Critical Review

Revising the Central Dogma: Regulated, Dynamic, and System-Dependent Information Coding and Decoding

Change L. Tan[1*] and Eric H. Anderson[2]

[1] Missouri Baptist University, St. Louis, Missouri, USA
[2] Engineering Research Group, Discovery Institute, Seattle, Washington, USA

Abstract

The central dogma of molecular biology, formulated by Francis Crick and popularized by James Watson, emphasizes the unidirectional transfer of genetic information from DNA to RNA to proteins. This principle has greatly influenced our scientific research and perspective of life. However, it fails to adequately account for the following discoveries: 1) there are different kinds and different levels of biological information; 2) the meaning and usage of biological information are system-dependent; 3) no information flow is possible without the cooperative function of DNA, RNA and proteins; 4) the coding system and the decoding system have to match; and 5) proteins, with the help of RNAs, control whether and how DNA is replicated as well as the stability, accessibility and usability of DNA. Thus, we propose updating the central dogma to the following: The central principle of molecular biology is regulated, dynamic, system-dependent information coding and decoding.

Cite as: Tan CL, Anderson EH (2024) Revising the Central Dogma: Regulated, Dynamic, and System-Dependent. BIO-Complexity 2024 (3):1-21. **doi:**10.5048/BIO-C.2024.3.

Editor: Douglas D. Axe

Received: June 7, 2023; **Accepted:** April 12, 2024; **Published:** July 20, 2024

Notes: A Critique of this paper, when available, will be assigned **doi:**10.5048/BIO-C.2024.3.c.

*Change.Tan@mobap.edu

INTRODUCTION

The central dogma of molecular biology has had a profound impact not only on the study of molecular biology, but also on our common perceptions about life, our understanding of the causes of diseases and our approach to treatments. As articulated by Francis Crick more than sixty years ago and subsequently reinterpreted by other scientists, the central dogma underwrote the common belief that identifying and manipulating certain genes would enable us to solve the twin problems of world hunger (e.g., via generating genetically modified organisms) and dreadful disease (e.g., via personalized medicine)[1,2]. A clear understanding of the true characteristics of molecular biology is both critical and urgent because the consequences of misunderstanding are severe and costly. In this paper, we will briefly review the history of and describe the problems with the central dogma and provide a revision that more accurately reflects our current understanding of molecular biology.

HISTORY OF THE CENTRAL DOGMA

Crick's Central Dogma

In a March 19, 1953, letter, Francis Crick told his 12-year-old son Michael about the discovery he and James Watson had made [3]:

> Jim Watson and I have probably made a most important discovery. We have built a model for the structure of de-oxy-ribose-nucleic-acid (read it carefully) called D.N.A. for short.… Now the exciting thing is that while there are 4 *different* bases, we find we can only put certain pairs of them together…only A with T and G with C.
>
> Now on one chain, as far as we can see, one can have the bases in any order, but if their order is fixed, then the order on the other chain is also *fixed*.… It is like a code. If you are given one set of letters you can write down the others.

> Now we believe that the D.N.A. *is* a code. That is, the order of the bases (the letters) makes one gene different from another gene (just as one page of print is different from another). You can now see how Nature *makes copies of the genes*. Because if the two chains unwind into two separate chains, and if each chain then makes another chain come together on it, then because A always goes with T, and G with C, we shall get two copies where we had one before. (Emphasis in original).

The discovery was published one month later [4]. Near the end of their famous one-page-long article, Watson and Crick observed: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material" [4].

Four years later at a symposium held at University College London, Crick described his "sequence hypothesis" and principles relating to the transfer of genetic information [5]. The latter was referred to in his notes and in later writings as the "Central Dogma" [6]. The sequence hypothesis states [6]:

> The specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and that this sequence is a (simple) code for the amino acid sequence of a particular protein.

The Central Dogma states that [6]:

> once 'information' has passed into protein it *cannot get out again*. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the *precise* determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein. (Emphasis in original)

In 1970, in response to a challenge against the central dogma based on the discovery of reverse transcriptases (i.e., RNA-dependent DNA polymerases), Crick explained, reaffirmed and clarified the central dogma [7]:

> The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid.

This later version is a combination of his sequence hypothesis and the original central dogma, even though Crick himself thought they were distinct, being a positive and a negative statement, respectively [7]. Furthermore, Crick emphasized that the central dogma [7]:

> says nothing about what the machinery of transfer is made of… says nothing about control mechanisms… and was intended to apply only to present-day organisms, and not to events in the remote past, such as the origin of life or the origin of the code.

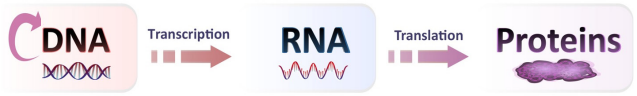

Figure 1. A schematic view of the central dogma. Graphically stylized version, based on Watson's original figure. **doi:**10.5048/BIO-C.2024.3.f1

Watson's Central Dogma

Watson popularized a simplified version of the central dogma via his widely used *Molecular Biology of the Gene* textbook (now in its eighth edition). He illustrated the dogma with a figure (on which Figure 1 is based) and included the following description [8,9]:

> The arrows indicate the directions proposed for the transfer of genetic information. The arrow encircling DNA signifies that DNA is the template for its self-replication. The arrow between DNA and RNA indicates that RNA synthesis (called transcription) is directed by a DNA template. Correspondingly, the synthesis of proteins (called translation) is directed by an RNA template. Most importantly, the last two arrows were presented as unidirectional; that is, RNA sequences are never determined by protein templates nor was DNA then imagined ever to be made on RNA templates.

Consequently, in most people's minds, the central dogma describes the unidirectional transfer of genetic information from DNA to RNA to proteins [5]. Indeed, while not from Crick himself, a popular extension of the central dogma is: DNA makes RNA, makes proteins, makes us. In molecular biologist and Nobel laureate Walter Gilbert's words: "Three billion bases of DNA sequence can be put on a single compact disc and one will be able to pull a CD out of one's pocket and say, 'Here is a human being; it's me!'" [10].

Pinning Down the Central Dogma

In spite of Crick's clarifications, and in no small part due to Watson's reformulation, significant confusion persists about precisely what the central dogma is and what it actually means in practice. Philosopher M. Polo Camacho observes that "the *Central Dogma* is not a unitary thesis with widely accepted meaning. The Dogma's interpretations abound..." [11]. Camacho identifies five interpretations of the Central Dogma [11]:

1. "DNA specifies RNA, which specifies protein. This view bears a close resemblance to a formulation often attributed to James Watson, which states roughly that DNA makes RNA makes protein." This view is the most common interpretation and is often used in textbooks.
2. "DNA [is] the most significant cause contributing to protein synthesis." This view is consistent with a gene-centric view of the organism.
3. The third version "concerns the transfer of information," saying "that DNA alone carries information for protein."

4. The fourth version is essentially Crick's formulation, stating that "the transfer of information from protein to protein and from protein to DNA is not possible."
5. The final version "concerns the heritability of traits and has been interpreted by many as synonymous with a rejection of the inheritance of acquired traits. This conception of the Central Dogma says that genes cannot be modified by environmental factors in a way that produces heritable traits in the organism."

Over the decades, many researchers have published findings challenging the central dogma. In response, defenders have emphasized the narrow scope of Crick's original version, accusing such researchers of misunderstanding the central dogma [12,13]. As well deserved as those responses may be, it remains a historical reality that Watson's version is far more well known, even among professional biologists, despite the common (and incorrect) attribution of Watson's formulation to Crick.

In our research we have observed, as does Camacho, that Crick's narrow version of the central dogma "has seen little discussion in the literature" [11]. Indeed, while defenders of Crick's narrow articulation correctly point out the discrepancy between Crick's and Watson's formulations,[1] they rarely offer any substantive analysis as to the relevance of Crick's narrow articulation to the actual workings of biology.[2] The ongoing defense of Crick's formulation with its narrow focus on the prohibited transfer of residue-by-residue information out of proteins, comes at the considerable expense of reduced biological relevance.

In addition to these multiple and often conflicting interpretations, some have suggested that Crick's version of the central dogma—despite other potential limitations—is technically correct from an information-theoretic standpoint and that the dogma should be understood only in this very limited sense. For example, Eugene Koonin, Senior Investigator at the National Center for Biotechnology Information, argued for the central dogma based on different kinds of information [15]. According to Koonin, the central dogma "emerges due to the transition from the digital information carriers, nucleic acids, to analog information carriers, proteins, which involves irreversible suppression of the digital information" [15].

In another approach, physicist and information theorist Hubert Yockey noted that the genetic code belongs to a class of codes known as non-isomorphic codes [17].[3] An isomorphic code has "a one-to-one mapping from one alphabet to the other"[17]. In contrast, in the case of the genetic code which uses a 64-position (4^3) table that is translated to an amino acid table with only 22 positions (20 plus start and stop)[4], it is usually not possible to reconstruct based solely on the resulting amino acid the precise 3-nucleotide codon that was originally present in the DNA sequence.[5] Figure 2 provides a simple example of an isomorphic code vs. a non-isomorphic code.

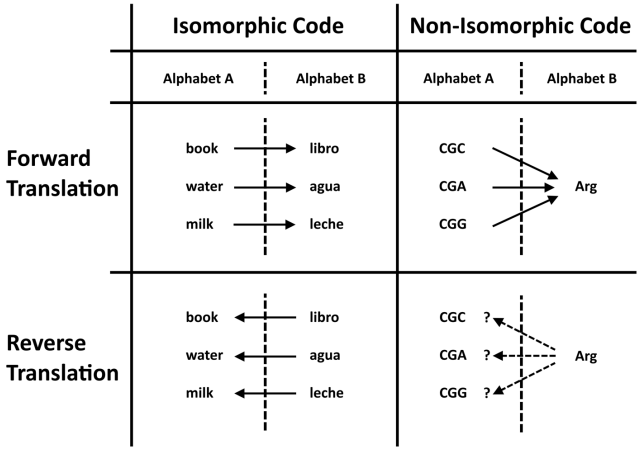

Figure 2. A simple example of an isomorphic code vs a non-isomorphic code. In the case of an isomorphic code, there is a one-to-one mapping of symbols from one alphabet to another, permitting a direct reverse translation to the original message. However, in the case of a non-isomorphic code, there is not a one-to-one mapping, thus preventing a direct reverse translation to the original message. Of course, this is an idealized example for illustrative purposes. Between any two natural languages there are many non-isomorphic correspondences, and even within the genetic code there are arguably one or two isomorphic sequences (e.g., AUG<-->Met). Nevertheless, it remains true that the genetic code is broadly non-isomorphic. **doi:**10.5048/BIO-C.2024.3.f2

The above discussion assumes a faithful translation process. However, there is an additional way in which information might be lost during any translation or transmission process. Any non-idealized system—meaning, we note, essentially all physical systems—will be subject to certain errors or "noise" in the process of translation, transmission, receipt and decoding.[6] Yockey's work, building on the pioneering work of Claude Shannon [20], emphasized the transmission of messages across a communication channel and, thus, focused heavily on the potential introduction of noise and the need for error correction.[7]

[1] Dan Graur provides a detailed and entertaining review of the historical aspects [13].

[2] In a subsequent paper, Camacho argues that Crick's narrow negative formulation of the central dogma undercuts its usefulness and that under such an approach "the Dogma amounts to a triviality" and "of no practical significance to science" [14]. In contrast, Eugene Koonin calls Crick's central dogma "the great biological exclusion principle" [15]. Koonin's opinion on Crick's central dogma seems to have changed over time. In 2012, he argued against it based on prions [16], while in 2015, he argued for it based on his understanding of information [15], which we address hereafter.

[3] If two alphabets are not isomorphic, then "no code exists such that the destination can send messages from alphabet B to the source in alphabet A. Thus, the genetic code, like all codes between probability spaces that are not isomorphic, has a Central Dogma" [17].

[4] While the genetic code is often thought of as representing 20 amino acids, plus start and stop, various exceptions have been discovered. For example, some organisms use the stop codons (normally used in other organisms to terminate translation) to code for amino acids [18,19].

[5] Note that the inability to reverse-translate a message in the case of a non-isomorphic code is based solely on examination of the code and the alphabets in question. It may be possible to reverse-translate the message if additional outside information is available. Yockey acknowledges that overlapping genes would provide additional information that could potentially enable the original genetic sequence to be recovered [17].

[6] This principle is underscored by the ubiquity of multiple error-correction mechanisms, in everything from our communication devices, to our computer storage systems, to living organisms.

[7] In addition to being non-isomorphic, the genetic code is also a block code, meaning that all "letters" (by which Yockey meant codons) of the genetic alphabet are of the same length. These equal length codon blocks are a key characteristic that enables important error correction capabilities to be employed in the face of genetic noise. Readers will be familiar with a similar block approach employed in modern computing: the 8-bit byte. Yockey's own work in this area is foundational to our understanding of the connection between information theory and biology but is beyond this scope of the present discussion.

Yockey notes that, "Although the genetic system is remarkably accurate, genetic noise…does cause some codons to be translated or decoded incorrectly" and that "in particular, that noise may be introduced in the form of a mistranslated mRNA or a mischarged tRNA" [17]. However, this type of information loss relates to communication of the message under the practical constraints of the physical system in question. It does not constitute a theoretical absolute that necessitates loss of the information as in the case of a non-isomorphic code as discussed above. It is one thing to observe that information *may* be lost in translation in a given instance due to system limitations and potential noise. It is another thing to categorically state, as Crick did, that as a matter of principle, "information cannot be transferred from protein to either protein or nucleic acid," [7] and that such a transfer is "impossible" [6].

Crick's formulation of the central dogma does not appear to have resulted so much from information-theoretic principles of non-isomorphic block codes, as from how biology was thought to work at the time. The proposition that the central dogma was intended to be understood as merely a narrow statement of the non-isomorphic nature of the genetic code is problematic, both practically and historically.

First, Crick seems not to have given serious consideration to the possibility of synonymous codons at the time he formulated the central dogma. Instead, together with John Griffith and Leslie Orgel, he proposed a version of the genetic code that had 20 'sense' and 44 'nonsense' combinations of nucleotides,[8] to provide a 1:1 match to the number of known amino acids in proteins—in short, an isomorphic code [6].

Second, even with our knowledge of 64 codons today, from a practical standpoint the alleged "loss" of information in going from 64 codons to 20 amino acids is something of a definitional point, rather than a substantive one. Setting aside new discoveries that are beginning to cast doubt on the long-held assumption that so-called "synonymous" codons are in fact truly synonymous[9] and granting for purposes of discussion that such codons are truly synonymous for all purposes in the cell, then replacing, for example, CGC (arginine) with CGA (arginine) would not result in any change of information, either loss or gain. Therefore, if we were to start with an arginine amino acid residue in a protein and reverse translate it back into a DNA codon, it would make no difference if it were reverse-translated as CGC or CGA (or any of the other 4 arginine codons), precisely because—on the assumption of fully synonymous codons—the only information contained in the DNA codon is whether it identifies arginine as the assigned amino acid. It would therefore make no difference which codon for arginine were used.

Thus, if we are interested in the substantive underlying information—the information that actually matters for cellular function—the alleged loss of information in reverse-translating a particular codon to its assigned amino acid is illusory. The only information of interest originally present (namely, which amino acid is assigned) can be easily recovered by reverse translating the amino acid into any of the relevant codons for that amino acid. The only information arguably "lost" to us would be the precise underlying DNA codon sequence that had originally been used to produce the amino acid—an interesting piece of historical data for the curious information theorist, perhaps, but irrelevant to the actual workings of the cell.[10]

Ironically, then, to the extent that codons long assumed synonymous turn out not to be synonymous after all, this narrow interpretation of the central dogma's proposed loss of information in the translation from DNA to proteins is called into question. At the same time, if the codons turn out to be truly synonymous, then such an interpretation of the central dogma is true as a triviality, without any real-world consequences for biology. In either case, this situation seems at odds with Crick's own view of the importance of the central dogma to biology. Crick said if any cell were found that could reverse the information flow of the central dogma, it "would shake the whole intellectual basis of molecular biology" [7]. Thus, Crick's view of the central dogma seems to be based on something other than the information-theoretic aspects considered above.

Third, at the time some believed that the information flow from DNA to RNA was unidirectional. Indeed, as quoted in the prior section, Watson emphasized this then-commonly-understood unidirectional flow of information. However, if the multi-directional flow of information were actually prohibited in nature, it would not be due to information-theoretic considerations. It is a straightforward matter to determine the underlying DNA sequence from a previously transcribed RNA transcript (setting aside modifications, such as alternative splicing and RNA editing, discussed later). Indeed, the discovery of reverse transcriptases was considered by many to challenge the central dogma, prompting Crick to respond by refocusing attention on the "residue-by-residue transfer of sequential information" from nucleic acids to proteins. Further, although he acknowledged the existence of certain prions, Crick explicitly excluded the possibility of protein-to-protein information transfer, although such transfer is not prohibited by information-theoretic considerations.[11]

[8] One such genetic code proposed by Crick, Griffith and Orgel in 1957 consisted of what we might call limited triplets. Specifically, they proposed three types of triplets, each one of which was limited to a certain number of possible 'sense' codons—2, 6 and 12, respectively—for a total of 20. The trick in reducing the allowed number of combinations to 20, lay in proposing various restrictions on which nucleotides could occupy each position within the three types of triplets. Of course, this turned out to be incorrect, but the ingenuity in this case, and Crick's appropriate modesty, are to be commended. In 1958, he wrote: "Thus we have deduced the magic number, twenty [amino acids], in an entirely natural way from the magic number four [nucleotides]. Nevertheless, I must confess that I find it impossible to form any considered judgment of this idea. It may be complete nonsense, or it may be the heart of the matter. Only time will show" [6].

[9] Recent discoveries suggest that in some cases so-called "synonymous mutations" can in fact have significant, even lethal, consequences for the organism. Much is still left to be discovered in this area, but possible reasons may include changes in the timing of polypeptide folding (and hence the resulting protein structure) due to time differences resulting from the stochastic availability of the relevant tRNAs, as well as possible overlapping or bi-directional reading frames for other genes that include the codon in question. See discussion in Section III.C.

[10] It should be noted that this point relates only to the theoretical question of information loss in translating from the 64-codon alphabet to the 22-amino-acid alphabet, assuming truly synonymous codons. We are not suggesting that no processes in the cell could be impacted by a change of codon assignment. Indeed, they sometimes are. See discussion in Section III.C.

[11] Yockey explicitly points out this disconnect between information theory and Crick's central dogma, observing that the mathematical properties of codes "does *not* forbid the protein-protein transfer of information, which is forbidden by the Central Dogma as stated by Crick" (emphasis in original) [17].

Thus, in the decades since Crick and Watson's famous discovery and despite some lack of clarity, as well as critical differences between Crick's initial formulation and Watson's later additions, the central dogma has most commonly been understood as codifying the *unidirectional flow of information*, from the sequential genetic information source, through sequential mRNA, to the physical instantiation of that information in functional proteins in the organism, rather than as a technical statement about a theoretical information loss from translating a non-isomorphic DNA sequence. Despite the interesting nuances of code tables and the potential loss of information whenever translating from a larger code table to a smaller one (as interesting as that may be), we believe a focus on the *directional flow of information* best represents the underlying substance of the central dogma, which was based more on the understanding of biology at the time than on information-theoretic considerations.

It is in this sense of information flow that we will discuss, critique and propose an update to the central dogma throughout this paper. We argue that: (i) Crick's concept of information needs to be expanded, (ii) what information can be transferred must be considered, and (iii) the context-dependent mechanisms of transfer are essential. Further, in light of the possible 'loss-of-information' approach to the central dogma just reviewed, we also provide a novel view of information gain and loss during the transcription and translation processes.

Impact of the Central Dogma

It seems impossible to measure the exact impact of the discovery of the double-helix structure of DNA and of Crick's formulation of his sequence hypothesis and the central dogma. Crick equated their discovery of the double helix to *the* "secret of life" [21]. With collectors recognizing the significance of this remarkable discovery, Crick's 1953 letter to his son Michael sold for six million dollars in 2013, becoming the most expensive letter ever sold at auction [3]. Professor of Zoology Matthew Cobb referred to Crick's subsequent 1957 symposium lecture as "one of the most significant lectures in the history of biology" and as "a lecture that changed how we think" [5]. Historian of molecular biology Horace Judson remarked that Crick's lecture "permanently altered the logic of biology" [22]. Koonin called the central dogma the only exception to the "'ubiquitous exception' rule" of biology in which "the only actual rule is that there are no rules, i.e. exceptions can be found to every 'fundamental' principle if one looks hard enough" [16]. In his molecular biology textbook, Burton Tropp declared that the central dogma provides *the* theoretical framework for molecular biology [23]. The double helix has become the icon of biology, even of science itself, with textbooks, science magazines, and popular science articles regularly adorned with an artistic rendition of the famous helix. It is currently widely accepted that, given the nucleotide sequence of one strand of DNA, we can write down that of the other, and that, with the genetic codon table at hand, we can spell out the amino acid sequence of the encoded protein.

Despite the unquestionable influence of the central dogma within both popular understanding and scientific research, sequence information encoded in DNA is only part of the information inside a cell. Furthermore, no information can flow without the integrated function of matching DNA, RNA and proteins, and whether a DNA sequence encodes for anything (and if so, what) depends not only on the sequence of DNA but also on what exists inside and outside of the cell. In short, information coding and decoding are interdependent and are organism, cell-status and environment specific.

AN EXPANDED BIOINFORMATION CONCEPT

Some challenges have been raised against the central dogma, first after the discovery of reverse transcriptases as mentioned above, then after the discovery of prions, and later after the discoveries of other forms of epigenetics and different means of cellular communication [16,24–31]. These challenges relate to the understanding of biological information and what is possible in biological systems.

Crick sought to address the first two of these in his 1970 clarification of the central dogma, in which he emphasized that one of the "two central concepts" was "sequential information." In order to simplify the analysis and home in on this "principal problem" of sequence information, Crick found it "necessary to put the folding-up process on one side" and also to assume that "there was probably a universal set of twenty [amino acids] used throughout nature." He therefore focused on "information transfer from one polymer with a defined alphabet to another…, the directional flow of detailed, residue-by-residue, sequence information from one polymer molecule to another" [7]. Thus, Crick's "information" is "sequence information," which refers to the nucleotide sequence in DNA or RNA and the amino acid sequence in proteins. Crick also clarified that what he proposed was prohibited was the transfer of protein-coding-sequence information from proteins to proteins or from proteins to nucleic acids (whether DNA or RNA).

In this 1970 piece, Crick acknowledged the existence of reverse transcriptases and pointed out that his central dogma did not prohibit the flow of information from RNA to DNA, although he felt this was something that "does not occur in most cells, but may occur in special circumstances" [7]. He also acknowledged that questions about inheritance had arisen when prions were discovered. After all, there did not seem to be a direct correlation between the coding-sequence information in DNA and the structure of the newly propagated prions because a prion and its protein counterpart share the same nucleotide and amino acid sequence [32,33]. Thus, Koonin would later regard prions as a special example of epigenetics [16]. A prion's ability to impact the shape of another protein began to create doubts about Crick's proposal that information could not be transmitted from protein to protein (although in the case of prions it should be noted that it is not the precise sequence of the amino acids which is altered, but the three-dimensional structure of the protein) [16,26,27].

While reverse transcriptases and prions did not directly refute Crick's narrow prohibition on the transfer of sequence information out of proteins, the general sense of the dogma's relevance as a "central" principle of biology was starting to be called

into question. Although not explicitly stated by either Crick or Watson, two major assumptions appear to underlie the central dogma: First, the protein-coding sequence information of DNA contains all the inheritable substance that determines the phenotype of an organism. In Crick's words, "the main function of the genetic material is to control (not necessarily directly) the synthesis of proteins.…Once the central and unique role of proteins is admitted *there seems little point in genes doing anything else*" [6](emphasis added). Second, the meaning of the code is independent of the decoding mechanism. Again in Crick's words, the central dogma "says nothing about what the machinery of transfer is made of…[and] nothing about control mechanisms" [7]. He also postulated that "the way in which [proteins] are *synthesized* is probably uniform and rather simple, and…gene action...is also likely to be uniform and rather simple" [6].

The first assumption, also known as the "gene-centric" or "genetic determinism" view of life, has come under increasing criticism with additional discoveries [29,31,34–41]. For example, many inheritable epigenetic factors have been discovered, which have complicated the simple relationship between genotype and phenotype hypothesized by the central dogma [29,31,42]. The second assumption, however, has received less attention and will be part of our analysis below.

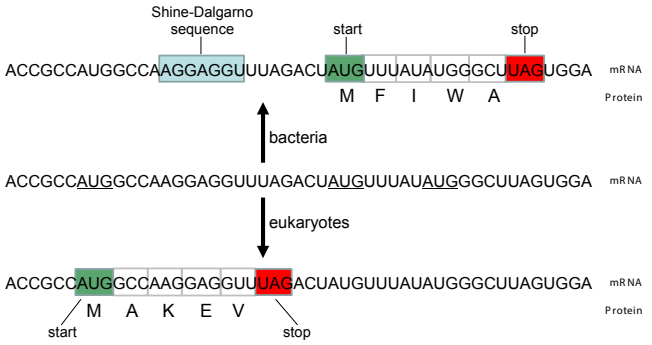

Figure 3. The same RNA may end up with two different proteins in bacteria and eukaryotes. Blue box: Shine-Dalgarno sequence; green box: translation initiation site; red box: translation stopping site. Top: The hypothetical mRNA would be used to code for a protein with amino acids MFIWA, based on a common mechanism of translation of bacteria like *E. coli*. The Shine-Dalgarno sequence is often important for translation initiation in bacteria. It hybridizes to an anti-Shine-Dalgarno sequence, which is reverse and complementary to the Shine-Dalgarno sequence, in the 16S rRNA. Bacteria use the AUG that is a few nucleotides downstream of the Shine-Dalgarno sequence as the translation initiation site. Bottom: The same hypothetical mRNA could be used to code for a protein with amino acids MAKEV, based on the mechanism of translation of eukaryotes like yeast. Eukaryotes normally use the first AUG from the 5′ end of an mRNA as the translation starting site. Note that the three AUGs (underlined in the middle panel) have different meanings: The first AUG is used as a translation starting site in eukaryotes but not in bacteria. In contrast, the second AUG is used as a translation starting site in bacteria but not in eukaryotes. The third AUG is part of two codons—the third base of the codon AUA and the first two bases of the codon UGG. Note also that the protein-coding regions are different in bacteria and in eukaryotes. **doi:**10.5048/BIO-C.2024.3.f3

The fluidity of biological information

Significantly, even though the "central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information," [7] there is, in fact, no fixed, one-to-one, biological "residue-by-residue transfer of sequential information." Instead, the meaning of a specific DNA segment depends on the system (i.e., the information transfer mechanism of a specific organism), the sequence context and the cellular context.

For example, the same RNA sequence can be translated into totally different proteins, depending on whether it is translated according to the bacterial or eukaryotic mechanisms, or not translated at all (Figure 3). Even in the same organism in the same gene, the same consecutive nucleotides (e.g., the underlined AUGs in Figure 3) can be used as a single triplet codon, parts of two separate codons or not as a codon at all, depending on its location within the RNA. Furthermore, what a particular AUG stands for can be changed by the deletion of a single base pair that is far away from that AUG, as is the case with the *dnaA* gene detailed later. Finally, the same mRNA can be translated as a functional or non-functional protein. For example, the mRNA encoding the bacterial release factor 2 (RF2), can be used to make a full length, functional RF2 or a truncated, non-functional RF2, depending on the concentration of RF2 [43]. In the latter case, a functional RF2 binds a normal UGA stop codon within the RF2 protein-coding region and stops translation, generating a truncated product. However, the ribosome translating the RF2 mRNA cannot be stopped at the UGA stop codon when the concentration of RF2 is low. Instead, the ribosome changes the reading frame by skipping the "U" part of the stop codon and making its "GA" into part of the next codon, resulting in the synthesis of the full-length protein. In this case, the normal translation "rule" of reading the nucleotides in a protein-coding region three-by-three is circumvented.

Different kinds of information

Five decades of biological research have revealed that there are different kinds of information inside cells. Koonin distinguished two kinds: (i) digital information, the one-dimensional sequence information contained in nucleic acids; and (ii) analog information, the three-dimensional structure of proteins [15].[12] Koonin's concept can, and we argue should, be extended since both nucleic acids and proteins have digital (sequence) information and analog (three-dimensional structure) information. The analog information should be extended to include not

[12] From an information-theoretic standpoint, information is fungible, in the sense that it can be encoded in any particular medium and subsequently retrieved, transmitted, translated into other symbolic systems, re-encoded in a different medium, and so on. Further, it is important to distinguish between (i) a physical object that *contains* information in the sense of an identifiable encoded symbolic representation of something outside of itself (such as we find in a book, in the daily newspaper, or in DNA) and (ii) a physical object that *performs a particular function* due to its physical characteristics (such as its three-dimensional structure, density, malleability, solubility, polarity, conductivity, position within a larger system, etc.). In a strict sense, the latter does not contain information merely because it possesses a particular physical structure, although information may of course have been used in the construction of a given physical structure. However, Koonin (like most other authors) fails to properly distinguish between these two, and a detailed analysis of these nuances is beyond the scope of this paper.

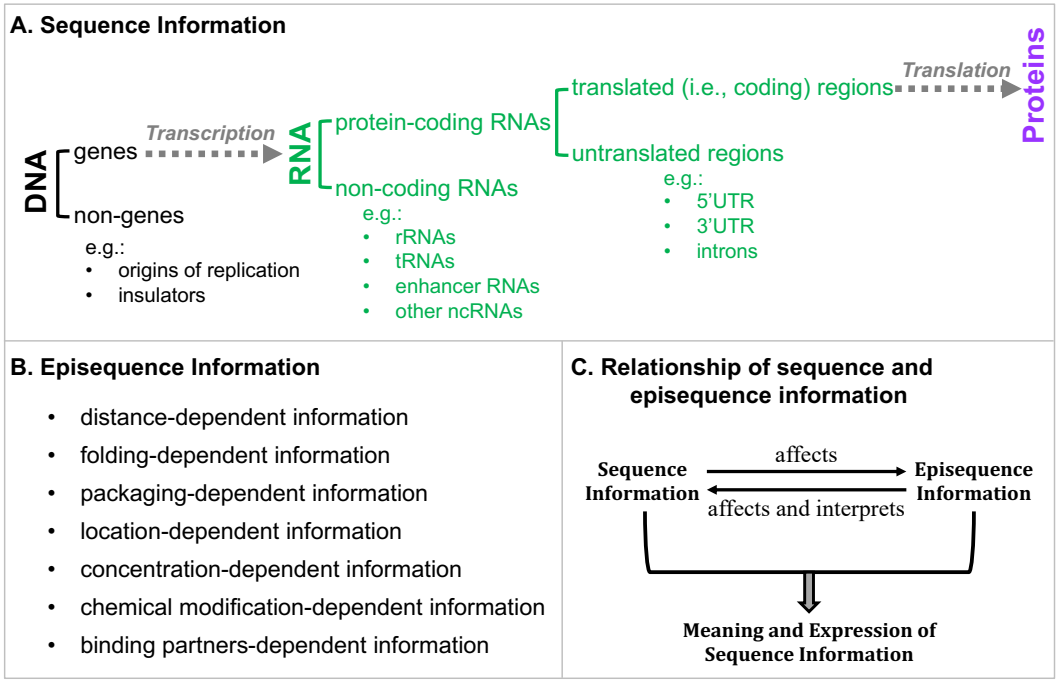

Figure 4. Different kinds of information inside cells. A: Coding-sequence-dependent information and its transfer. B: Some examples of episequence (coding-sequence-independent) information. C: Relationship of sequence and episequence information. **doi:**10.5048/BIO-C.2024.3.f4

only the three-dimensional structure of a given protein, but also protein localization, post-translational modification, network components (i.e., available binding partners in a specific cell) and cell metabolites—in essence, the broader cellular context.

We suggest naming such extended analog information "episequence information." This episequence information consists of all relevant cellular context information, including DNA, RNAs, proteins, metabolites and other molecules inside a cell at the specific moment under consideration, plus their chemical modifications, localization, structures, concentration and interactions. While the term "epigenetics" could theoretically refer to anything beyond DNA ('epi' (upon, above, beyond) and 'genetic' (DNA sequence)), the term has been something of a moving target and often refers to DNA or histone modifications in the context of chromatins [44–47]. As a result, it is necessary to have a more expansive and precise concept that includes other information within the cellular context outside of the actual sequence information of DNA, RNA and proteins.

Therefore, cells contain both sequence and episequence information (Figure 4A and 4B). As discussed below, the episequence information interprets or decodes the sequence information (Figure 4C). Correct interpretation occurs only when the coding and the decoding systems match. Together they determine whether the sequence information is meaningful, and, if it is, what it means and whether and how it should be expressed (i.e., transcribed or translated). On the flip side, sequence information also affects episequence information. For example, the structure of a protein can be greatly affected by its sequence. A specific DNA or protein modification may only occur to a nucleotide (for DNA) or an amino acid (for protein) within a specific sequence motif.

Different levels of information

In addition to these different kinds of information, cells also contain information at different levels. Researchers have discovered that different levels of information are embedded within DNA and RNA. The protein-coding level is only one of these. One remarkable discovery in recent years has been the discovery that so-called "silent mutations" can have vital impacts on cellular function. Under the common view implicitly enshrined in the central dogma that the only role of DNA was to code for a particular protein, it was long assumed that any codon for a particular amino acid was equivalent to any other codon for the same amino acid. Thus, for example, because GGA and GGG both code for glycine, it was assumed that these codons were equivalent and that a mutation from, say, "A" to "G" in the third position would have no possible consequence in the organism. Such a mutation would be a "silent" mutation, invisible to the internal workings of the organism and to its outward appearance.

This inherent "redundancy" of genetic codons has been a staple of biology education for decades, with much ink spilt debating the reasons for such redundancy and much speculation about its potential role in evolution. While it is far too early to suggest that each codon within the genetic code performs a unique role (and variations may exist between different organisms), it came as something of a shock to the received wisdom that some of the "silent mutations" were not, in fact, so silent after all (e.g., [48–53]). There is now good evidence

that, at least in some cases, recoding an amino acid with codons that code for the same amino acid but have different nucleotide sequences can disrupt the function of a segment of DNA or RNA at a level other than the mere designation of the particular translated amino acids. Such changes, previously thought to be completely neutral, can in fact be lethal, as several of the laboratories involved in yeast chromosome engineering have discovered with surprise [54–57]. These "silent" mutations can affect RNAs, including their structures, stabilities, localization and translatabilities [58–63]. They can also affect the properties of the proteins encoded by these RNAs since the translation speed and localization of an RNA can affect the folding, processing and function of the proteins it encodes [62–66]. These "silent" mutations can also impact the rate, location and characteristics of further mutation [67].

Another example of coding-sequence-independent information is information that is sequence independent but distance dependent. For instance, DNA between the upstream and the core promoter elements of the human ribosomal RNA gene appears to tolerate nucleobase substitutions but not significant alterations of its length. Researchers have observed that a removal of 44 bp (base pairs) between the two promoter elements reduces the promoter strength by 90% compared to the wild-type and an addition of 49 bp reduces promoter strength by 70% [68,69].

The dynamic nature of information

Our understanding of cellular information has extended beyond different kinds and levels of information to include the dynamic nature of information. While the changes of gene transcription and translation within the conditions of cell physiology, cell pathology and external conditions are more well known, the dynamics of DNA contents are often underappreciated. The sequence of DNA (or its code) and how this encoded information is expressed (or decoded), if at all, can be modified in cells in particular circumstances. DNA can be altered by point mutations, regional duplication or deletions, translocation, recombination, mobile DNA element insertions, whole genome degradation (as during apoptosis) or whole nuclear removal (as during the formation of human red blood cells). Biologist James Shapiro refers to some of these processes as "natural genetic engineering" and regards genomes as read-write instead of read-only systems [70–72]. Further, the decoding of DNA codes can be modified, in that the genes encoded within a genome can be silenced via dynamic epigenetic modifications of DNA and histones or DNA packaging (i.e., formation of heterochromatins)[73–75].

In addition to their synthesis, the stability of DNA, RNAs and proteins is tightly regulated within cells to meet the cell's needs, which can change in correspondence to its intra- and extra-cellular conditions. While many proteins are involved in DNA-damage repair, unwanted or damaged RNAs and proteins are often degraded. For example, while there are only two human protein synthesizers (ribosomes, one for translating nuclear-encoded genes and the other for translating mitochondrial-encoded genes), some 600 specialized human proteins (proteases) that break down proteins have been identified [76]. We will further address the dynamic nature of information later in this paper.

The system-dependent nature of biological information

The foregoing considerations lead us to a fundamental aspect of biological information; namely, it is system-dependent. Understanding the meaning and use of biological information is not just a matter of finding more information in more places (e.g., epigenetics, the sugar code, etc.). Instead, the meaning and use of information within a complex functional system—whether a cell, a tissue, an organ, or an entire organism—is inextricably linked to that particular system. This will become more obvious as we analyze the interdependence of the coding and the decoding systems in the following sections.

An inclusive concept of biological information

Figure 4 summarizes some of the information aspects that have been identified within a cell. As mentioned previously, cells contain both sequence and episequence information,. yet even genomic DNA contains both genes and non-genes. While a common misconception is that genes always code for proteins, many experts apply a broader definition and refer to a gene as any segment of DNA that is transcribed into a functional RNA molecule [77], while a non-gene is any DNA segment that does not encode a gene. Next, there are protein-coding RNAs and non-protein-coding RNAs (in short, non-coding RNAs or ncRNAs). Then there are translated and untranslated regions in a protein-coding RNA.

Crick's "residue-by-residue transfer of sequential information" covers only the protein-coding part of genomic DNA and only its sequence information. Therefore, the DNA-to-RNA-to-proteins view fails to account for much of the genetic information, especially in complicated organisms like humans. As detailed below, not all regions of genomic DNA encode genes, not all genes are protein coding, and not all regions of a protein-coding gene code for amino acids of that protein [78].

It has long been known that the regions of DNA that encode the most abundant and most stable RNAs—the ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs)—are not protein-coding. However, rRNAs and tRNAs were regarded as rare anomalies and were not directly addressed in the central dogma. Consequently, the protein-coding regions are generally referred to as coding regions, while the rest of the genomic regions are referred to as non-coding regions because they do not encode proteins. Crick and many others after him viewed the non-coding regions as "junk DNA" that had little or no effect on the organism [79–81].

Strikingly, numerous genome-scale studies have revealed that, instead of being a rare anomaly, a significant amount of DNA does not code for proteins (although its transcribed products may be involved in protein production and protein function), especially in higher organisms. For example, it is estimated that only 1.1–1.5% of the human genome encodes proteins [82–84]. Even though recent research has shown that protein-coding regions are more pervasive than previously

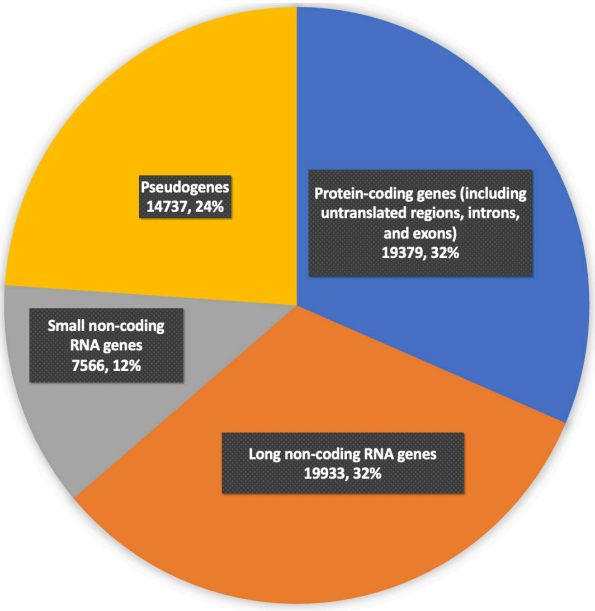

Figure 5: Relative percentages of different types of genes within the human genome. The numbers represent estimates of gene numbers and their *percentages* of the total number of transcribed regions, but not the *length* of the DNA encoding those genes. Data are based on GENCODE version 42 (GRCh38.p13, release date: October 2022, https://www.gencodegenes.org/human/stats_42.html). Note that not all human genomic DNA encodes genes, and so the above numbers do not represent percentages of the overall genome. Even though most regions of the human genome are transcribed, its exact gene-coding regions and their protein-coding status are not entirely known [87]. Indeed, researchers even differ in how a gene should be defined [88]. Note also that most nucleotides of protein-coding genes are located within the genes' untranslated regions and introns and, thus, do not code for amino acids in proteins. **doi:**10.5048/BIO-C.2024.3.f5

thought [85,86], there are many more non-protein-coding genetic sequences than protein-coding genetic sequences in the human genome [77,78](see Figure 5). Note that determining the exact percentage of gene-encoding regions and of each gene category in the human genome is still a challenge [87]. This is partly due to the complicated organization of genes: some genes are located within or overlap with other genes, on the same or on the opposite strand of DNA [88].

The non-protein-coding regions can be essential to the viability or reproduction of the organism. For instance, although they were previously widely regarded as junk DNA, introns can be vital for an organism. Deletions of certain introns are lethal for yeast [54,57] and humans [89], and failures in sex-specific, alternative-intron splicing prevent proper male and female differentiation and cause infertility in *Drosophila melanogaster* [90–92]. Mutations in the 5' or 3' untranslated regions of a protein-coding gene can alter its functions and cause diseases [62,63]. Furthermore, most of the functional elements identified by the ENCODE project and of the single nucleotide polymorphisms (SNPs) associated with disease by genome-wide association studies (GWAS) are localized within the non-protein-coding regions of the human genome [84].

Therefore, the central dogma, with its focus on the sequence information that codes for proteins, must be updated to account for the greater proportion of DNA that is functional, though not protein coding, as well as the different kinds, different levels, and dynamic and system-dependent nature of information. Furthermore, episequence information must be included as an integral part of biological information to truly reflect the molecular workings of cells.

INFORMATION CARRIERS EXAMINED

DNA

One widespread and persistent popular misconception is that DNA self-replicates, perhaps due to the influence of Watson's textbook and his description of the central dogma. As quoted in the historical section of this paper [8,9], he stated that "DNA is the template for its *self-replication*" (emphasis added). Although it is not clear whether Watson actually thought DNA could self-replicate or whether this was merely an unfortunate slip of the tongue, this misconception is often repeated, even though the fact that DNA cannot self-replicate is one of the most certain truths discovered by molecular biology studies. Therefore, we would like to emphasize that *DNA does not and cannot self-replicate.*

Shapiro used 12 Boolean propositions to illustrate various cellular information transfer events, and his first proposition is:

"DNA + 0 ==> 0" (Figure 1 of [31]).

Shapiro's point was not that DNA was "0" or "nothing" in the sense of being unimportant or of no value. Rather, he was emphasizing that DNA, on its own, does not provide biological function, whether that of replicating itself or even of producing RNAs or proteins. In fact, as we will show in the section on the interdependence of the coding and decoding systems, DNA does not carry out its life-critical function without specific RNAs and proteins geared toward working with that DNA—which can be termed "cognate" or "matching" RNAs and proteins. In other words, Shapiro's proposition could even be expanded as follows:

DNA + non-cognate RNAs + non-cognate proteins ==> 0.

As Denis Noble observes, "the cell, not its DNA, is the real replicator." DNA "replicates accurately only in a complete cell containing all the objective functionality that enable cells to be alive," relying "on an army of specialized proteins and on the lipid membranous structures for which there are no DNA sequences. Outside a living cell, DNA is inert, dead…. Active causation lies at the level of the cell, or of multicellular structures and organisms" [29]. A similar comment is made by Johannes Jaeger: "There is no privileged control by replicator genes: genetic causation always has to be interpreted in its organismic context" [39].

Indeed, the double helix is a double-edged sword; it not only "suggests a possible copying mechanism for the genetic material" [4] but also makes DNA replication difficult. In part, this is because cellular DNA molecules are very long, and the two strands are tightly wrapped around each other.

For example, the *Escherichia coli* genomic DNA is 4.6 million base pairs long and separation of the two strands, even for a few base pairs, requires a specialized enzyme (a helicase) and ATP. Furthermore, when helicase unwinds the two strands, the DNA ahead of the opening will become overwound and needs to be untangled by another specialized protein enzyme known as topoisomerase. Left unabated, the torsion resulting from the overwinding would quickly stop the ability of DNA polymerase or RNA polymerase to continue down the DNA strand during DNA replication or transcription, respectively. It could also result in permanent breakage or damage to the DNA strand. All told, more than 25 different proteins are required for minimally regulated replication of *E. coli* genomic DNA in vitro [93]. Regulated, faithful replication inside cells requires more proteins. A search in the OrthoInspector database [94] reveals that 56 and 133 different proteins have been identified to be involved in DNA replication in the bacterium *E. coli* and the eukaryote *Saccharomyces cerevisiae*, respectively. More significantly, most of the proteins involved in bacterial DNA replication are unrelated in amino acid sequence to those involved in eukaryotic DNA replication, a conclusion consistent with other studies (e.g., [95–100]).

In addition to this sophisticated suite of protein machinery required for proper DNA replication, the timing and extent of DNA replication is tightly regulated. A cell may replicate part of or the whole genome and may make multiple copies. A cell may use high fidelity polymerases or error-prone polymerases during replication and vary the extent to which replication errors will be corrected by the DNA repair system [101]). The cell's ability to sense its internal and external conditions, the decision about whether, when, how much, and how accurately to replicate DNA and the very execution of DNA replication all depend on the integrated functions of numerous RNAs and proteins in the cell and the protein-loaded cell membranes.

It is worth repeating that the genome of a cell is much more dynamic than expected—certainly far more dynamic than the central dogma had led us to believe. There are multiple ways cells can manipulate their DNA contents, and, in a growing number of cases, genomic DNA is turning out to be a read-write rather than a read-only system [102]. For example, Zhang and colleagues deleted all 100 copies of endogenous *S. cerevisiae* ribosomal DNA (rDNA), replaced them with a DNA fragment containing 1.2 or 2 copies of the rDNA unit carrying a hygromycin B resistance mutation and cultured the resulting strains in medium containing increasing amounts of hygromycin B [57]. After two weeks, a new rDNA cluster had been regenerated and the copy number was comparable to that of the wild type. This demonstrates that the cells have a mechanism to detect the copy number of rDNA and maintain the desired copy number. In *S. cerevisiae*, this can be accomplished by the upstream activating factor (UAF) for RNA polymerase I. In other words, UAF helps ensure ribosomal RNA (rRNA) production not only by the transcription of rDNA but also by controlling its copy number in the genome [103]. *D. melanogaster* has also been found to be able to adjust its rDNA copy numbers [104,105]. Van Hofwegen and colleagues found that aerobic citrate-utilizing *E. coli* (Cit^+) could be rapidly and repeatedly produced when wild type *E. coli* was cultured in a minimal medium supplemented with citrate, resulting from an active internal cellular process that generated additional *citT* and *dctA* loci followed by rearrangement of the DNA [106]. Specifically, the *E. coli* cells rearrange their chromosome in the presence of oxygen, moving around open reading frames and promoter elements to allow for expression of these genes and, therefore, synthesize enzymes that are needed for citrate metabolism that are normally synthesized only in the absence of oxygen. Strikingly, *E. coli* cells that lacked functional *citT* or *dctA* were not able to respond to the same environmental challenges to become Cit^+.

A growing number of studies of molecular mechanisms reveal that nucleotide changes are often part of a highly regulated process (meaning that these "mutations" are not mistakes, as previously thought) and are either activated or up-regulated temporarily when cells/organisms are stressed, as part of a proactive control system (reviewed in [101] and [107]). Furthermore, proteins actively survey DNA with the help of RNAs and either maintain the DNA intact or orchestrate needed alterations–even its total degradation (such as in the case of programmed cell death or in the development of anucleated human red blood cells).

In summary, it appears true that DNA is not synthesized using protein as a template (i.e., no reverse translation), and as such, proteins presumably played no role in determining the nucleotide sequence of the first strand of DNA in the first cell. However, proteins play a central role in determining the DNA contents of the descendants of that cell, as well as in the coding potential and usefulness of the DNA, as discussed below. As we will see, every protein functions in the context of other proteins, RNAs, and, in fact, in the whole cellular context of a cell.

RNA

Just as it is worth emphasizing that DNA cannot self-replicate, it is worth emphasizing that without a matching decoding system in place or the desired cellular conditions, genes that would otherwise be transcribed into an RNA product may be silenced or simply ignored (i.e., not be transcribed into an RNA product). About the same time when Crick formulated his central dogma of molecular biology, Jacob and Monod demonstrated that the transfer of genetic information could be blocked [108–110]. Many studies have shown that large regions of a genome (up to whole chromosomes) can be silenced by epigenetic information (e.g., [74,75]).

Basically, the cell, including its RNAs and proteins, determines whether a segment of DNA will be used as a template to make an RNA molecule, whether an RNA will be used to direct protein synthesis or whether an RNA will be actively and rapidly degraded, based on the internal and external conditions detected by the cell. In other words, whether a segment of DNA will be recognized as a gene and whether that gene will be used to generate a protein depends, among other factors, on the RNAs and proteins present in the cell at that moment. It often also depends on what is present in the surroundings outside the cell. Indeed, it is the overall cell (including the RNAs and

proteins inside the cell and in its membranes and metabolites) that determines whether a segment of DNA will be treated as a gene or a non-gene and whether that gene will be used to direct the production of any RNA or proteins.

The common tendency to refer to a given segment of DNA as a "gene" because it happens to code for a protein in a specific instance in one particular organism belies a simplistic view of the richness of biology—a view exacerbated by the central dogma's tidy emphasis on DNA to RNA to proteins. The study of molecular biology would be much simpler (although no doubt less interesting) if the sequence identification of a particular segment of DNA could tell us all we needed to know about what protein would be produced (if any), when and to what extent it would be produced and the function of the protein product. In reality, regulation of gene expression (i.e., transcription and translation) accounts for much of molecular biology research.

Studies of the ENCODE Consortium and others have uncovered that there are more genes that do not encode any protein than those that do in the human genome, as noted above [84,111,112]. Not only can a given genetic sequence code for multiple RNA transcripts that code for different proteins—resulting from different transcription or translation starting sites or stopping sites or from alternative splicing—but genes often overlap with each other. These studies have unveiled unexpected challenges in delineating genes. It seems that genes are having an "identity crisis" [88].

In short, molecular biology has uncovered a rich and complex array of components and systems that underlie the production of proteins from DNA. Rather than a simple, inevitable flow of information from DNA to RNA to proteins, it is now clear that no information can flow or be transferred without the interdependent, integrated, functions of the DNA, RNAs and proteins of the cell. The sequence-dependent information of DNA is somewhat like a recipe book. Its value and usefulness depend on the user. It is not a book to be read from cover to cover, conveying the same information to every reader. The chef can choose which recipe to use and modify the recipe as needed. Unfortunately, the DNA "book" does not contain punctuations or paragraph breaks that can be easily recognized by us humans, and its sentences (genes) often overlap and can sometimes be read backwards (i.e., encoded on the opposite strands of DNA) and sometimes need to be skip-read (e.g., RNA back-splicing [113], RNA trans-splicing [114], programmed frame-shifting [115], and programmed translational bypassing [116]).

Protein

One fact we would like to draw specific attention to regarding proteins is that synthesizing a polypeptide is not equal to making a protein that performs a desired function. A protein can perform an opposite function or no function at all depending on its location, posttranslational modifications or binding partners. For example, DnaA, the bacterial origin-of DNA-replication recognition protein, is active when it is ATP bound but inactive when it is ADP bound [117]. Furthermore, a protein without correct binding partners may be actively degraded as soon as it is made [118,119].

Discussions about the purpose of proteins in the cell tend to focus only on the *construction* of proteins for active cellular functions, especially when one is thinking of the origin of life. Therefore, it came as something of a revelation that each known genome also encodes many proteases, which are proteins that *break down* proteins. For example, 4.74% of the 24,194 human protein-coding genes are proteases, according to the renowned peptidase database MEROPS [120,121]. In addition, in several genomes studied to date, there are more nucleases encoded in each genome to break down DNA and RNA than there are polymerases to synthesize DNA and RNA [122–126].

These proteases and nucleases are like the cell's self-destructive molecular demolishers. Fortunately, inside cells they are kept under tight control so that they only break down excess, malformed, damaged or no-longer-needed DNA, RNAs and proteins to ensure that only those needed at the specific time and in the specific intra- and extra-cellular conditions are generated or maintained at the right levels. Some other well-known molecular demolishers are lysosomes, macrophages, osteoclasts and digestive organs.

Thus, proteins have many facets. Some of them help determine whether, when and how to synthesize DNA, RNAs, proteins, and many other molecules inside cells.Some of them break down what has been synthesized. Proteins help to interpret the messages encoded by DNA and RNAs; they help to alter the messages and messengers (e.g., through RNA editing, alternative splicing and alternative translation); they can silence genes epigenetically; they work with DNA, RNAs, proteins, and other molecules inside the cells; they are, in turn, subject to control by molecules inside the cells; and they are encoded in genomes whose protein-coding ability cannot be realized inside cells apart from proteins.

In summary, each genome encodes not only proteins to synthesize cellular components (including DNA, RNAs, proteins, lipids, and sugars) but also proteins to break them down. Both the builders and the demolishers are essential for cell viability and proliferation. If not synthesized or degraded at the correct time, correct levels, and in the correct locations, the products of genes, including those proteins that are builders, can kill the cell.

INFORMATION TRANSFER: A REALISTIC VIEW

If we limit the discussion of information in living systems to sequence information as Crick did, then the many discoveries about biological information in the intervening years do not technically challenge Crick's version of the central dogma [7,15]. We could still observe, for example, that when DNA is being replicated accurately, "we shall get two copies where we had one before," as Crick said. And when a gene is translated (assuming we ignore additional processing steps, such as alternative splicing and RNA editing), once we know the protein-coding region (i.e., the open reading frame (ORF)) of a gene, we can spell out the amino acid sequence of its encoded

protein, and that the resulting sequential amino acid "information cannot be transferred from protein to either protein or nucleic acid" [7].

Such a limited view of cellular information might allow us to maintain the perception that the central dogma is true, but it would be true only in an increasingly limited way and with an increasing number of exceptions. In light of new discoveries, we might be able to hold onto the proud tradition of the central dogma but only at the expense of accuracy and relevancy.

A more current understanding would recognize that the viability and reproduction of a cell depends on its ability to determine those conditional aspects, namely, (i) "when DNA is being replicated accurately," (ii) "when a gene is translated," and (iii) "once we know the protein-coding region of a gene." These are what the central dogma, in Crick's own words, "says nothing about" [7].

However, what the central dogma "says nothing about" matters. Countless experiments have shown that no information can be transferred without the coordinated interaction of DNA, RNA and proteins, as mentioned above and further detailed below. More importantly, both the meaning and the usefulness of a code depend on the decoding systems. The situation is similar to human languages. The four-letter word "gift" means a present in English, a poison in German and nothing but gibberish in Chinese. "Your room is on the first floor" points to a very different location in England (one level above the ground level) than in the United States (the ground level), even though both countries speak English.

Interdependence of the coding and the decoding systems

A key issue that has often been ignored by researchers who focus on the sequence information transfer underlying the central dogma is that the coding and the decoding systems need to match. That is, the DNA to be replicated and the molecular machines that replicate the DNA must match, and the genes to be transcribed and translated and the molecular machines that transcribe and translate the genes also must match.

For example, Craig Venter's team synthesized the entire one-megabase (Mb) genome of *Mycoplasma mycoides* in yeast, but the yeast cannot produce a *M. mycoides* cell using that cloned bacterial genome [127]. This is because the genes encoded in the cloned genome need to be transcribed and translated using the molecular machines that can recognize the encoded genes as such and that are present in *M. mycoides* and its close relative *Mycoplasma capricolum* [127–130].

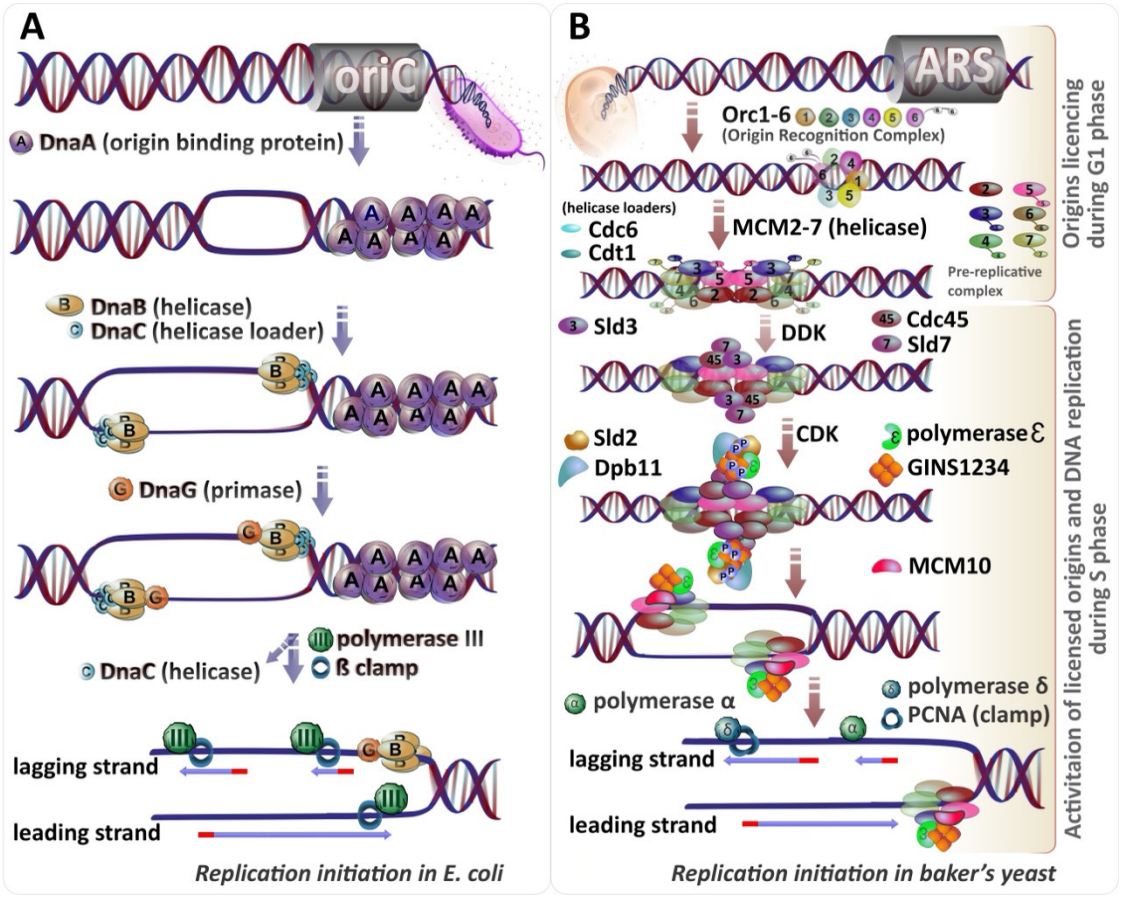

Figure 6. A comparison of DNA replication initiation in *E. coli* and *S. cerevisiae*. A: Initiation in *E. coli.* B: Initiation in *S. cerevisiae.* Note that in each case the proteins involved are unique to either *E. coli* or *S. cerevisiae.* **doi:**10.5048/BIO-C.2024.3.f6

The inability of a yeast cell to decode the bacterial *M. mycoides* genetic code is a consequence of the domain-specific information processing systems, including DNA replication, transcription and translation [97,100,131]. Figure 6 provides a comparison of DNA replication initiation in the bacterium *E. coli* and the eukaryote yeast *S. cerevisiae*. What is striking is not so much that the number of proteins involved are different (as important as that is) but that the identities of these proteins are different [97,100,131]. The proteins used for bacterial DNA replication are mostly bacteria specific and do not have known homologs in eukaryotes. Likewise, the proteins used for eukaryotic DNA replication are mostly eukaryote specific and do not have known homologs in bacteria. Consequently, to clone a bacterial genome in a eukaryotic cell, a eukaryotic origin of replication had to be artificially incorporated into the bacterial genome [127,128].

The transcription and translation machineries of bacteria and eukaryotes are also very different. For a piece of DNA to be recognized as a gene and transcribed by a bacterial cell, that DNA segment must be sandwiched between a bacterial promoter and a bacterial transcription terminator. Since the bacterial promoter will not be recognized as a promoter by the eukaryotic transcription machinery, the bacterial promoter must be replaced with a eukaryotic promoter to have the same stretch of DNA recognized as a gene and transcribed by a eukaryotic cell. The bacterial transcription terminator must also be replaced with a eukaryotic one to ensure proper termination of transcription. Furthermore, the same RNA transcript may be interpreted as encoding totally unrelated proteins by a bacterial cell and a eukaryotic cell due to the differences of bacterial and eukaryotic translation machineries [100,132], as discussed above (see Figure 3).

Interestingly, not only would a eukaryote cell have trouble decoding a bacterial genetic code (i.e., reading, interpreting and executing the instructions encoded in a bacterial genome), but even one particular bacterial cell may not be able to read the genetic instructions of another bacterial cell. For instance, adding the whole 3.5-Mb genome[13] of the photosynthetic bacterium *Synechocystis* PCC6803 into the 4.2-Mb genome of the mesophilic bacterium *Bacillus subtilis* did not enable *B. subtilis* to perform photosynthesis. The resultant cells could not even be cultured in the medium culturing *Synechocystis*, indicating that the added *Synechocystis* genome was not able to be used successfully by the host cell [133], despite the clear benefit the added genome might have provided in that medium. Although from our outside perspective we might be tempted to think that the added *Synechocystis* genome contained all the information necessary to enable the host cells to thrive in the culturing medium, the extensive sequence information contained in the *Synechocystis* genome seems to have been unrecognizable and of no value to the host cell. Similarly, a eukaryotic cell may not be able to read the genetic instructions of another eukaryotic cell. For example, mouse ribosomal RNA genes cannot be transcribed by the human transcription machinery, and vice versa [134-136].

It is worth mentioning that small differences in genetic codon meanings can have dramatic effects on the protein products a cell can produce from even the same RNA and whether the products can support the life of the cell. For example, four organisms were used to generate Venter's synthetic cell: *E. coli*, *S. cerevisiae*, *M. mycoides* and *M. capricolum* [127]. *E. coli* and *S. cerevisiae* use the "standard" genetic codon table, while *M. mycoides*, and *M. capricolum* use a slightly different genetic codon table in which the codon "UGA" is used as a tryptophan codon rather than as a stop codon. Consequently, even if a cloned *M. mycoides* gene (e.g., the *M. mycoides* DNA replication initiation gene *dnaA*) could be transcribed in *E. coli* or in *S. cerevisiae* and the resulting mRNA could be used to start translation at the same translation starting site, the translation would be stopped when the translation machinery reaches the *M. mycoides* tryptophan codon "UGA" because the same "UGA" indicates a translation stopping site in *E. coli* and *S. cerevisiae* (Figure 7A and 7B). If *M. mycoides* used the standard genetic codon table, then it would not be able to produce a functional DnaA protein to replicate its genome, and the organism would not be able to propagate.

To further illustrate the effect of subtle differences in genetic codon meanings among different organisms, we analyzed the open reading frames (ORFs), i.e., potential protein-coding regions, of both strands of the DNA covered by the protein-coding region of the *M. mycoides dnaA* gene. Using the default setting (minimal ORF length: 75 nucleotides; start codon: ATG) of the National Center for Biotechnology Information's Open Reading Frame Finder[14] program, five ORFs (four in the forward strand and one in the reverse strand) can be identified if the standard genetic codon table is used (Figure 7A). In contrast, with the codon table that *M. mycoides* uses, seven ORFs (three in the forward strand and four in the reverse strand) can be identified (Figure 7B).

The necessity of a functional DnaA protein for the life of *M. mycoides* and the dependence of the meaning of a specific nucleotide triplet on the presence of other nucleotides, even those from far away, is demonstrated by the *M. mycoides* cloning experiment of Venter's team [127]. Near the beginning of their cloning process, somehow a nucleotide deletion in the coding region of *dnaA* was introduced when they generated their 1-kb (kilobase) fragments. That mistake, unfortunately, was carried throughout the rest of the cloning process, leading to their failure to generate their desired synthetic cell. They had to add back the missing base pair to succeed.

Since they did not report which nucleotide was missed, we decided to remove one of the seven consecutive As of the 1353-nucleotide-long *M. mycoides dnaA* ORF from 319 to 325 to illustrate the potential effect of deleting one nucleotide (Figure 7C–E). This single nucleotide deletion causes a frame shift of *dnaA* (ORF1 in Figure 7B). Consequently, the combination of nucleotide triplets in the protein-coding region of *dnaA* and, thus, their encoded amino acids are changed. Furthermore, a

[13] The two ribosomal RNA genes of *Synechocystis* were not included because they are toxic to *Bacillus*. The toxicity may be due to the fact that they are close enough to the ribosomal RNA genes of *Bacillus* to be transcribed but different enough that it would interfere with the translational machinery of the host cell.

[14] https://www.ncbi.nlm.nih.gov/orffinder/

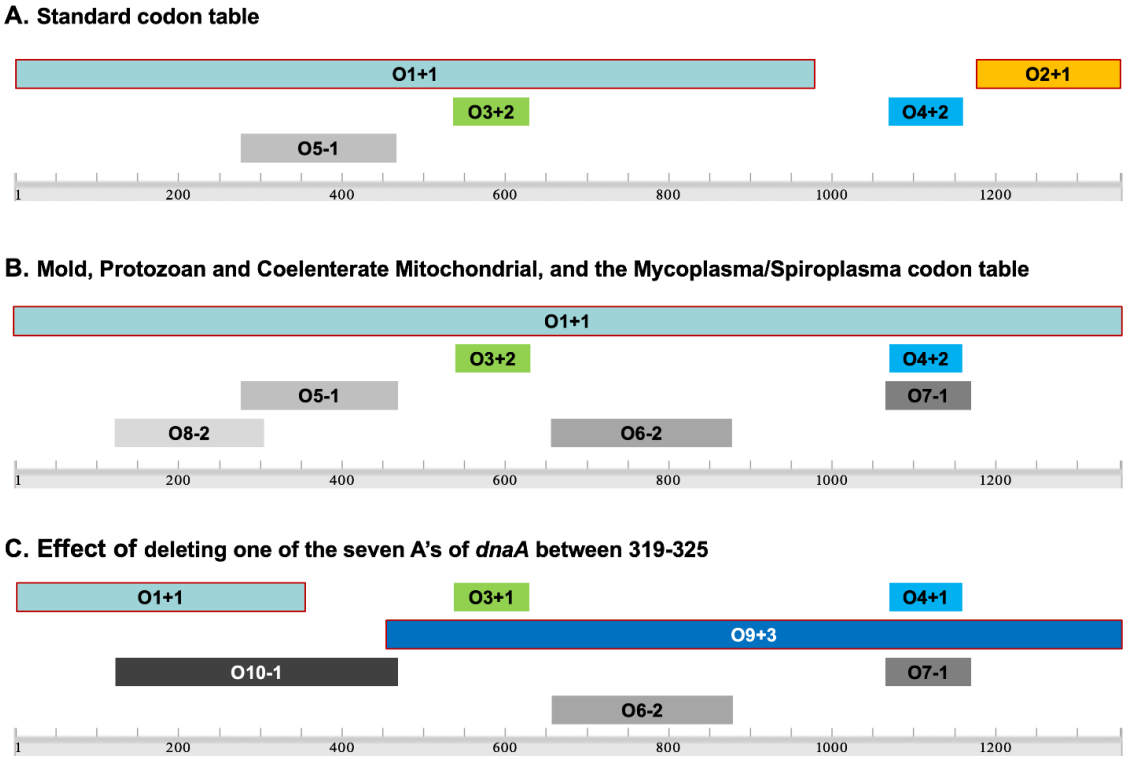

Figure 7. Changes of meaning of *M. mycoides' dnaA* protein-coding region. A 1353-nucleotide-long ruler is included at the bottom of Panels A–C. ORFs in six reading frames are indicated. ORFs on the plus strand (or forward strand) are color coded and their three reading frames are indicated with +1, +2, or +3. ORFs on the minus strand (or the reverse strand) are indicated with different shades of gray and their reading frames are indicated with -1, -2, or -3. ORFs that code for DnaA protein or part of it are outlined with red lines. A: ORFs according to the standard genetic codon table. B: ORFs according to the genetic codon table used by mold, protozoan and coelenterate mitochondria, and the mycoplasma/spiroplasma. C–E: Effect of deleting one of the seven As of *dnaA* between 319–325 on all the ORFs (C), on ORF1 (D), and on ORFs 5 and 8 (E). **doi:**10.5048/BIO-C.2024.3.f7

premature stop codon is introduced so that only a truncated DnaA protein product can be made, assuming the mRNA with this stop codon is translated instead of being degraded (Figure 7 C and D). Interestingly, this deletion also causes the loss of the stop codon of ORF5, resulting in the fusion of ORF5 and ORF8 and generation of ORF10 (Figure 7C and 7E). The deletion also resulted in another new ORF (ORF9). This ORF encodes the same amino acids as does the C-terminus of DnaA, but it is in reading frame 3, instead of reading frame 1 (Figure 7C).

In short, the meaning of three consecutive nucleotides is not fixed, as the well-known single standard genetic codon table might lead us to believe, because they can belong to the same codon or be divided into parts of multiple codons, depending

on the reading frame. Furthermore, if the nucleotide triplet is located outside of a protein-coding region, it will not be translated into any amino acid. Last, but not least, whether the triplet is within a protein-coding region, and, if it is, in which reading frame, may differ depending on the organism—how the organism determines whether a gene is protein-coding or not, how it determines the translation starting and stopping sites, and which genetic codon table it uses. This demonstrates the system-dependent nature of biological information and its lack of fixed, one-to-one, "residue-by-residue, [directional flow of] sequence information from one polymer molecule to another" [7] that Crick's central dogma envisioned.

Effects of cellular contexts

In addition to the species-specific match required for proper coding and decoding, a special match is often necessary within a species, such as a match between the stages (or cell cycle) of a cell or the match that exists between the type of cell and its genome and the molecules that decode the genome. In fact, we might think of the decoding system as the entire cell, with its many RNAs, proteins and metabolites. This is because it is the function of the cell as a whole, in the context of a specific tissue or environment, that determines whether the genomic DNA will be replicated in the first place and, if so, whether only part of the genome or the whole genome will be replicated, whether error-prone DNA polymerases will be allowed to participate in the replication (as in a stress-response situation) or only DNA polymerases with high fidelity, and what parts of the genome will be transcribed or translated. This cell-type and cell-status match occurs every day in every living organism, although we normally do not think of it that way. Evidence now suggests that mismatched tissue- or cell-specific transcription and translation is an important contributing factor for many diseases, including cancer and diabetes [137,138]. Imagine what would happen if muscle fibers were made in a nerve cell instead of a muscle cell. Or consider the pain of having bones grow in a place where they should not be.

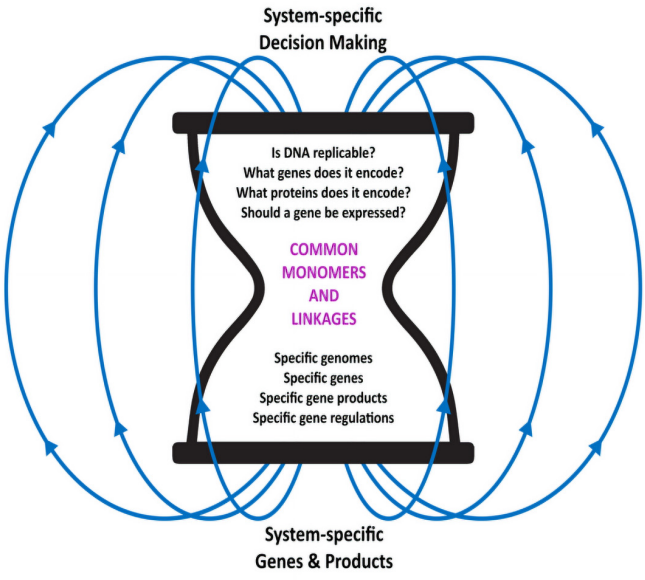

Figure 8. An Information Transfer Hourglass. Top: System-specific decision making; bottom: System-specific genes and gene products; middle: non-system-specific monomers, their linkages and linking chemical reactions. **doi:**10.5048/BIO-C.2024.3.f8

In summary, the presence of a particular DNA gene sequence does not guarantee the making of an RNA transcript. In fact, it is vital that this is so, since unregulated gene expression not only wastes resources, but can be deadly to the survival or reproduction of the organism in some cases. In addition, the presence of an RNA transcript does not guarantee the making of a protein, and the presence of a protein does not guarantee that it will perform an expected function. Thus, knowledge of the sequence of a genome does not enable one to predict the transcriptome (all the RNAs in a cell) or the proteome (all the proteins in a cell). Both a cell's transcriptome and proteome can change based on cell type and status and what is present in the environment. The only way of precisely knowing the transcriptome and the proteome of a cell is to independently examine them. In addition, it is well known that while determining the raw genome sequence of an organism is relatively easy now, annotating the genome (i.e., determining which parts of the genome actually encode genes) is quite challenging [139,140].

Effects of environmental conditions

In addition to the important match that must exist between a genome and the proper species of organism and the additional coordination that must exist within the same species between its genome and the relevant stages of cell development and the various molecules within the cell, an organism's coding-decoding system must also be coordinated to operate properly within a given environment. This includes an organism's effect on and actions within an environment, based on which molecules are inside the organism and which molecules are embedded within its membranes and in contact with the outside environment. This has been observed again and again since the dawn of molecular biology. For example, in a culture medium with both glucose and lactose, *E. coli* will normally not make galactosidase, a protein needed for lactose metabolism, until all the glucose is used up [110]. Then, after the lactose is used up, galactosidase production will again be stopped.[15]

Further, organisms are not necessarily passive responders to the environment. They can actively change the environment. Examples of an organism's effects on the environment include niche construction and the generation of wastes. Note that the wastes of one organism may be nutrients for another (i.e., the production and use of oxygen and carbon dioxide of plants and animals). Examples of environmental factors that an organism may have to deal with include nutrients, temperatures, pH and other organisms, such as pathogens, predators and siblings.

An information transfer hourglass

Combining the above discussions with the known chemical steps of DNA replication and gene transcription and translation, we propose an "Information Transfer Hourglass" (Figure 8

[15] Galactosidase may be generated abnormally but not be usable by the cell if the lac-repressor gene has been mutated and is no longer functioning properly or has been artificially modified and controlled.

Table 1. System-dependent and system-independent aspects of information transfer.

Processes	System-dependent	System-independent
Replication	• Whether a segment of DNA can be replicated • When and how a segment of DNA is replicated	• Using the parental DNA as a template • Linking deoxyribonucleotides together via phosphodiester bonds • Base pairing during DNA replication
Transcription	• Whether a segment of DNA is a gene • Whether the gene is transcribed • Locations of transcription starting and stopping sites • When and how the gene is transcribed	• Using the antisense strand of DNA as a template • Linking ribonucleotides together via phosphodiester bonds • Base pairing during transcription
Translation	• Whether an RNA molecule is protein-coding • Whether an RNA molecule is translated • Locations of translation starting and stopping sites • Codon status of a specific nucleotide (whether it belongs to a codon triplet within a protein-coding region and, if so, whether it is the 1st, 2nd, or 3rd nucleotide in that codon) • When and how the gene is translated	• Protein sequence determined by the sequence of mRNA and the translation machinery • Linking activated amino acids via peptide bonds • Each codon is three nucleotides long • Codons are consecutive and non-overlapping (exceptions: programmed frame shift and programmed translational bypassing, which are system-dependent) • The codon tables of organisms are almost the same

and Table 1). At one end is the determination of whether a DNA molecule is replicable, whether a segment of DNA encodes any gene, whether a gene is protein-coding, and whether the gene should be expressed. At the other end are specific gene products, including specific RNAs and proteins. Both ends vary with organisms, tissues, cell types, cell status and environmental conditions. The processes and decision points at one end of the hourglass influence and impact the resulting genes and gene products at the other end, and vice versa. Further, the constituents at the one end must match—biochemically, physically and in information-transfer capability—those at the other end, forming a single integrated, coordinated and coherent system.

In contrast, the middle of the information transfer hourglass appears to be the same for all currently known organisms.

Table 2. Information content changes during transcription and translation.

Processes	Information loss	Information gain
Transcription	• Epigenetically silenced genes • Intergenic regions (non-genes) • Enhancers and promoters of a gene that are not part of the main transcribed region of the gene and their connections with the gene	• Status of chromosomes • Availability of transcription factors, their concentration, localization, modification and interactors • Availability of specific metabolites and their concentration and localization • Presence or absence of specific environmental conditions
RNA processing	• Introns • External transcribed spacers • Internal transcribed spacers (e.g., the external and internal transcribed spacers of pre-ribosomal RNA)	• New combinations via RNA splicing • RNA editing • RNA nucleotide modifications • Non-sequence information (folding, localization, formation of complexes)
Translation	• 5′ untranslated regions • 3′ untranslated regions • Non-coding (or non-protein-coding) RNA	• Structures of RNAs • Availability of RNA-binding factors • Availability of translation factors, their concentration, localization, modification and interactors • Availability of specific metabolites and their concentration and localization • Presence or absence of specific environmental conditions
Protein processing or maturation	• Cleavage of signal peptides • Intein splicing • Deleting of other regions that are not in the mature proteins	• Non-sequence information (folding, localization, formation of complexes) • New combination via intein splicing • Posttranslational modification

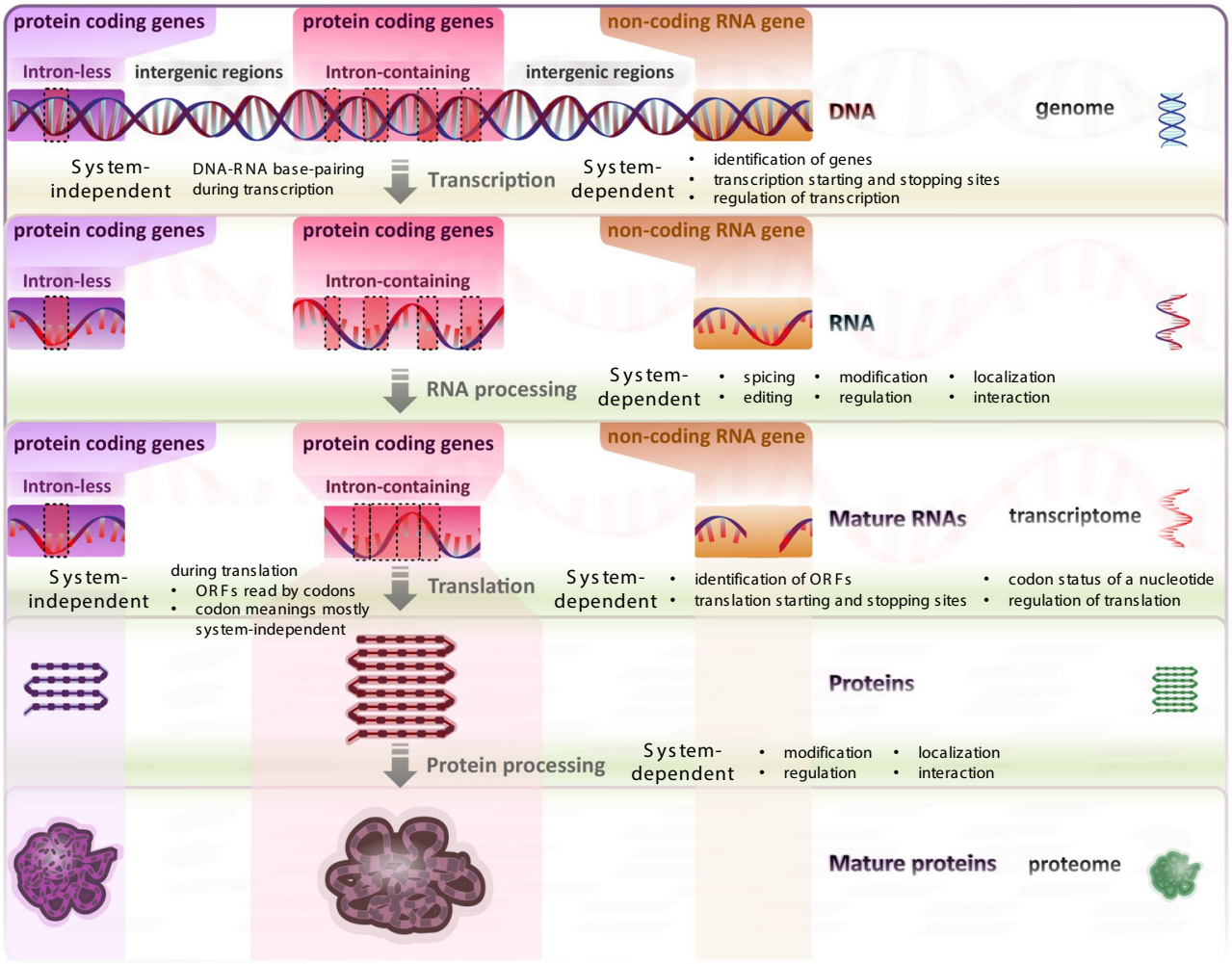

Figure 9. A schematic view of information transfer from DNA to RNA to proteins. Only the regions boxed with dashed lines are protein coding.
doi:10.5048/BIO-C.2024.3.f9

This includes compositional monomers of nucleic acids or proteins, the chemical linkages among the monomers in their corresponding polymers and the chemical reactions involved in monomer activation and polymerization. These chemical formation reactions of nucleic acids and proteins both include ATP-dependent activation of monomers and water-generating condensation. Interestingly, a similar information transfer hourglass has been observed on higher levels, such as during an organism's formation of morphological characters [141].

INFORMATION GAIN & LOSS: A NEW VIEW

Figure 9 shows some of the changes and processes involved in turning a protein-coding gene sequence into a functional protein. In addition, Table 2 provides an alternative way to view the standard transcription and translation decoding process, in light of the above-discussed concepts and the transfer of biological information, indicating the aspects of information loss and information gain alongside the basic steps in the process of protein production.

Note that by "information loss" in Table 2 we are not suggesting that information is somehow irretrievably lost within the cell. Rather, in the narrow context of coding-sequence information, as discussed by Crick in formulating the central dogma, the sequence in the next stage of the process (e.g., an mRNA vs. the underlying DNA sequence; or an amino acid sequence vs. the underlying mRNA) is missing information, in that it does not allow for reverse transfer of the earlier sequence from the later one. While this can be termed "information loss" and we have followed this convention in Table 2, in fact what is occurring during protein production is that *additional information* is being brought to bear by the cell *from outside the relevant coding-sequence* in order to complete the next stage of the production process. For example, while it is true that we cannot start with a protein and recreate the full genomic sequence underlying that protein, as reflected above in Figure 9, the reason is not so much that information has been "lost" as is often described, but rather that additional information has been brought to bear by the cell in order to read, decode and act upon that underlying genomic sequence in order to produce

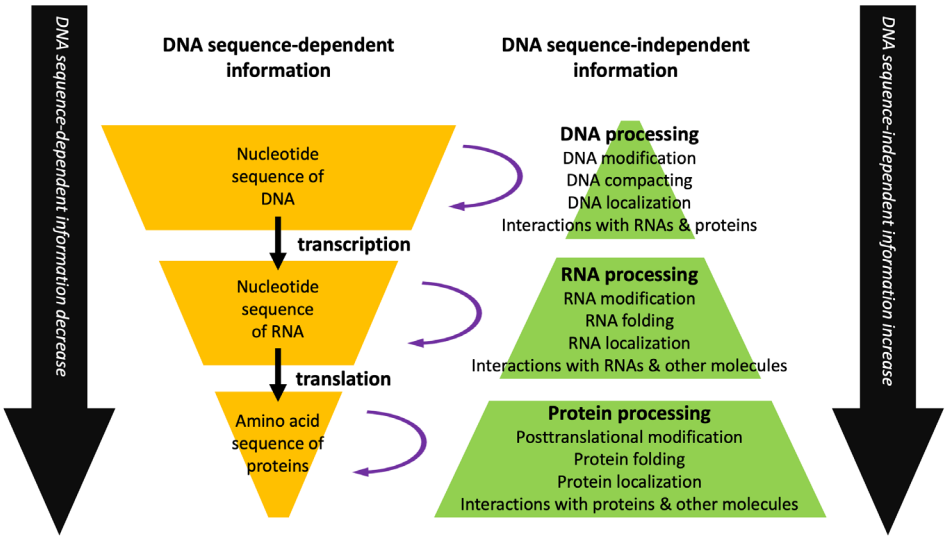

Figure 10. The information gain and loss funnels. Left: Loss of sequence information from DNA to RNA to proteins during transcription and translation. Right: Gain of episequence information from DNA to RNA to proteins during or after transcription and translation.
doi:10.5048/BIO-C.2024.3.f10

what is needed by the cell, in the right quantity, at the right time, and in the relevant context, as discussed throughout this paper. Note that each step of the information transfer can be interrupted artificially (as during molecular cloning) or naturally (as a physiological or pathological response).

On the flip side of this "loss" of sequence information from DNA to RNA to proteins (Figure 10 left), we also have an inverted information funnel in which episequence information and even additional sequence information is brought to bear during RNA processing and protein processing (Figure 10, right). For example, alternative splicing of intron-containing RNAs can produce new combinations of RNA segments, resulting in RNA molecules that encode different proteins. In addition, RNA editing can dramatically change the sequence of an RNA molecule and, hence, the amino acid sequence of the protein encoded by the corresponding DNA [142]. Further, the presence and/or concentration of a specific protein or RNA can alter the translation potential of an mRNA or its encoded protein products. For example, the production of full-length, functional *E. coli* release factor 2 depends on a translation frame shift that is controlled by the concentration of the release factor 2 itself, as described earlier [43]. In another case, the translation of *E. coli* translation initiation factor IF3 is autogenously controlled via a negative feedback loop acting at the level of initiation codon detection by IF3 [43]. We anticipate that many more such cases will be discovered.

However, Crick's central dogma completely ignores the episequence and epigenetic information of cells and their effects. As discussed above, it is the episequence information inside cells that determines how the sequence information is interpreted and whether it will be used to generate any RNA or protein products. This includes determining whether a segment of DNA encodes any genes, whether a gene is protein-coding, where the transcription or translation starting and ending sites are located, where a transcription or translation product should be transported, how the product should be processed and with which molecules the product should interact.

A REVISED CENTRAL DOGMA

More than 60 years ago at the dawn of the genetic age, a model of genetic information transfer was proposed that emphasized protein-coding nucleotide sequences in DNA as the fundamental source of information in the cell. Extensions of the model emphasized a one-way flow of information from DNA to RNA to proteins. For decades, the central dogma, as formulated by Crick and as modified by Watson, has influenced biological research and has impacted views of how information processing occurs in the cell and of which aspects of DNA are deemed functional or worthy of study.

Yet despite its originators' remarkable insights and contributions to modern biology, the central dogma is inadequate to account for either of the different kinds and levels of information inside a cell or the requirements and complexity of information transfer. In other words, it is an oversimplified model that cannot adequately describe the flow of biological information. Although perhaps not intended, the resulting reductionist, static, DNA-centric view of life has become a hindrance to our understanding of life. The exceptions and contradictions have finally reached the point where they can no longer be dismissed as occasional anomalies or be explained away with definitional clarifications of the central dogma. The central dogma's underlying and normally unstated assumptions of the primacy of protein-coding DNA sequence information within the cell and of the independence of the coding and decoding systems can no longer be considered a viable way of understanding the activity and role of information in biology.

Recently, Jafari et al. provided a rate-independent Boolean mathematical framework to model the information flow of the central dogma, based on present-day knowledge of molecular biology that includes additional molecular components (e.g., microRNA, as well as RNA- and protein-degrading factors) and the interconnected relationships between the various components [143]. They showed that this more detailed enhancement of the dogma is much more complex than the original central dogma and is also more consistent with the actual states of biological systems, such as the rarity of steady-state cell systems with high transcription and low translation in nature. This finding was later supported in a study by Hausser et al. using high-throughput measurement evidence focusing on two of the four essential rates of what the central dogma "says nothing about," namely, the rates of transcription, translation, mRNA decay, and protein decay [7,144]. Hausser et al. demonstrated that genes with high transcription and low translation are rare in four model organisms (*S. cerevisiae*, *E. coli*, *Mus musculus* and *Homo sapiens*). They proposed that this observation can be explained due to the trade-off between precision and economy in biological systems[16]. These studies highlight that more complete conceptual models (beyond the basic DNA-to-RNA-to-proteins approach) are required to enable us to understand and predict the dynamic behaviors of biological systems.

As a result of the insights of biology research over the past decades and the informational aspects we have discussed herein, we propose an update to the central dogma as follows (Figure 11):

> The central principle of molecular biology is regulated, dynamic, and system-dependent information coding and decoding.

Specifically, 1) no information transfer can occur without the interdependent and integrated function of cognate (or matching) DNA, RNA, and proteins; 2) proteins, with the help of RNAs, determine the maintenance, propagation and coding potential of DNA and RNA, as reflected in the Information Transfer Hourglass; 3) nucleotide sequence information can be lost but episequence information incorporated during transcription, translation and RNA or protein processing; and 4) information transfer is an active response of a cell to its internal and external conditions.

IMPLICATIONS: DISEASE & ORIGIN OF LIFE

The purpose of this paper has been to stimulate critical thinking about the central dogma of molecular biology to improve our understanding of biological information and information flow. More importantly, we proposed new ways of understanding the information processes at work in living systems, including the concepts of episequence information, an Information Transfer Hourglass, information gain and loss funnels and a revised central dogma of molecular biology. We trust this will serve as a more complete framework than the original central dogma to help facilitate our understanding of life and stimulate further research into the causes and treatment of diseases, as well as the origin of life.

For research on diseases, instead of focusing only on gene mutations, we need systemic analyses of episequence information. Also, we need to address the health effects of our diets, micro-organisms inside and outside our bodies, and the relationships among different organisms and our environment: in short, a more holistic view of life. In origin-of-life studies, the problem of the origin of the genetic code (or codons) should instead be recognized as the problem of the origin of entire genetic coding and decoding systems. The genetic code could not have been a "frozen accident" as Crick famously said [145], because the meaning of a nucleotide triplet is not fixed or "frozen." Instead, it is system- and context-dependent, as we have detailed throughout this paper.

Like Crick's central dogma, our revised version is also limited by our current understanding of the molecular mechanisms of life. As future research continues to expand our understanding of these processes, we look forward to learning and appreciating more about the intricacy, function, transfer and regulation of information in biology.

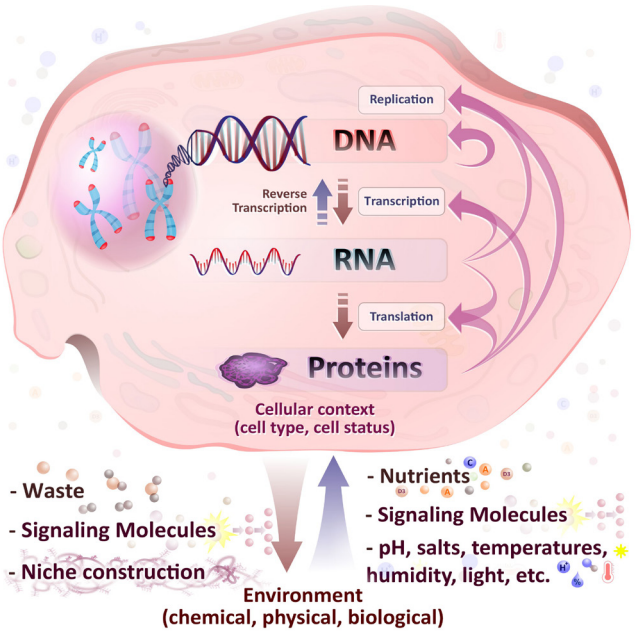

Figure 11. A schematic view of the revised central dogma. Note the cell-type- and cell-status-specific, environment-responsive, interdependence of DNA, RNA and proteins. Dashed arrows in the middle: sequence information transfer. Curved purple arrows at the right: kinds of molecules needed for the corresponding information transfer. Triangular arrows at the bottom: interactions between a cell and its environment. **doi:**10.5048/BIO-C.2024.3.f11

[16] Hausser et al. noted that an equivalent steady-state protein abundance in the cell could be achieved through either (i) a high transcription rate coupled with a low translation rate, or (ii) a low transcription rate coupled with a high translation rate. However, these two approaches impact cellular function differently. Specifically, a high transcription rate coupled with a low translation rate results in lower precision control of activity in the cell, but with better economy (due to the lower quantity of translation machinery required). Meanwhile, a low transcription rate coupled with a high translation rate results in higher precision, but at a higher economic cost. While most proteins they studied tended to lie within a predicted range of this trade-off between precision and economy, the authors noted several exceptions and suggested that these outliers might be due to bet hedging, greater need for precision vs. economy for response to specific growth conditions, or faster response times.

Acknowledgements

We would like to thank Rob Stadler, John Calvert, Ruth Wu, Mohieddin Jafari, and two anonymous reviewers for feedback, and Ehsan Zangeneh (https://sites.google.com/view/scimage) for helping with figures 1, 6, 9, and 11.

1. Karalis DT, Karalis T, Karalis S, Kleisiari AS, (2020) Genetically modified products, perspectives and challenges. Cureus. 12(3): e7306. **doi:**10.7759/cureus.7306
2. Pixley KV, Falck-Zepeda JB, Giller KE, Glenna LL, Gould F, et al. (2019) Genome editing, gene drives, and synthetic biology: will they contribute to disease-resistant crops, and who will benefit? Annu Rev Phytopathol. 57: 165-188. **doi:**10.1146/annurev-phyto-080417-045954
3. Copied letter from Francis Crick to Michael Crick (1953). Wellcome Collection, Francis Crick (1916-2004): archives. https://wellcomecollection.org/works/v3ndr2ux
4. Watson JD, Crick FH (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature. 171(4356): 737-8. **doi:**10.1038/171737a0
5. Cobb M (2017) 60 years ago, Francis Crick changed the logic of biology. PLoS Biol. 15(9): e2003243. **doi:**10.1371/journal.pbio.2003243
6. Crick FH (1958) On protein synthesis. Symp Soc Exp Biol. 12: 138-63.
7. Crick FH (1970) Central dogma of molecular biology. Nature. 227(5258): 561-3. **doi:**10.1038/227561a0
8. Watson JD, et al. (2008) Molecular Biology of the Gene. 6th ed..: Cold Spring Harbor Laboratory Press (Cold Spring Harbor, N.Y) p 32
9. Watson JD, et al. (2013) Molecular Biology of the Gene. 7th ed. Pearson p 33
10. Dobbs D Long Genome, Lively Book. 2012. Last accessed March 21, 2023 https://www.wired.com/2012/07/long-genome-lively-book/.
11. Camacho MP (2019) The Central Dogma Is Empirically Inadequate. Philosophy, Theory, and Practice in Biology. 11(6): 1-15. **doi:**10.3998/ptpbio.16039257.0011.006
12. Moran LA, Basic Concepts: The Central Dogma of Molecular Biology. 2007. Last accessed June 11, 2024. https://sandwalk.blogspot.com/2007/01/central-dogma-of-molecular-biology.html
13. Graur D The Fallacious Commingling of Two Unrelated Hypotheses: "The Central Dogma" and "DNA Makes RNA Makes Protein".2018. Last accessed June 11, 2024. https://judgestarling.tumblr.com/post/177554581856/the-fallacious-commingling-of-two-unrelated
14. Camacho MP (2021) Beyond descriptive accuracy: The central dogma of molecular biology in scientific practice. Studies in History and Philosophy of Science Part A. 86: 20-26. **doi:**10.1016/j.shpsa.2021.01.002
15. Koonin EV (2015) Why the Central Dogma: on the nature of the great biological exclusion principle. Biol Direct. 10: 52. **doi:**10.1186/s13062-015-0084-3
16. Koonin EV (2012) Does the central dogma still stand? Biology Direct. 7. **doi:**10.1186/17456150-7-27
17. Yockey HP (1992) Information Theory and Molecular Biology. Cambridge University Press (New York) pp. 96,97,106–107,113–114.
18. Alkalaeva E, Mikhailova T (2017) Reassigning stop codons via translation termination: How a few eukaryotes broke the dogma. Bioessays. 39(3). **doi:**10.1002/bies.201600213
19. Chen W, Geng Y, Zhang B, Yan Y, Zhao F, et al. (2023) Stop or not: genome-wide profiling of reassigned stop codons in ciliates. Mol Biol Evol. 40(4). **doi:**10.1093/molbev/msad064
20. Shannon CE (1948) A mathematical theory of communication. The Bell System Technical Journal. XXVII: 379-423, 623-56. **doi:**10.1002/j.1538-7305.1948.tb00917.x
21. BBC, 1953: Scientists describe 'secret of life', in *On This Day*. BBC.
22. Judson HF (1996) The Eighth Day of Creation: Makers of the Revolution in Biology, Commemorative Edition. Cold Spring Harbor Laboratory Press.
23. Tropp BE (2012) Molecular Biology: Genes to Proteins. 4th ed. Jones & Bartlett Learning, LLC (Sudbury, MA) p 22.
24. Nature (1970) Central dogma reversed. Nature. 226(5252): 1198-9. **doi:**10.1038/2261198a0
25. Peedicayil J (2005) DNA methylation and the central dogma of molecular biology. Med Hypotheses. 64(6): 1243-4. **doi:**10.1016/j.mehy.2004.12.022
26. Bussard AE (2005) A scientific revolution? The prion anomaly may challenge the central dogma of molecular biology. EMBO Rep. 6(8): 691-4. **doi:** 10.1038/sj.embor.7400497
27. Biro JC (2004) Seven fundamental, unsolved questions in molecular biology: cooperative storage and bi-directional transfer of biological information by nucleic acids and proteins: an alternative to "central dogma". Med Hypotheses. 63(6): 951-62. **doi:**10.1016/j.mehy.2004.06.024
28. Noble D (2012) A theory of biological relativity: no privileged level of causation. Interface Focus. 2(1): 55-64. **doi:**10.1098/rsfs.2011.0067
29. Noble D (2018) Central dogma or central debate? Physiology (Bethesda). 33(4): 246-249. **doi:**10.1152/physiol.00017.2018
30. Noble D (2008) Genes and causation. Philos Trans A Math Phys Eng Sci. 366(1878): 3001-15. **doi:**10.1098/rsta.2008.0086
31. Shapiro JA (2009) Revisiting the central dogma in the 21st century. Ann N Y Acad Sci. 1178: 6-28. **doi:**10.1111/j.1749-6632.2009.04990.x
32. Gibbons RA Hunter GD (1967) Nature of the scrapie agent. Nature. 215(5105): 1041-3. **doi:**10.1038/2151041a0
33. Griffith JS (1967) Self-replication and scrapie. Nature. 215(5105): 1043-4. **doi:**10.1038/2151043a0
34. de Lorenzo V (2014) From the selfish gene to selfish metabolism: revisiting the central dogma. Bioessays. 36(3): 226-35. **doi:**10.1002/bies.201300153
35. Noble D (2011) Differential and integral views of genetics in computational systems biology. Interface Focus. 1(1): 7-15. **doi:**10.1098/rsfs.2010.0444
36. Noble D (2017) Evolution viewed from physics, physiology and medicine. Interface Focus. 7(5). **doi:**10.1098/rsfs.2016.0159
37. Noble D (2015) Evolution beyond neo-Darwinism: a new conceptual framework. J Exp Biol. 218(Pt 1): 7-13. **doi:**10.1242/jeb.106310
38. Yaffe MB (2019) Why geneticists stole cancer research even though cancer is primarily a signaling disease. Sci Signal. 12(565). **doi:**10.1126/scisignal.aaw3483
39. Jaeger J (2021) The Fourth Perspective: Evolution and Organismal Agency. In: Organization in Biology, M. Mossio, Editor. 2021, Springer: Berlin.
40. DiFrisco J Jaeger J (2020) Genetic causation in complex regulatory systems: An Integrative Dynamic Perspective. Bioessays. 42(6): e1900226. **doi:**10.1002/bies.201900226

41. Laland KN, Uller T, Feldman MW, Sterelny K, Müller GB, et al. (2015) The extended evolutionary synthesis: its structure, assumptions and predictions. Proc Biol Sci. 282(1813): 20151019. **doi:**10.1098/rspb.2015.1019
42. Liu J, Dou X, Chen C, Chen C, Liu C, et al. (2020) N (6)-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. Science. **doi:**10.1126/science.aay6018
43. Betney R, de Silva E, Krishnan J, Stansfield I (2010) Autoregulatory systems controlling translation factor expression: thermostat-like control of translational accuracy. RNA. 16(4): 655-63. **doi:**10.1261/rna.1796210
44. Greally JM (2018) A user's guide to the ambiguous word 'epigenetics'. Nat Rev Mol Cell Biol. 19(4): 207-208. **doi:**10.1038/nrm.2017.135
45. Ung CY, Correia C, Billadeau DD, Zhu S, Li H (2023) Manifold epigenetics: A conceptual model that guides engineering strategies to improve whole-body regenerative health. Front Cell Dev Biol. 11: 1122422. **doi:**10.3389/fcell.2023.1122422
46. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. Genes Dev. 23(7): 781-3. **doi:**10.1101/gad.1787609
47. Deichmann U (2016) Epigenetics: The origins and evolution of a fashionable topic. Dev Biol. 416(1): 249-254. **doi:**10.1016/j.ydbio.2016.06.005
48. Patel UR, Gautam S, Chatterji D (2019) Unraveling the role of silent mutation in the omega-subunit of Escherichia coli RNA polymerase: structure transition inhibits transcription. ACS Omega. 4(18): 17714-17725. **doi:**10.1021/acsomega.9b02103
49. Ando H, Miyoshi-Akiyama T, Watanabe S, Kirikae T (2014) A silent mutation in mabA confers isoniazid resistance on Mycobacterium tuberculosis. Molecular Microbiology. 91(3): 538-547. **doi:**10.1111/mmi.12476
50. Ballard A, Bieniek S, Carlini DB (2019) The fitness consequences of synonymous mutations in Escherichia coli: Experimental evidence for a pleiotropic effect of translational selection. Gene. 694: 111-120. **doi:**10.1016/j.gene.2019.01.031
51. Singh R, et al. (2018) A Synonymous Mutation at Bovine Alpha Vitronectin Domain of Integrin Host Receptor (ITGAV) Gene Effect the Susceptibility of Foot-and-Mouth Disease in Crossbred Cattle. In: Advances in Microbiology, Infectious Diseases and Public Health: Volume 9, G. Donelli, Editor. 2018, Springer International Publishing: Cham. 41-45. **doi:**10.1007/5584_2017_47
52. Lupino KM,Romano KA, Simons MJ, Gregg JT, Panepinto L, et al. (2018) A recurrent silent mutation implicates fecA in ethanol tolerance by Escherichia coli. BMC Microbiol. 18(1): 36. **doi:**10.1186/s12866-018-1180-1
53. Toscano C, Raimundo S, Klein K, Eichelbaum M, Schwab M, et al. (2006) A silent mutation (2939G>A, exon 6; CYP2D6*59) leading to impaired expression and function of CYP2D6. Pharmacogenetics and Genomics. 16(10). **doi:**10.1097/01.fpc.0000236331.03681.24
54. Mitchell LA, Wang A, Stacquadanio G, Kuang Z, Wang X, et al. (2017) Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. Science. 355(6329). **doi:**10.1126/science.aaf4831
55. Shen Y, Wang Y, Chen T, Gao F, Gong J, et al. (2017) Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. Science. 355(6329). **doi:**10.1126/science.aaf4791
56. Wu Y, Li B, Zhao M, Mitchell LA, Xie ZX et al. (2017) Bug mapping and fitness testing of chemically synthesized chromosome X. Science. 355(6329). **doi:**10.1126/science.aaf4706
57. Zhang W, Zhao G, Luo Z, Lin Y, Wang L, et al. (2017) Engineering the ribosomal DNA in a megabase synthetic chromosome. Science. 355(6329). **doi:**10.1126/science.aaf3981
58. Frumkin I, Lajoie MJ, Gregg CJ, Hornung G, Church GM, et al. (2018) Codon usage of highly expressed genes affects proteome-wide translation efficiency. Proc Natl Acad Sci U S A. 115(21): E4940-E4949. **doi:**10.1073/pnas.1719375115
59. Kristofich J, Morgenthaler AB, Kinney WR, Ebmeier CC, Synder DJ, et al. (2018) Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. PLoS Genet. 14(8): e1007615. **doi:**10.1371/journal.pgen.1007615
60. Lebeuf-Taylor E, McCloskey N, Bailey SF, Hinz A, Kassen R (2019) The distribution of fitness effects among synonymous mutations in a gene under directional selection. Elife. 8. **doi:**10.7554/eLife.45952
61. Kudla G, Murray A, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. Science. 324(5924): 255-8. **doi:**10.1126/science.1170160
62. Mayr C (2019) What Are 3' UTRs Doing? Cold Spring Harb Perspect Biol. 11(10). doi:10.1101/cshperspect.a034728
63. Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, et al. (2020) Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. Nat Commun. 11(1): 2523. **doi:**10.1038/s41467-019-10717-9
64. Trovato F, O'Brien EP (2017) Fast protein translation can promote co- and posttranslational folding of misfolding-prone proteins. Biophys J. 112(9): 1807-1819. **doi:**10.1016/j.bpj.2017.04.006
65. Zhao F, Yu CH, Liu Y (2017) Codon usage regulates protein structure and function by affecting translation elongation speed in *Drosophila* cells. Nucleic Acids Res. 45(14): 8484-8492. **doi:**10.1093/nar/gkx501
66. Stein KC Frydman J (2019) The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. J Biol Chem. 294(6): 2076-2084. **doi:**10.1074/jbc.REV118.002814
67. Horton JS, Flanagan LM, Jackson RW, Priest NK, Taylor TB (2021) A mutational hotspot that determines highly repeatable evolution can be built and broken by silent genetic changes. Nat Commun. 12(1): 6092. **doi:**10.1038/s41467021-26286-9
68. Haltiner MM, Smale ST, Tjian R (1986) Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. Mol Cell Biol. 6(1): 227-35. **doi:**10.1128/mcb.6.1.227-235.1986
69. Learned RM,Learned TK, Haltiner MM, Tijan RT (1986) Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. Cell. 45(6): 847-57. **doi:**10.1016/00928674(86)90559-3
70. Shapiro J, Noble D (2021) What prevents mainstream evolutionists teaching the whole truth about how genomes evolve? Prog Biophys Mol Biol. 165: 140-152. **doi:**10.1016/j.pbiomolbio.2021.04.004
71. Shapiro, J.A. (2011) Evolution: A View from the 21st Century. FT Press.
72. Shapiro JA (1992) Natural genetic engineering in evolution. Genetica. 86(1-3): 99-111. **doi:**10.1007/BF00133714
73. Allshire RC, Madhani HD (2018) Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol. 19(4): 229-244. **doi:**10.1038/nrm.2017.119
74. Burgio E, Piscitelli P, Colao A (2018) Environmental carcinogenesis and transgenerational transmission of carcinogenic risk: From genetics to epigenetics. Int J Environ Res Public Health. 15(8). **doi:**10.3390/ijerph15081791
75. Dossin F, Heard E (2022) The molecular and nuclear dynamics of X-chromosome inactivation. Cold Spring Harb Perspect Biol. 14(4). **doi:**10.1101/cshperspect.a040196
76. Puente XS, Sánchez LM, Overall CM, López-Otín C (2003) Human and mouse proteases: a comparative genomic approach. Nat Rev Genet. 4(7): 544-58. **doi:**10.1038/nrg1111
77. Salzberg SL (2018) Open questions: How many genes do we have? BMC Biol. 16(1): 94. **doi:**10.1186/s12915-018-0564-x

78. Aspden JL, Wallace EWJ, Whiffin N (2023) Not all exons are protein coding: Addressing a common misconception. Cell Genom. 3(4): 100296. **doi:**10.1016/j.xgen.2023.100296
79. Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature. 284(5757): 604-7. **doi:**10.1038/284604a0
80. Bernardi G (2021) The "Genomic Code": DNA pervasively moulds chromatin structures leaving no room for "junk". Life (Basel). 11(4). **doi:**10.3390/life11040342
81. Ariel FD, Manavella PA (2021) When junk DNA turns functional: transposon-derived non-coding RNAs in plants. J Exp Bot. 72(11): 4132-4143. **doi:**10.1093/jxb/erab073
82. Lander ES, et al. (2001) Initial sequencing and analysis of the human genome. Nature. 409(6822): 860-921. **doi:**10.1038/35057062
83. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science. 291(5507): 1304-51. **doi:**10.1126/science.1058040
84. Consortium, EP (2012) An integrated encyclopedia of DNA elements in the human genome. Nature. 489(7414): 57-74. **doi:**10.1038/nature11247
85. Ruiz-Orera J, Albà MM (2019) Translation of small open reading frames: Roles in regulation and evolutionary innovation. Trends Genet. 35(3): 186-198. **doi:**10.1016/j.tig.2018.12.003
86. Liu Q, Peng X, Shen M, Quian Q, Xing J, et al. (2023) RibouORF: a comprehensive data resource of upstream open reading frames (uORFs) based on ribosome profiling. Nucleic Acids Res. 51(D1): D248-D261. **doi:**10.1093/nar/gkac1094
87. Amaral P, Carbonell-Sala S, De La Vega F, Faial T, Frankish A, et al. (2023) The status of the human gene catalogue. Nature. 622(7981): 41-47. **doi:**10.1038/s41586-023-06490-x
88. Mudge JM, Frankish A, Harrow J (2013) Functional transcriptomics in the post-ENCODE era. Genome Res. 23(12): 1961-73. **doi:**10.1101/gr.161315.113
89. Szafranski P, Yang Y, Nelson MU, Bizzarro Mj, Morotti RA, et al. (2013) Novel FOXF1 deep intronic deletion causes lethal lung developmental disorder, alveolar capillary cysplasia with isalignment of pulmonary veins. Hum Mutat. 34(11): 1467-71. **doi:**10.1002/humu.22395
90. Sosnowski BA, Belote JM, McKeown M (1989) Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. Cell. 58(3): 449-59. **doi:**10.1016/0092-8674(89)90426-1
91. Venables JP, Tazi J, Juge F (2012) Regulated functional alternative splicing in Drosophila. Nucleic Acids Research. 40(1): 1-10. **doi:**10.1093/nar/gkr648
92. Salz HK, Erickson JW (2010) Sex determination in Drosophila: The view from the top. Fly. 4(1): 60-70. **doi:**10.4161/fly.4.1.11277
93. Su'etsugu M, Takada H, Katayama T, Tsujimoto H (2017) Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. Nucleic Acids Res. 45(20): 11525-11534. **doi:**10.1093/nar/gkx822
94. Nevers Y, Kress A, Defosset A, Ripp R, Linard B, et al. (2019) OrthoInspector 3.0: open portal for comparative genomics. Nucleic Acids Res. 47(D1): D411-D418. **doi:**10.1093/nar/gky1068
95. Yao NY, O'Donnell ME (2016) Evolution of replication machines. Crit Rev Biochem Mol Biol. 51(3): 135-49. **doi:**10.3109/10409238.2015.1125845
96. Yao NY, O'Donnell ME (2021) The DNA replication machine: structure and dynamic function. Subcell Biochem. 96: 233-258. **doi:**10.1007/978-3-030-58971-4_5
97. Tan C, Tomkins JP (2015) Information processing differences between bacteria and eukarya—Implications for the myth of eukaryogenesis. Answers Research Journal. 8: 143–162.
98. Leipe DD, Aravind L, Koonin EV (1999) Did DNA replication evolve twice independently? Nucleic Acids Res. 27(17): 3389-401. **doi:**10.1093/nar/27.17.3389
99. Forterre P (2013) Why are there so many diverse replication machineries? Journal of Molecular Biology. 425(23): 4714-4726. **doi:**10.1016/j.jmb.2013.09.032
100. Tan CL (2022) Facts Cannot be Ignored When Considering the Origin of Life #3: Necessity of Matching the Coding and the Decoding Systems. Answers Research Journal. 15: 49–58.
101. Fitzgerald DM, Rosenberg SM (2019) What is mutation? A chapter in the series: How microbes "jeopardize" the modern synthesis. PLoS Genet. 15(4): e1007995. **doi:**10.1371/journal.pgen.1007995
102. Shapiro JA (2016) The basic concept of the read-write genome: Mini-review on cell-mediated DNA modification. Biosystems. 140: 35-7. **doi:**10.1016/j.biosystems.2015.11.003
103. Iida T, Kobayashi T (2019) RNA Polymerase I activators count and adjust ribosomal RNA gene copy number. Mol Cell. 73(4): 645-654 e13. **doi:**10.1016/j.molcel.2018.11.029
104. Lu KL,Nelson JO, Watase GJ, Warsinger-Pepe N, Yamashita YM (2018) Transgenerational dynamics of rDNA copy number in *Drosophila* male germline stem cells. Elife. 7. **doi:**10.7554/eLife.32421
105. Nelson JO, Watase GJ, Warsinger-Pepe N, Yamshita YM (2019) Mechanisms of rDNA copy number maintenance. Trends Genet. 35(10): 734-742. **doi:**10.1016/j.tig.2019.07.006
106. Van Hofwegen DJ, Hovde CJ, Minnich SA (2016) Rapid evolution of citrate utilization by *Escherichia coli* by direct selection requires *citT* and *dctA*. J Bacteriol. 198(7): 1022-34. **doi:**10.1128/JB.00831-15
107. Gottesman S (2019) Trouble is coming: Signaling pathways that regulate general stress responses in bacteria. J Biol Chem. 294(31): 11685-11700. **doi:**10.1074/jbc.REV119.005593
108. Pardee AB, Jacob F, Monod J (1958) The role of the inducible alleles and the consrtutive alleles in the synthesis of beta-galactosidase in zygotes of Escherichia coli. C R Hebd Seances Acad Sci. 246(21): 3125-8.
109. Pardee AB, Jacob F, Monod J (1959) The genetic control and cytoplasmic expression of "Inducibility" in the synthesis of β-galactosidase by E. coli. J Mol Biol. 1(2): 165-178. **doi:**10.1016/S0022-2836(59)80045-0
110. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology. 3(3): 318-356. **doi:**10.1016/S0022-2836(61)80072-7
111. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, et al. (2012) GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22(9): 1760-74. **doi:**10.1101/gr.135350.111
112. Consortium, EP, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 447(7146): 799-816. **doi:**10.1038/nature05874
113. Xu B, Meng Y, Jin Y (2021) RNA structures in alternative splicing and back-splicing. WIREs RNA. 12(1): e1626. **doi:**10.1002/wrna.1626
114. Lasda EL, Blumenthal T (2011) *Trans*-splicing. WIREs RNA. 2(3): 417-434. **doi:**10.1002/wrna.71
115. Tinoco I, Jr., Kim HK, Yan S (2013) Frameshifting dynamics. Biopolymers. 99(12): 1147-66. **doi:**10.1002/bip.22293
116. Lang BF, Jakubkova M, Hegedusova E, Daoud R, Forget L, et al. (2014) Massive programmed translational jumping in mitochondria. Proc Natl Acad Sci U S A. 111(16): 5926-31. **doi:**10.1073/pnas.1322190111
117. Katayama T, Kasho K, Kawakami H (2017) The DnaA cycle in *Escherichia coli*: activation, function and inactivation of the initiator protein. Front Microbiol. 8: 2496. **doi:**10.3389/fmicb.2017.02496
118. Gomez-Fabra Gala M, Vogtle FN (2021) Mitochondrial proteases in human diseases. FEBS Lett. 595(8): 1205-1222. **doi:**10.1002/1873-3468.14039

119. Verhamme IM, Leonard SE, Perkins RC (2019) Proteases: pivot points in functional proteomics. Methods Mol Biol. 1871: 313-392. **doi:**10.1007/978-1-4939-8814-3_20
120. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, et al. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46(D1): D624-D632. **doi:**10.1093/nar/gkx1134
121. Rawlings ND, Bateman A (2021) How to use the MEROPS database and website to help understand peptidase specificity. Protein Sci. 30(1): 83-92. **doi:**10.1002/pro.3948
122. Costa SM, Saramago M, Matos RG, Arraiano CM, Viegas SC (2022) How hydrolytic exoribonucleases impact human disease: Two sides of the same story. FEBS Open Bio. **doi:**10.1002/2211-5463.13392
123. Mohanty BK, Kushner SR (2018) Enzymes involved in post-transcriptional RNA metabolism in Gram-negative bacteria. Microbiol Spectr. 6(2). **doi:**10.1128/microbiolspec.RWR-0011-2017
124. Sun D, Han C, Sheng J (2022) The role of human ribonuclease A family in health and diseases: A systematic review. iScience. 25(11): 105284. **doi:**10.1016/j.isci.2022.105284
125. Jarrous N, Liu F (2023) Human RNase P: overview of a ribonuclease of interrelated molecular networks and gene-targeting systems. RNA. 29(3): 300-307. **doi:**10.1261/rna.079475.122
126. Lange H, Gagliardi D (2022) Catalytic activities, molecular connections, and biological functions of plant RNA exosome complexes. Plant Cell. 34(3): 967-988. **doi:**10.1093/plcell/koab310
127. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 329(5987): 52-6. **doi:**10.1126/science.1190719
128. Lartigue C, Vashee S, Algire MA, Chuang RY, Benders GA, et al. (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science. 325(5948): 1693-6. **doi:**10.1126/science.1173759
129. Venter C Watch me unveil "synthetic life". 2010, TED (Technology, Entertainment and Design). Last accessed June 11, 2024. https://www.ted.com/talks/craig_venter_watch_me_unveil_synthetic_life.
130. Matuscak, S.T. and C.L. Tan (2016) Who are the parents of Mycoplasma mycoides JCVI-syn1.0? BIO-Complexity. 2: 1-5. **doi:**10.5048/BIO-C.2016.2
131. Tan C, Tomkins JP (2015) Information processing differences between archaea and eukaraya—Implications for homologs and the myth of eukaryogenesis. Answers Research Journal. 8: 121–141.
132. Holm M, Natchiar SK, Rundlet EJ, Myasnikov AG, Watson ZL, et al. (2023) mRNA decoding in human is kinetically and structurally distinct from bacteria. Nature. **doi:**10.1038/s41586-023-05908-w
133. Itaya M, Tsuge K, Koizumi M, Fujita K (2005) Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. Proc Natl Acad Sci U S A. 102(44): 15971-6. **doi:**10.1073/pnas.0503868102
134. Eberhard D, Grummt I (1996) Species specificity of ribosomal gene transcription: A factor associated with human RNA polymerase I prevents transcription of mouse rDNA. DNA and Cell Biology. 15(2): 167-173. **doi:**10.1089/dna.1996.15.167
135. Heix J, Zomerdijk JCBM, Ravanpay A, Tijan R, Grummt I (1997) Cloning of murine RNA polymerase I-specific TAF factors: Conserved interactions between the subunits of the species-specific transcription initiation factorTIF-IB/SL1. Proceedings of the National Academy of Sciences. 94(5): 1733. **doi:**10.1073/pnas.94.5.1733
136. Murano K, Okuwaki M, Momose F, Kumakura M, Ueshima S, et al. (2014) Reconstitution of human rRNA gene transcription in mouse cells by a complete SL1 complex. J Cell Sci. 127(Pt 15): 3309-19. **doi:**10.1242/jcs.146787
137. Friday RP, Pietropaolo SL, Profozich J, Trucco M, Pietropaolo M (2003) Alternative core promoters regulate tissue-specific transcription from the autoimmune diabetes-related *ICA1* (ICA69) gene locus. J Biol Chem. 278(2): 853-63. **doi:**10.1074/jbc.M210175200
138. Teng S, Li YE, Yang M, Qi R, Huang Y, et al. (2020) Tissue-specific transcription reprogramming promotes liver metastasis of colorectal cancer. Cell Res. 30(1): 34-49. **doi:**10.1038/s41422-019-0259-z
139. Lobb B, Tremblay BJM, Moreno-Hagelsieb G, Doxey AC (2020) An assessment of genome annotation coverage across the bacterial tree of life. Microb Genom. 6(3). **doi:**10.1099/mgen.0.000341
140. de Crécy-lagard V, de Hegedus RA, Arighi C, Babor J, Bateman A, et al. (2022) A roadmap for the functional annotation of protein families: a community perspective. Database (Oxford). 2022. **doi:**10.1093/database/baac062
141. DiFrisco J, Love AC, Wagner GP (2020) Character identity mechanisms: a conceptual model for comparative-mechanistic biology. Biology & Philosophy. 35(4). **doi:**10.1007/s10539-020-09762-2
142. KnoopV (2011) When you can't trust the DNA: RNA editing changes transcript sequences. Cell Mol Life Sci. 68(4): 567-86. **doi:**10.1007/s00018-010-0538-9
143. Jafari M, Ansari-Pour N, Azimzadeh S, Mirzaie M (2017) A logic-based dynamic modeling approach to explicate the evolution of the central dogma of molecular biology. PLoS One. 12(12): e0189922. **doi:**10.1371/journal.pone.0189922
144. Hausser J, Mayo A, Keren L, Alon U (2019) Central dogma rates and the trade-off between precision and economy in gene expression. Nat Commun. 10(1): 68. **doi:**10.1038/s4146701807391-8
145. Crick FH (1968) The origin of the genetic code. J Mol Biol. 38(3): 367-79. **doi:**10.1016/00222836(68)90392-6