José Dávila-Velderrain* and Elena Álvarez-Buylla Roces**

Linear Causation Schemes in Post-genomic Biology: The Subliminal and Convenient One-to-one Genotype-Phenotype Mapping Assumption

Abstract | In this essay we question the validity of basic assumptions in molecular biology and evolution on the basis of recent experimental data and through the lenses of a systems and nonlinear perspective. We focus our discussion on two well-established foundations of biology: the flow of information in molecular biology (i.e., the central dogma of molecular biology), and the "causal" linear signaling pathway paradigm. Under both paradigms the subliminal assumption of a one-to-one genotype-phenotype mapping (GPM) constitutes an underlying working hypothesis in many cases. We ask if this is empirically sustainable in post-genomic biology. We conclude that when embracing the notion of complex networks and dynamical processes governing cellular behavior—a view now empirically validated—one-to-one mapping can no longer be sustained. We hypothesize that such subliminal and sometimes explicit assumption may be upheld, to a certain degree, because it is convenient for the private appropriation and marketing of scientific discoveries. Hopefully, our discussion will help smooth the undergoing transition towards a more integrative, explanatory, quantitative and multidisciplinary systems biology. The latter will likely also yield more preventive and sustainable medical and agricultural developments, respectively, than a reductionist approach.

Keywords | post-genomic biology – genotype-phenotype mapping – genetic determinism – flow of genetic information

Introduction

SCIENCE IS MOSTLY PRACTICED out of consensus. Scientific progress, however, is also sustained by the continual challenge to accepted ideas. Unstated agreements break from time to time, and then—some say—a transition, a so-called paradigm

* Instituto de Ecología-Universidad Nacional Autónoma de México. **E-mail:** jdjosedavila@gmail.com

** Centro de Ciencias de la Complejidad-Universidad Nacional Autónoma de México. **E-mail:** eabuylla@gmail.com

Dávila-Velderrain, José and Elena Álvarez-Buylla. «Linear Causation Schemes in Post-genomic Biology: The Subliminal and Convenient One-to-one Genotype-Phenotype Mapping Assumption.» *Interdisciplina* 3, nº 5 (2015): 267-280.

shift, occurs (Kuhn 2012 [1962]). In the last decades, several authors have discussed the possibility of a paradigm shift in biology, given the apparent crisis of some of its foundational principles. (Wilkins 1996; Strohman 1997; O'Malley and Boucher 2005). In this paper, we would instead like to substantiate that a large portion of mainstream biological research subliminally embraces particular assumptions that are empirically unsustainable in this post-genomic era. Some of these assumptions are so deeply rooted that they still permeate the design, interpretation and description of a wide range of biological research at the molecular level, although, if explicitly confronted, anyone would dismiss them. Routinely we look for single, "causal" mutations responsible for complex phenotypes and assume that by finding the molecular basis of a mutation that is correlated to a particular condition, the emergence of the latter is explained. Importantly, such rationale implies that in most cases a one-to-one relationship will be possible. By extending such assumptions we define signaling pathways as autonomous entities instructing the cell how to behave under a particular condition. If pathological behavior arises, we look for the source of incorrect instructions: the mutated component or pathway. We automatically interpret any manifestation of a *learned* feature, such as drug resistance, as the consequence of the optimization principles of (*Darwinian*) adaptation by means of "random" mutation and selection. Is this recurrent bias towards *ad hoc* explanations based solely on plausibility given the evidence, or is it the mere consequence of a naively inherited tradition? We consider that an explicit presentation of some of the assumptions in light of post-genomic empirical data, and through the lenses of a systems, nonlinear perspective to biology, will clarify this question. This may prove useful for current biology students and scientists interested in multidisciplinary research.

We also include in the term post-genomic *several features that characterize modern biology: (1) abundance of experimental molecular data, (2) access to systematic ways of characterizing cellular phenotypic states, and (3) a tendency to produce quantitative data and to formulate mathematical/computational models. Consequently, in our view,* post-genomic biology *is necessarily multidisciplinary, integrative, formal, and quantitative*

A first necessary detour: *What do we mean by post-genomic biology?* The availability of complete genome sequences (and also transcriptomes, proteomes, metabolomes, etc) obviously impacted biological research, enabling new levels of interrogation –as well as unmasking new sources of empirical support (rejection) for otherwise assumed facts. Here, however, besides access to genome-wide data, we also include in the term *post-genomic* several features that characterize modern biology: (1) abundance of experimental molecular data, (2) access to systematic ways of characterizing cellular phenotypic states, and (3) a tendency to produce quantitative data and to formulate mathematical/computational models. Consequently, in our view, *post-genomic biology* is necessarily multidisciplinary, integrative, formal, and quantitative.

The Most Basic, Naive Assumption: The One-to-One GPM

Nowadays, it is common to think about the relationship between genotypes and phenotypes in terms of some kind of complex mapping (Kauffman 1993; Mendoza and Álvarez-Buylla 1998; Wagner and Zhang 2011; Davila-Velderrain and Álvarez-Buylla 2014; Ho and Zhang 2014). The concept of a "genotype-phenotype map" can be traced back to Alberch, who elegantly proposed a model based on the principles of systems dynamics to express the inadequacy of what some call (molecular) *genetic determinism*, i.e., the assumption that genes directly determine phenotypes (Alberch 1991). Equally limited would be to assume an *epigenetic determinism*. Importantly, such A gene-centered assumption is the conceptual basis of the often invoked metaphors of a 'genetic blueprint' or a 'genetic program' (Pigliucci 2010). Furthermore, it also implies a linear relationship between genotypes and phenotypes; in other words, a *one-to-one* mapping. This simplistic model is attractive, since it naturally embraces a cause-and-effect interpretation, which makes it intuitively appealing. But if we think about this assumption of *one genotype specifically producing a particular phenotype*, we have to address how such a simplistic view can fit any observation. Nonetheless, this one-to-one model is still at the basis of most mainstream programs of biomedical or biotechnological developments (e.g., transgenic crops).

A second necessary detour: w*hat genotype and phenotype?* In the epistemology of evolution and biology, in general, it is common to talk about genotype and phenotype as absolute terms. But these can be defined at different levels, and in practice genotype and phenotype distinctions are just partial and dynamical (Lewontin 2011). In post-genomic biology this distinction is commonly aided by the use of simple GPM models (see, for example Soyer 2012). Consequently, there is not only one type of genotype and phenotype. A GPM model can be specified in different ways. For the sake of this essay we establish

that the genotype will be represented by a gene regulatory network (GRN) and the phenotypes by a gene expression profile or configuration (see below). Nevertheless, it is noteworthy that in the current era of next-generation sequencing (NGS) and single-cell biology, the empirical characterization of the complete genotypes of multiple individual cells is becoming feasible. Unfortunately, for both conceptual and technical reasons, the same cannot be said for phenotypes—although specific systematic phenotyping strategies are under development (see, for example Houle et al. 2010; Hancock 2014).

One-to-One Genotype-Phenotype Mapping and the Central Dogma

Crick declared "the central dogma of molecular biology" first in 1958 and then it was reiterated once again in 1970 (Crick 1958, 1970). In simple terms, the *dogma* posits that information flows within cells from DNA to RNA to proteins; and, as a result, the cellular phenotype is determined (Shapiro 2009). The simplifications involved in the model have been already questioned from an information viewpoint, concluding that discoveries in the last decades have made the dogma untenable (Shapiro 2009). Here we focus instead on the cemented role of the dogma regarding the implicitly assumed linear and unidirectional scheme of causation of molecular phenotypes. According to an explicit interpretation of the dogma one gene encodes for one protein, which somehow determines one observable trait (i.e. phenotype). This simplistic view can be framed effectively into a one-to-one GPM model (see Figure 1a). How do we define a phenotype? Here a phenotype is assigned to a molecule, a protein, because it is said to have a *function*. This function should be then an observable characteristic of the cell (organism). Therefore, the first one-to-one GPM to discuss would be: a gene (i.e., the genotype) codes for a protein, which performs a specific function that determines an observable characteristic (i.e., the phenotype).

Is this One-to-One (Gene-to-Function) Model Empirically Sustainable in Post-Genomic Biology?

A first difficulty that we can think of is conceptual in nature. What do we mean by *function*? Defining a function in biology is not trivial (Huang 2000; Huneman 2013; Brunet and Doolittle 2014, Doolittle *et al.* 2014). First of all, the function assignment can be given to entities at multiple levels of molecular organization; such as gene, protein, protein domain, protein complex, or pathway (Huang 2000). In the last years, researchers in the areas of genomics and epigenomics are even advocating the mapping of function at genome-wide level and single-nucleotide resolution (Kellis *et al.* 2013). For the sake of concreteness, let us

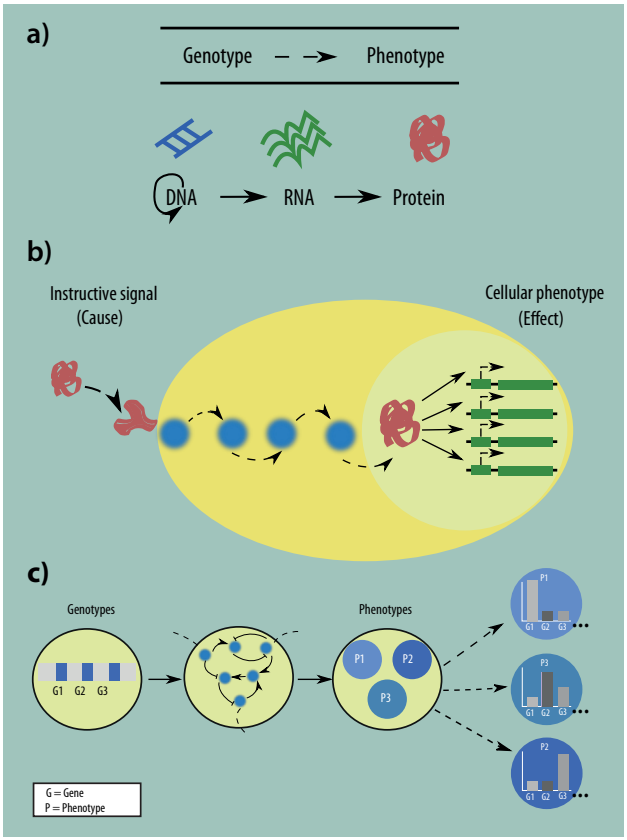

Figure 1. Schematic representation of the GPM exposed in the main text. a) One-to-one GPM model representing the central dogma of molecular biology: a gene (i.e., the genotype) codes for a protein, which performs a specific function that determines an observable characteristic (i.e., the phenotype). b) One-to-one GPM model representing the causal linear signaling pathway paradigm: genes code the proteins involved in the pathway (genotype), and these map one specific molecular signal (instruction) to a one specific cellular phenotype. c) A non-linear GPM representing cell phenotype specification by GRN dynamics: genes in a single genome (genotype) interact in complex GRNs whose regulatory interactions ultimately determine observable cell phenotypes.

just focus on function at the protein level. Although what we define as protein function is most of the times conditional on the context –i.e., cellular environment– (Huang 2000), for the purpose of our discussion, let us also assume that a protein function can be invariably assigned. Thus, in the simple one-to-one model, one gene is invariably linked to a specific function through the action of a protein.

According to the most recent assembly version of the human genome in Ensembl database (http://www.ensembl.org/), humans have 20,389 coding genes, 9,656 small noncoding genes and 14,470 long non-coding genes. A first obvious observation is that not all genes code for proteins. Two post-genomic facts: (1) most of the human genome is non-protein-coding (Alexander *et al*. 2010), and (2) transcription occurs much more often than anticipated (Carninci *et al*. 2005; Cheng *et al*. 2005). Do the genes that do not encode proteins also define a phenotype? Well, probably, in some way; but surely not by means of a one-to-one GPM, given the emerging view that non-coding transcription is tightly linked to gene regulation and cell-type specification (Natoli and Andrau 2012). For example, it was recently shown that RNA transcribed from enhancers, the so-called eRNA, is able to regulate transcription (Plosky 2014). As we will see below, gene regulation in itself is the core mechanism behind the definition of gene regulatory networks; it is also fundamental for understanding network collective behavior. Conceptualizing cell behavior in terms of molecular networks, in turn, represents a complete deviation from a one-to-one GPM (see below).

Recent assembly version of the human genome in Ensembl database, humans have 20,389 coding genes, 9,656 small noncoding genes and 14,470 long non-coding genes. A first obvious observation is that not all genes code for proteins. Two post-genomic facts: (1) most of the human genome is non-protein-coding and (2) transcription occurs much more often than anticipated. Do the genes that do not encode proteins also define a phenotype?

Besides (non)coding genes, the number of proteins coded in the human genome and represented by transcript modifications has been estimated to be between 50,000 and 500,000 (Uhlen and Ponten 2005). Considering the now known number of both genes and (estimated) proteins in other organisms, several authors have pointed out that genomic (and proteomic) complexity are not correlated with phenotypic complexity (see, for example Huang 2002). This empirical fact again is not consistent with what we would expect by extension of the dogma.

Beyond curiosity awakened by newly generated genomic data, a more serious drawback of the one-to-one GPM associated with the *central dogma* is that it completely ignores gene interactions (Tyler *et al*. 2009). Epistasis refers to the phenomenon in which the functional effect of one gene is conditional on other

genes (Phillips 2008), whereas *Pleiotropy* refers to one function being affected by multiple genes (Stearns 2010); these two phenomena are well-established facts (and concepts) in classical and modern genetics (Lehner 2011; Wagner and Zhang 2011). Nowadays such genetic interactions are being studied systematically at a genomic scale. For example, it is now possible to test millions of different combinations of double mutants and to evaluate their effects on a quantifiable function, as Costanzo and colleagues did using the budding yeast, *Saccharomyces cerevisiae* (Costanzo *et al.* 2010). Studies such as this one have clearly shown that the effect of one gene on a specific phenotype depends on the activity (or lack thereof) of many other genes. In this sense, a *genetic* interaction is defined on the base of this conditional functional effect. Although a careful discussion of epistasis and pleiotropy is beyond the scope of this paper, it is noteworthy that such mechanisms are closely related with two undeniable types of experimental evidence: (1) very different results can be produced from a nearly identical set of genes or the same genotype can produce contrasting phenotypes, and (2) virtually identical phenotypic end points can be reached by using extremely different genotypes. Evidently, these facts do not fit a one-to-one GPM. Although seemingly paradoxical, both statements can be perfectly reconciled by considering a many-to-many GPM model in which interactions among genetic and non-genetic components are explicitly considered; a view much more consistent with how living, adaptable systems behave and evolve.

One-to-One Mapping and Signaling Pathways

Extending the one-to-one view to a higher level, molecular biologists apply it to associating an altered signaling pathway to a particular phenotypic condition. Extracellular signals are transmitted by intermediary to effector proteins; which eventually activate the sets of genes responsible for the establishment of "*appropriate*" phenotypes. Note that the term pathway by itself makes reference to a group of events that occur orderly along a single *line*. Thus, in a sense, this multi-molecular model continues the *dogmatic* idea of linear, unidirectional information transfer. Thereby, in our view, it also effectively constitutes a one-to-one GPM (see figure 1b). Genes encode the proteins involved in the pathway (genotype), and these map unto one specific molecular signal (instruction) to one specific cellular phenotype. The linear property of signaling pathways also implies unidirectional cause-and-effect: a given instructional signal is thought to directly cause a phenotypic manifestation. Biologists have traditionally taken this simple pathway picture as a valid explanation at the molecular level for many cellular phenotypes. Not even a one-to-one approach to associate a network with a phenotype is valid (see below).

Is this One-To-One (Signal-to-Phenotype) Model Empirically Sustainable in Post-Genomic Biology?

Similar questions as the ones raised above can be posed here. For instance, are there enough signaling pathways for the number of possible extracelullar cues? Is there a direct, one-to-one, relationship among signals and phenotypes? If so, why do cellular phenotypes (i.e. cell types) seem to be discrete while, for example, signals carried by soluble growth factors display concentrations subject to continuous variation? And, more importantly, how and why are cellular phenotypes maintained after the signal has ceased? As we will explain below, rethinking cell behavior as the result of constraints imposed by regulatory interactions of complex molecular networks is useful to address these questions.

The genomic explosion has led to the brute-force characterization of molecular components and their interactions, which are now being integrated in large databases (Chatr-aryamontri *et al.* 2013). As expected, efforts have also tried to classify such components in genome-wide collections of signaling pathways in multiple organisms (Schaefer *et al.* 2009, Croft *et al.* 2010). What has been learned? Does the exhaustive characterization of pathways enable understanding of cellular phenotypes and their plasticity? In analogy to the failure of the pre-genomic prediction that by characterizing all the genes of an organism one will understand the genome-encoded rules instructing its behavior; listing molecular components and their interactions in pathways has only uncovered a picture that is much more complex than anticipated. But phenotypic manifestations are far from being explained by means of linear chains of molecular causation (Huang 2011)—or, in other words, of linear associations rather than explanatory models.

Decades of experimentation have shown that there is extensive crosstalk between the individually characterized signaling pathways. Accordingly, the phenomena of epistasis and pleiotropy explained above are naturally extended at the pathway level. While several different pathways can converge to specific phenotypes, one specific pathway and molecular signal can also produce different phenotypes depending on the context (Huang 2000). These observations suggest cross interactions beyond linear cascades. On the other hand, an effect similar to the one "caused" by a specific molecular signal can be produced by nonspecific stimuli or even in a stimulus-independent manner. For example, mechanical stimuli such as those induced by cell shape alterations can induce specific cell phenotypes without any molecular elicitor or genetic change (Huang 2000). On the other hand, given the intrinsic stochasticity of both extra- and intra-cellular biochemical reactions, cells in a lineage-specific manner can assume different and heritable phenotypes either in the absence of an associated genetic or environmental difference or by processing stochastic, nonspecific

environmental cues (Perkins and Swain 2009; Balázsi *et al*. 2011). These facts render a mechanistic explanation by means of the one-to-one GPM at the pathway level untenable, as well. The inevitable plasticity of cell behavior and the robustness of observed phenotypic manifestations call for an alternative explanatory model. We argue below that the formal perspective of cell behavior as an emergent property of the constraints imposed by gene regulatory networks provides an alternative view to how genotypes map unto phenotypes, providing a starting point for addressing otherwise highly complex processes.

Beyond the One-to-One GPM: A Network Dynamics Perspective

How do the two views (gene and signaling pathway to function one-to-one mapping) above stand in post-genomic, systems biology? Genes, encoded proteins, and linear signaling pathways are actually embedded in complex networks of genetic and non-genetic components which generally have various positive and negative feedback loops and dynamical behavior. We focus here on gene regulation, which is the basis for conceptualizing gene interactions, the fundamental property underlying nonlinear, gene regulatory networks. The concept of gene regulation itself, which is nothing new, is not consistent with a one-to-one GPM, because it implies that the phenotypic effect of one gene function will depend on the activity of other genes regulating it. Although explicit awareness of the fact that the genes coding for all the proteins in the cell are necessarily regulated by some other regulatory proteins, which are themselves also regulated, seems overwhelming; such realization can be succinctly represented in qualitative gene regulatory network (GRN) models. These are becoming very useful to follow and understand the concerted action of multiple interacting components.

A common working model in systems biology is that in which the genome is mapped directly to a GRN, and the cellular phenotype is represented by the activity of each of its genes, its expression pattern. Thus in a genotype-phenotype distinction based on GRN dynamics, a network represents effectively the genotype of the cell, while its associated expression profile represents its phenotype (Davila-Velderrain and Álvarez-Buylla 2014). The structure of the genome (and network) remains virtually constant through development while the cellular phenotype changes. Why are phenotypic changes observed through development in such robust and reproducible patterns?

The genomic nature of the GRN implies a physically coded structure, by means of which the network naturally constrains the permissible temporal behavior of the activity of each gene. For example, a specific gene *a* is regulated by a specific set of genes. Given the activity state of these regulators and the functional form of the regulation, each time gene *a* will be channelled to take

specific future states. This simple regulatory rule applies simultaneously to all the genes producing a self-organizing process that would inevitably lead to the establishment of only those cellular states (phenotypes), which are logically consistent with the underlying regulatory logic. Hence, the GRN imposes constraints on the behavior of the cell. The observed robustness and reproducibility of cell behavior emerges naturally as a self-organizing process. Any source of extracellular (non) specific inductive stimulus would inevitably converge to one of the phenotypic states which are logically consistent with the underlying regulatory logic of the network being considered.

The rationale briefly exposed above has been exploited to propose GRNs grounded on experimental data for understanding how cell-fate specification occurs during, for example, early flower development (See Mendoza and Álvarez-Buylla 1998; Espinosa-Soto *et al.* 2004; and an update in Sanchez-Corrales *et al.* 2010), and root stem cell patterning (Azpeitia *et al.* 2010); and it is now supported by a wealth of consolidated theoretical and experimental work (see, for example Huang *et al.* 2005; Azpeitia *et al.* 2014).

Importantly, in contrast to the assumptions implicit in the one-to-one GPM, interactions in the network are fundamental to the establishment of the phenotype, and thus the effect of a mutation on the manifested phenotype will be conditional on the network context of the gene under consideration (Davila-Velderrain *et al.* 2014). Given that the multitude of observed robust cellular phenotypic states would depend on network constraints due to gene regulatory interactions, the orchestrating role of GRNs effectively constitutes a many-to-many (*non-linear)* GPM, in which most components can, at the same time, constitute both causes and effects (Figure 1c).

Blind, Indifferent or market-oriented Biomedical and Biotechnological Research?

Notwithstanding all the evidence produced by almost two decades of post-genomic research, the subliminal presence of the over-simplified one-to-one GPM, although most of the time it is not credited, is undeniable. It is implicitly assumed as a main goal driving mainstream biomedical research that genes cause, for example, cancer; for they cause phenotypes by coding proteins (Huang 2013). This is also the case in biotechnological research, where it is acknowledged that a particular gene from one species in which a particular "function" is produced, can be readily put into another species expecting the same "function" (Vaeck *et al.* 1987). Considering that a myriad of studies search for "causal" mutations, apparently this gene-centric assumption is rarely noticed—or, alternatively, it is just ignored. Despite the huge amount of resources invested in

genome sequencing projects, such thing as a universal (causal) mutation for a degenerative disease has not been successfully identified (Huang 2013). Nevertheless, having specific molecules as candidate causal factors of particular diseases enables companies to develop new drugs for the market. Given the limited nature of the underlying simplistic one-to-one GPM, this approach is likely to fail. It may reproduce only based on its limited effectiveness—and mostly on marketing strategies—instead of deep explanations or much needed solutions. Importantly, such continuing search for potential molecular targets in therapeutics or single-gene golden bullet solutions to complex agricultural threats evidences the prevalence of the one-to-one GPM, i.e., by assuming that there is a protein for every disease or for any environmental challenge in agriculture.

The potential for therapy also complicates matters, for it may be a perfectly acceptable research goal regardless of its impact on improving understanding or on actually proving causation. Thus, it could be the case that biomedical research itself has not naturally evolved to such a naive state; it might be instead that the market driven technocentric character of modern "*science*" happens to stimulate the inheritance of old ideas that continue to be convenient—unfortunately for science, though, the rate of increase in conceptual understanding seems not to be following the fast-paced technological evolution.

To summarize, the prevailing paradigm implicitly assumes that genes determine cell behavior through a one-to-one GPM. Specifically, genes code proteins which directly determine phenotypes, and consequently, mutations in the genes should by themselves alter phenotypes. Therefore, targeting altered proteins produced from mutated genes seems to be the best strategy to "correct" a pathological phenotype—the same can be said of epigenetic alterations, altered pathways or even networks. However, a multitude of post-genomic evidence makes the one-to-one GPM untenable. In contrast, a GPM in terms of the orchestrating role of molecular regulatory networks, which constitutes a many-to-many GPM, naturally explains paradoxical observations and provides a formal framework for the interpretation of ever-growing post-genomic molecular data.

Acknowledgements

This work was supported with ERAB grants: Conacyt (Mexico) 180098 and 180380; and UNAM-DGAPA-PAPIIT: IN203113.

References

Alberch, P. «From genes to phenotype: dynamical systems and evolvability.» *Genetica* 84, nº 1 (1991): 5-11.

Alexander, R. P., G. Fang, J. Rozowsky, M. Snyder and M. B. Gerstein. «Annotating non-coding regions of the genome.» *Nature Reviews Genetics* 11, nº 8 (2010): 559-571.

Azpeitia, E., J. Davila-Velderrain, C. Villarreal and E. Álvarez-Buylla. «Gene regulatory network models for floral organ determination.» *Flower Development* (2014): 441-469.

———, M. Benítez, I. Vega, C. Villarreal and E. Álvarez-Buylla. «Single-cell and coupled GRN models of cell patterning in the Arabidopsis thaliana root stem cell niche.» *BMC systems biology* 4, nº 1 (2010): 134.

Balázsi, G., Van Oudenaarden, A. and J. J. Collins. «Cellular decision making and biological noise: from microbes to mammals.» *Cell* 144, nº 6 (2011): 910-925.

Brunet, T. D. and W. F. Doolittle. «Getting "function" right.» *Proceedings of the National Academy of Sciences* 111, nº 33 (2014): E3365-E3365.

Carninci, P., *et al.* «The transcriptional landscape of the mammalian genome.» *Science* 309, nº 5740 (2005): 1559-1563.

Chatr-aryamontri, A., *et al.* «The BioGRID interaction database: 2013 update.» *Nucleic acids research* 41 nº D1 (2013): D816-D823.

Cheng, J., *et al.* «Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution.» *Science* 308, nº 5725 (2005): 1149-1154.

Costanzo, M., *et al.* «The genetic landscape of a cell.» *Science* 327, nº 5964 (2010): 425-431.

Crick, F. H. «Central dogma of molecular biology.» *Nature* 227, nº 5258 (1970): 561-563.

———. «On protein synthesis.» *Symposia of the Society for Experimental Biology* 12 (1958): 138.

Croft, D., *et al.* «Reactome: a database of reactions, pathways and biological processes.» *Nucleic acids research*, (2010): gkq1018.

Davila-Velderrain, J., A. Servin-Marquez and E. Álvarez-Buylla. «Molecular evolution constraints in the floral organ specification gene regulatory network module across 18 angiosperm genomes.» *Molecular biology and evolution* 31, nº 3 (2014): 560-573.

——— and E. Álvarez-Buylla. «Bridging genotype and phenotype.» In *Frontiers in Ecology, Evolution and Complexity*, edited by Octavio Miramontes, Alfonso Valiente-Banuet and Mariana Benítez. CopIt ArXives, 2014.

Doolittle, W. F., T. D. Brunet, S. Linquist and T. R. Gregory. «Distinguishing between "function" and "effect" in genome biology.» *Genome biology and evolution* 6, nº 5 (2014): 1234-1237.

Espinosa-Soto, C., P. Padilla-Longoria and E. Álvarez-Buylla. «A gene regulatory network model for cell-fate determination during Arabidopsis thaliana

flower development that is robust and recovers experimental gene expression profiles.» *The Plant Cell Online* 16, nº 1 (2004): 2923-2939.

Hancock, J. M. (Ed.). *Phenomics.* CRC Press, 2014.

Ho, W. C. and J. Zhang. «The Genotype-Phenotype Map of Yeast Complex Traits: Basic Parameters and the Role of Natural Selection.» *Molecular biology and evolution* 31, nº 6 (2014): 1568-1580.

Houle, D., D. R. Govindaraju and S. Omholt. «Phenomics: the next challenge.» *Nature Reviews Genetics* 11, nº 12 (2010): 855-866.

Huang, S., G. Eichler, Y. Bar-Yam and D. E. Ingber. «Cell fate sas high-dimensional attractor states of a complex gene regulatory network.» *Physical Review Letters* 94, nº 12 (2005): 128701.

———. «Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells.» *Cancer and Metastasis Reviews* 32, nº 3-4 (2013): 423-448.

———. «Rational drug discovery: what can we learn from regulatory networks?» *Drug discovery today* 7, nº 20 (2002): s163-s169.

———. «Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. Philosophical Transactions of the Royal Society B.» *Biological Sciences* 366, nº 1575 (2011): 2247-2259.

———. «The practical problems of post-genomic biology.» *Nature biotechnology* 18, nº 5 (2000): 471-472.

Huneman, P. *Functions: selection and mechanisms.* Springer, 2013.

Kellis, M., *et al.* «Defining functional DNA elements in the human genome.» *Proceedings of the National Academy of Sciences* 111, nº 17 (2014): 6131-6138.

Kuhn, T. S. *The structure of scientific revolutions.* University of Chicago Press, 2012 [1962].

Lehner, B. «Molecular mechanisms of epistasis within and between genes.» *Trends in Genetics* 27, nº 8 (2011): 323-331.

Lewontin, R. «The genotype/phenotype distinction.» In *Stanford Encyclopedia of Philosophy*. 2011.

Mendoza, L. and E. Álvarez-Buylla. «Dynamics of the genetic regulatory network for arabidopsis thaliana flower morphogenesis.» *Journal of Theoretical Biology* 193, nº 2 (1998): 307-319.

Natoli, G. and J. C. Andrau. «Noncoding transcription at enhancers: general principles and functional models.» *Annual review of genetics* 46 (2012): 1-19.

O'Malley, M. A. and Y. Boucher. «Paradigm change in evolutionary microbiology.» *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 2005: 183-208.

Perkins, T. J. and P. S. Swain. «Strategies for cellular decision-making.» *Molecular systems biology* 5, nº 1 (2009).

Phillips, P. C. «Epistasis-the essential role of gene interactions in the structure and evolution of genetic systems.» *Nature Reviews Genetics* 9, nº 11 (2008): 855-867.

Pigliucci, M. «Genotype-phenotype mapping and the end of the 'genes as blueprint'metaphor.» *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, nº 1540 (2010): 557-566.

Plosky, Brian S. *eRNAs Lure NELF from Paused Polymerases.Molecular Cell.* 2014.

Rose, M. R. and T. H. Oakley. «The new biology: beyond the Modern Synthesis.» *Biology direct* 2, nº 1 (2007): 30.

Sanchez-Corrales, Y. E., E. Álvarez-Buylla and L. Mendoza. «The Arabidopsis thaliana flower organ specification gene regulatory network determines a robust dif-ferentiation process.» *Journal of Theoretical Biology* 264, nº 3 (2010): 971-983.

Schaefer, C. F., *et al.* «PID: the pathway interaction database.» *Nucleic acids research*, 2009: D674-D679.

Shapiro, J. A. «Revisiting the central dogma in the 21st century.» *Annals of the New York Academy of Sciences* 1178, nº 1 (2009): 6-28.

Soyer, O. S. (Ed.). *Evolutionary systems biology* 751 Spring 2012.

Stearns, F. W. «One hundred years of pleiotropy: a retrospective.» *Genetics* 186, nº 3 (2010): 767-773.

Strohman, R. C. «The coming Kuhnian revolution in biology.» *Nature biotechnology* 15, nº 3 (1997): 194-200.

Stuart A., Kauffman. *The origins of order: Self-organization and selection in evolution.* Oxford, UK: Oxford University Press, 1993.

Tyler, A. L., F. W. Asselbergs, S. M. Williams and J. H. Moore. «Shadows of complexity: what biological networks reveal about epistasis and pleiotropy.» *Bioessays* 31, nº 2 (2009): 220-227.

Uhlen, M. and F. Ponten. «Antibody-based proteomics for human tissue profiling.» *Molecular & Cellular Proteomics* 4, nº 4 (2005): 384-393.

Vaeck, M., *et al.*«Transgenic plants protected from insect attack.» *Nature* 328 (1987): 33-37.

Wagner, G. P. and J. Zhang. «The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms.» *Nature Reviews Genetics* 12, nº 3 (2011): 204-213.

Wilkins, A. S. «Are there 'Kuhnian'revolutions in biology?» *BioEssays* (1996): 695-696.