The principle of recursive genome function.

Pellionisz AJ.

The Principle of Recursive Genome Function

Andras J. Pellionisz

Abstract Responding to an open request, the principle of recursive genome function (PRGF) is put forward, effectively reversing two axioms of genomics as we used to know it, prior to the Encyclopedia of DNA Elements Project (ENCODE). The PRGF is based on the reversal of the interlocking but demonstrably invalid central dogma and "Junk DNA" conjectures that slowed down the advance of sound theory of genome function, as far as information science is concerned, for half a century. PRGF illustrates the utility of the class of recursive algorithms as the intrinsic mathematics of post-ENCODE genomics. A specific recursive algorithmic approach to PRGF governing the growth of the Purkinje neuron is sketched, building the structure in a hierarchical manner, starting from primary genomic information packets and in each recursion using auxiliary genomic information packets, cancelled upon perusal. The predictive power of the principle and its experimental support are indicated. It is argued that genomics is no longer an exceptional instance of the applicability of recursion throughout the sciences.

Keywords Cerebellar · Circuit · DNA · RNA · Proteins

Introduction: Scope of the Call for Rethinking the Axioms of Genomics

Upon publication of the results of the project Encyclopedia of DNA Elements (ENCODE), a 4-year research effort led by the US Government, its architect issued a mandate: "the scientific community will need to rethink some long-held views" [1].

What views require our revision? The sequencing of the human genome engendered the idea of genomics as information science [2]. New avenues must be explored that are opening up to scientific research once breakthroughs from decades-old theoretical cul-de-sac lead to theoretical and experimental advances. "DNA has two types of digital information—the genes that encode proteins, which are the molecular machines of life, and the gene regulatory networks that specify the behavior of the genes" [2]. This paper reviews the dichotomy of genomics regarding historical conflicts regarding gene and gene regulation and offers a guiding principle for their synthesis. The introduction of this principle is made possible by a long-delayed but now respectful removal of two pragmatic dogmas, replacing them by a sound information-theoretical axiom.

In response to the open request in the post-ENCODE era that welcomes increasingly rigorous theoretical and mathematical foundations, this note comes forward with a principle learned from data in the course of an attempt to mathematize biology. In early efforts,[1] recursive algorithms came to the fore in explaining the function of neural networks.[2] However, since such an idea ran counter to the central dogma of molecular biology [10] that then prevailed in genetics, wherein only a "forward growth" mindset obtained (see Figs. 1, 2, and 3), this author subdued his claim, stating that "establishing a rigorous relation of these 'code sequences' to the genetic code that underlies the

A. J. Pellionisz (✉)
HelixoMetry, HQ,
935 Rosette Court,
Sunnyvale, CA 94086, USA
e-mail: pellionisz@junkdna.com

---

[1] *Neuronal Modeling* (cf. [3]) and *Artificial Intelligence to Neural Networks* (cf. [4]).

[2] See a specific recursion on page 365 [4] of a paradigm of *Neural Growth: Structural Manifestation of Repeated Access to Genetic Code* ([5], paragraph 3.1.3), and a collection of algorithms [6–9] in [4].
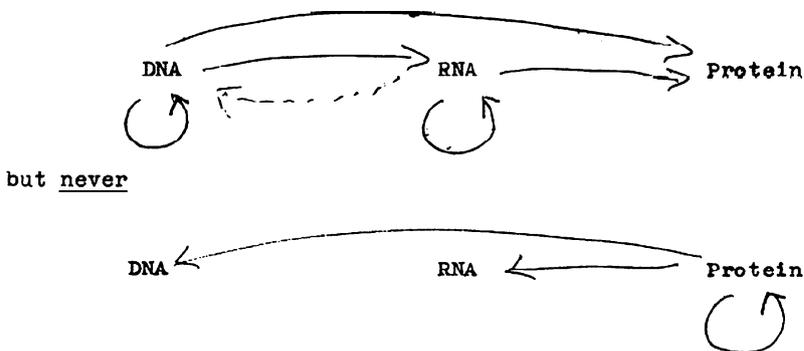
Ideas on Protein Synthesis (Oct. 1956)

The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it.

That is, we _may_ be able to have

DNA — RNA — Protein

but _never_

DNA ← RNA ← Protein

where the arrows show _the transfer of information._

morphogenesis of differentiated neurons may be far in the future" [5]. Now, however, with the ENCODE report in hand and with the field of neural networks now blossoming, the time seems ripe to advance a recursive principle whereby the genome governs growth of organelles, organs, and organisms.

In post-ENCODE genomics, the issue of "genome regulation" takes its long-deserved place. The concept, aimed at *controlling elements*, began with Barbara McClintock in the 1940s. Before the double helix was revealed [11], she discovered transposable elements in maize. She called them controlling elements because they altered gene expression. She published in the same year [12] when the Crick's central dogma was conceived (1956), wherein key feedback pathways were arbitrarily excluded. Gene regulation advanced to the *operon theory* of Jacob and Monod [13], for which they received the Nobel Prize in 1965. Jacob reminisced in his *Memoirs* about "...one of the oldest problems in biology: in organisms made up of millions, even billions of cells, every cell possesses a complete set of genes: how then, is it that all the genes do not function in the same way in all tissues?" [14]. This profound question is examined below.

Fig. 2. Watson's simplified rendering of Crick's central dogma states what is certainly a fact and strips the dogma of its controversial prohibitions ([62], p. 298)

By 1969, the field of gene regulation was growing at a healthy rate, to result in the work of Britten and Davidson [15]. However, in 1970, a wound opened up in genomics that has not yet healed. On one hand, the first major failure on Crick's doctrine was revealed [16]. Both Crick and Watson responded but in different ways [17, 18]. To keep the establishment together, Ohno declared the same year that all but the genetic DNA was *garbage* DNA [19], along with the slightly modified term of *junk* DNA [20]—a notion which prevailed for a generation.

The objective of this paper is to provide an historical review of the bifurcation and to offer a theoretical synthesis to remedy. With post-ENCODE genomics now removing obsolete impediments, principle of recursive genome function (PRGF) is expected to rapidly evolve, especially since some workers have been laboring for years in a clandestine fashion, quietly disregarding obsolete views (Pellionisz 2002; see in [21]).

In a general sense, the profound impact of changing axioms should be outlined, such as in medicine, bioenergy, nanotechnology, and synthetic biology—and even in philosophy. It is possible, as it was in physics [22], and even in neuroscience (*Neurophilosophy*, [23]) that the reversal of long-held views may have philosophical implications, giving rise to what might be called genome philosophy [24]. For instance, based on the ENCODE results, a synthesis may be necessary to integrate some goal-directed Lamarckian notions of evolution [25]—not confronting but rather surpassing simplification of original Darwinian notions [26], where, as we all learn in school,
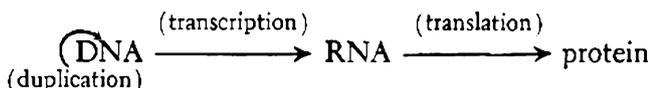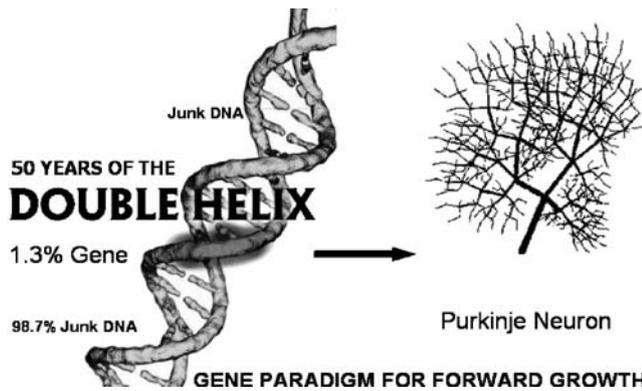
Fig. 3. Gene Paradigm for Forward Growth as of 2003. The *left diagram* of the double helix with a "genic" region *highlighted* is modified from the cover of Scientific American (April, 2003). *Diagram on the right side* is a brain cell. The diagram depicts the oversimplification, as if 1.3% of the DNA could determine, in a forward-growth manner, not only the Purkinje neuron shown but, given enough genes, all phenotypes resulting from separate genotypes. The 98.7% of junk DNA is a no man's land to which there is no recursion

natural selection suffices for the emergence of species. So saying, we are led to a few key conclusions concerning algorithmic approaches vis-à-vis the ENCODE study [27].

Core Idea: The Principle of Recursive Genome Function Reverses the Double Lock on Our Understanding of the Double Helix

The main body of this essay aims at accomplishing two goals: The first and far lesser task entails a brief review of history, since it has been attempted and to different extents attained numerous times before (see brief reviews below), to put to rest a long-standing but increasingly controversial theoretical *double lock* on the understanding of the function of "double helix" from the viewpoint of what we might call *genome informatics*. The second, far more important and difficult goal is, since "data never kill theories, only better theory can kill less tenable theories," that this note should leave no vacuum by removal of the key dogma but should replace two discarded, obsolete conjectures—regarding the *central dogma* of molecular biology, along with the notion of *junk* DNA—and, by means of their reversal, synthesize into a single principle (PRGF), more completely grounded in empirical data and withstanding more scrutiny from the viewpoint of information theory.

The recursive genome function is expressed by a process of already-built proteins, iteratively accessing sets of first primary and ensuing auxiliary information packets of DNA to build hierarchies of protein structures.

In abstraction, recursion is meant as a process of defining functions in which the function being defined is applied within its own definitions.

Applying these postulates to the genome, the most concise formulation is as follows (see Eq. 1).

Every $1-m$ finite state (Z) of the protein system (e.g., the $n+1$st state, denoted by Z_{n+1}) relies on the previous state of the protein system (e.g., Z_n) by applying a recursive function (f). The process is bounded by the limitation to the maximal number of states (m), where there is a function (f) from the nth state Z_n to be executed on Z_n to yield Z_{n+1}:

$$Z_{n+1} = \sum_{n=1}^{m} f(Z_n) \tag{1}$$

The diagram below (see also Fig. 4) pictures, in simplest terms, the principle of recursion of genomic function. In this equation, we see the cardinal role of the main path of recursive processes play in the construction of protein systems:

$$DNA \rightarrow RNA \rightarrow PROTEIN \rightarrow DNA \rightarrow RNA \rightarrow$$
$$PROTEIN \rightarrow DNA \rightarrow \dots$$

These recursive feedback processes then snowball into evolving (protein) structures, governed by DNA.

Purkinje brain cells provide an illuminating example of building a protein structure by means of an *L-string replacement* recursive algorithm [28]. The application of that algorithm is given elsewhere [5]. Experimental support of the quantitative predictions of the recursive approach is also readily available [21].

Readers will note that the PRGF is consistent not only with the recursive algorithms used in neural networks but is conceptually akin to a particular recursive formula, viz., the Mandelbrot set (see Eq. 2, from [29]):

$$Z_{n+1} = Z_n^2 + C \tag{2}$$

Further, fractal sets (see also the Julius Ruis set [30]) are representatives of just one of the class of recursive algorithms.

The postulated principle of recursive genomic function opens new avenues by way of a class of recursive algorithmic functions. Just one (fractal) example is given to show that the formulation of experimentally testable hypotheses for genomic function is plausible and supported by experimental results [21, 31].

Deep Background: Recursion is a Well-accepted Process in Science; Why Should it Not Become a Principle of Genome Function?

Recursion is a well-established concept in the sciences, ranging from pure mathematics to biological neural nets (cited in "Introduction"). A common linchpin between the two given domains is the *least squares* algorithm, which

minimizes errors; see the recursive mathematical and neural network basics [32]. From the viewpoint of the information theory, it is particularly notable that informational recursion from proteins (that are exposed to the external world) drastically alters the conception of genomic function as a *closed* system (to which the second principle of thermodynamics applies, with entropy increasing) to a system *open* to the world. Note that such recursion—involving outside factors—helps resolve the paradox between random mutation and the natural selection theory, where it is questioned whether the genome, featured as a closed system, can cope with an out-of-bounds increase in genomic entropy [33], once we consider that entropy can be regularized, given an open system [34]. Genomic function has hitherto been a strange exception to the widespread modeling of living and nonliving systems in terms of recursion. This singularity is especially peculiar since great physicists of the last century already predicted that our times would become the century of biology and that their physics-minded thinking processes, as given in Wiener's *Cybernetics* [35], explicitly invoked feedback as a primary principle in animal and machine. Schrödinger's *What is Life?* [36], von Neumann's *The Computer and the Brain* [37], and Szilard's *A theory of aging* [38] argued in unison that information-theoretic aspects would become key to a future understanding of biology. However, biology is a very young science—it is a mere 231 years since its coinage [39]. Genetics, as we knew it in pre-ENCODE genomics, just slightly exceeds a single century [40, 41]. Thus, the mathematical rigor that has characterized physics for over two millennia since Aristotle ([42], ca. 400 B.C.) could not be hastily enforced on unripe subjects—who were, moreover, for a long time somewhat unready and occasionally unwilling.

The above does not mean, however, that recursive algorithms have not been applied in conjunction with genomic systems, e.g., for extrinsic description and construction. In fact, the plethora of *extrinsic applications* makes it curious that the ideas forming PRGF have not heretofore overcome resistance and declared a breakthrough—a fact attributable to doctrines protected by the bulwark of scientists who underwrote pre-ENCODE genomics.

Recursive algorithms have encircled genomics and not only in respect of neural networks. Noted representative examples are genetic algorithms [43], recursive PCR [44], algorithms using DNA sequences as templates for encryption [45], construction of DNA structures by recursive algorithms [46, 47], and reconstruction of the genome by recursive assembly [48–51].

The above encirclement of genomics by recursive approaches greatly facilitates a dignified removal of both the central dogma and junk DNA conjectures. A breakthrough from Fig. 4a to b completes the liberation of post-ENCODE genomics, to better and more fully embrace the principle of recursion. Significantly, the inherent mathematics of genomic informatics would no longer be perceived as running against the establishment.

Obituaries of the central dogma and junk DNA are offered elsewhere. In brief, the demise of the obsolete axioms has been a yearly event in recent times, summarily refuted by leaders [52–57]. Specific factual anomalies
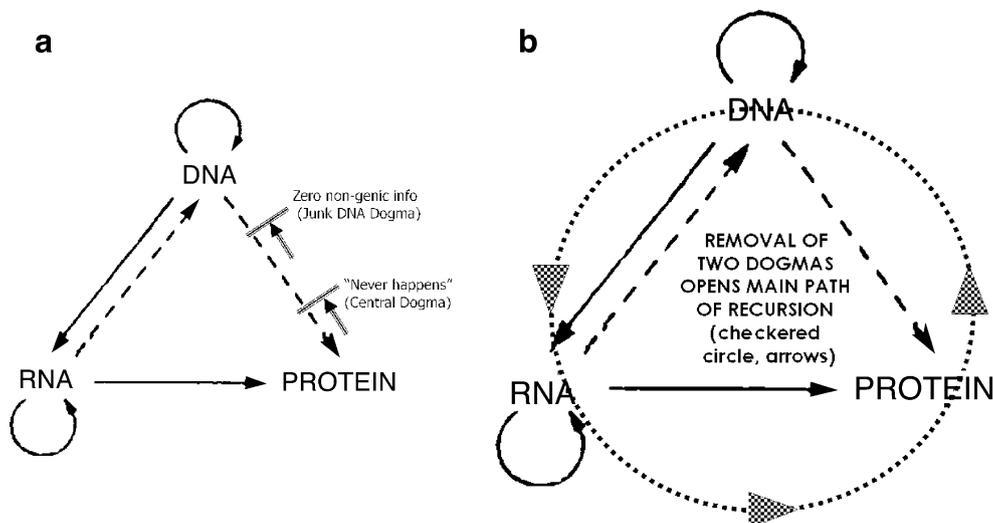

Fig. 4. PRGF breaks through the Double Lock of central dogma and junk DNA barriers (shown in **a** by *triple lines*), to yield PFGF (shown in **b** by *checkered circle*). The background figures in both **a** and **b** is from Fig. 2 from Crick (1970, see in this paper as the *right side* of Fig. 5), which permits only a "forward growth" from DNA that dead ends in proteins. By removal of both the central dogma that arbitrarily forbids information feedback from proteins to DNA as well as disposing the "Junk DNA" conjecture that claimed that (even if there was a path back to DNA), zero information would be found in the "Junk DNA," a main recursive path (PRGF, *checkered circle* and *arrows*) is not only available in principle, but it *is* the principle of recursive genome function

contradicting the doctrine have been reported for decades (see in a separate section). One should appreciate that, even at its outset, serious reservations were voiced; see Jacobs, e.g., in his *Memoirs* (pp. 288 in [14]). Jacob and Monod [13] provided Nobel Prize-winning evidences within half a decade of Crick's concept that operon regulation exerts a demonstrable feedback on DNA gene activity. Likewise, the *junk DNA* misnomer was summarily voided of its scientific validity as recently as in the suggested formal abandonment of the term as a scientific notion (International PostGenetics Society, 2006 —paper of 20 Founders rejected without review) and later in the ENCODE report stating that "the DNA is pervasively transcribed" [58]. The conjecture was finally put to rest by Mattick [59].

In all fairness, upholding vague and even controversial axioms in the nascent stage of biology (compared to more than ten times older and thus much more mathematical physics) were necessities dictated by practical constraints. This is perhaps best described by Brenner in his Nobel Lecture [60]:

> In 1985, when the first suggestions were made to sequence the human genome, I thought that the sequencing techniques, even with incremental improvements, would not be equal to the task, and would require a factory scale operation to do it. I had also come to the conclusion that most of the human genome was junk, a form of rubbish which, unlike garbage, is not thrown away. My view at the time was that we should treat the human genome like income tax and find every legitimate way of avoiding sequencing it.... I was puzzled by the enormous variations in the amounts of DNA in different organisms. Indeed, whereas most physicists thought that organisms did not have enough DNA to specify their complexity, it was clear to me that many organisms had too much. I discovered from

Hinegardner that one group of fish, the Tetraodontidae, which included the Japanese pufferfish, Fugu, had very small genomes, with a haploid content of about 400 megabases as opposed to the 3000 megabases of mammalian genomes. Although teleost fish are distant from humans they are still vertebrates, with the same body plans, development and physiological systems as ourselves. Because of these basic similarities it seemed unlikely that Fugu, with a haploid DNA content one eighth that of mammals would have eight times fewer genes, making it much more probable that what was missing in Fugu was junk DNA. [60]

This pragmatic consideration is also explicit in Crick's 1970 "revision" ([17], cf. his Fig. 1, reproduced in Fig. 5a of this paper).

> The principal problem could then be stated as the formulation of the general rules for information transfer from one polymer with a defined alphabet to another. This could be compactly presented by the diagram of Fig. 1. [of Crick, 1970] (which was actually drawn at that time [1958], though I am not sure that it was ever published) in which all possible simple transfers were represented by arrow. The arrows do not, of course, represent the flow of matter but the directional flow of detailed, residue-by-residue, sequence information from one polymer molecule to another. *Now if all possible transfers commonly occurred it would have been almost impossible to construct useful theories.* [Emphasis added, Pellionisz]. Nevertheless, such theories were part of our everyday discussions. This was because it was being tacitly assumed that certain transfers could not occur. It occurred to me that it would be wise to state these preconceptions explicitly. [17]
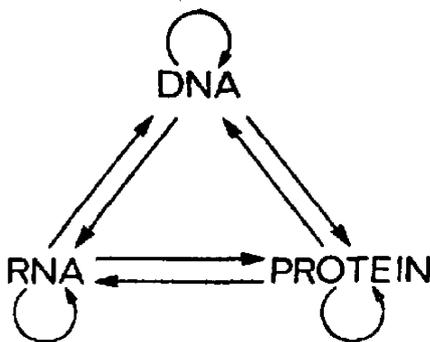

Fig. 1. The arrows show all the possible simple transfers between the three families of polymers. They represent the directional flow of detailed sequence information.
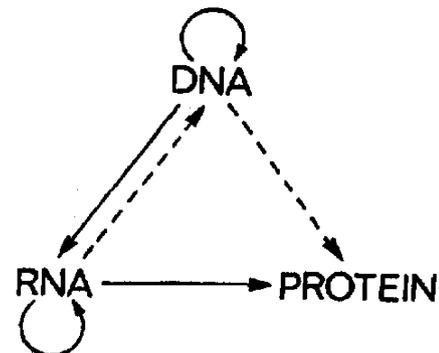

Fig. 2. The arrows show the situation as it seemed in 1958. Solid arrows represent probable transfers, dotted arrows possible transfers. The absent arrows (compare Fig. 1) represent the impossible transfers postulated by the central dogma. They are the three possible arrows starting from protein.

Fig. 5. (From Figs. 1 and 2 of Crick 1970). *Left side* permits an infinite number of possible recursive paths. *Right side* arbitrarily prohibits the main path of recursion by "dead-ending" proteins. The

DNA→RNA→PROTEIN→DNA recursion is just a single obvious recursive path, and the fractal approach already elaborated to some extent is just one of the possible recursive algorithms

The saga (life, death, and obituary) of junk DNA is treated elsewhere. Suffice it here that both the central dogma and the interlocking junk DNA dogma were finally and officially put on hold only by the conclusions of the ENCODE pilot project, on June 14th, 2007 [58].

Historical Recount: Specific Review of the Nascence and Demise of the Central Dogma and Junk DNA Conjectures

The central dogma in its various renderings held that transfer of information from proteins and RNA back to DNA *never* happens. The central dogma of molecular biology was put forward in Francis Crick's talks from 1956 (Fig. 4, cf. recollection of Jacob [14], p. 286) and published 2 years later [61].

This concept might be called "don't look back" or "no feedback permitted" postulate. Further, proponents of junk DNA [20] claimed that, even if a process could be found, a recursion from proteins and/or RNA to DNA could not retrieve information from functionless junk DNA (that is 98.7% of the human DNA). This conjecture may be called "even if you look back, you find only junk" postulate. Recursion for information was not only forbidden but in addition was prejudged as useless because of an assumed void of information.

It may be Watson's reduced version of the central dogma [62], wherein he emphasized what was surely found in genomic function and avoided needless and unsupported prohibitions, which helped the central dogma receive common acceptance from 1965 to 1969–1970, skirting sharp personal criticisms that were already present at its conception (e.g., questions as to its dogmatic stance; see Jacob's *Memoirs* [14], pp. 288).

A factor in the prevalence of the central dogma during this period might be that, since the 1965 Nobel Prize for Jacob and Monod's work on operon regulation [13], it had to be evident to Watson that, given the operon regulation of gene expression (as a function of the level of produced proteins), the *prohibition* of a protein-to-DNA information channel need not be in Watson's textbooks. His simplified and convenient view (avoiding controversial prohibitions and their refutation by data) is called here the concept of forward growth—which is, without recursion, only half the loop.

This view, backed by Watson, prevailed so strongly that even the 50th anniversary issue of *Scientific American*, celebrating the discovery of the structure of DNA by Watson and Crick (1953), depicted as the general understanding that gene expression of DNA results, through RNA, in the construction of protein structures (see Fig. 3, a composite illustration of the modified cover

page figure of *Scientific American* of April, 2003, where on the right side the model of a Purkinje brain cell is pictured, from [5]).

Reality was, as might be expected, much more complicated than any simplification. As early as 1969, in the Britten and Davidson theory of gene regulation [15], and by 1970 (cf. philosophical reflection by Darden [63]), the central dogma was squarely confronted by the discovery of reverse transcription, later called retroviruses, from RNA to DNA [16, 64].

Both Watson and Crick responded promptly but separately to the challenge posed by the discovery of the new enzyme that flagrantly violated their views. In the June 27, 1970 issue of *Nature*, the reverse transcriptase discovery was announced, and an anonymous "News and Views" article claimed: "Central Dogma Reversed."

> (Quote from Darden [63]): Watson, in the 1970 second edition of his *Molecular Biology of the Gene* [18], said: "The concept of a DNA provirus for an RNA virus is clearly a radical proposal. *It overturns the belief that flow of genetic information always goes in the direction of DNA to RNA and never RNA to DNA.* [Emphasis added, AJP] On the other hand, it offers an even greater variety of ways for cells to exchange genetic information. Considering the enormous complexity of biological systems, it would not be surprising if this device were uniquely advantageous in some situations. ([18], pp. 621–622)

Crick (1970) also responded immediately to the challenge [17] but in a different way (unfortunately, the dogma was not allowed to gracefully expire). Crick published a paper in *Nature*. His version of the central dogma, he contended, had not been reversed, as the anonymous *Nature* article had claimed. Crick stated, correctly, that in 1958, he had framed the central dogma in terms of the general transfer of information from nucleic acids to protein—but not the reverse (Crick 1958). That abstract claim had not yet been challenged. *If it were shown that information could flow from proteins to nucleic acids, he said, then such a finding would "shake the whole intellectual basis of molecular biology"* ([17], p. 563; quote from Darden [63], emphasis added, Pellionisz).

Thus, by 1970, the intellectual split between Crick, the originator of "The Central Dogma" (1956, [61]), and its promoter Watson (1965, [62]), threatened the collapse of the genomics establishment. The shaky ground of "The Central Dogma" was not really firmed up by the confession: "Dogma was just a catch phrase" (Crick, quoted in [65]).

In the same year of the split, Ohno's junk DNA idea came to the rescue. Ohno first referred to garbage DNA in the human genome ([19], p. 62). Meaning that, even if there was recursive information access to DNA—from proteins

or from RNA, e.g.—there was supposedly no information in the intronic and intergenic regions to be found and retrieved. Although the term "garbage DNA" floated in 1970, it did not take hold, but by 1972, in his presentation, he began using the more suitable term "junk DNA," which did stick [20]. One should appreciate that, immediately after his presentation, the first person to rise in the discussion vehemently objected to the basis of the junk DNA conjecture (see "Discussion" by Boyer, in [20]): "It thus seems to me that the permissible number of structural loci is—as yet—a somewhat suspect way to arrive at figures of 1% structural utility to 99% junk."

Why is it that Nobel Prize-winning experimental work (cf. [66], Jacob's Nobel Lecture on his Prize with Monod, 1965) was available as early as within 5 years after Crick's conception of his dogma—yet no theoretical confrontation developed? Their operon regulation [13] clearly demonstrated that the protein level, viewed as a result of genetic activity, did have an information-feedback mechanism on the genes in the DNA—such that down- or upregulated DNA–RNA activity resulted, in accordance with the amount of protein already generated:

> Experiments on genetic transfer by conjugation not only led to a revision of the concepts on the mechanisms of information transfer which occur in protein synthesis; they also made it possible to analyze the regulation of this synthesis.... the operator is not transcribed into messenger and repression can be exerted only at the level of DNA.... Gene expression was then usually believed to consist in the accumulation of stable structures in the cytoplasm, probably the RNA of ribosomes, which were assumed to serve as templates specifying protein structures... Such a scheme, which can be summarized by the aphorism "one gene-one ribosome-one enzyme", was hardly compatible with an immediate protein synthesis at maximal rate." [13]

In retrospect and judging from Jacob's *Memoirs* [14], it seems evident that Jacob was fully aware of the intellectual conflict between the Jacob–Monod finding and Crick's central dogma, even at its birth (not shying away from direct criticism of the label "dogma," however):

> In an acute sense of publicity, to baptize Central Dogma— that is to say, incontestable truth—a hypothesis that was unsupported by any serious argument. ([14], pp. 288)

However, it appears that Jacob did not directly confront Crick on the latter's conception of the dogma [61], since that was prior to the publication of the Jacob–Monod operon concept. Jacob and Monod (1961) published their "operon regulation" work soon after but did not receive their Prize until 1965, whereas in the following year, Watson and Crick (1962) received their award. Thus, it was

arguably more politic to avoid a direct conflict among the foursome. Almost simultaneously, however, in 1965, Watson [62] distanced himself from the central notion by putting forward his "simplified version," emphasizing what was undeniably true, although he did not dwell upon issues already controversial (see Fig. 2).

The point of this paper, however, is neither iconoclastic (discarding pragmatic doctrines while simply leaving a theoretical void in their place) nor to merely cite evidence for the widely reported means of feedback processes from proteins to DNA (and RNA to DNA). The point is to fill the void. PRGF is proposed to break through the "double ceiling" of the central dogma and junk DNA that impeded theoretical advances for half a century.

Theoretical and Factual Breakdown of the Central Dogma and Junk DNA

Further flogging of two dead horses is avoided as much as possible. This abbreviated section merely supplies evidence, gives credit where most needed, and points to the most powerful reviews.

From the theoretical viewpoint of informatics, the "double lock" on recursive information has long been suspect. First, there is no more information for hereditary material than that present in the DNA. In humans, if 98.7% of the DNA is arbitrarily closed to access and in addition its information voided, the remaining information that 1.3% of the (human) genome harbors is deemed simply insufficient to govern development of such advanced organisms as vertebrates. It was painfully experienced, for instance, that when constructing a computer model of 1.68 million brain cells of the frog cerebellum, algorithmic approaches had to be invoked, rather than pretending that an impossible amount of information was available to specify the vast neural network in its every detail [67].

As for factual contradictions, the array of evidence— against both the central dogma and junk DNA conjectures— is staggering; see respective reviews [52, 53, 56].

Beyond the early factual evidence belying the validity of the central dogma ("Operonic Regulation by Feedback from Proteins to DNA"), another large assortment of facts is available, as follows.

For an account of the (forbidden) information transfer from RNA back to the DNA, see the very recent review by Mattick [59], with background about the RNA world [68–72].

Major issues are *gene silencing* (or "turning genes on and off," e.g., by so-called LINE way stations, and switching via SINE-s). It is noteworthy that the PRGF is fully consistent with the currently vague notion of "turning genes on and off" but goes further by invoking recursion not only the sign(s) of the "parallel feedback" are

meaningful but much more important information is the set of algorithmic values of recursive signals [73–76].

For the "forbidden" protein-to-DNA interaction, see the work on protein binding with the DNA and *methylation* of the DNA by proteins—rendering DNA transcription reversibly or permanently impossible [77–82].

For the (also forbidden) protein–protein interactions, see *Prions* [83] and a detailed and philosophical review [52].

Specific Process for Purkinje Neuron Growth Governed by a Recursive Genomic Function

In its simplest form, PRGF can be metaphorically described as the manner whereby an assembler employs a user's manual with respect to a streaming supply of parts. First, the assembler looks at step 1 of the manual, and, operating according to instructions in the primary information packet, the assembler puts together the indicated components, taken from the supply of parts.

Next, the assembler compares the emerging structure to step 1 as depicted in the instruction manual by referring to the primary source of information (in our case, the DNA, the "genes"). It is noteworthy that it is useful for the assembler to mark as *done* the just completed instruction step, to avoid its repetition by mistake.

Next, the assembler proceeds to step 2 in the instructions. Accessing the next *auxiliary information packet*, the assembler puts together the indicated components into the next layer of the hierarchy. Comparing the emerging structure in the second hierarchy with step 2 by looking back at the manual, the assembler marks step 2 as *done.*

The process goes on in a recursive fashion through the manual, until all the finite steps are taken. The assembler then runs out of instructions (all instructions are marked *done*).

For the DNA, RNA, and protein (and so on), in the circular chain depicted in Figs. 4 and 6, the term "recursion" means that, by reversing the central dogma, a main path opens for recursive algorithms to be applied as the intrinsic mathematics of genome function. That is, not only an information transfer back from proteins to DNA is allowed, but this feedback mechanism is (again) relied upon as the enabling feature of a circular process, thus:

$$DNA \rightarrow RNA \rightarrow PROTEIN \rightarrow DNA \rightarrow RNA \rightarrow PROTEIN \rightarrow \ldots$$

Further, by reversing the notion of junk DNA, the principle of recursion postulates that the above main recursive path accesses functional (as opposed to junk)
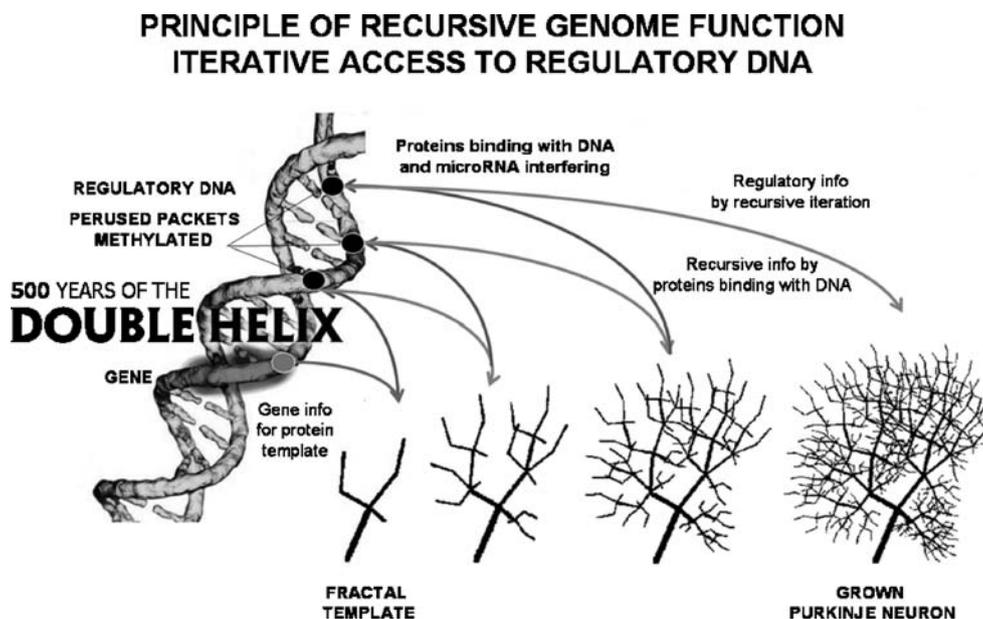

Fig. 6. Sketch of recursive genomic government of the growth of Purkinje neuron. Starting from a primary information packet *highlighted*, a Y-shaped protein template is built by the "forward growth" process in accordance with the simplified (Watson) picture through transcription of DNA to RNA and, in turn, RNA building nucleic acids that form a structural protein. During the construction of the Y-shaped template, the primary gene is in a "turned on" condition. Thus, the most primitive primary part of the process retains Watson's simplified scheme. In other words, the postulated process does not contradict to the process of "DNA makes RNA that makes proteins" but goes beyond it, by violating both the forbidden feedback mechanisms and the notion of junk DNA. In each recursive step, the perused auxiliary information packet (formerly "junk DNA" or "regulatory DNA") is cancelled (methylated) upon perusal

Fig. 7. An example of a recursive-looking organism (Cauliflower Romanesca). It is possibly grown by a Lindenmayer L-string replacement recursive algorithm, e.g., governed by the DNA→RNA→PROTEIN→DNA recursion, a massively parallel process executed repeatedly

DNA, viz., information from intronic and intergenic regions that were formerly regarded as useless.

PRGF claims that the overall DNA function is expressed by a recursive process determined and governed by repetitive access to information packets contained in the DNA through the channel depicted in the schema above. The postulated process is active and bounded. By *active*, it is meant that information packets accessed may be rendered inaccessible (in reversible and/or a permanent manner by de novo methylation) and that the consumption of information governing growth leads to an eventual death of the organism.

Despite tremendous advances, the full genomic government of Purkinje cell assembly still remains largely unknown [84, 85]. This paper illustrates the applicability of PRGF in a sketch of the development of this brain cell (cf. [5, 21]). It is expected that the recursive framework provided will contribute to further advances in revealing genomic government of developments of Purkinje neurons and other structures, making it an eminent platform for post-ENCODE genomics.

To recapitulate and expand, Fig. 6 pictures the recursive process as follows.

Structural proteins are generated by a DNA primary information packet (pre-ENCODE "gene"), growing a Y-shaped template. Completion of proteins of step 1, however, is not a dead end, as formerly asserted by the dogma. Completion may be reported by a *completion marker protein* that, in its simplest version, binds with the DNA. Specifically, the completion marker protein can "turn off the gene."

Completion of this first step shuts down the first stage of growth. Evidence is available for micro-RNAs (and interfering micro-RNA-s, small inhibitory RNA-s; see [86, 87]) that can signal completion by turning off the primary information packet (formerly "gene").

The auxiliary packet of information (formerly "junk" DNA or regulatory DNA) is turned off by de novo methylation upon perusal of retrieved information; each such auxiliary information packet, once perused, is rendered temporarily or permanently unreadable. This provides a framework to explain the oldest problem in biology, the fact that the differentiated cells of an organism are no longer omnipotent. Their methylation pattern permits specific and limited further growth, as much as permitted by the remainder of unmethylated auxiliary information to be accessed by further recursion. This framework is consistent with the argument of the theory of aging [38], the proposition that genetic damage leads to progressive degradation of the ability to make necessary proteins.

Generalizations of the Principle of Recursive Genome Function

PRGF obviously transgresses the once forbidden feedback mechanism and also relies on auxiliary genomic information packets—previously regarded as junk. It is noteworthy that, prior to the publication of *Cybernetics* [35], which made feedback mechanisms a most conspicuous aspect of biological processes, most philosophies in biology (including traditional evolution, [26]) assumed a similar, rudimentary forward-growth process via random mutations and natural selection. In view of the interaction of protein structures (organisms) with the environment, recursion would seem to enable nonrandom development.

Another comment toward generalization is that the depicted recursion that—not unlike neural reflex arcs in the early history of neuroscience—neural networks at first glance looked like a chain [3]. However, just as with neural networks, genomic recursion is inherently parallel, since the recursion is not limited to a single primary information packet and its auxiliary information packets. In the case of genomic function, the three main layered sets of elements (DNA, RNA, and proteins) operate in a massively parallel manner, inviting neural network algorithms (see in "Introduction").

In this context, it is noteworthy that there is no separate or even separable operating system [55] as the recursive genome is self-governed (unsupervised). In fact, the view here is much in agreement with the concept first touched upon in "What is life?" [36]:

But 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. [36]

Von Neumann [37] also stated his principle that there is no difference between the two kinds of information (code and data) so far as its repository (memory) is concerned. Primary information packets ("genes") and auxiliary information packets are all nucleotide sequences; it is the PRGF recursive process of access that distinguishes them.

Readers familiar with the "fractal approach" [21] may note the applicability of PRGF to that particular paradigm pursued by this author to describe fractal development of a Purkinje cell in the "Pre-ENCODE wilderness years." The fractal approach is especially encouraged by the measurements that "pykon-like elements" [88] of the whole genome

of *Mycoplasma genitalium* [89, 90] apparently follow the Zipf–Mandelbrot parabolic fractal distribution (personal communication by Pellionisz in [31, 91] also pointing out a found Pareto-distribution, that is, a truncated Taylor series to the Zipf–Mandelbrot parabolic fractal distribution.

Readers will also note the generalization that PRGF opens a way not only for fractal recursive iterations (e.g., a fractal DNA resulting in the fractal structure of the Romanesca vegetable photographed for illustrative purposes on Fig. 7) but to an entire class of recursive algorithms, with the recursion in principle certainly not limited to the main recursion via DNA→RNA→PROTEIN→DNA→ (etc) but possibly involving an infinite number of recursive loops pictured in Fig. 5a (Fig. 1 of [17] reproduced).

Just for one more intricate example:

DNA → RNA → DNA → RNA → PROTEIN → PROTEIN → RNA → DNA → RNA → PROTEIN →

From the viewpoint of an overriding pragmatism of genomics, one may note that the theoretically infinite possible variations of pathways of recursions in Fig. 1 of Crick [17] justifiably frightened workers (including Crick [17]) away from the theoretical problem at the time of 1960–1970, rushing to reduce genomic function into a two-step procedure (transcription and translation; [62]), and also putting a second lock on recursion by way of the junk DNA conjecture. It has long been known in physics that the *two-body problem* of two masses interacting can be described exactly, while the *three-body problem* presents formidable mathematical challenges. Crick, who later ventured into neural networks himself, did not seem to favor heavy use of mathematics [23].

Indeed, early in the second century of genetics (now in its post-ENCODE era, i.e., PostGenetics), the rudimentary recursive sketch of the fractal growth of a single Purkinje cell is a far cry from fully defining (and, through experimentation, verifying in every detail) a complete mathematical model of even the genomic governance of even one of the best known single-cell platform of the well-familiar multicellular organ of the cerebellum, the single Purkinje neuron.

Our best immediate hope is to gather support for revealing first the most rudimentary genomic regulation present in the smallest DNA of a free-living organism (*Mycoplasma genitalium*, where intergenic sequences total a mere 50,000 nucleotides, and thus not only actual fractal structures could be revealed in the DNA but "fractal defects" were corroborated with glitches in regulatory intergenic sequences; [92]). Next, we can target more complex (multicellular) organisms, to start with Purkinje

neurons, later proceeding for instance to the quite obviously fractal-looking Cauliflower Romanesca, which visibly evolves in a massively parallel manner (Fig. 7).

To emphasize that the class of both massively parallel and recursive algorithms of neural nets led toward this school of thinking [6, 5] and will increasingly be applied in implicit denial of the central dogma, the most recent article from the flagship of NN R&D is cited [93].

The utilization of PRGF is expected to lead to predictable implications—as well as others, unforeseen. As indicated at the outset, new advances in applications of PRGF are likely to include epigenetic medicine (for diagnosis, identification of factors that block PRGF—such as defects in regulatory sequences—and for therapies that block PRGF to create new types of antibiotics, etc.). For bioenergy, nanotechnology, and synthetic biology, it seems essential to first understand genomic function, including genes but also their regulatory mechanisms. As suggested by Fig. 7, our plate is full for the second century of (postmodern) genomics.

References

1. Collins F (2007) New findings challenge established views on human genome. June 14. http://www.genome.gov/25521554
2. Hood L, Galas D (2003) The digital code of DNA. Nature 421 (6921):444–448
3. Pellionisz A (1979) Modeling of neurons and neuronal networks. In: Schmitt FO, Worden FG (eds) The Neurosciences, IVth Study Program. MIT, Boston, MA, pp 525–546
4. Anderson J, Pellionisz A, Rosenfeld E (1989) Neurocomputing-2. MIT, Boston, MA

5. Pellionisz AJ (1989) Neural geometry: towards a fractal model of neurons. In: Cotterill RMJ (ed) Models of brain function. Cambridge University Press, Cambridge, pp 453–464

6. Werbos PJ (1994) The roots of backpropagation. Wiley, New York (includes Werbos's Harvard Ph.D. thesis: Beyond Regression, 1974)

7. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79(8):2554–2558

8. Kohonen T, Barna G, Chrisely R (1989) Statistical pattern recognition with neural networks: benchmarking studies. In: Anderson J, Pellionisz A, Rosenfeld E (eds) Proceedings of the IEEE International Conference on Neural Networks, San Diego, 1988. Neurocomputing-2. MIT, Boston, MA, pp I–61–I–68

9. Carpenter G, Grossberg S (1990) ART-2: Self-organization of stable category recognition codes for analog input patterns. Appl Opt 26:4919–4930

10. Crick FHC (1958) On protein synthesis. Symp Soc Exp Biol 12:138–163 (early draft version at http://profiles.nlm.nih.gov/SC/B/B/F/T/_/scbbft.pdf)

11. Watson JD, Crick FHC (1953) Molecular structure for deoxyribose nucleic acid. Nature 171:737–738

12. McClintock B (1956) Controlling elements and the gene. Cold Spring Harb Symp Quart Biol 21:197–216

13. Jacob F, Monod JJ (1961) Genetic regulatory mechanisms in the synthesis of proteins. Mol Biol 3:318–356

14. Jacob F (1995) The statue within—an autobiography. Cold Spring Harbor Laboratory, New York

15. Britten RJ, Davidson EH (1969) Gene regulation for higher cells: a theory. Science 165(3891):349–357

16. Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226:1209–1211

17. Crick FHC (1970) Central dogma of molecular biology. Nature 227:561–563

18. Watson JD (1970) Molecular biology of the gene, 2nd edn. Benjamin Cummings, Reading, MA

19. Ohno S (1970) Evolution by gene duplication. Springer, New York

20. Ohno S (1972) So much junk in the human DNA. In: Brookhaven Symposia, pp 366–369

21. Simons MJ, Pellionisz AJ (2006) Genomics, morphogenesis and biophysics: triangulation of Purkinje cell development. The Cerebellum 5(1):27–35

22. Heisenberg W (1927) Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. Z Phys 43:172–198 (English translation: J. A. Wheeler and H. Zurek, Quantum Theory and Measurement Princeton Univ. Press, 1983, pp. 62–84)

23. Churchland PS (1986) Neurophilosophy: toward a unified science of the mind–brain. Bradford/MIT, Cambridge, MA

24. Dupre J (2005) Darwin's legacy: what evolution means today. Oxford University Press, Oxford

25. Lamarck JB (1801) Système des animaux sans vertèbres, ou tableau général des classes, des ordres et des genres de ces animaux; présentant leurs caractères essentiels et leur distribution, d'après la considération de leurs, Paris, Detreville, 8:1–432

26. Darwin C (1859) On the origin of species. Murray, London, UK

27. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR et al (1970) Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc Natl Acad Sci USA 104:18613–18618

28. Lindenmayer A (1968) Mathematical models for cellular interaction in development I. Filaments with one-sided inputs. J Theor Biol 18:280–289

29. Mandelbrot BB (1977) The fractal geometry of nature. WH Freeman, New York

30. Ruis J (2004) The Julius Ruis Set. http://www.fractal.org/Beelden/Julius-Ruis-Set.jpg

31. Simons MJ, Pellionisz AJ (2006) Implications of fractal organization of DNA on disease risk genomic mapping and immune function analysis, In: Proceedings of the Australasian and Southeast Asian Tissue Typing Association 30th Scientific Meeting, November, Chiangmai, Thailand, pp 22–24

32. Widrow B, Stearns SD (1985) Adaptive signal processing. Prentice-Hall, Englewood Cliffs, NJ

33. Sanford JC (2005) Genetic entropy. Elim, New York

34. Erdogmus D, Yadunandana N, Principe RJC, Fontenla-Romero O, Alonso-Betanzos A (2003) Recursive least squares for an entropy regularized MSE cost function. In: ESANN'2003 proceedings—European Symposium on Artificial Neural Networks Bruges (Belgium), 23–25 April, pp 451–456

35. Wiener N (1949) Cybernetics or control and communication in the animal and the machine. Wiley, New York

36. Schrödinger E (1944) What is life? The physical aspect of the living cell. Based on lectures delivered under the auspices of the Dublin Institute for Advanced Studies at Trinity College, Dublin, in February 1943

37. von Neumann J (1958) The computer and the brain. Yale University Press, New Haven

38. Szilárd L (1959) Theory of ageing. Nature 184(4691):957–958

39. Hanov MC (1766) Philosophiae naturalis sive physicae dogmaticae: geologia, biologia, phytologia generalis et dendrologia

40. Bateson W (1928) Letter to Sedgwick, April 18, 1905. In: Bateson B (ed) William Bateson, F.R.S.: his essays and addresses. Cambridge University Press, Cambridge, p 93

41. Bateson W (1906) A text-book of genetics. Nature 74:146–147

42. Aristotle (1989) De memoria et reminiscentia (Aristotle on memory) (ca. 400 B.C.; English translation). In: Anderson J, Pellionisz A, Rosenfeld E (eds) Neurocomputing-2. MIT, Boston, MA, pp 1–10

43. Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. MIT, Boston, MA

44. Prodromou C, Pearl LH (1992) Recursive PCR: a novel technique for total gene synthesis. Protein Eng 8:827–829

45. Barua R, Misra J (2003) Binary arithmetic for DNA computers. Lect Notes Comput Sci 2568:124–132

46. Carbone A, Seeman NC (2003) Coding and geometrical shapes in nanostructures: a fractal DNA-assembly. Nat Comput 2:133–151

47. Winfree E (2000) Algorithmic self-assembly of DNA: Theoretical motivations and 2D assembly experiments. J Biol Mol Struct 11(2):263–270

48. Fields CA, Soderlund CA (1990) A practical tool for automating DNA sequence analysis. Comput Appl Biosci 6:263–270

49. Dong Q, Wilkerson MD, Brendel V (2007) Tracembler—software for in-silico chromosome walking in unassembled genomes. BMC Bioinformatics 8:151

50. Lander ES, Linton LM, Birren B, Nusbaum C et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

51. Venter JC, Adams MD, Myers EW, Li PW et al (2001) The sequence of the human genome. Science 291:1304–1351

52. Boussard AE (2005) A scientific revolution? The prion anomaly may challenge the central dogma of molecular biology. EMBO J 6(8):691–694

53. Goodman AF, Bellato CM, Khidr L (2005) The uncertain future of the central dogma. Scientist 19(12):20

54. Mattick JS (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. BioEssays 25:930–939

55. Mattick JS (2004) The hidden genetic program of complex organisms. Sci Am 291:60–67

56. Henikoff S (2002) Beyond the Central Dogma. Bioinformatics 18:223–225

57. Thieffry D (1998) Forty years under the central dogma. Trends Biochem 23:312–316

58. Brinley E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Marguiles EH et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE Pilot Project (14 June 2007). Nature 447:799–816

59. Mattick JS (2007) The human genome: RNA machine—contrary to current dogma most of the genome may be functional. Scientist 21(10):61

60. Brenner S (2002) Nobel lecture 2002. Available at: http://nobelprize.org/nobel_prizes/medicine/laureates/2002/brenner-lecture.html

61. Crick FHC (1956). Ideas on protein synthesis. Available at: http://profiles.nlm.nih.gov/SC/B/B/F/T/_/scbbft.pdf (acknowledged by Crick in 1958)

62. Watson JD (1963) The molecular biology of the gene, 1st edn. W. A. Benjamin, New York

63. Darden L (2006) Reasoning in biological discoveries. Cambridge University Press, Cambridge

64. Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226:1211–1213

65. Judson HF (2004) The eighth day of creation: makers of the revolution in biology. Cold Spring Harbor, New York

66. Jacob F (1965) Nobel lecture (on the Prize with Monod). Available at: http://nobelprize.org/nobel_prizes/medicine/laureates/1965/jacob-lecture.pdf

67. Pellionisz A, Llinás R, Perkel DH (1977) Computer model of the cerebellar cortex of the frog. Neuroscience 2:19–35

68. Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* genome. Curr Biol 11:1017–1027

69. Carthew RW (2001) Gene silencing by double-stranded RNA. Curr Opin Cell Biol 13:244–248

70. Eddy SR (2001) Non-coding RNA genes and the modern RNA world. Nat Rev Genet 2:919–929

71. O'Gorman W, Akoulitchev A (2006) What is so special about Oskar wild? Sci STKE 365:51

72. Woese CR (2001) Translation: in retrospect and prospect. RNA 7:1055–1067

73. Lyon MF (1998) X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet 80:133–137

74. Bailey JA, Carrel L, Chakravarti A, Eichler EE (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. PNAS 97:6634–6639

75. Mlynarczyk SK, Panning B (2000) X inactivation: Tsix and Xist as yin and yang. Curr Biol 10:R899–R903

76. Matzke M, Matzke AJM, Kooter JM (2001) RNA: guiding gene silencing. Science 293:1080–1083

77. Lunyak VV, Prefontaine GG, Núñez E, Cramer T, Ju BG, Ohgi KA et al (2007) Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science 317 (5835):248–251

78. Lilley DMJ (1995) DNA-protein: structural interactions. IRL Press at Oxford University Press, Oxford

79. Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–340

80. Heard E, Rougeulle C, Amaud D, Avner P, Allis CD, Spector DL (2001) Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell 107:727–738

81. van Steensel B, Delrow J, Henikoff S (2001) Chromatin profiling using targeted DNA adenine methyltransferase. Nature Gen 27:304–308

82. Weidman JR, Dolinoy DC, Maloney KA, Cheng JF, Jirtle RL (2006) Imprinting of opossum Igf2r in the absence of differential methylation and air. Epigenetics 1(1):49–54

83. Lindquist S (1997) Mad cows meet psychotic yeast: the expansion of the prion hypothesis. Cell 89:495–498

84. Barski JJ, Lauth M, Meyer M (2002) Genetic targeting of cerebellar Purkinje cells: history, current status and novel strategies. Cerebellum 1(2):111–118

85. Oberdick JD, Jankowski J, Holst MI, Liebig C, Baader SL (2004) Engrailed-2 negatively regulates the onset of perinatal Purkinje cell differentiation. J Comp Neurol 472:97–99

86. Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:S.806–811

87. Mattick JS (2005) The functional genomics of noncoding RNA. Science 309(5740):1527–1528

88. Rigoutsos I, Huynh T, Miranda K, Tsirigos A, McHardy A, Platt D (2006) Short blocks from the non-coding parts of the human genome have instances within nearly all known genes and relate to biological processes. PNAS 103:6605–6610

89. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD et al (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270(5235):397–403

90. Peterson S, Camella C, Bailey S, Jorgen S, Jensens S, Martin B et al (1995) Characterization of repetitive DNA in the *Mycoplasma genitalium* genome: possible role in the generation of antigenic variation. PNAS 92:11829–11833

91. Csuros M, Noe L, Kucherov G (2006) Reconsidering the significance of genomic word frequency. arXiv:q-bio.GN/0608022v1. Available at: http://arxiv.org/PS_cache/q-bio/pdf/0609/0609022v1.pdf

92. Pellionisz A (2006) PostGenetics: genetics beyond genes. The journey of discovery of the function of Junk DNA. In: BCII2006 Proceedings

93. Xu R, Ganesh K, Venayagamoorthy K, Wunsch DC (2007) Modeling of gene regulatory networks with hybrid differential evolution and particle swarm optimization. Neural Netw 20:917–927