The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology?

Sui Huang

The Neo-Darwinian concept of natural selection is plausible when one assumes a straightforward causation of phenotype by genotype. However, such simple 1:1 mapping must now give place to the modern concepts of gene regulatory networks and gene expression noise. Both can, in the absence of genetic mutations, jointly generate a diversity of inheritable randomly occupied phenotypic states that could also serve as a substrate for natural selection. This form of epigenetic dynamics challenges Neo-Darwinism. It needs to incorporate the non-linear, stochastic dynamics of gene networks. A first step is to consider the mathematical correspondence between gene regulatory networks and Waddington's metaphoric 'epigenetic landscape', which actually represents the quasi-potential function of global network dynamics. It explains the coexistence of multiple stable phenotypes within one genotype. The landscape's topography with its attractors is shaped by evolution through mutational re-wiring of regulatory interactions – offering a link between genetic mutation and sudden, broad evolutionary changes.

Keywords:
attractor; epigenetics; gene regulatory network; Neo-Darwinism; systems biology

DOI 10.1002/bies.201100031

Institute for Systems Biology, Seattle, WA, USA

Corresponding author:
Sui Huang
E-mail: sui.huang@systemsbiology.org

Abbreviation:
GRN, gene regulatory network.

Supporting information online

Introduction

In the 1940s and 1950s the Modern Synthesis in biology reconciled Darwin's theory of natural selection of random, gradual variants with the Mendelian concept of discrete genes as the vehicles of inheritance [1]. This synthesis, which we owe to quantitative population genetics [2–4], gave Darwin's theories a mechanistic basis and led to what has become known, loosely, as 'Neo-Darwinism' (Fig. 1). A decade later, the central dogma of molecular biology [5], now remembered by its compact scheme 'gene → mRNA → protein (=trait)', established the molecular basis of genetic inheritance. This linear scheme of causation (one gene - one trait) further cemented the Neo-Darwinian principle that implicitly assumes a straight and deterministic genotype-to-phenotype mapping. Such a linear causation scheme is necessary for the enrichment of a gene by selection of a phenotype to be plausible.

But what if a genotype does not translate in an obvious manner into a corresponding phenotype – or worse, if genes are not even the sole basis of inheritance? This is of course not a new proposal. It was first articulated by C. Waddington when he coined the term 'epigenetics', at roughly the same time as the consolidation of the Modern Synthesis. However, the idea of non-genetic contributions to determination and inheritance of traits has been sidelined as Neo-Darwinism and the central dogma has risen to biological orthodoxy. But the arrival of genome-wide analyses of genes and gene expression and the entry of systems dynamics thinking into molecular biology have now exposed, with inescapable clarity, the complexity in the relationship between genotype and phenotype. It has also brought the role of higher level organization that lies beyond direct genetic instruction into sight. Specifically, experimental characterization of physical interactions between genes and the emergence of the concept of gene regulatory networks (GRNs) [6–8], through which genes act not as soloists but in concert, have begun to shed light on the population geneticists' black box where gene-gene interactions were acknowledged, at best, in the form of the phenomenological concept of epistasis [9].

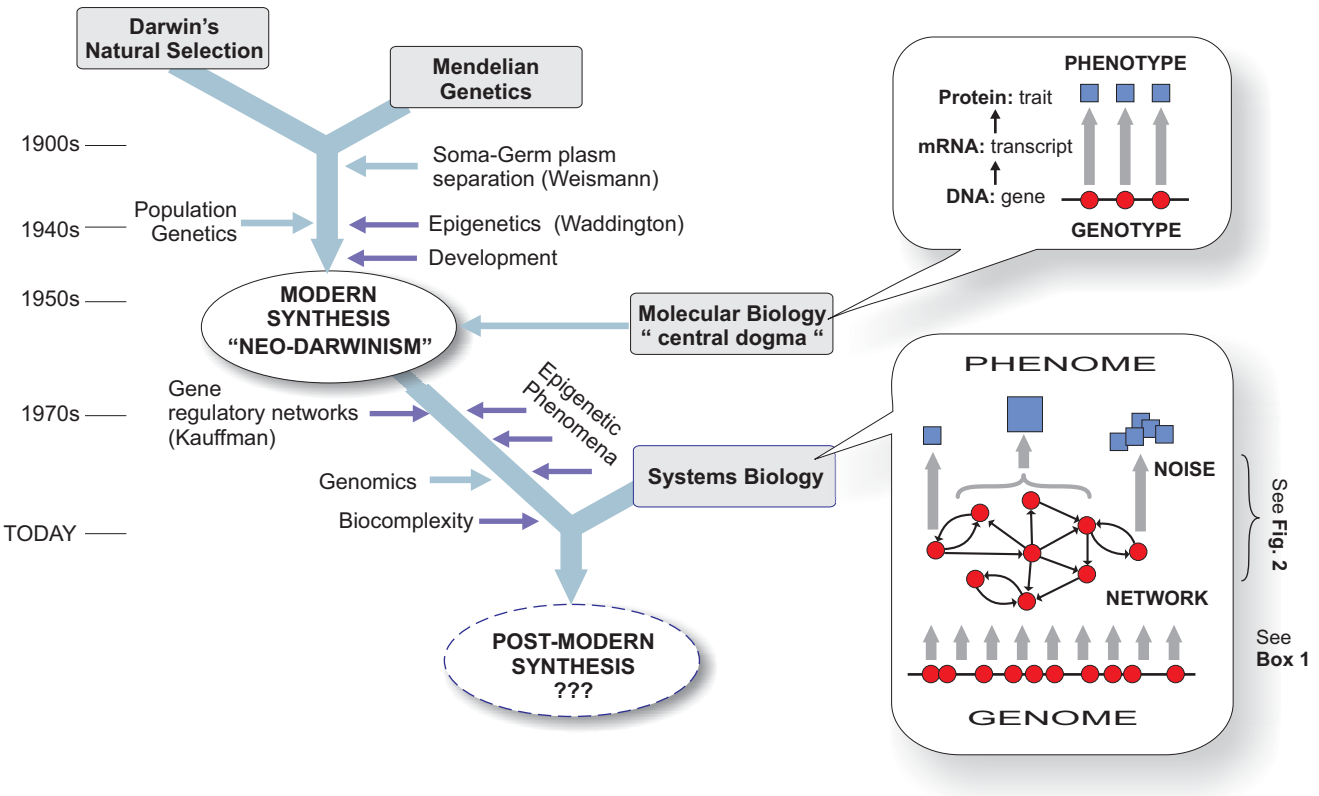


Figure 1. Historical overview of the 'Modern Synthesis' of evolutionary biology, Neo-Darwinism and the possibility of a similar future synthesis of the latter with Systems Biology in post-Darwinian biology ('Post-Modern Synthesis'). Small arrows on the side indicate theories, concepts or observations that influenced the development of Darwinism. The proposed new synthesis would have to explain findings and incorporate theories (blue arrows) that are not considered in the central dogma of molecular biology and appear to defy Neo-Darwinism.

Both central pillars in the edifice of Neo-Darwinism, natural selection and the linear mono-causal genotype-phenotype mapping, more recently referred to as 'genetic determinism' [10], have received ample criticism ever since they were erected, long before the rise of the modern-day formal notions of complexity [11–16]. Among the diverse alternative or complementary ideas, the old notion of 'constraints' epitomizes a departure from Neo-Darwinian thinking that is relevant here.

Waddington was one of the first to point to 'developmental constraints' in the physical implementation of the instructions of the genes in forming an organism. He captured the idea of constraints in the now famous metaphor of 'epigenetic landscape' [10–12], to which I turn my prime attention later. But it is perhaps owing to the work of Gould and Lewontin [17] as well as Goodwin [18, 19] that many contemporary biologists are now aware of the fallacy of pure 'selectionism', according to which natural selection is the sole, almighty sculptor of all phenotypic traits. Natural constraints in organismal design, emanating from the inescapable laws of chemistry, physics and even mathematics, as well as from history [20], present prefabricated modules of high complexity for natural selection to choose from. But the complexity itself is not the work of natural selection. The fractal, optimally space-filling structure of branching tissues, such as the lung, or the appealing stripe and spot patterns of animal coats are the most lucid examples of the creative force of self-organization that can be reduced to the laws of physics [21] and whose genesis does not require natural selection. More likely, it is the incorporation and polishing, but not the initial design, of such modules in 'building' complex organisms that benefited from selection. Constraints through self-organization offer Darwinian selection 'order for free', as Kauffman, the pioneer of biocomplexity, aptly put it [22]. I show that constraints play a central role in how GRNs produce, almost for free, the stable gene expression patterns needed to govern coherent cellular behaviours.

Towards a theory of epigenetics based on GRNs

A recurring theme in the voices cautioning against unfettered Neo-Darwinism and genetic determinism, while calling for more 'holistic' approaches that consider complexity theory, is a set of observations that I summarize here under '*epigenetic phenomena*'. Since they pertain to all that is beyond ('epi') the 'genetic' mechanisms of phenotype specification and inheritance, such phenomena most prosaically defy the paradigm of genetic determination. (The confusing semantics of 'epigenetic' is discussed in the online Supporting Information, Text I) But what 'epigenetic' apparatus can universally provide information encoding and storage without altering the DNA sequence? The answer is: the GRN ([6], or briefly, 'gene network') – the network established by the fact that genes influences the expression of other genes via a web of molecular

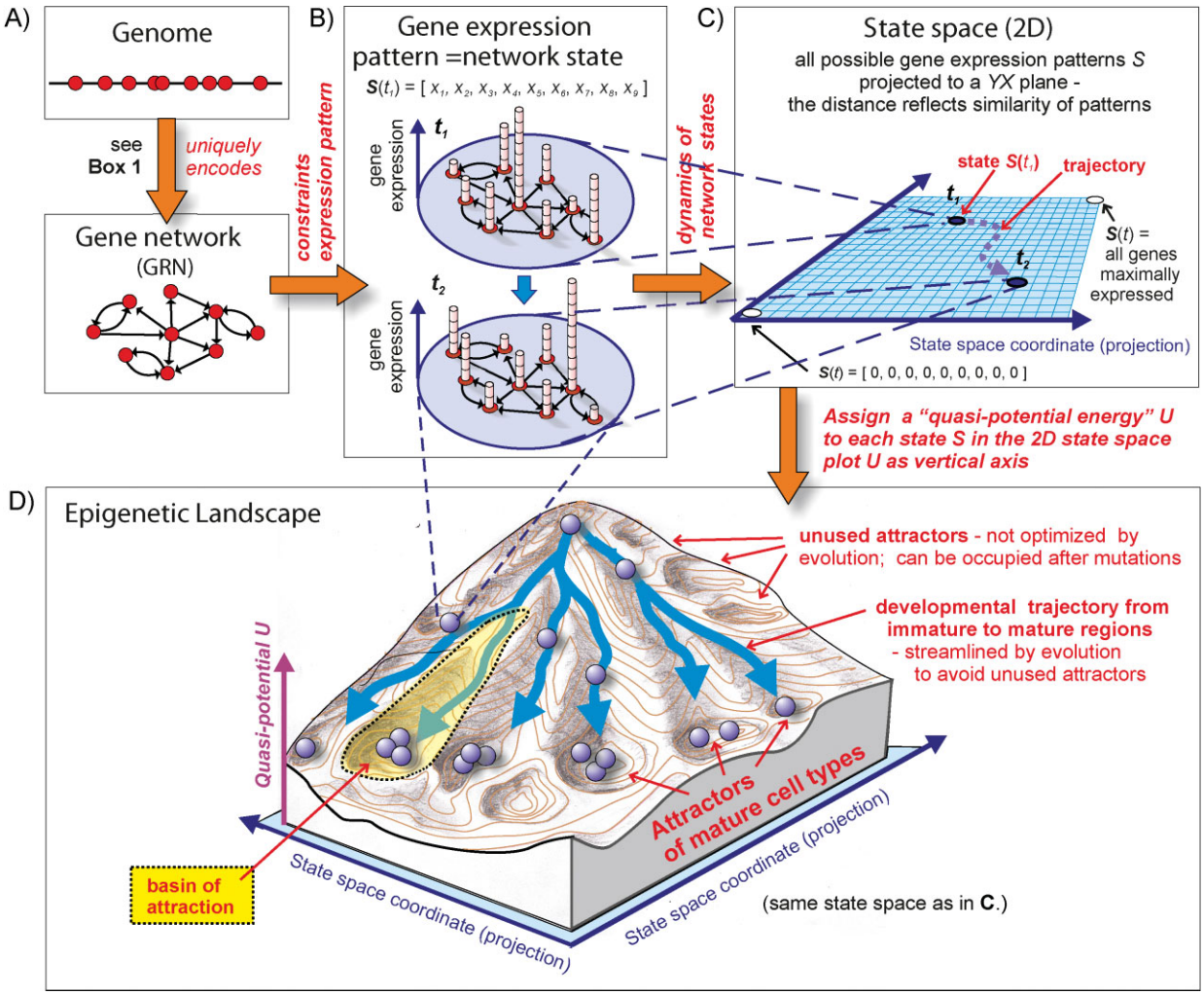


Figure 2. The unequivocal correspondence (unique mapping) between genome (**A**) and associated network architecture and the epigenetic landscape (**D**) via the dynamics of the expression patterns (**B**) in state space (**C**) controlled by the GRN. The schematic representation is for a 9-gene GRN. The central concept to understanding the landscape is that each network state *S* (gene expression pattern, hence cell state = blue discs in **B** and **C**) maps into a point (=blue balls in **D**) on the landscape. The position of the point (network state) *S*, is determined as follows: The *N* gene expression values defining a given state *S* act as the coordinates in defining its position in that *N*-dimensional space, where each dimension (axis) represents the expression level of a gene. Each step (orange arrow) in entering a new (more abstract) conceptual domain (boxes A, B, C, D) can be formalized in terms of mathematical principles [8, 45]. For **A**, see also Box 1. In **B** the two time points t_1 and t_2 represent the dynamics and the constrained change of gene expression pattern. Note that the quasi-potential is not a true potential energy since the gene network dynamics is a non-equilibrium, typically non-integrable system. The value of *U* can be intuitively (but formally not correctly) approximated by the negative logarithm of the steady-state probability *P*(*S*) to find the network in state *S*, i.e. $U \sim -\ln[P(S)]$, or by decomposing the vector field that contains the forces *F*(*S*) that drive *S* into two perpendicular components, one of which is a gradient of some quasi-potential function *U* [53]. Red circles = genes; blue axes = state space coordinates after hypothetical dimension-reduction to two dimensions, permitting the projection of the state space into an XY-plane (light blue in **C**, **D**), so that it can be used to display *U* as a third dimension.

regulatory interactions encoded in the genome (Fig. 2A). The GRN concept is the pea under the mattress of the comfortable paradigm of a direct genotype-phenotype mapping that has served Neo-Darwinian thinking so well. But commentators have begun to explicitly articulate the need to integrate gene networks in evolution theory [9, 23, 24].

Since the 1990s the amount of literature on 'gene networks' has exploded. It covers a broad range of forms: from the use as graphical representation of regulatory interactions between genes collected by high-throughput experiments or as toy models in computer simulations of in silico evolution [25–29] to the analysis of the static structure of gene networks and its evolution [30–34]. However, the dynamic behaviour of networks is rarely considered as the object per se of evolution.

Structure and dynamics of gene networks – a critical distinction

For a fruitful discussion of the dynamics of networks and its role in the genotype-phenotype mapping, a set of shared, unambiguous definitions of terms is critical. But sadly, many biologists are not explicitly aware of the ontological dichotomy between network structure (topology) and network

Box 1

On the terminology of 'gene regulatory network': A tacit dualism in meaning

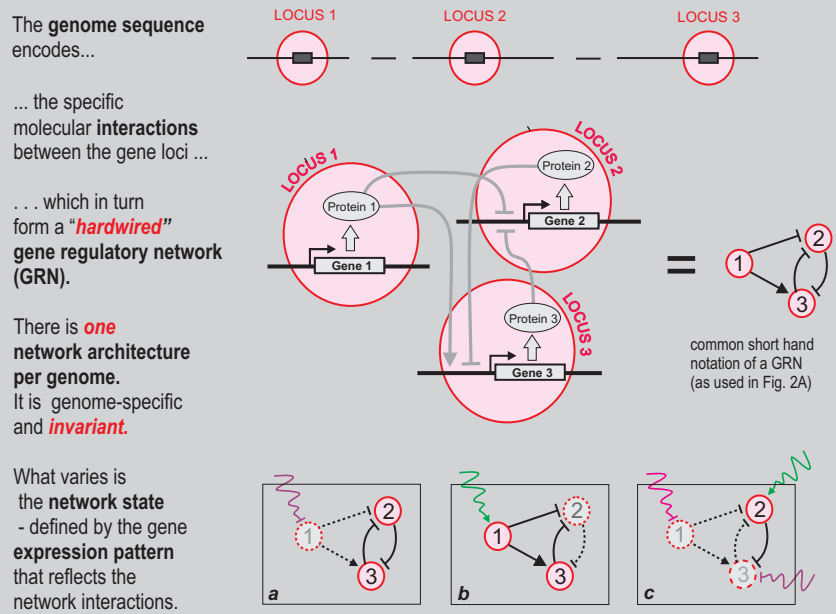
A widespread but unarticulated misunderstanding of the concept of 'gene regulatory network' (GRN) warrants a clarifying remark. In current discussions there is a dichotomy of notions of 'network' among biologists.

One group (A) holds that the term 'network' refers to a hard-wired (i.e. time-invariant) architecture. Here, the links in the network represent gene regulatory relationships, which are akin to the law: carved in stone and thus invariant, but applied to situations only when relevant. In this definition the GRN is fully specified by the genome sequence (top), which through encoding protein structure and cis-regulatory sequences defines which gene locus regulates which one. This definition also implies that the network does not change in a life time (except as a consequence of mutations). What changes is the expression level for the genes, which collectively manifest as the change of the gene expression pattern (bottom panel) and is referred to as 'network dynamics'.

In the view of a second group of biologists (B) the network is not fixed. Instead, regulatory interactions change with time. For instance, *Gene 2* may inhibit *Gene 3* (bottom panel *b* and *c*) or vice versa depending on the condition (e.g. cell type). Consequently, the network is not hard-wired in the genome but continuously changes its connections during life time [71]. This use of the term 'network' captures the loose idea of a web of instantaneous causal relationships. Some authors go as far as equating an entire gene expression pattern, as measured by microarrays, to a 'network' to imply the massively parallel, correlated behaviour. In this view any two genes that exhibit highly correlated changes over time or across conditions are connected by a line in a diagram – giving rise to the impression of a 'network' [72].

I argue that viewpoint (A) has a more solid semasiological foundation. Its consistency is illustrated in the bottom panel example, which displays the behaviour (network dynamics) of the very same 3-gene network: In condition *a* where *Gene 1* is not expressed (due to inhibitory signals, symbolized by the wiggling arrow) the *Genes 2* and *3* balance each other and are expressed at intermediate levels. Once *Gene 1* is expressed (*b*) this balance of mutual check is tilted in favour of expression of *Gene 3* – as prescribed by the molecularly specified (=genomically hard-wired) interactions. 'The existing law that '*Gene 1* suppresses *Gene 2* but activates *Gene 3*' now is applied because *Gene 1* is expressed'. This constitutes a change of conditions (external inputs allowing the presence of *Gene 1* activity) that triggers a change of the network state (now *Gene 3* is high, *Gene 2* is low) and not a change of the network wiring (='the law'). The network states in *b*, *c* illustrate the behaviour of a 'toggle' switch. The very same network architecture of mutual inhibition between *Genes 2* and *3* allows the network (genome) to produce two alternative stable expression patterns, either [Gene 2] >> [Gene 3] or vice versa, depending on the perturbation of the gene expression levels that could even have occurred transiently in the past [35, 64]. Thus, the very same genome can generate multiple enduring network states (phenotypes) – herein lies its flexibility and not in the change of its actual wiring.

I recommend adhering to the 'network' definition of (A) as it is conceptually clean and consistent with that used by pioneers of GRN long before the idea of 'networks' entered mainstream biology [8, 63] and it is also the basis for the increasing mathematical treatment of networks as dynamical systems. The usage (B) is an ad hoc descriptive term, warranted in some context, but it may perpetuate confusion and hamper communication.



dynamics [8]. This indifference has so far prevented meaningful discourse. (The problem surrounding the very concept of 'gene network' is explained in Box 1.) Whereas the 'structure' of a network is a static entity, a collection of interconnected nodes representing the genes X_i that warrants the oft used analogy of a 'wiring diagram', the 'dynamics' of a network pertains to the collective behaviour of genes that is relevant only if one considers that each gene, X_i, changes its level of expression, $x_i(t)$, over time. Network dynamics is, then, the coordinated change of the expression levels, $x_1(t)$, $x_2(t)$, $x_3(t)$ . . ., of all the genes in a GRN as a consequence of the entirety of the regulatory interactions displayed as connections (wiring) in the fixed 'wiring diagram'. In other words: strictly speaking, when talking about GRNs, there are no 'dynamic networks'. Instead, 'network dynamics' is a legitimate term indicating the characteristic dynamic behaviour exhibited by a given network.

Thus, it is crucial to distinguish between two separate timescales: network dynamics (the coordinated change of the expression levels $x_i(t)$ of the genes of a given network) takes place within one organism's lifetime, during which the wiring diagram of that network does not change. By contrast, a change of the structure, or 'rewiring' of the network, is the result of a genomic mutation (which may, for instance, affect how a regulatory gene controls its target gene) and occurs at the evolutionary time scale. Obviously, a change in the structure of the network will affect its dynamics in a particular way. Precisely herein lies the central principle for comprehending how a mutation maps, not directly but via the network, into a new phenotype.

Perhaps network dynamics and its characteristics are rarely considered as a phenotype in evolution biology [7] because their description is too abstract, although the concept dates back to the 1960s when Monod and Jacob found that genes can regulate each other's expression [35], which inspired Kauffman to build the first computational gene networks [36] more than 40 years ago to study the global dynamics of complex GRNs.

Network dynamics has recently gained popularity among molecular biologists who realize that the phenotype cannot be understood as a linear superposition of the isolated action of individual genetic pathways. This insight, accompanied by the arrival of new technologies, was a major driving force behind the rise of systems biology more than a decade ago [37]. But systems biology should aspire to more than being the exhaustive categorization and characterization of the molecular parts list of the genome and the static interactions between these parts. To understand how the phenotype (the whole) is more than the sum of the genes (the parts) [38], entirety of analysis must be followed by analysis of entirety [39]. This entails embracing the collective dynamics of the genes coordinated by the GRN.

In the Neo-Darwinian scheme, if the genes provide the molecular medium for inheritance, and mutations deliver the random variability required for natural selection, the following two epigenetic phenomena may rival genes and mutations as the substrate of natural selection (Fig. 1):

(i) The dynamics of the GRN can generate a multitude of stable (persisting) states of the system and, thus, act as a source of variability and as a medium of information storage – both capabilities traditionally attributed to mutations and genes [8];
(ii) Gene expression noise, i.e. the random fluctuations of gene expression levels that propagate through the GRN, permits the random occupation of the various stable states, which results in (non-genetic) phenotypic heterogeneity of (even clonal) cell populations – and, thus, is a source of phenotypic randomness [40, 41].

In brief, networks and noise are non-genetic sources of diversity and stochasticity – the chief ingredients of Darwinian evolution. They may claim a spot on the pedestal of evolutionary mechanisms that has long been the uncontested domain of genetic mutations. If properly incorporated into Neo-Darwinian thinking, so to speak in a 'Post-Modern Synthesis', these two epigenetic phenomena will not threaten its fundament but broaden it, reconciling natural selection with observations that have stimulated quite imaginative stretching of the narrative of natural selection.

Here I present a concise, permissively simplified explanation of the formal concepts of gene network dynamics. This will hopefully provide a more tangible notion of network behaviour, hence allowing us to study its evolution, and clear the view on the obscure term 'epigenetics', the loose usage of which has stifled communication in biology and made its integration with Neo-Darwinism difficult (see Supporting Information, Text I).

From molecular interaction to gene networks to systems dynamics

How can a network of regulatory interactions between genes create distinct phenotypic states with memory of themselves and which can be occupied in response to a signal or just by chance? What Waddington called developmental constraints and epigenetics [42–44] can now be identified as the layers of molecular regulatory networks and cell-cell communication networks – a web of interactions through which genomic information must percolate to produce the macroscopic phenotype. Viewing through the lens of dynamical systems theory, this blurry layer of dense biological interactions will appear in our focal plane with new clarity. Here I discuss the layer of GRNs to illustrate the basic principles.

The key ideas of GRN dynamics are summarized in Fig. 2 (for more details see [8, 45]). Let us start here by declaring, for the sake of argument, the premise that a given phenotypic state of a cell can be viewed, in a first approximation, as encoded by its momentary genome-wide gene expression pattern. ('Expression' is here understood as the active state of a gene locus manifested as the presence of the protein that it encodes.) A given gene expression pattern is one among the hyperastronomically vast number of combinatorial configurations of the expression status of the N gene loci in the entire genome. Now, the expression of an individual gene is not independent of that of other genes, precisely because of the presence of the GRN, through which the gene loci (in a locus-specific manner) influence each other's expression and, hence, contribution to the phenotype. This inter-loci

communication is best known to occur via transcriptional control of gene expression and via protein-protein interactions. However, one can also include nucleosome and chromatin conformation [46–48] as well as chromosome-mediated interaction between distant loci as exerting regulatory effects [49–52].

The GRN is thus the inter-gene loci control network and is commonly plotted as a wiring diagram indicating which gene has the potential to regulate which other, under what conditions and by what modality (inhibition or induction of expression). Here I reiterate the fundamental notion that the structure of the GRN diagram (Fig. 2A) is fully encoded ('hard-wired') in the genome because each single interaction is ultimately molecular in nature and its specificity and effect is determined by the structure of proteins and by the cis-regulatory DNA sequence. Therefore, the network structure has been wired by evolution and hard-coded into the genomic sequence. It is intrinsic to and, in the absence of mutations, invariant for a given genome (organism). Thus, each cell in metazoa has the same GRN.

Here, I also emphasize again that the GRN structure changes in phylogenetic time scale only. This central property is often misrepresented in expressions like 'the changes in the GRN during development' (see Box 1 for elaboration of this misunderstanding). There is ONE, and only ONE, static network topology (=structure = wiring diagram) in a given genome. What changes in development and physiology is not the network structure but the gene expression patterns, i.e. the configuration of the expression states $x_i(t)$ of the individual genes in the very same (invariant) genome. This configuration constitutes a network state S. A change of the network states S takes place in ontogenetic time scale. This is what is commonly referred to by the term 'network dynamics'.

In summary, a change in network state (i.e. network dynamics) is caused by the coordinated alteration in the expression status of the genes within the same genome, whereas a change in network structure is caused by a genetic mutation – thus, naturally separating ontogeny from phylogeny. This dichotomy is central for a formal understanding of the following explanation.

Network dynamics reflects constraints on the realization of expression patterns

As a consequence of the inter-dependence of expression of individual gene loci, the genome is not free to choose any possible gene expression pattern. Each of the zillions of combinatorially possible configurations of expression status at all the gene loci across the genome establish a gene expression pattern but they are not created equal (Fig. 2B): some gene expression patterns S are much more 'stable' than others. Some are so unstable as to be barely realizable because doing so would violate the inter-dependencies of gene expression determined by the regulatory relationships. For instance, if a gene A (unconditionally) inhibits the expression of the genes B and C, as determined by the network structure, then any gene expression pattern in which gene A is highly active but genes B and C are also highly expressed would be 'forbidden' by the network. Such a pattern would spontaneously adopt a more stable configuration (e.g. by downregulating genes B and C) that complies with the interaction rules.

Thus, gene-gene regulatory interactions introduce massive constraints on the realization of possible configurations of gene activities and this is manifest in the distinct stabilities of gene expression patterns. The coordinated change of gene expression underlies the change of the network state as an integrated entity. Mathematically, the network state S of a network with N genes can be represented by a point at a particular position S, whose coordinates are defined by the N values of the gene expression levels in S, within the space of all possible gene expression patterns. This abstract space is referred to as the N-dimensional 'state space' (Fig. 2C). When the gene expression pattern changes, the position of S changes accordingly and the state S moves along a predestined path ('trajectory', Fig. 2C). This movement is driven by the network's search for more stable expression patterns – or network states.

Attractors as quasi-potential wells: multistability

Since network states have distinct stabilities, one needs to know the relative stability of each state S, which is a consequence of the particular 'wiring' of the regulatory interactions; this will allow one to predict a phenotype change. The relative stabilities can indeed be computed (with some fundamental restrictions – see legend of Fig. 2) and represented as a quasi-potential energy U of each state S [53]. Thus, the specific network interactions impart on each state S the urge to move on a specified trajectory towards a more stable state until all regulatory influences are satisfied. In such a balanced, 'force-free' state S^* the gene expression pattern is stationary and also 'locally stable'. Such stability implies that it is surrounded by less stable states – thus, it will re-establish its characteristic gene expression pattern when slightly perturbed and displaced to a close neighbourhood. A stable state S^* is called 'attractor state' as it attracts the nearby, less-stable states that are in its 'basin of attraction'. Trajectories within a basin of attraction converge to the attractor state of that respective basin – the state where the quasi-potential U exhibits a local minimum – akin to being at the bottom of a 'potential energy well' in classical physics [53, 54].

If complex regulatory networks are rather sparse and contain non-linear regulatory relationships and positive feedback loops [8], as is typically the case with GRNs, it turns out that they readily give rise to a multitude attractor states within one network. This coexistence of multiple attractors ('multistability') is the chief reason why a quasi-potential U offers additional information not available in traditional dynamical systems theory that examines only *local* stabilities; the landscape affords a comparison of the relative stabilities ('depths') of attractors [53–55].

Since a stable attractor state S^* defines a particular stable expression pattern of all the N genes of a network, it also, according to the premise, represents an observable cellular phenotype [8, 45]. It is now accepted that the distinct cell types in metazoa, which are characterized by specific, self-stabilizing gene expression profiles, are attractor states – a general

idea first proposed by Delbrück [56] and by Monod and Jacob [35] more than half a century ago, and in its modern form by Kauffman in the 1970s [22, 57]. Experimental support has been obtained by measuring the convergence of trajectories of genome-wide transcript profile [58, 59], although a nominal cell type may actually be a heterogeneous mixture of sub-cell types – represented accordingly by multiple small adjacent attractors [60]. Accordingly, a coherent developmental change of cell phenotype, equivalent to the coordinated alteration of the expression status of genes across the genome, is a movement in state space towards a more stable state (among multiple alternative choices) – a journey along a path predestined by the wiring of the GRN.

The broader biological significance of the attractor concept is that one single GRN can produce a diversity of attractor states – discretely distinct, stable gene expression profiles, hence, phenotypes. This represents a first departure from the one-to-one genotype-to-phenotype mapping.

Towards the epigenetic (quasi-potential) landscape

To achieve the familiar three-dimensional 'potential landscape' picture accessible to human imagination, the *N* dimensions of the gene expression state space is compressed (projected) into a two-dimensional *XY* plane (Fig. 2d), in which the neighbourhood relationship is preserved, and in which two nearby points (states) represent similar expression patterns. The third, Z-dimension then can be utilized to depict the quasi-potential energy *U* of each gene expression pattern at the given *XY* position (state *S*) as an 'elevation' that reflects the stability of respective state *S*. A higher elevation, by convention, represents higher quasi-potential and, hence, lower stability. The resulting quasi-potential landscape can be depicted as a geographical landscape with a characteristic topography – precisely as Waddington envisioned in his epigenetic landscape [9, 43, 44, 61]. The cell state *S* is defined by a position on that landscape, marked by Waddington's marble which rolls down the valleys as *S* changes, seeking the lowest points available (local minimum of *U*), which corresponds to the most stable gene expression pattern in a given region. The stable gene expression pattern of the attractor state that the cell adopts appears to self-organize across the genome, when in reality it merely seeks stability of the network state. Again, 'order for free'. By contrast, hills represent 'energy barriers' that constrain the course of the marble to the trajectory of *S* in state space. This trajectory represents the succession of gene expression patterns and, hence, represents a developmental path that Waddington called 'chreods'.

In summary, the specific landscape topography (Fig. 2D) is a mathematical manifestation of the constraints imposed by the particular network architecture (Fig. 2A) on the global dynamics of a GRN. From GRN studies in silico it has been shown that a network of the architecture type found in evolved GRNs almost inevitably generates a 'rugged landscape' with multiple attractor states [8, 22, 28]. Thus, one genome specifies one particular GRN wiring diagram, which in turn specifies one particular landscape that captures that genome's entire developmental potential. It is through the landscape topography, as an 'extended genotype', that the genome determines developmental trajectories and terminal gene expression profiles of cell types and thus, ultimately, the organismal phenotype. This principle is in line with the idea that Davidson has long championed: that embryonic development is hard-wired in the genome [62, 63]. But there is room for some malleability as discussed next.

Epigenetic landscape and network perturbations

Beyond the intrinsic downward force imposed by the interactions hard-wired in the network, external control or perturbation of the expression of individual gene loci also affects the dynamics of the state *S*. An alteration of the expression levels of genes will, by definition, displace a state *S*, thus changing its position on the landscape, by working against the intrinsic topography-driven movement. Perturbations, akin to a wind gust blowing onto Waddington's descending marble, can derail the marble from its fated trajectory and, if large enough, cause it to switch valleys and hence destination phenotype [53, 64]. Such an attractor transition represents an externally induced, discrete phenotype switch. Once in an attractor, the network (or cell) will stay there long after the external perturbation has vanished. Thus, attractor states convey a memory – allowing the cell to remember a particular gene expression profile induced by a previously encountered signal. It stores information of past perturbations (which can be a random event) – but without the need for mutations. The genome and, hence, the topography (specifically, heights of separating hills) controls the susceptibility to such perturbation-induced transitions.

A perturbation of a particular kind is caused by gene expression noise, due to continuous intrinsic thermal fluctuations of the expression activity of each gene locus or to random external events that impinge of the expression mechanism [41]. Because the amplitudes of the fluctuations in the former case are small, they are typically seen as a 'wiggling' movement of *S* within the valleys, which usually does not deviate much from its course because of the 'attracting' forces around attractor states. Its trajectory is 'buffered' as Waddington said. But the intrinsic noise or, more likely, random external perturbations, may occasionally allow a cell to 'jump' out of a basin of attraction over a sufficiently low hill into a neighbouring valley (attractor).

It is important to note that the landscape is not a true energy potential as in classical physics, which gives rise to 'path independence' (see legend of Fig. 2). It arises from the mathematical analysis of the global stability of states of the GRN treated as a dynamic system.

To sum up, the landscape manifests the forces emanating from the intrinsic regulatory constraints of the network that extrinsic perturbations need to work against [54]. For all practical purposes, it has to be remembered that each genome, which uniquely determines which gene regulates which one and how, uniquely maps into a particular landscape topography. This purist definition of the endogenous, network-determined constraints allows us to erect a clean conceptual separation between nature and nurture and facilitates the

discussion of epigenetics and its reconciliation with Neo-Darwinism. With the above narrow formalism, it is now plausible that mutations change the network architecture and hence reshape the landscape topography, whereas non-mutagenic perturbations change the gene expression pattern and hence displace the marble on the landscape.

Conceptual implications for evolutionary biology

The formalism for mapping a GRN into an epigenetic, quasi-potential landscape sheds new light on the black box of the genotype-phenotype correspondence. (The Supporting Information Text II lists the most salient corollaries of this concept.) The quasi-potential offers a solid conceptualization of the much-sought departure from the one-to-one mapping implied by the central dogma (Fig. 1) that fails to capture complex, non-linear relationships, and opens a new perspective for examining the evolution of network behaviour itself.

A single-gene mutation will only alter a phenotype in a predictable manner if the phenotype is directly encoded by the mutated gene – as is the case for peripheral effector genes (e.g. encoding haemoglobin chain or enzymes involved in melanin synthesis) that do not control expression of other loci and hence do not contribute to network dynamics [65]. Mutations in such effector genes are plausibly apparent as monogenetic inherited disease (e.g. albinism, thalassemia). If, in contrast, the gene subjected to a mutation is a regulator embedded in the GRN, the network will be rewired and the phenotypic consequences are indirect and non-obvious [66]. But the landscape concept now affords some general explanatory principles (if the entire network architecture is known) [28]. As explained above, the mutation will alter the landscape, either by inflicting a graduated change, such as distorting the contours of a valley (basin of attraction) and altering the height of separating hills (hence modulating accessibility), or, more dramatically, by destroying or creating an attractor [67].

The latter type of mutation could explain phenomena in evolution characterized by sudden, rare and discontinuous events that have been difficult to reconcile with the gradual nature of evolution. When a new attractor suddenly becomes accessible and occupied following a mutational landscape distortion (a rather rare event [28]), not only does it open an entire pre-organized stable, but previously inaccessible, gene expression configuration (possibly, a new cell type) but also a spurt of evolutionary progress could be triggered. This is because the gene expression patterns encoded by the newly occupied attractors in uncharted land have not been subjected to evolution and thus, have not undergone optimization for contributing to organismal fitness. Therefore, these still possess much optimization potential – allowing for a sudden rapid climb to a new fitness peak (e.g. the evolution of the effector functions of a new cell type). This formal explanation obviously differs fundamentally from that of the Neo-Darwinian school that depends on a linear genotype-phenotype mapping and has struggled with the question of how small mutations in the conserved homeotic genes cause large, discrete jumps in phenotype space [68].

Conclusions

Many of the conceptual difficulties of Neo-Darwinian theory can be traced to its failure to embrace the dynamic consequences of gene regulatory interactions in their entirety – a much neglected source of self-organizing constraints, variability and randomness with persistent effects. The quasi-potential landscape with attractors – a mathematical entity that has a molecular basis and is not a mere metaphor – must be considered as a key intermediate layer in the genotype-to-phenotype correspondence that underlies Neo-Darwinian theory. Gene network dynamics readily accounts for Waddington's genetic assimilation, the related Baldwin effect [69], Neo-Lamarckism and other epigenetic phenomena [70] presented by critics of Neo-Darwinism. It is plausible that these non-genetic mechanisms could accelerate evolutionary change by allowing the temporary selection of a metastable, epigenetically encoded phenotype while it 'waits' for the appropriate genetic mutation to occur (by chance) – as if acting as a lubricant for Darwinian natural selection. This would help to explain observations that hitherto required faith in some unlimited power of natural selection of random mutants. Ironically, both critics as well as defenders of Neo-Darwinian orthodoxy carry out their arguments often with a temperament of well-articulated, colourful metaphoric ideas instead of using the formal, inescapable principles of complex dynamic systems. But both groups would benefit from incorporating them in their thinking.

Acknowledgments

The author would like to thank Stuart Kauffman for inspiring discussions over the past decade and Adam Wilkins for thoughtful comments, and to acknowledge the following agencies for funding: iCore/Alberta Innovates the Future (AITF); Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. **Ayala FJ**, **Fitch WM.** 1997. Genetics and the origin of species: an introduction. *Proc Natl Acad Sci USA* **94**: 7691–7.
2. **Dobzhansky T.** 1937. *Genetics and the Origin of Species*. New York: Columbia University Press.
3. **Huxley JS.** 1942. *Evolution, the Modern Synthesis*. London: Allen and Unwin.
4. **Mayr E.** 1942. *Systematics and the Origin of Species*. New York: Columbia University Press.
5. **Crick FH.** 1958. On protein synthesis. *Symp Soc Exp Biol* **12**: 138–63.
6. **Davidson EH**, **Erwin DH.** 2006. Gene regulatory networks and the evolution of animal body plans. *Science* **311**: 796–800.
7. **Huang S.** 2004. Back to the biology in systems biology: what can we learn from biomolecular networks. *Brief Funct Genomics Proteomics* **2**: 279–97.
8. **Huang S**, **Kauffman S.** 2009. *Complex gene regulatory networks – from structure to biological observables: cell fate determination*. In Meyers RA, ed; *Encyclopedia of Complexity and Systems Science*. Berlin: Springer.
9. **Wilkins AS.** 2008. Waddington's unfinished critique of neo-darwinian genetics: then and now. *Biol Theor* **3**: 224–32.
10. **Strohman RC.** 1997. The coming Kuhnian revolution in biology. *Nat Biotechnol* **15**: 194–200.
11. **Pigliucci M.** 2009. An extended synthesis for evolutionary biology. *Ann NY Acad Sci* **1168**: 218–28.

12. **Lynch M.** 2007. Colloquium papers: the frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA* **104**: 8597–604.
13. **Noble D.** 2008. Genes and causation. *Philos Trans A Math Phys Eng Sci* **366**: 3001–15.
14. **Rose S**, **Kamin LJ**, **Lewontin RC.** 1985. *Not in Our Genes: Biology, Ideology, and Human Nature*. New York: Pantheon.
15. **Lewontin R.** 2000. *The Triple Helix: Gene, Organism, and Environment*. Cambridge: Harvard University Press.
16. **Morange M.** 2001. *The Misunderstood Gene Cambridge*. Harvard University Press.
17. **Gould SJ**, **Lewontin RC.** 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* **205**: 581–98.
18. **Goodwin B.** 1993. *How the Leopard Changed Its Spots: The Evolution of Complexity*. Princeton: Princeton University Press.
19. **Goodwin BC**, **Webster GC.** 1999. Rethinking the origin of species by natural selection. *Riv Biol* **92**: 464–7.
20. **Autumn K**, **Ryan MJ**, **Wake DB.** 2002. Integrating historical and mechanistic biology enhances the study of adaptation. *Q Rev Biol* **77**: 383–408.
21. **Nicolis G**, **Prigogine I.** 1989. *Exploring Complexity: An Introduction*. San Francisco: W.H. Freeman & Company p. 328.
22. **Kauffman SA.** 1993. *The Origins of Order*. New York: Oxford University Press.
23. **Pigliucci M.** 2010. Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philos Trans R Soc Lond B Biol Sci* **365**: 557–66.
24. **Bard J.** 2010. A systems biology view of evolutionary genetics. *BioEssays* **32**: 559–63.
25. **Berg J**, **Lassig M**, **Wagner A.** 2004. Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evol Biol* **4**: 51.
26. **Draghi J**, **Wagner GP.** 2009. The evolutionary dynamics of evolvability in a gene network model. *J Evol Biol* **22**: 599–611.
27. **Siegal ML**, **Promislow DE**, **Bergman A.** 2007. Functional and evolutionary inference in gene networks: does topology matter? *Genetica* **129**: 83–103.
28. **Aldana M**, **Balleza E**, **Kauffman S**, **Resendiz O.** 2007. Robustness and evolvability in genetic regulatory networks. *J Theor Biol* **245**: 433–48.
29. **Wagner A.** 2008. Neutralism and selectionism: a network-based reconciliation. *Nat Rev Genet* **9**: 965–74.
30. **Babu MM**, **Luscombe NM**, **Aravind L**, **Gerstein M**, et al. 2004. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14**: 283–91.
31. **Bornholdt S**, **Rohlf T.** 2000. Topological evolution of dynamical networks: global criticality from local dynamics. *Phys Rev Lett* **84**: 6114–7.
32. **Barabasi AL**, **Oltvai ZN.** 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**: 101–13.
33. **Alon U.** 2003. Biological networks: the tinkerer as an engineer. *Science* **301**: 1866–7.
34. **Wilkins AS.** 2007. Colloquium papers: between ''design'' and ''bricolage'': genetic networks, levels of selection, and adaptive evolution. *Proc Natl Acad Sci USA* **104**: 8590–6.
35. **Monod J**, **Jacob F.** 1961. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* **26**: 389–401.
36. **Kauffman SA.** 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* **22**: 437–67.
37. **Hood L**, **Rowen L**, **Galas DJ**, **Aitchison JD.** 2008. Systems biology at the institute for systems biology. *Brief Funct Genomic Proteomic* **7**: 239–48.
38. **Aristotle**. 350 BC. Metaphysica. Book VIII, Part 6.
39. **Huang S**, **Wikswo J.** 2006. Dimensions of systems biology. *Rev Physiol Biochem Pharmacol* **157**: 81–104.
40. **Huang S.** 2009. Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**: 3853–62.
41. **Eldar A**, **Elowitz MB.** 2010. Functional roles for noise in genetic circuits. *Nature* **467**: 167–73.
42. **Waddington CH.** 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150**: 563–5.
43. **Waddington CH.** 1957. *The Strategy of the Genes*. London: Allen and Unwin.
44. **Slack JMW.** 2002. Timeline – Conrad Hal Waddington: the last renaissance biologist? *Nat Rev Genet* **3**: 889–95.
45. **Huang S.** 2009. Reprogramming cell fates: reconciling rarity with robustness. *BioEssays* **31**: 546–60.
46. **Fuks F.** 2005. DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* **15**: 490–5.
47. **Mellor J.** 2006. Dynamic nucleosomes and gene transcription. *Trends Genet* **22**: 320–9.
48. **Kouzarides T.** 2007. Chromatin modifications and their function. *Cell* **128**: 693–705.
49. **Lee JS**, **Smith E**, **Shilatifard A.** 2010. The language of histone crosstalk. *Cell* **142**: 682–5.
50. **Deng W**, **Blobel GA.** 2010. Do chromatin loops provide epigenetic gene expression states? *Curr Opin Genet Dev* **20**: 548–54.
51. **Jackson DA.** 2010. Spatial epigenetics: linking nuclear structure and function in higher eukaryotes. *Essays Biochem* **48**: 25–43.
52. **Li G**, **Reinberg D.** 2011. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* **21**: 175–86.
53. **Wang J**, **Xu L**, **Wang E**, **Huang S.** 2010. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys J* **99**: 29–39.
54. **Freidlin M**, **Wentzell A.** 1984. *Random Perturbations of Dynamical System*. New York: Springer-Verlag.
55. **Nicolis G.** 1986. Dissipative systems. *Rep Prog Phys* **49**: 873–949.
56. **Delbrück M.** 1949. Discussion. Unités biologiques douées de continuité génétique. Colloques Internationaux du Centre National de la Recherche Scientifique. CNRS, Paris. pp. 33–55.
57. **Kauffman S.** 1969. Homeostasis and differentiation in random genetic control networks. *Nature* **224**: 177–8.
58. **Huang S**, **Eichler G**, **Bar-Yam Y**, **Ingber DE.** 2005. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* **94**: 128701.
59. **Chang HH**, **Hemberg M**, **Barahona M**, **Ingber DE**, et al. 2008. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**: 544–7.
60. **Huang S.** 2011. Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. *Philos Trans R Soc Lond B Biol Sci* **366**: 2247–59.
61. **Siegal ML**, **Bergman A.** 2002. Waddington's canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci USA* **99**: 10528–32.
62. **Arnone MI**, **Davidson EH.** 1997. The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**: 1851–64.
63. **Davidson EH.** 2006. *Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. San Diego: Academic Press.
64. **Zhou JX**, **Huang S.** 2011. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends Genet* **27**: 55–62.
65. **Guo Y**, **Feng Y**, **Trivedi NS**, **Huang S.** 2011. Medusa structure of the gene regulatory network: dominance of transcription factors in cancer subtype classification. *Exp Biol Med (Maywood)* **236**: 628–36.
66. **Prud'homme B**, **Gompel N**, **Carroll SB.** 2007. Emerging principles of regulatory evolution. *Proc Natl Acad Sci USA* **104**: 8605–12.
67. **Huang S**, **Ernberg I**, **Kauffman S.** 2009. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev Biol* **20**: 869–76.
68. **Merabet S**, **Hudry B.** 2011. On the border of the homeotic function: re-evaluating the controversial role of cofactor-recruiting motifs. *BioEssays* **33**: 499–507.
69. **Crispo E.** 2007. The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution* **61**: 2469–79.
70. **Jablonka E**, **Lamb MJ.** 2002. The changing concept of epigenetics. *Ann NY Acad Sci* **981**: 82–96.
71. **Lu R**, **Markowetz F**, **Unwin RD**, **Leek JT**, et al. 2009. Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* **462**: 358–62.
72. **Butte AJ**, **Kohane IS.** 2006. Creation and implications of a phenome-genome network. *Nat Biotechnol* **24**: 55–62.