bistable circuit consisting of TFs, *X1* and *X2*, which inhibit each other, is a general motif of network architecture that controls binary branch points between two mutually exclusive cell lineages produced by the a common multipotent progenitor cell. For instance, differentiation of the common myeloid progenitor (CMP) into two lineages, the megakaryocyte-erythroid progenitors (MEP, or "erythroid" lineage) and the granulocyte-monocyte progenitors (GMP, or "myeloid" lineage), is controlled by two mutually inhibiting fate-determining transcription factors: *X1* = GATA1 which promotes MEP commitment and *X2* = PU.1 which promotes GMP commitment.[73] Accordingly, the expression patterns in these two lineages are as predicted by the state attractors: $\boldsymbol{S}^{\mathbf{MEP}}$ $[x_1 \gg x_2] = [\text{GATA1}^{\text{high}}, \text{PU.1}^{\text{low}}]$ for MEP and $\boldsymbol{S}^{\mathbf{GMP}}$ $[x_1 \ll x_2] = [\text{GATA1}^{\text{low}}, \text{PU.1}^{\text{high}}]$ for GMP,[74] see Fig. 1E. Similarly, the dichotomy between neutrophils and monocytes is also controlled by two mutually inhibiting TFs, Egr2 and Gfi1.[60] Many other pairs of mutually regulating, fate determining TFs have been described, mostly in the hematopoietic system[73] but also in early embryonic development.[75]

The bistable system and the state space idea now provide a conceptual framework for explaining the intuitively anticipated transdifferentiation between neighboring lines by overexpression of one of the fate-determining factors.[76] We can now equate "reprogramming" with an attractor transition. Then, given the dynamical properties of an attractor as a valley in a landscape, we can derive several key properties for reprogramming:

(i) Some sloppiness in manipulating the gene expression levels of *X1* and *X2* is tolerated: As long as alteration of either *X1* or *X2*, or both, lands the circuit state $\boldsymbol{S}$ in the basin of the target attractor, the 'correct' expression profile will "self-organize" as $\boldsymbol{S}$ is attracted to the attractor.
(ii) The attractors and the memory property predict that once the transition is made, the external overexpression is no longer needed to maintain the new state.
(iii) Even in this simple model it can be formally shown that perturbation of a combination of genes, rather than a single gene, can synergistically enhance the transition efficiency. For instance, traveling from $\boldsymbol{S_A}$ to $\boldsymbol{S_B}$ along the diagonal of the state space (Fig. 2B, C) requires the

manipulation of both *X1* and *X2*. This trajectory corresponds roughly to the shortest path going through the lowest barrier (''notch''). [However, note that the landscape is not a *true* ''energy'' landscape – see bottom of Box 3]

Because of the above characteristics, reprogramming in general is more likely to be successful than we may expect in view of the highly specific configuration of gene expression that defines the target state. Later we will see why despite this robustness, a reprogramming event is still a rare event.

Multipotent state: an attractor with balanced expression patterns

If attractors provide a model for the stable gene expression profiles of committed cell fates, what about the uncommitted multipotent stem cells, their ''ground state'' nature, and what about reprogramming back to the ES-like state? It has long been proposed that the state of indeterminacy in pluri- or multipotent cells is characterized by a ''promiscuous gene expression'' of opposing fate-determining factors.[77,78] In the case of the afore-discussed GATA1 – PU.1 circuit, the CMP which faces the binary decision to commit to either MEP (triggered by high GATA1), or GMP (triggered by high PU.1), indeed expresses both GATA1 and PU.1 at equal but low levels[74] (Fig. 1E).

Thus, the multi-potent state is literally a balanced, undecided state: its expression pattern, $[x_1 \approx x_2]$ is poised at the center between the two attractors. But this would suggest that multipotency is an unstable steady state ($\boldsymbol{s}^{\boldsymbol{c}}$) or is elsewhere on the separatrix—a delicate balance on a mountain crest, prone to fall toward the attractors on either side followed by the slightest disturbance (Fig. 2B, C). While multipotent stem or progenitor cells indeed have a high tendency to spontaneously differentiate, they are to some extent stable entities that can be isolated and maintained in culture over extended periods of time. What stabilizes the undecided, multipotent progenitor or even the pluripotent stem cell?

From state space to the epigenetic landscape

The flow-like vector field (Fig. 2B, 2E) gives rise to the intuition of a physical flow toward lowest points (attractors). In fact, one can construct a ''*quasi potential energy landscape*'' in which the 3rd dimension above the state space plane is the elevation $V(\boldsymbol{s})$, which assigns each point $\boldsymbol{s}$ the elevation or ''potential'' *V*. Such a landscape would be the formal equivalent of Waddington's metaphoric ''epigenetic landscape (Fig. 2G) and would afford the intuition of a kind of gravity that would drive the stability-seeking movement of $\boldsymbol{s}$ ''downwards.'' Then, in this quasi-potential landscape, stable attractor states are represented by the lowest points in the ''potential wells'' or by ''valleys,'' which themselves are separated by hills. Accordingly, he unstable steady states are represented by the ''hill tops'' (in at least one direction).[74,96] The hills now provide an intuitive picture of the ''epigenetic barriers''.

It is, however, important to note that the elevation *V* does not constitute a true ''potential energy'' in the classical sense as proposed for systems like protein folding,[85] because the system equations for the regulatory networks (which underlie the vector fields in Fig. 2B and 2E) are not integrable. Nevertheless, unlike the traditional analysis of steady-states that examines their local stability (response to minimal perturbations), landscapes help conceptualize the global dynamics of non-equilibrium systems (such as gene regulatory networks) by providing information on the ''relative weights'' of the various valleys.[79]

Without going into the details[79,96] the relative ''depth'' of attractors and ''height'' of hills, that is, the elevation contours of the landscape, are computed as follows. The central notion needed is that gene expression is ''noisy,''[100] due, in part to the small numbers of molecules in the cell, so that random, short-time-scale local molecular events are not averaged out.[84] This results in random fluctuations of the expression level of a gene, $x_i(t)$ around a mean – the value measured biochemically in entire cell populations. The random fluctuations of x_i in time translate into the ''wiggling'' of the trajectory of $\boldsymbol{s}$ in state space, which will describe a random walk like the Brownian motion of a diffusing particle. Hence, $\boldsymbol{s}$ literally diffuses in the state space, and accordingly can be statistically described: There is a *probability* of finding $\boldsymbol{s}$ at a given position when a steady (equilibrium) state is reached. Thus, a *stable* attractor state is a ''*probable*'' state, analogous to a ''low energy state,'' whereas *unstable* states are less likely (''improbable'') to be occupied, and correspond to a ''high energy state.'' At the steady-state of the system we find, by numerical simulation, a *probability distribution* over the state space $P(\boldsymbol{s})$ that assigns to each $\boldsymbol{s}$ a probability *P*. To allow the intuitive equivalence ''low energy (stability) = high probability'' an inverse function of the probability $P(\boldsymbol{s})$ for each state $\boldsymbol{s}$ can then be used to plot an elevation (''quasi-potential'') $V(\boldsymbol{s})$ over each point $\boldsymbol{s}$ to generate a landscape. For formal reasons, *V* is calculated by taking the negative logarithm of the probability: $V(\boldsymbol{s}) = -\ln(P(\boldsymbol{s}))$. It is important to recall that the network is a so-called non-equilibrium system and the driving force of $\boldsymbol{s}$ is not simply related to the gradient (''down-hill direction'') in terms of the *V* landscape.

An answer may be offered by the finding that mutually inhibitory fate determining TFs, including GATA1 and PU.1, often also stimulate their own expression[74] (see other examples in Fig. 2H). Adding this self-activation (positive feedback) loop to the bistable circuit architecture (see Fig. 2A vs. 2D) changes the dynamics in a way that is difficult to comprehend intuitively. The analysis of the associated dynamical model,[74] however, shows that the presence of self-activation converts the central unstable steady state $\boldsymbol{s}^{\boldsymbol{C}}$ that displays ''promiscuous gene expression'' into a locally stable attractor state!

Thus, the state of indeterminacy with a balanced expression pattern $[x_1 \approx x_2]$ is itself an attractor, located between the two ''asymmetric attractors'' with $[x_1 \gg x_2]$ and $[x_1 \ll x_2]$ as shown schematically in Fig. 2E, F. This behavior manifests the *typical* dynamics of such circuits, which means that it is produced by a wide range of parameters in the underlying equations. Importantly, it explains why a balanced, pluri- or multi-potent state is to some extent self-stabilizing and why reprogramming pluripotency by rather crude over-expression of genes is relatively robust. As a local attractor in a potential landscape, it thus has the property of a ''ground state'' that is self-maintaining.[11] It is nevertheless globally situated at a ''high altitude,'' which affords the state a strong urge to ''differentiate away'' and populate all other attractors situated at a lower ''altitude.''[79]

Cell fate decision: leaving the ground state

What happens during fate decisions, for instance, when a growth factor, such as Epo triggers differentiation of the multipotent CMP toward the erythroid lineage (MEP)? A simple mechanistic model postulates that Epo tilts the GATA-PU.1 balance, since Epo signaling induces GATA1 expression and activation.[80] This constitutes a perturbation of the circuit, which kicks it out of the $\boldsymbol{s}^{\boldsymbol{C}}$ attractor into the basin of the $\boldsymbol{s}^{\boldsymbol{MEP}} = [\text{GATA1}^{\text{high}}, \text{PU.1}^{\text{low}}]$ attractor. This example may help define the role of signaling pathways in development: Signal transduction may have evolved to coordinate the ''combinatorially unlikely'' set of changes in gene expression in the state vector $\boldsymbol{s} = [x_1, x_2, \ldots]$ that is most efficient in overcoming an epigenetic barrier to accomplish a transition between high-dimensional attractors.

Reality, however, may again be more complex. In brief, there is evidence suggesting that differentiation signals might also cause a change in the interaction parameters of the circuit (strength of the interactions), such that the central progenitor attractor $\boldsymbol{s}^{\boldsymbol{C}}$ is destabilized, that is, converted from a local valley to a hilltop state as seen in the pure bistable system[74] (Fig. 2A–C). An event that qualitatively alters the landscape structure due to changes in network parameters is called a *bifurcation*.[74,81] Once the $\boldsymbol{s}^{\boldsymbol{C}}$ attractor is destabilized and the ''ground state'' character is gone (transforming the landscape from Fig. 2F to 2C), any small, temporal random asymmetry of the balanced configuration $[x_1 \approx x_2]$ caused by gene expression noise will randomly push the cell down to either one of the two attractors $\boldsymbol{s}^{\boldsymbol{A}}$ and $\boldsymbol{s}^{\boldsymbol{B}}$ representing the prospective cell fates. This is in line with the observation that when multipotent progenitor cells or embryonic stem cells are placed in culture conditions unfavorable to the stem cell state, they will spontaneously and stochastically differentiate into multiple lineages without the need of an ''instructive'' signal to convey the lineage-specific gene expression pattern.[82,83] Thus, let us briefly consider the stochastic aspect.

Stochasticity: pluripotency as cloud in state space

The dynamics discussed in the models so far is purely deterministic: a state $\boldsymbol{s}$ is a 'sharp', precise point in state space, a trajectory is a crisp line. But gene expression is ''noisy,''[84] i.e., the expression level fluctuates randomly in an individual cell (see Box 3). Thus, a snapshot of a clonal population of nominally identical cells that are supposed to be in the same cellular state will not map onto a single point in state space, but will appear as a *cloud of points*, each representing the state of an individual cell (Fig. 3).[68] Because of the fluctuations, an individual cell in an attractor will randomly wiggle around in the basin of attraction, bouncing from one edge of the basin to another but on average remaining around the attractor state.

This was demonstrated using a hematopoietic cell line that possesses the multi-potency of the CMP stage. When cells were isolated by FACS from the border of a cloud, i.e., from the ''tail'' of the population histogram with respect to the expression of PU.1 and GATA1, these noise-perturbed, ''outlier cells,'' which exhibit unbalanced PU.1, GATA1 levels will return, over several days, back to the center of the distribution with respect to these genes, hence, re-establishing the [GATA1 $\sim$ PU.1] balanced state. This relaxation of ''outliers'' back to the undecided CMP corroborates the idea that a multipotent/pluripotent state is an attractor or a ''ground state.'' In the case of ES cells, this would correspond to a balanced attractor maintained by the gene circuit around Oct4, Sox2, Nanog and other proteins.[59]

Importantly, the relaxation back to the ground state was unexpectedly slow (over a week)[68] and mathematical models of its kinetics suggested that the attractor is not a simple potential well surrounded by smoothly ascending slopes, but rather embedded in a *rugged* epigenetic landscape with multiple sub-attractors.[85] The presence of the latter leads to potential wells with a ''wash-board'' surface, as shown in Fig. 3B, so that individual outlier cells persist for

extended times near attractor borders. At the population level in equilibrium, gene expression noise maintains, via local noise-induced state transitions, an equilibrium distribution of occupancy of theses sub-states by individual cells. The population then appears as the familiar stable histograms of flow cytometry measurements for a protein *X*. These are population snapshots of all the present sub-states within an attractor projected to one state space axis, *X*. Rugged landscapes with sub-attractors are typical for complex high-dimensional dynamical systems.[18]

The cloud in the stem cell attractor offers a model for uniting the stochastic nature of cell fate commitment with the role of deterministic, instructive signals[86]: the cells that at a given time happen to be near the rim of the basin are most responsive to differentiating signals that kick them out of the stem cell attractor or destabilize the latter.[68]

Whether the well-documented noisy heterogeneity of expression levels x_i of master TFs, such as Oct4, Nanog, Cdx2 in individual ES cells[87,88] reflects a flat attractor basin that is particularly wide in pluri/multipotent cells, allowing for large fluctuations, as opposed to narrowes attractors of terminally differentiated cells, remains to be seen. It has been suggested that some signaling pathways may actually control noise, and hence, the stability of the pluripotency ground state, by regulating the dispersion of the cells in it.[89]

In summary, in the dynamics perspective, individual cells in a clonal population slowly bounce back and forth within the attractor basin and, since state space position translates into gene expression profile, they continuously ''scan'' a broad region of the state space for gene expression patterns that may prime them for a particular prospective fate while waiting for the commensurate external differentiation signal. Thus, multi- or pluripotency may merely be a manifestation of the fact that an apparently uniform clonal population of stem cells is actually a heterogeneous mixture of ''microstates,'' each primed for a distinct fate, but transitioning into each other in a dynamic equilibrium (within the attractor basin) when no fate-committing external cue is present. This concept establishes a formal basis for the metaphor of ''multi-lineage priming.''[77,78]

Reprogramming pluripotency: robust yet rare

Since the pluripotent state is an attractor state with a rather large basin of attraction, it is robust—a ground state. Inaccuracies in gene expression levels are tolerated. This property facilitates the reprogramming of a cell toward pluripotency by increasing the target size (summarized in Fig. 3). But we now also can explain why reprogramming events are rare despite the robustness of the target cell type. More generally, attempts to direct a stem cell to differentiate into any particular lineage or to reprogram a committed cell to another lineage are notoriously inefficient. The desired transition events are rare and slow.

We propose that the main culprit for this inefficiency is the ruggedness of the attractor landscape. Cells are dispersed and occupy multiple metastable sub-attractors within a ''wash-board potential'' attractor—the basis for heterogeneity of cell populations. The large number of such microstates within any cell population receiving the (re)programming perturbation also disperses the response profiles: only a small fraction of the cells in the population, namely those whose fluctuating microstate map into a gene expression pattern that fulfils some particular priming requirement, may actually be responsive to the nature of the reprogramming signals (Fig. 3B). Furthermore, the sub-attractor states in the ''potential wall'' represent distinct metastable intermediate states that need to be traversed when moving between attractors. Such intermediate stages, often manifest in non-observed state space dimensions of a higher-dimensional trajectory, have been characterized during DMSO-induced neutrophil differentiation[90] and reprogramming of fibroblasts to iPS cells.[91,92] In addition, the numerical computation of the landscape structure (Box 3) expectedly suggests a high elevation for the pluripotency attractor. This implies that re-establishing pluripotency is an ''uphill battle''[79] against the global slope in the landscape, which accounts for the arrow of time of development.

Conclusion

An epistemological habit in molecular biology is to explain a biological observable, such as pluripotency, by reducing it to the existence of a molecular entity, such as a gene, protein or pathway, as lucidly manifest in the long quest for THE ''stemness gene.''[93] Here, we have attempted to encourage a different type of reasoning, namely, one based on the fundamental, formalizable principles of a dynamical system as epitomized by a gene regulatory network. Then ''stemness'' is an emergent dynamical state rather than the direct consequence of the activity of a particular ''stemness'' gene. We also explained how, in such a formal conceptual framework, apparently surprising observations, including the possibility of reprogramming pluripotency and its ''ground state'' character, are naturally explained, without hand-waving or *ad hoc* metaphors. In doing so, we covered these key points:

1. Gene-gene interactions within the GRN constrain the dynamics of the network state as a whole. The constraints channel the movement of the network state in the state space, away from unstable expression patterns toward stable steady states or attractor states. This behavior

can formally be represented by a quasi-potential landscape which may have a fundamental correspondence to Waddington's epigenetic landscape.

2. In this landscape, an attractor state is represented by the bottom of a valley (''potential''-well) and the associated gene expression pattern is ''self-stabilizing.'' It resists (minor) perturbations in expression levels of the genes. A central idea is that attractors represent cell types. On the other hand, hilltops and crests are unstable states. They separate the attractors, and hence, represent the ''epigenetic barriers'' between the discrete cell phenotypes.
3. *Reprogramming* is not the *de novo* gene-by-gene instruction to build the expression pattern of a desired phenotype, but the activation of a pre-existing coherent gene expression ''program'' by stimulating a transition into the attractor that encodes such a program. The challenge is to find a walkable path that connects the attractors (see 6.).
4. The multi- or pluri-potent states are also attractors; but they are located on hills between the valleys that represent their prospective fates, the more differentiated cell types. Such ''high-altitude attractors'' are metastable, akin to a mountain lake, because, on the one hand, they are locally self-stabilizing, giving rise to the ''*ground state*'' property of stem cells, but on the other hand, upon perturbations the network state will ''flow down'' the valleys to the much lower attractors of differentiated cells. Located centrally in the state space, stem cell attractors naturally exhibit ''promiscuous gene expression'' and have access to the attractors of various lineages.
5. The heterogeneity of clonal populations of cells reflects the slow, random fluctuations of gene expression patterns in individual cells that appear to hop between subattractors in the rugged epigenetic landscape. These coherent fluctuations of entire expression patterns allow a stem cell to ''scan'' the state space and to temporally approach to the pattern of a prospective lineage. Such transient excursions may constitute the reversible ''priming'' of stem cells.
6. Since the stability of cell type-specific gene expression patterns emerges from the dynamics of reversible regulatory interactions, *no transition between any two such expression patterns is in principle impossible*. However, some paths between attractors are easier to walk and are used in normal development, whereas others are rarely or never used naturally. One should hence neither be surprised by the very possibility of reprogramming any cell type nor be disturbed by the discovery of new, unorthodox inter-cell developmental paths under experimental conditions.[94]
7. Finally, for pragmatic purposes, knowing the ''epigenetic landscape'' inferred from the GRN architecture as well as the spread of the ''clouds'' of cells within attractors may one day pave the way for predicting the epigenetic barriers and designing reprogramming strategies along the best paths. This possibility will likely focus our efforts in generating desired cell types on the transitions between related cell lineages (''neighboring valleys''), instead of taking the detour via the pluripotent iPS (''the summit'') as demonstrated by the recent generation of insulin-secreting pancreatic cells from closely related exocrine cells.[10] Moreover, the self-stabilizing and memory properties of attractors allows for the possibility of using transient perturbations,[23] perhaps with small molecule drugs, to achieve lasting reprogramming, thus obviating the need for viral vectors. Drugs that enlarge the attractor basins and thus promote the ground-state character of the target cell state[11] may be used to increase reprogramming efficiency.

Acknowledgments: The author gratefully acknowledges the support by the Air Force Office of Scientific Research (USAF/AFOSR), the National Institutes of Health/National Cancer Institute (NIH/NCI) and the Canadian Foundation for Innovation (CFI). The author wishes to thank Dr. S. Kauffman for inspiring discussions and reading of the manuscript.

References

1. **Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T. and Yamanaka, S.,** Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007. **131**: 861–872.
2. **Nishikawa, S., Goldstein, R. A. and Nierras, C. R.,** The promise of human induced pluripotent stem cells for research and therapy. *Nat Rev Mol Cell Biol* 2008. **9**: 725–729.
3. **Rodolfa, K. T. and Eggan, K.,** A transcriptional logic for nuclear reprogramming. *Cell* 2006. **126**: 652–655.
4. **Raff, M.,** Adult stem cell plasticity: fact or artifact? *Annu Rev Cell Dev Biol* 2003. **19**: 1–22.
5. **Hochedlinger, K. and Jaenisch, R.,** Nuclear reprogramming and pluripotency. *Nature* 2006. **441**: 1061–1067.
6. **Li, W. C., Yu, W. Y., Quinlan, J. M., Burke, Z. D. and Tosh, D.,** The molecular basis of transdifferentiation. *J Cell Mol Med* 2005. **9**: 569–582.
7. **Querfurth, E., Schuster, M., Kulessa, H., Crispino, J. D., Doderlein, G. and Graf, T.,** Antagonism between C/EBPbeta and FOG in eosinophil lineage commitment of multipotent hematopoietic progenitors. *Genes Dev* 2000. **14**: 2515–2525.
8. **Iwasaki, H., Mizuno, S., Wells, R. A., Cantor, A. B., Watanabe, S.,** et al. GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. *Immunity* 2003. **19**: 451–462.
9. **Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. and Graf, T.,** Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. *Immunity* 2006. **25**: 731–744.
10. **Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. and Melton, D. A.,** In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* 2008. **455**: 627–632.
11. **Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B. and Cohen, P.,** The ground state of embryonic stem cell self-renewal. *Nature* 2008. **453**: 519–523.
12. **Kauffman, S.,** Homeostasis and differentiation in random genetic control networks. *Nature* 1969. **224**: 177–178.
13. **Huang, S.,** Cell fates as attractors - stability and flexibility of cellular phenotype. Endothelial Biomedicine, 1st ed. New York, Cambridge University Press, 2007. pp 1761–1779.

14. **Furusawa, C. and Kaneko, K.,** Morphogenesis, plasticity and irreversibility. *Int J Dev Biol* 2006. **50**: 223–232.
15. **Hass, R.,** Retrodifferentiation and cell death. *Crit Rev Oncog* 1994. **5**: 359–371.
16. **Joshi, C. V. and Enver, T.,** Plasticity revisited. *Curr Opin Cell Biol* 2002. **14**: 749–755.
17. **Frisen, J.,** Stem cell plasticity? *Neuron* 2002. **35**: 415–418.
18. **Kauffman, S. A.,** The Origins of Order, New York, Oxford Univ. Press, 1993.
19. **Pera, M. F.,** Stem cells. A new year and a new era. *Nature* 2008. **451**: 135–136.
20. **Egli, D., Birkhoff, G. and Eggan, K.,** Mediators of reprogramming: transcription factors and transitions through mitosis. *Nat Rev Mol Cell Biol* 2008. **9**: 505–516.
21. **Kim, J. B., Zaehres, H., Wu, G., Gentile, L., Ko, K.,** et al. Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* 2008. **454**: 646–650.
22. **Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K. and Chiba, T.,** Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* 2008. **321**: 699–702.
23. **Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. and Hochedlinger, K.,** Induced pluripotent stem cells generated without viral integration. *Science* 2008. **322**: 945–949.
24. **Meissner, A., Wernig, M. and Jaenisch, R.,** Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol* 2007. **25**: 1177–1181.
25. **Mali, P., Ye, Z., Hommond, H. H., Yu, X., Lin, J. and Zou, J.,** Improved efficiency and pace of generating induced pluripotent stem cells from human adult and fetal fibroblasts. *Stem Cells* 2008. **26**: 1998–2005.
26. **Kloc, M. and Zagrodzinska, B.,** Chromatin elimination–an oddity or a common mechanism in differentiation and development? *Differentiation* 2001. **68**: 84–91.
27. **Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A.,** et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* 2004. **101**: 6062–6067.
28. **Zhang, W., Morris, Q. D., Chang, R., Shai, O., Bakowski, M. A.,** et al. The functional landscape of mouse gene expression. *J Biol* 2004. **3**: 21.
29. **Bruno, L., Hoffmann, R., McBlane, F., Brown, J., Gupta, R. and Pearson, S.,** Molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells in vitro. *Mol Cell Biol* 2004. **24**: 741–756.
30. **Akashi, K., He, X., Chen, J., Iwasaki, H., Niu, C. and Zhang, J.,** Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* 2003. **101**: 383–389.
31. **Laiosa, C. V., Stadtfeld, M. and Graf, T.,** Determinants of lymphoid-myeloid lineage diversification. *Annu Rev Immunol* 2006. **24**: 705–738.
32. **Maniatis, T., Goodbourn, S. and Fischer, J. A.,** Regulation of inducible and tissue-specific gene expression. *Science* 1987. **236**: 1237–1245.
33. **Tinbergen, N.,** Derived activities; their causation, biological significance, origin, and emancipation during evolution. *Q Rev Biol* 1952. **27**: 1–32.
34. **Khorasanizadeh, S.,** The nucleosome: from genomic organization to genomic regulation. *Cell* 2004. **116**: 259–272.
35. **Fisher, A. G.,** Cellular identity and lineage choice. *Nat Rev Immunol* 2002. **2**: 977–982.
36. **Arney, K. L. and Fisher, A. G.,** Epigenetic aspects of differentiation. *J Cell Sci* 2004. **117**: 4355–4363.
37. **Fuks, F.,** DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* 2005. **15**: 490–495.
38. **Kouzarides, T.,** Chromatin modifications and their function. *Cell* 2007. **128**: 693–705.
39. **Richards, E. J.,** Inherited epigenetic variation–revisiting soft inheritance. *Nat Rev Genet* 2006. **7**: 395–401.
40. **Ptashne, M.,** On the use of the word 'epigenetic'. *Curr Biol* 2007. **17**: R233–R236.
41. **Bird, A.,** Perceptions of epigenetics. *Nature* 2007. **447**: 396–398.
42. **Kubicek, S. and Jenuwein, T.,** A crack in histone lysine methylation. *Cell* 2004. **119**: 903–906.
43. **Trojer, P. and Reinberg, D.,** Histone lysine demethylases and their impact on epigenetics. *Cell* 2006. **125**: 213–217.
44. **Mellor, J.,** Dynamic nucleosomes and gene transcription. *Trends Genet* 2006. **22**: 320–329.
45. **Silva, A. J., Ward, K. and White, R.,** Mosaic methylation in clonal tissue. *Dev Biol* 1993. **156**: 391–398.
46. **Waddington, C. H.,** The epigenotype. *Endeavour* 1942. **1**: 18–20.
47. **Grass, J. A., Boyer, M. E., Pal, S., Wu, J., Weiss, M. J.,** et al. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A* 2003. **100**: 8811–8816.
48. **de la Serna, I. L., Ohkawa, Y., Berkes, C. A., Bergstrom, D. A., Dacwag, C. S. and Imbalzano, A. N.,** MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. *Mol Cell Biol* 2005. **25**: 3997–4009.
49. **Tapscott, S. J.,** The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* 2005. **132**: 2685–2695.
50. **Krysinska, H., Hoogenkamp, M., Ingram, R., Wilson, N., Tagoh, H. and Singh, H.,** A two-step, PU.1-dependent mechanism for developmentally regulated chromatin remodeling and transcription of the c-fms gene. *Mol Cell Biol* 2007. **27**: 878–887.
51. **Metzger, E., Wissmann, M. and Schule, R.,** Histone demethylation and androgen-dependent transcription. *Curr Opin Genet Dev* 2006. **16**: 513–517.
52. **Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., Hara, E. and Hozumi, N.,** Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. *Oncogene* 2006. **25**: 2477–2488.
53. **Bonifer, C.,** Epigenetic plasticity of hematopoietic cells. *Cell Cycle* 2005. **4**: 211–214.
54. **Madhani, H. D., Francis, N. J., Kingston, R. E., Kornberg, R. D., Moazed, D. and Panning, B.,** Epigenomics: a roadmap, but to where? *Science* 2008. **322**: 43–44.
55. **Delbrück, M.** Discussion. Unités biologiques douées de continuité génétique Colloques Internationaux du Centre National de la Recherche Scientifique: CNRS, Paris; 1949.
56. **Monod, J. and Jacob, F.,** Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* 1961. **26**: 389–401.
57. **Thomas, R.,** Logical analysis of systems comprising feedback loops. *J Theor Biol* 1978. **73**: 631–656.
58. **Roeder, I. and Glauche, I.,** Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors GATA-1 and PU.1. *J Theor Biol* 2006. **241**: 852–865.
59. **Chickarmane, V., Troein, C., Nuber, U. A., Sauro, H. M. and Peterson, C.,** Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol* 2006. **2**: e123.
60. **Laslo, P., Spooner, C. J., Warmflash, A., Lancki, D. W., Lee, H. J. and Gantner, B. N.,** Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* 2006. **126**: 755–766.
61. **MacArthur, B. D., Please, C. P. and Oreffo, R. O.,** Stochasticity and the molecular mechanisms of induced pluripotency. *PLoS ONE* 2008. **3**: e3086.
62. **Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S.,** et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005. **122**: 947–956.
63. **Zhou, Q., Chipperfield, H., Melton, D. A. and Wong, W. H.,** A gene regulatory network in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 2007. **104**: 16438–16443.
64. **Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W.,** et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 2006. **38**: 431–440.
65. **Mathur, D., Danford, T. W., Boyer, L. A., Young, R. A., Gifford, D. K.,** et al. Analysis of the mouse embryonic stem cell regulatory networks obtained by ChIP-chip and ChIP-PET. *Genome Biol* 2008. **9**: R126.
66. **Chickarmane, V. and Peterson, C.,** A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PLoS ONE* 2008. **3**: e3478.

67. **Huang, S., Eichler, G., Bar-Yam, Y. and Ingber, D. E.,** Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* 2005. **94**: 128701.
68. **Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. and Huang, S.,** Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 2008. **453**: 544–547.
69. **Huang, S. and Ingber, D. E.,** Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp Cell Res* 2000. **261**: 91–103.
70. **Waddington, C. H.,** The strategy of the genes, London, Allen and Unwin, 1957.
71. **Reik, W. and Dean, W.,** Back to the beginning. *Nature* 2002. **420**: 127.
72. **Goldberg, A. D., Allis, C. D. and Bernstein, E.,** Epigenetics: a landscape takes shape. *Cell* 2007. **128**: 635–638.
73. **Orkin, S. H. and Zon, L. I.,** Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 2008. **132**: 631–644.
74. **Huang, S., Guo, Y. P., May, G. and Enver, T.,** Bifurcation dynamics of cell fate decision in bipotent progenitor cells. *Dev Biol* 2007. **305**: 695–713.
75. **Ralston, A. and Rossant, J.,** Genetic regulation of stem cell origins in the mouse embryo. *Clin Genet* 2005. **68**: 106–112.
76. **Graf, T.,** Differentiation plasticity of hematopoietic cells. *Blood* 2002. **99**: 3089–3101.
77. **Enver, T. and Greaves, M.,** Loops, lineage, and leukemia. *Cell* 1998. **94**: 9–12.
78. **Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M. and Enver, T.,** Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* 1997. **11**: 774–785.
79. **Wang, J., Xu, L., Wang, E. K. and Huang, S.,** The ''potential'' landscape of genetic circuits imposes the arrow of time in stem cell differentiation. Manuscript in review.
80. **Palani, S. and Sarkar, C. A.,** Positive receptor feedback during lineage commitment can generate ultrasensitivity to ligand and confer robustness to a bistable switch. *Biophys J* 2008. **95**: 1575–1589.
81. **Kaplan, D. and Glass, L.,** Understanding Nonlinear Dynamics, New York, Springer, 1995.
82. **Fairbairn, L. J., Cowling, G. J., Reipert, B. M. and Dexter, T. M.,** Suppression of apoptosis allows differentiation and development of a multipotent hemopoietic cell line in the absence of added growth factors. *Cell* 1993. **74**: 823–832.
83. **Keller, G. M.,** In vitro differentiation of embryonic stem cells. *Curr Opin Cell Biol* 1995. **7**: 862–869.
84. **Kaern, M., Elston, T. C., Blake, W. J. and Collins, J. J.,** Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 2005. **6**: 451–464.
85. **Frauenfelder, H., Sligar, S. G. and Wolynes, P. G.,** The energy landscapes and motions of proteins. *Science* 1991. **254**: 1598–1603.
86. **Enver, T., Heyworth, C. M. and Dexter, T. M.,** Do stem cells play dice? *Blood* 1998. **92**: 348–352.
87. **Dietrich, J. E. and Hiiragi, T.,** Stochastic patterning in the mouse preimplantation embryo. *Development* 2007. **134**: 4219–4431.
88. **Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B. and Vrana, J.,** Nanog safeguards pluripotency and mediates germline development. *Nature* 2007. **450**: 1230–1234.
89. **Hayward, P., Kalmar, T. and Arias, A. M.,** Wnt/Notch signalling and information processing during development. *Development* 2008. **135**: 411–424.
90. **Chang, H. H., Oh, P. Y., Ingber, D. E. and Huang, S.,** Multistable and multistep dynamics in neutrophil differentiation. *BMC Cell Biol* 2006. **7**: 11
91. **Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M. and Jaenisch, R.,** Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* 2008. **2**: 151–159.
92. **Stadtfeld, M., Maherali, N., Breault, D. T. and Hochedlinger, K.,** Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* 2008. **2**: 230–240.
93. **Vogel, G.,** Stem cells. 'Stemness' genes still elusive. *Science* 2003. **302**: 371.
94. **Murre, C.,** Defining the pathways of early adult hematopoiesis. *Cell Stem Cell* 2007. **1**: 357–358.
95. **Eichler, G. S., Huang, S. and Ingber, D. E.,** Gene Expression Dynamics Inspector: for integrative analysis of expression profiles. *Bioinformatics* 2003. **19**: 2321–2322.
96. **Kim, K. Y. and Wang, J.,** Potential energy landscape and robustness of a gene regulatory network: toggle switch. *PLoS Comput Biol* 2007. **3**: e60.
97. **Lin, Y. F.**, **Jing, W.**, **Wu, L.**, **Li, X. Y.**, **Wu, Y.** and **Tang, W.**, Identification of osteo-adipo progenitor cells in fat tissue. *Cell Prolif Cell* 2008. **41**: 803–812.
98. **de Jong, H.,** Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 2002. **9**: 67–103.
99. **Huang, S. and Kauffman, S.,** Complex gene regulatory networks - from structure to biological observables: cell fate determination. *Encyclopedia of Complexity and Systems Science (Springer)* 2009. in press.
100. **Nicolis, G.,** Dissipative systems. *Rep Prog Phys* 1986. **49**: 873–949.