Cognitive functions of metaphor in the natural sciences.

Evelyn Fox Keller

Abstract: Today's genome looks very different from the one with which the science of genetics began. Rather than a set of genes initiating causal chains leading to the formation of traits, it looks far more like an exquisitely sensitive reactive system – a device for regulating the production of specific proteins in response to the constantly changing signals it receives from its environment. True, the signals it detects come most immediately from its intra-cellular environment, but these, in turn, reflect input from the external environments of the cell and of the organism. Humans are especially reactive systems, and they are so on every level at which they are capable of interacting: cultural, interpersonal, cellular, and even genetic. The reconceptualization of the genome that I propose allows us – indeed obliges us – to abandon the dichotomies between genetics and environment, and between nature and culture, that have driven so much fruitless debate, for so many decades. If much of what the genome 'does' is to respond to signals from its environment, then the bifurcation of developmental influences into the categories of genetic and environmental makes no sense. Similarly, if we understand the term *environment* as including cultural dynamics, neither does the division of biological from cultural factors. We have long understood that organisms interact with their environments, that interactions between genetics and environment, between biology and culture, are crucial to making us what we are. What research in genomics seems to show is that, at every level, biology itself is constituted by those interactions – even at the level of genetics.

Keywords: genes; genomes; gene action; reactive systems.

Introduction

By profession, I am not a philosopher, and so something of an outlier in this volume. Which is not to say that the questions that concern me are not philosophical, only that they tend not to neatly align with the questions that preoccupy professional philosophers, even those who focus on the role of metaphor in science. They veer off in different directions. To be sure, this is largely due to the fact that my professional training is *in* the natural sciences, rather than *in* philosophy, and the questions I ask tend (however provocatively) to point in the direction of that early disciplinary training. My interest

ISSN (print) 2281-8618-ETS

in metaphor provides a clear case in point.

Just as for so many others, this interest arose from simple observation of the ubiquity of metaphor in scientific discourse, judging by the apparent impossibility of dispensing with it, its seeming necessity. But necessary for what? In the first instance, for generating knowledge about a world not yet known. How else is one to seek understanding of the new, of the not yet intelligible, if not by comparison with that which is already familiar? But still, what we confront is something new, not quite contained in the world we already know, and for that reason, the analogical relation is not quite sufficient. As Hesse writes, what is lost in the effort to reduce metaphor to simple analogy or paraphrase are "the elements of surprise, tension, and creativity [that] are essential to the metaphor but missing from the paraphrase. Everyone recognizes a distinction between live and dead metaphor: metaphor is interesting only when it is alive – provoking surprise and shock, indicating new thought". (1988: 4). Indeed, the essence of a live metaphor is precisely the juxtaposition of similarity and difference, the manifest untruth of equating source and target. And what gives it its value in the logic of scientific discovery is precisely the instability it generates by virtue of its insistence on both similarity *and* difference, its insistence that, at one and the same time, man both *is* and *is not* a wolf. Lose this duality, and one loses the vitality of the metaphor.

Reading Hesse's 1988 article was important to me because of its emphasis on both the ubiquity and the function of metaphor in scientific thought, and above all, for the legitimacy and importance it conferred on the need to study that function. But beyond that, our questions begin to part ways. What first drew me to the phenomenon of metaphor in science was not its significance for questions about theories of meaning, the philosophy of language, or the architecture of thought, nor even about the irreducibility of metaphor, or the precise relation between metaphor and analogy. Rather, it was the far more mundane question of how the particular metaphors scientists invoke shape our scientific view of the world. If we take the relation between source and target to play an essential role in the construction of scientific knowledge, how would that knowledge be different starting with a different choice of source? Far more powerfully than analogical similarity, metaphoric identity shifts our scientific attention away from differences, stressing instead similarities, between source and target and a number of obvious questions arise: What other features of the system under study might an alternative metaphor focus attention on? Are some features made salient by one metaphor that would lose significance in the framework invited by another? How is our scientific picture of the world shaped by our choices of metaphor? To what other discoveries might different choices have led us? My own thesis is that the central metaphors of classical and molecular genetics contributed greatly to the successes of these

fields, while at the same time also fostering the neglect of crucial questions about embryonic organization and regulation.

These were the sort of questions that initially guided my inquiry into the role of metaphor in science, and I found no shortage of examples to pursue. Three that loomed especially large in my early inquiry were: Nature as "red in tooth and claw"; "Science as "a chaste and lawful marriage between Mind and Nature"; DNA as the Master Molecule of Life; and so on. When new, each of these metaphors was undoubtedly both evocative and creative, and indeed provided crucial incentive for new lines of research that attempted to both illustrate and explicate the aggressive, competitive aspects of ecology, the passivity of Nature, the causal powers of DNA, etc. Of course, the reigning metaphors of a discipline would also, and inevitably, draw attention away from such phenomena and/or dynamics that suggested different characterizations.

Over time, repetition lent ever greater familiarity to these metaphors --indeed, to the point of robbing them of their liveliness, often even of their very metaphoricity. Over time, and again, inevitably, it would become more and more difficult (even unnatural) to attend to those aspects of species interaction that derive from cooperation and mutual dependence; to those features of scientific inquiry (or of the relation between Mind and Nature) that require an inversion of conventional gender metaphors; or to those attributes of DNA more suggestive of a role as servant of life than as its master. The early metaphors may have lost their initial vitality, but in the wake of that vitality, they left a conceptual framework that continued to shape how scientists actually perceived their objects of study.

Part I: Metaphor in the making of classical genetics:

In *Making Sense of Life,* I sought more specifically to understand the ways in which genes were used to craft explanations of biological development during the 20th century. Here I argued that "much of the theoretical work involved in constructing explanations of development from genetic data is linguistic – more specifically, that it depends on productive use of the cognitive tensions generated by multiple meanings, ambiguity, and more generally, by the introduction of novel metaphors" (Keller 2002: 117). Indeed, I claimed that the very notion of 'gene' was built out of a marriage of two distinct (and not obviously reconcilable) metaphors: first, it was taken to be the "atom" of life, a particulate element that would serve for biology much as the atom had done for physics and chemistry, and second, as an agent capable of directing the formation of

particular traits. As Hugo de Vries wrote in 1889, "Just as physics and chemistry are based on molecules and atoms, even so, the biological sciences must penetrate to these units in order to explain by their combinations the phenomena of the living world". At the same time, such a unit must also be able, as he put it, to "impress its character upon the cell" (de Vries [1889] 1910: 194). If it is either to 'represent' the properties of an adult organism or cause their coming into being, it must obviously be something larger and more complex than a chemical molecule, much less an atom. "These minute granules," he concluded, "are more correctly to be compared with the smallest known organism" (de Vries [1889] 1910: 4). De Vries had referred to these units as *pangens*; Mendel as *elemente*, Weismann as *determinants*, but in each case, a single term was invoked to bridge the manifestly disparate functions the unit would have to serve. Finally, in 1909, Johannsen introduced the *gene*, and from that time on, this was the term that prevailed. Of course, no one knew what a gene was, but it was generally taken for granted that it must be able to both endure (to guarantee intergenerational stability) and to guide individual development in a single generation; simultaneously immortal and agentic.

Just what kind of entity might serve this dual role remained a critical issue (especially difficult was the question of how a gene might 'determine' a phenotypic trait). In the absence of a research program that might lead to an answer, some kind of narrative would clearly have to suffice – a narrative, i.e., that would somehow bracket that question while at the same time providing a framework within which other (perhaps more readily tractable) questions could be productively pursued. In this sense, I suggest that narratives can perform a function similar to that of Hacking's "styles of reasoning": Not themselves candidates for truth or falsity, they nonetheless establish the means by which certain kinds of statements can become candidates for truth or falsehood. One might say that such narratives provide a way of endowing questions with scientific sense, of formulating questions in ways that make them answerable. The cognitive function they serve is thus not so much that of guiding scientific research towards closer "correspondence" with nature, but rather of establishing the criteria according to which scientific truth claims are to be evaluated.

I can think of no better illustration than that provided by Alfred Sturtevant In his classic paper on "the developmental effects of genes," presented at the 1932 International Congress of Genetics. Here Sturtevant explained that

> One of the central problems of biology is that of differentiation – how does an egg develop into a complex many-celled organism? That is, of course, the traditional major problem of embryology; but it also appears in genetics in the form of the question, "How do genes produce their effects?" (304).

Moreover, he claimed that

> in most cases there is a chain of reactions between the direct activity of a gene and the end-product that the geneticist deals with as a character"; genetic experiments can therefore approach the problem through the "analysis of certain chains of reactions into their individual links (307).

In this reformulation, Sturtevant was invoking a narrative, inspired by Johannsen's powerful little word, that was rapidly coming to dominate the thinking of classical geneticists. Elsewhere (Keller 1995; 2000; 2002), I have referred to this narrative as "the discourse of gene action," arguing that it provided a spatial and temporal map of biological development that had a clearly identifiable effect on the course of research. This map defined the onset of development (T=0) as the moment of fertilization, and the structure causally responsible as the nucleus. As such, it not only guided the ways in which geneticists posed (and researched) questions about development, but it also effectively barred scientific sense-making from the ways in which embryologists had traditionally formulated their own questions about development. Almost certainly, its most conspicuous effect was on the sidelining of questions about regulation. By 1934, T. H. Morgan – himself an early advocate of the notion of *gene action* – was already having second thoughts, and he offered alternative framing of the problem of development. He wrote:

> The implication in most genetic interpretation is that all the genes are acting all the time in the same way. This would leave unexplained why some cells of the embryo develop in one way, some in another, if the genes are the only agents in the results. An alternative view would be to assume that different batteries of genes come into action as development proceeds. […] The idea that different sets of genes come into action at different times [… requires that] some reason be given for the time relation of their unfolding. The following suggestion may meet the objections. It is known that the protoplasm of different parts of the egg is somewhat different […] and the initial differences may be supposed to affect the activity of the genes. The genes will then in turn affect the protoplasm […]. In this way we can picture the gradual elaboration and differentiation of the various regions of the embryo (Morgan 1934: 9).

As it happened, Morgan's critique went largely unheeded, as did his proposed reformulation. For the next 40 years, the field of genetics grew steadily in size and influence, and research in that discipline continued to make remarkable strides. But relatively little of this work shed light on the question of how genes actually contributed to, guided, or shaped biological development. Not surprisingly, the field of embryology fared rather less well over this period, suffering a steady decline in funding, in institutional support, and in the recruitment of new blood. By the time of the molecular revolution, its own

central questions, far from having been elucidated, had largely faded from attention. Moreover, their absence was scarcely noticed. The discourse of gene action had cast an aura of effective invisibility over the problems that Morgan had attempted to explode.

Global Narratives

The science of developmental genetics is hardly unique in its reliance on global narratives. As I argued in *Making Sense of Life,* many sciences have a clearly recognizable need for

> global narratives that guide scientific research in ways that do not directly depend on either "test-and-feedback" or verification – for narratives that make productive use of the imprecision of metaphor and other linguistic tropes not so much as a way of guiding us toward a more precise and literal description of phenomena, be they local or global, but rather as a way of providing explanatory satisfaction where not available either from descriptions of local phenomena or from general laws. ... Just because we have no access to rules or laws describing the role of genes in development, gene-based accounts of these processes need to make use of the associations generated by metaphor, by ambiguity, and by the dynamic interplay among the different meanings a given term may connote. For these purposes, even the absence of a clear and univocal meaning of the concept of the gene, i.e. of the basic explanatory unit, can be a positive resource for drawing different experimental systems and different research programs into a coherent scientific agenda (Keller 2002: 120).

Indeed, one of the main points I wanted to stress in that account was the essential role played by the conceptual fluidity and versatility allowed by the lack of a clear and precise definition of the constituent terms – especially, in this case, in the definition of the gene.

In line with this argument, I claimed more specifically that, because molecular genetics brought with it (at least initially) a concrete and well-defined reference for the term *gene* – namely, a specific region of a molecule of double-stranded DNA that "coded" for a protein – the figure of *gene action* could no longer serve the particular function that had depended so critically on uncertainty and fluidity in the definition of the classical gene: in particular, it could no longer provide the illusion of explanation; it could no longer bridge the gap left by the absence of a scientific account of development. Geneticists required a new narrative, and the figure of speech introduced by molecular biology, the *genetic program*, seemed to take over much of the work that had earlier been done by *gene action*. Or so, at least, I argued in 2002.

Since then I have come to reconsider. In fact, the discourse of gene action did not in fact disappear quite so precipitously. Although seemingly superced-

ed by talk of programs and information, I now realize that it continued (even if less conspicuously) to play a critical role in shaping the new molecular narratives. Furthermore, an examination of this role brings to light a somewhat different cognitive role for metaphor (and other forms of linguistic imprecision) in the natural sciences – a conservative rather than innovative role, stabilizing rather than destabilizing.

In particular, examination reveals a pervasive duality pervading the entire discourse of Molecular Biology, self-reinforcing and self-stabilizing through the network of meanings upon which that discourse depends. The impact of Molecular Biology was revolutionary, but it failed to disrupt the essential structure of the discourse of gene action on which classical genetics had previously been built. That disruption seems to have become possible only with the startling findings of Genomics. I turn, therefore, to Part Two of this essay:

Part II: From Gene Action to Reactive Genomes:

Last year, Philip Ball, a former editor of Nature, published a commentary celebrating the current state of (or, as he emphasizes, current gaps in) our understanding of DNA. Indeed, his title, "Celebrate the unknowns", suggests that it is above all the gaps which he wants us to celebrate. As he wrote:

> This week's diamond jubilee of the discovery of DNA's molecular structure rightly celebrates how Francis Crick, James Watson and their collaborators launched the 'genomic age' by revealing how hereditary information is encoded in the double helix. Yet the conventional narrative ... is as misleading as the popular narrative of gene function itself, in which the DNA sequence is translated into proteins and ultimately into an organism's observable characteristics, or phenotype (Ball 1980).

A bit later, he added:

> A student referring to textbook discussions of genetics and evolution could be forgiven for thinking that the 'central dogma' devised by Crick and others in the 1960s – in which information flows in a linear, traceable fashion from DNA sequence to messenger RNA to protein, to manifest finally as phenotype – remains the solid foundation of the genomic revolution. In fact, it is beginning to look more like a casualty of it (*Ibid*.).

In other words, we celebrate the Watson-Crick revelation of "how hereditary information is encoded in the double helix" while at the same time admitting the utterly misleading nature of the "conventional narrative" of their discovery – a narrative "as misleading as the popular narrative of gene function

itself, in which the DNA sequence is translated into proteins and ultimately into an organism's observable characteristics, or phenotype".

But what exactly is it that is so misleading? Ball actually gives us two narratives – one he refers to as the conventional, the other as the popular narrative; one is a claim about hereditary information, the other a claim about the genetic code. These are not the same, but does he actually mean to distinguish them? Unclear. And is the realization that it/they are misleading a new realization? Not really. Then how is it that such "misleading" narratives are so routinely perpetuated in the teaching of Molecular Biology, indeed in so much of the technical, the lay, and even the philosophical literature?

Part of the answer to this question is to be found in the replication of this ambiguity throughout the discourse of Molecular Biology (preceded by a parallel set of ambiguities in the discourse of classical genetics) that has worked for 60 years to simultaneously sustain and obscure what Ball now sees as misleading.

Let me begin with the two narratives that Ball invokes:

a) DNA sequence codes for proteins which ultimately form an organism's observable characteristics, or phenotype. In popular lingo, DNA makes RNA, RNA makes proteins, and proteins make us, and,

b) hereditary information is encoded in the double helix,

Accordingly, we can conclude that hereditary information is encoded in the sequence of nucleotides in the DNA.The concept of code figures crucially in both, but in fact, in two quite different senses of the term. In one (the first), the meaning of *code* (or encode) is quite clear. It derives from telegraphy and cryptography and is in fact the first definition given by the Oxford English Dictionary: to encode is "to translate into cipher or code; to express information by means of a *code*. Colloquially, 'to code'". As in the Morse code. Indeed, Crick himself was explicit about this being the sense in which he used the term code in his sequence hypothesis. The genetic code referred to the process of translation from a text written in nucleotide sequences to one written in amino acid sequences. (Incidentally, he was also careful to distinguish the sequence hypothesis from what he called the central dogma (the hypothesis that "Once information has got into a protein it can't get out again")).

The deciphering of the genetic code was a tremendous achievement in the history of biology and it well deserves to be celebrated. Furthermore, it has held up remarkably well. Perhaps more surprisingly, so too has the central dogma, at least as Crick understood it. Moreover, there is nothing to suggest anything misleading in either claim. What is it then that *has* the reader led astray?

The difficulty to which Ball refers arises when people speak of the heredi-

tary information encoded in the double helix, for, in this formulation, another quite different sense of encode is commonly invoked, namely the information required not for a set of proteins, but for an organism. More in the sense of Schroedinger's *codescript*, a notion that preceded Crick's concept of code by 15 years – i.e., in the sense, as Schroedinger himself wrote, that

> Every complete set of chromosomes contains the full code; so there are, as a rule, two copies of the latter in the fertilized egg cell, which forms the earliest stage of the future individual. In calling the structure of the chromosome fibres a code-script we mean that the all-penetrating mind, once conceived by Laplace, to which every causal connection lay immediately open, could tell from their structure whether the egg would develop, under suitable conditions, into a black cock or into a speckled hen, into a fly or a maize plant, a rhododendron, a beetle, a mouse or a woman (1944: 21).

But, as Schroedinger also acknowledged, "the term code-script is, of course, too narrow. The chromosome structures are at the same time instrumental in bringing about the development they foreshadow. They are law-code and executive power – or, to use another simile, they are architect's plan and builder's craft-in one" (*Ibid*.: 22).

In other words, unlike that of the Morse code, Schroedinger could not say what a code-script *is*. Indeed, its meaning had to remain open-ended, characterized not by what it is but by what it is expected to do, by the answer it is hoped to provide – but which, for now, could only be indicated by a cloud of evocative analogy.[1] The question that arises is this: are these two different narratives, one referring to hereditary information and the other to sequence information, one to the hereditary code-script and other to the genetic code, or are they two different versions of the same narrative, linked by a common vocabulary?

I suggest that we have here two distinct narratives, both of which have played crucial (perhaps even essential) roles in molecular biology, and that what Ball characterizes as misleading results from the collapse of these two narratives into one. As illustrated below, this collapse is both effected and sustained by a pervasive duality running through the discourse of Molecular Biology as a whole – a duality that misleads by obscuring the central problem that had *not* been addressed, namely that of accounting for the organization of proteins into a functioning organism.

But this duality is not limited to the meanings of code. Indeed, it recurs in

[1] In this sense, code-script is not unlike another key term of molecular discourse, namely, that of *program* – especially as François Jacob introduced that term when he described the organism as "the realization of a programme prescribed by its heredity", and in his subsequent claim that "when heredity is described as a coded program..., the paradox [of development] disappears", or, "the concept of programme has made an honest woman of teleology".

virtually all the new terms imported into biology with the molecular revolution: think, e.g., of genetic information, genetic instructions, genetic blueprint, etc.. Not surprisingly, the same elisions also recur in the links commonly assumed between genes, information, codes and code-script. Take, e.g., the definition of code-script as it appeared in *New Biology* in 1952: "All the genes together constitute the 'code-script' of the cell; that is, the complete set of controlling factors which carry in condensed form all the 'information' necessary for an organism to develop the distinctive properties of its species". Or take Ball's own reference to Crick's 'central dogma' as the thesis that "information flows in a linear, traceable fashion from DNA sequence to messenger RNA to protein, to manifest finally as phenotype", conflating Crick's own version of the central dogma as the thesis that «Once information has got into a protein it can't get out again» (DNA → RNA → protein) with the colloquial version of DNA makes RNA, RNA makes proteins, and proteins make us (DNA → RNA → Protein→ Phenotype). The difference, namely the link between a list of proteins and the organization of these proteins into a functioning organism, is of course crucial, and the locus of much of the most critical slippage. For example, is a genome the full complement of an organism's genes or of its DNA? Is Genetics the study of genes or the study of heredity? Is the genetic code the mechanism for translating nucleotide sequence to amino acid sequence or to phenotype? Does the Central Dogma refer to the information in proteins or in phenotype? Similarly, does "genetic information" refer to the sequences coding for proteins or to all DNA sequences?

Each of these questions stems from a collapse of meanings, from an elision between one, concrete, meaning, and another open-ended and ambiguous. Such elision invites the illusion that the ambiguity of the open-ended term has been resolved, and by implication, that the gap between actual achievement and promise has been closed. The fact remains however that, despite the phenomenal progress Molecular Biology has made since the cracking of the genetic code, we remain to this day without an adequate account of the organization of proteins into an organism.

Two points seem worth noting: First is that these elisions are not casual but systematic. Second is their simultaneous transparency and opacity. Once identified, they seem (or ought to seem) crystal clear, plain for anyone to see; recognition depends neither on special expertise nor on new findings. Yet this style or habit of chronic slippage from one set of meanings to the other has prevailed for over 50 years; it has become so deeply ensconced as to have been effectively invisible to most readers of the biological literature. This feature might qualify it as a Foucauldian discourse – by which I mean a discourse that operates by

historically specific rules of exclusion, a discourse that is constituted by what can be said and thought, by what remains unsaid and unthought, and by who can speak, when, and with what authority.

For me, the questions of primary interest are: (1) How have these elisions affected the research trajectory of Molecular Biology? And (2) What makes it possible now for Ball to write that "the conventional narrative […] is as misleading as the popular narrative of gene function itself"? Has Ball in fact escaped the confines of the prevailing discourse? And if he has, what has made it possible to do so? To address these questions, I focus on the tacit equation that underlies the very definition of genetics, namely, the equation of the totality of an organism's genes with its genomes, its chromosomes, and its genetics – an equation, inherited directly from the earlier language of classical genetics, that has been a staple of the discourse of Molecular Biology since its beginning.

In the early days of genetics, there would have been no obvious reason to question the understanding of the study of genetics as the study of genes, and indeed, the frequent elision between genes and mutations in that literature suggests the expectation that there was no other chromosomal locus in which heritable mutations could arise. But from the 1970's on, especially as the focus of molecular genetics shifted to the study of eukaryotic organisms, and as the study of regulation assumed increasing centrality to that science, the relation between genes and genetics has become far less straightforward. To the extent that regulation is a property of DNA, it is surely genetic, but the question arises, is regulation always attributable to genes?[2] Clearly, the answer to this question depends on the meaning attributed to the word *gene*, so let me rephrase, taking genes in their most commonly invoked sense, and ask: Is regulation always attributable to protein-coding sequences?

A related challenge to the equivalence between genes and genetic material came from a series of discoveries of substantial expanses of non protein-coding ("non-genic" or "extra") DNA sequences in eukaryotic genomes. Of particular importance were the discoveries of (1) large amounts of repetitive DNA in 1968, and later, of transposable elements; (2) the wildly varying relationship between the amount of DNA in an organism and its complexity (the "C-value paradox", Thomas 1971); and (3) split genes (protein-coding sequences interrupted by non-coding "introns" (identified in 1977). However, that challenge was soon blunted by the designation of such DNA as "junk" (Ohno 1972). After 1980, with the appearance of two extremely influential papers published back-to-back in *Nature* (Doolittle and Sapienza 1980; Orgel and Crick 1980), the idea of "junk DNA" seemed to become entrenched.

[2] The operon model of Monod and Jacob assumed that it was.

Borrowing Dawkins' notion of selfish DNA, Orgel and Crick were explicit about their intention to use that term

> in a wider sense, so that it can refer not only to obviously repetitive DNA but also to certain other DNA sequences which appear to have little or no function, such as much of the DNA in the introns of genes and parts of the DNA sequences between genes. […] The conviction has been growing that much of this extra DNA is 'junk', in other words, that it has little specificity and conveys little or no selective advantage to the organism […] (*Ibid.*: 604).

Until the early 1990's, the assumption that the large amounts of non-coding DNA found in eukaryotic organisms had "little or no function," that it contributed nothing to their phenotype and could therefore be ignored, remained relatively uncontested. For all practical purposes, genomes (or at least the interesting parts of genomes) could still be thought of as collections of genes. Indeed, when the Human Genome Project first announced its intention to sequence the entire human genome, much of the opposition to that proposal was premised on this assumption. Thus, e.g., Bernard Davis of the Harvard Medical School complained that "blind sequencing of the genome can also lead to the discovery of new genes […], but this would not be an efficient process. On average, it would be necessary to plow through 1 to 2 million 'junk' bases before encountering an interesting sequence" (Davis 1990). And in a similar vein, Robert Weinberg of MIT argued,

> The sticky issue arises at the next stage of the project, its second goal, which will involve determining the entire DNA sequence of our own genome and those of several others. Here one might indeed raise questions, as it is not obvious how useful most of this information will be to anyone. This issue arises because upwards of 95% of our genome contains sequence blocks that seem to carry little if any biological information. […] In large part, this vast genetic desert holds little promise of yielding many gems. As more and more genes are isolated and sequenced, the arguments that this junk DNA will yield great surprises become less and less persuasive (Weinberg, 1991).

Weinberg's assumptions were widely shared in the molecular biology community, and inevitably, they had consequences. Take, e.g., work in medical genetics. For decades, it has been commonplace for medical geneticists to regard the significance of mutations in non-coding DNA solely in terms of their value in locating the main actors of interest, i.e., the genes considered responsible for disease, ignoring any regulatory role they might play even when far removed from protein-coding sequences. Thus, e.g., the goal of the International HapMap Project (launched in 2003) is officially described as follows: "to develop a public resource that will help researchers find genes that are associ-

ated with human health and disease" (The International HapMap Consortium 2004: 468; see also The International HapMap Consortium 2003). This confounding of all "DNA sequence variation" with "genes" that are associated with human health and disease recurs throughout the official discussions, and indeed, is reinforced by the explanations of the methodology of the project.

These explanations make use of the term "allele", routinely defined as "an alternative form of a gene that is located at a specific position on a specific chromosome" (Biology-Online), to help consolidate the equation between single nucleotide polymorphisms (SNP's) and genes. For example, on the home page of the HapMap website (http://hapmap.ncbi.nlm.nih.gov/abouthapmap.html), we learn that "Sites in the genome where the DNA sequences of many individuals differ by a single base are called single nucleotide polymorphisms (SNPs). Some people may have a chromosome with an A at a particular site where others have a chromosome with a G. Each form is called an allele". And just in case the message has still not gotten through, the website continues:

> For geneticists, SNPs act as markers to locate genes in DNA sequences. Say that a spelling change in a gene increases the risk of suffering from high blood pressure, but researchers do not know where in our chromosomes that gene is located. They could compare the SNPs in people who have high blood pressure with the SNPs of people who do not. If a particular SNP is more common among people with hypertension, that SNP could be used as a pointer to locate and identify the gene involved in the disease (http://snp.cshl.org/abouthapmap.html).

A similar story can also be told about the neglect of non-coding (or non-genic) DNA in molecular evolution. For reasons partly technical and partly conceptual, work of molecular evolutionary biologists has traditionally focused on changes in the protein coding sequences of DNA, with conclusions based on the assumption that such sequences can be taken as a stand in for the entire genome. But such stories of neglect – in medical genetics, of the medical implications of non-genic DNA, in Evolutionary Biology, of its evolutionary implications – can be told now only because the assumptions on which they were based have now begun to be noticed, and accordingly, to be challenged. So what happened that made this possible (that made Ball's article possible)?

The launching of the Human Genome Project (HGP) in 1990 was almost certainly the most significant moment in the history of our understanding of the relations between genes and *genomes*. With the rise of genomics our view of the genome as simply a collection of genes has all but collapsed.

Of particular shock value were the discoveries, first, of how few genes the human genome contained, and (2) of how small a portion of the genome's structure is devoted to protein coding sequences. In a review article published

in 2004, John Mattick published a graph displaying the ratio of non-coding to total genomic DNA as an increasing function of developmental complexity (2004: 317). For humans, he estimated that ratio as 98.5%; since then, estimates have increased to 99%. The obvious question is, what is all that non-coding DNA for? Can it possibly all be junk?

The notion of junk DNA handily accommodates the classical view of genomes as collections of genes (be they trait forming or protein-coding entities). But the recent accumulation of genomic data has brought that accommodation to a breaking point. In 2003, a new metaphor came into use, one that has by now largely replaced the older one (Gibbs 2003). Instead of "junk", non-genic DNA has become "the dark matter of the genome".

This was also the year in which the research consortium ENCODE (ENCyclopedia Of DNA Elements) was formed, charged with the task of identifying all the functional elements in the human genome. Early results of that effort (based on the analysis of 1% of the genome) were reported in *Nature* in 2007, and they effectively put the kibosh on the hypothesis that non-coding DNA lacked organismic function. They confirmed that the human genome is "pervasively transcribed" even where non-coding; that regulatory sequences of the resulting non-coding RNA (ncRNA) may overlap protein coding sequences, or that they may be far removed from coding sequences; and finally, that non-coding sequences are often strongly conserved under evolution. Furthermore, they showed not only that non-coding DNA is extensively transcribed, but also that the transcripts are involved in forms and levels of genetic regulation that had heretofore been unsuspected.

In his summary of the latest findings of ENCODE, Mark Blaxter identifies three interacting systems that coordinate gene expression in space and time:

> transcription factors that bind to DNA in promoters of genes, ncRNA that modifies gene expression posttranslationally, and marking of the histone proteins on which the DNA is wound with chemical tags to define regions of the genome that are active or silent (2010: 1758).

Of particular interest is the strong correlation between chromatin marks and gene expression and the connectivity between and among the different regulatory systems that are now being studied (correlations and connectivity themselves mediated by ncRNA).

*

The take home message would seem to be clear. Genetics is not just about genes and what they code for. It is also about how the DNA sequences that give rise to proteins are transcribed, spliced, and translated into amino acid

sequences, in the appropriate amounts at the appropriate time and place; about how these, once assembled into proteins, navigate or are transported to the sites where, and when, they are needed, and much more. All of this requires coordination of an order of complexity only now beginning to be appreciated. And it is now evident that the ncRNA transcripts of the remaining 98-99% of the genome are central to this process. These transcripts come in many sizes and are associated with a number of different mechanisms. Small RNA's can destabilize messenger RNA, influence the formation of chromatin and chromatin marks, and have even been linked to cancer. Now another class of ncRNA transcripts has been identified – "large intervening noncoding RNAs" (or "lincRNAs") – that can operate across long distances and may prove as important to cell function as protein-coding sequences (Pennisi 2010). It is now clear that ncRNA's are crucial to the regulation of transcription, alternative splicing, chromosome dynamics, epigenetic memory, and more. They are even implicated in the editing of other RNA transcripts, and of modulating the configuration of the regulatory networks these transcripts form. In short, they provide the means by which gene expression can respond to both immediate and longer range environmental context and adapt appropriately.

Behavioral (or physiological) adaptation does not require the direct alteration of DNA sequences: environmental signals trigger a wide range of signal transduction cascades that routinely lead to short-term adaptation. Moreover, by lending to such adaptations the possibility of inter-generational transmission, epigenetic memory works to extend short-term to long-term adaptation. As Mattick explains, "the ability to edit RNA […] suggests that not only proteins but also – and perhaps more importantly – regulatory sequences can be modulated in response to external signals and that this information may feed back via RNA-directed chromatin modifications into epigenetic memory" (2010: 551).

Finally, environmental signals are not restricted to the simple physical and chemical stimuli that directly impinge: organisms with central nervous systems have receptors for forms of perception that are both more complex and longer range. Humans have especially sophisticated perceptual capacities, enabling them to respond to a wide range of complex visual, auditory, linguistic, and behavioral/emotional signals in their extended environment. Recent research has begun to show that responses to such fundamentally social signals also extend way down to the level of gene expression (see, e.g., the work Steven Cole and his colleagues 2007; 2009).

*

The gap between a collection of protein-coding sequences and the full complement of genetic material (or DNA) of an organism is not only huge, but is

also promising to be of immense importance. There is of course debate about just how important; in particular *Encode's* attribution of functionality to virtually *all* transcribed sequences has been hotly disputed. But I doubt that anyone today would claim that non-coding DNA is without function; the question under debate is more one of how much and what kinds of function can be tied to ncDNA. Nor is there consensus about the implications of this shift. A relatively conservative response is simply to rename all transcribed sequences of the DNA as genes, attempting thereby to hold onto the view of these entities (and hence of genomes) as effectively autonomous formal agents, containing within themselves the blueprint for an organism's life – i.e., all of the biological information needed to build and maintain a living organism.

The important point is that current research allows for – even demands – a more radical change in our view of the nature of the genome, and in good part, it does so by obliging us to focus attention on features that have been missing from our conceptual framework. In addition to providing information required for building and maintaining an organism, the genome also provides a vast amount of information enabling it to adapt and respond to the environment in which it finds itself. Fortunately so, for without such responsivity how is an organism to develop and maintain itself in the face of environmental vicissitudes? Another problematic notion often employed in the early days of Molecular Biology was that of a genetic program-- an image that, like the code-script, though inspired by computer programs, necessarily remained ill-defined. That notion attracted a great deal of criticism from philosophers of biology, but I think for the wrong reason. Like most biologists, the philosophical critics tended not to think of programs as conditional if-then structures, structures built to respond, but rather like a set script – more like a concert program than a computer program. For those who did think of it as a computer program, the image was more one of a "when - then" than an "if - then" – allowing the genome to respond to its internal temporal sequence but still not to the external world. In fact, if we keep the conditional nature of the program metaphor at the fore, we might even resurrect it to help us think about how Natural Selection has shaped the evolution of a genome that is "programmed" to respond and adapt.[3]

[3] As Prof. van der Meer has suggested (personal communication), "Another way to characterize the radical change would be to view the genome as organized in space and time […] In other words, the genome has almost come to be seen as an organism itself". His comment reminds me of the remarkably prescient remarks of Barbara McClintock in her Nobel acceptance speech where she described the genome "as a highly sensitive organ of the cell that monitors genomic activities and corrects common errors, senses unusual and unexpected events, and responds to them, often by restructuring the genome" (McClintock 1984).

Today's genome looks very different from the one with which the science of genetics began. Rather than a set of genes initiating causal chains leading to the formation of traits, it looks far more like an exquisitely sensitive reactive system – a device for regulating the production of specific proteins in response to the constantly changing signals it receives from its environment. True, the signals it detects come most immediately from its intra-cellular environment, but these, in turn, reflect input from the external environments of the cell and of the organism.

Humans are especially reactive systems, and they are so on every level at which they are capable of interacting: cultural, interpersonal, cellular, and even genetic. The reconceptualization of the genome that I propose allows us – indeed obliges us – to abandon the dichotomies between genetics and environment, and between nature and culture, that have driven so much fruitless debate, for so many decades. If much of what the genome 'does' is to respond to signals from its environment, then the bifurcation of developmental influences into the categories of genetic and environmental makes no sense. Similarly, if we understand the term *environment* as including cultural dynamics, neither does the division of biological from cultural factors. We have long understood that organisms interact with their environments, that interactions between genetics and environment, between biology and culture, are crucial to making us what we are. What research in genomics seems to show is that, at every level, biology itself is constituted by those interactions – even at the level of genetics.

But this reformulation leaves us with an obvious question: if the genome is so responsive to its environment, how is it that the developmental process is as reliable as it is? This is a question of major importance in biology, and it is rapidly becoming evident that the answer must be sought not only in the structural (sequence) stability of the genome, but also in the relative constancy of the environmental inputs, and, most importantly, in the dynamic stability of the system as a whole (see, e.g., Keller 2000). Genomes are responsive, but far from infinitely so; the range of possible responses is severely constrained, both by the organizational dynamics of the system in which they are embedded and by their own structure.

It is good that Ball calls on us to "celebrate the unknowns", but if, as I claim, recent work in genomics has finally disrupted the narratives of developmental genetics that have prevailed for over a century, geneticists will now need a new narrative to help guide them through the thickets that lie before them. I have suggested that thinking of genomes as "reactive systems" might be a good place to start, but only time and the creative elaboration of such a framework will tell. One thing however seems certain: scientists working

in this field have an enormous – and immensely exciting – challenge before them. I only wish I was just starting rather than ending my career.

References

Ball, Philip, 2013, "DNA: Celebrate the unknowns", in *Nature*, 496: 419-420.

Blaxter, Mark, 2010, "Revealing the Dark Matter of the Genome", in *Science*, 330: 1758-1759.

Cole, Steve W, 2009, "Social Regulation of Human Gene Expression", in *Current Directions in Psychological Science*, 18, 3,: 132-137.

Cole, Steve W., *et al.*, 2007, "Social regulation of gene expression in human leukocytes", in *Genome biology*, 8, 9: R189.

Davis, Bernard D., 1990, "The Human Genome and Other Initiatives", in *Science*, 249: 342-343.

De Vries, Hugo, 1889 [1910], *Intracellular Pangenesis*, Open Court, London.

Doolittle, W. Ford and Carmen Sapienza, 1980, "Selfish genes, the phenotype paradigm and genome evolution", in *Nature*, 284 (5757): 601-603.

Gibbs, W. Wayt, 2003, "The Unseen genome: Gems among the Junk", in *Scientific American*, Nov. Scientific American, 289: 46-53.

Hesse, Mary B., 1988, "The Cognitive Claims of Metaphor", in *Journal of Speculative Philosophy*, 2, 1:1-16.

Johannsen, Wilhelm, 1923, "Some Remarks about Units in Heredity", in *Hereditas*, 4: 133-141.

Keller, Evelyn F., 1995, *Refiguring Life: Metaphors of Twentieth Century Biology*, Columbia University Press.

Keller, Evelyn F., 2000, *The Century of the Gene*, Harvard University Press, Cambridge.

Keller, Evelyn F., 2002, *Making Sense of Life,* Harvard University Press, Cambridge.

Mattick, John S., 2004, "RNA regulation: a new genetics?", in *Nature Reviews Genetics*, 5, 316-323.

Mattick, John S., 2010, "RNA as the substrate for epigenome-environment interactions: RNA guidance of epigenetic processes and the expansion of RNA editing in animals underpins development, phenotypic plasticity, learning, and cognition", in *Bioessays*, 32: 548-552.

McClintock, Barbara, 1984, "The Significance of Responses of the Genome to Challenge", in *Science*, 226: 792-801.

Monod, Jacques, Jacob François, 1961, "Teleonomic mechanisms in cellular metabolism, growth, and differentiation", in *Cold Spring Harbor symposia on quantitative biology*, 26: 389-401.

Morgan ,Thomas H., 1934, *Embryology and Genetics*, Columbia University Press, New York.

Orgel, Leslie E. and Francis H. Crick, 1980, "Selfish DNA: the ultimate parasite", in *Nature*, 284(5757): 604-607.

The International HapMap Consortium, 2003, "The International HapMap Project", in *Nature*, 426: 789-794.

Pennisi, Elizabeth, 2001, "The Human Genome", in *Science* 291, 5507: 1177-1180.

Pennisi, Elizabeth, 2003, "A Low Number Wins the GeneSweep Pool", in *Science*, 300, 5625: 1484-1485.

Pennisi, Elizabeth, 2010, "Shining a light on the genome's 'dark matter'", in *Science*, 330: 1614.

Schroedinger, Erwin, 1944, *What is Life?*, Cambridge University Press, Cambridge.

The ENCODE Project Consortium, 2007, "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project", in *Nature*, 447: 799-816.

The International HapMap Consortium, 2004, "Integrating ethics and science in the International HapMap Project", in *Nature Reviews Genetics*, 5: 467-475.

Weinberg, Robert A., 1991, "There are two large questions", in *Debate*, 5: 78.

Evelyn Fox Keller
Masachussetts Institute of Technology
efkeller@mit.edu