*A third way to the selected effect/causal role distinction in the great ENCODE debate*

Ehud Lamm, Sophie Juliane Veigl

**Abstract**

Since the ENCODE project published its final results in a series of articles in 2012, there is still no consensus on what its implications are. ENCODE's central and most controversial claim was that there is essentially no junk DNA: most sections of the human genome believed to be "junk" are functional. This claim was met with many reservations. If researchers disagree about whether there is junk DNA, they have first to agree on a concept of function and how function, given a particular definition, can be discovered. The ENCODE debate centered on a notion of function that assumes a strong dichotomy between evolutionary and non-evolutionary function and causes, prevalent in the Modern Evolutionary Synthesis. In contrast to how the debate is typically portrayed, “both” sides share a commitment to this distinction. This distinction is, however, much debated in alternative approaches to evolutionary theory, such as the EES. We show that because the ENCODE debate is grounded in a particular notion of function, it is unclear how it connects to broader debates about what is the correct evolutionary framework. Furthermore, we show how arguments brought forward in the controversy, particularly arguments from mathematical population genetics, are deeply embedded in their particular disciplinary contexts, and reflect substantive assumptions about the evolution of genomes. With this article, we aim to provide an anatomy of the ENCODE debate that offers a new perspective on the notions of function both sides employed, as well as to situate the ENCODE debate within wider debates regarding the forces operating in evolution.

**Keywords**: function, evolutionary causation, junk DNA, evolutionary theory, genomics, EES, population genetics; proximate-ultimate

**1 Introduction**

In 2013 one of the most notorious scientific articles of recent years was published, titled *On the Immortality of Television Sets: "Function" in the Human Genome According to the Evolution-Free Gospel of ENCODE,* authored by Dan Graur. In the article, Graurcrusaded against the conclusions of the ENCyclopedia Of DNA Element project– in short, ENCODE – and was subsequently criticized for causing an "undignified academic squabble" (1) and being "angry, dogmatic, scattershot, sometimes inaccurate" (1). Purportedly, the "ENCODE debate" has been about function and whether the authors of the largest sequencing project to date have used the wrong definition of function when arguing that 80% of the human genome was functional. It was at bottom also a symptom of struggles to conceptualize and understand what genomes are.

The publications of the ENCODE consortium were in tension with canonical views on junk DNA. The term "junk DNA" was developed in the 1970s, primarily in the work of Japanese-American geneticist Susumu Ohno (2).[[1]](#footnote-1) Based on a growing understanding of the molecular evolution of genomes at the time and models of intragenomic conflict and selection strength, it has become increasingly recognized that a large part of genomes consists of sequences that do not encode functional proteins. Today, "junk DNA" is often used in a broad sense to refer to any DNA sequence that does not play a functional role in development, physiology, or some other organism-level capacity (3). Based on many different lines of analysis, it has become common wisdom among many biologists that large, eukaryotic genomes, in particular, contain large amounts of junk DNA (4). This perspective is grounded in addition in the theory of neutral evolution (3), the notion of selfish genetic elements (5), and drift as a factor in genome evolution (1) (4). Significantly, the notion of junk DNA is not specifically about the human genome but a more general observation. The ENCODE project, or at least the publicity surrounding it, seemed to challenge this common wisdom.

That junk DNA is not functional has never been claimed by evolutionary biologists in such a stark and simplistic way. Very early ideas about transposons being important regulators of gene expression date back to the first half of the 20th century. Barbara McClintock, who discovered transposable elements, considered them important regulators and showed that they became active during periods of stress on the organism (6) (7) (8) (9). While transposons could never be demonstrated to be master regulators of gene expression, discussions of how, for instance, transposons are functional can be found in Dawkins (10), Doolittle and Sapienza (11), and Orgel and Crick (12). What several authors in this tradition questioned was not the functionality of transposons per se, but the functionality on the organism level of junk DNA, of which a large part consists of transposons, in particular inactive transposons and remnants of viruses (13). Other researchers studying various biological systems continued to stress that the activation of TE during stress may benefit the (cellular and/or organismal) lineage while others are skeptical that this effect was selected for (13,14). This rich discussion brings with it questions about levels of selection and the levels of selection relevant for understanding the genome, as well as questions regarding levels of biological organization and causation.

The ENCODE debate is often taken to be the result of the incompatibility of two different notions of function, a causal, mechanistic notion and an evolutionary notion of selected effect functions. In this article, however, we will argue that both ENCODE and its detractors are on common ground in accepting this distinction when doing biology: with ENCODE using a causal notion and its detractors a selected effect notion of function. In contrast to both, researchers identifying with the Extended Evolutionary Synthesis or Third Way endeavors reject precisely this separation of concerns.

We shall argue that an up-to-now missed aspect of the ENCODE debate is whether it is even possible (and useful) to distinguish different types of functioning in the way both the proponents and the opponents of ENCODE assume. And in particular, the implications of this distinction for trying to understand genomes better. We will discuss how this outlook on functioning, prevalent within population genetics-focused evolutionary biology constrained the ways participants in the ENCODE debate conceptualize function. Take one example – in a more peaceful article titled *"An Upper Limit on the Functional Fraction of the Human Genome,"* written in 2017, Graur used population genetics models to derive an upper bound on the fraction of the human genome that can possibly be functional. For illustration, we discuss this paper at some length in the next section to set the stage for the intellectual and disciplinary context of the ENCODE debate. We take this discussion as a vantage point to return to ENCODE in the following sections. We will particularly focus on the conceptual issues tied to different notions of functioning and whether they facilitate or impede connecting ENCODE to other recent approaches towards genomes, their functioning, and their evolution. In doing so, we will, however, also touch on the disciplinary situatedness of such debates since such configurations considerably influence the issues at stake.

**2. Deriving a theoretical upper bound on the amount of functional DNA**

Graur (15) calculates how much of the human genome can potentially be functional using empirical data on genome size, mutation rates, the fraction of deleterious mutations in functional regions, and fertility rates. A theoretical deduction such as this is impressive, particularly in biology. The idea of the article is to define the upper limit of the functional fraction in the genome based on the genetic load – the reduction in the mean fitness of a population relative to the individual with the maximal fitness, caused by processes such as mutations, recombination, migration or inbreeding, amongst other factors. The notion of genetic load, with its eugenicist undertones, was introduced by H. J. Muller in 1950 and remains one of the field's most powerful theoretical ideas (16). It is directly connected to the "classical vs. balance" debate between Muller and Dobzhansky about whether selection in natural populations is primarily a hill-climbing process and about the role of heterozygosity in evolution (17). It was thus intimately related to controversies about the best way to conceptualize the genome (18).

The genetic load determines the mean fitness of a population. The mean fitness of a population, in turn, determines the mean fertility required to maintain the population at a particular size. With this calculation, Graur intended to mathematically demonstrate why ENCODE's claim that 80% of the genome is functional cannot be maintained. Graur demonstrates that given very conservative estimations of mutation rates of the human genome, the functional fraction of the genome cannot exceed 15%. Assuming, as ENCODE concluded, that 80% of the genome is functional, each couple would have to conceive on average 102 children to maintain the size of the population, a number that is impossible to attain for human, monogamous couples.

Graur's claims are situated in a very specific discourse. The first task is to clarify what Graur means when he states that the functional fraction cannot exceed 15%. Graur operates with a concept of function denoting *selected effect* (se) function – strictly those functions that have been shaped and maintained by natural selection (i.e., a function that has not only been selected to perform this particular function but is still kept to perform that very same function). Graur operationalizes his definition of function even further: a genomic segment possesses se function if at least one mutation in that segment is deleterious (15). Examining Graur's model illustrates the theoretical assumptions concerning genome evolution that come together with this notion of function.

Graur's model (15) is based on a small number of theoretical constructs and a few empirically determined parameters. The model allows Graur to be explicit about his reasoning. It allows him to note where he uses approximate data or data ranges.

The empirical data on which the estimate is based are:

1. The size of the human genome: 6.114∗10^9 nucleotides.

2. Mutation rate: varies by a factor of 2.5, from 1.0∗10^−8 to 2.5∗10^−8 mutations per nucleotide site per generation.

3. Fraction of mutations in functional regions that are deleterious: estimate of 40% of total mutation rate. This is a low estimate, the actual value is probably higher.

4. By multiplying (2) & (3), we get that the rate of deleterious mutation ranges between 4∗10−9 and 1∗10−8.

Model:

1. The genetic load $L≈C\overline{μ}\_{del}$,

where $\overline{μ}\_{del}$ is the mean deleterious mutation rate and C is constant between 1 for completely recessive mutations, and 2, for completely dominant mutations.

2. $\overline{w}=(1-2\overline{μ}\_{del})^{n}$

where n is the number of functional nucleotide sites (*n*) in the genome.

3. $\overline{F}=\frac{1}{\overline{w}}$

Where $\overline{F}$ is replacement level fertility.

The idea is that when the empirically grounded parameters are plugged into this simple model, we get a value for replacement level fertility. This value depends on *n*, the number of functional sites in the genome, allowing Graur to conclude how many sites lead to impossibly high fertility values (values above 1.8). In other words, the argument is a *reductio* with the following *modus tollens* structure: if the number of functional sites is high, and given the mutation rate, the replacement level fertility has to be high; the fertility in humans is much lower; hence, the number of functional sites in not this high (the mutation rate is an empirically grounded number). Given the mathematical relationships established by the model, the number of functional sites has to be lower.

There are three kinds of assumptions in the argument: approximations and estimates for the empirical values and their justifications; theoretical assumptions, built into equations (1)-(3); and substantive assumptions about the evolution of the human genome and, with them, assumptions about the factors that should play a role in accounts of the evolution of the genome. The assumptions of the first kind are discussed explicitly by Graur. Discussion of the second kind of assumptions, while encapsulated in the equations, is to be found in the population genetics literature where they are derived. Assumptions of the third kind are not discussed in Graur (15). We discuss each in turn.

**Empirical values**: The model makes explicit the data used to support the estimate of the percentage of the genome that is functional. This allows Graur to justify the use of data that is approximate or provides a wide range of possible values for the model parameters. Among the available values, he plugs into the model such values that, by the sheer mathematical logic of the model, would predict the smallest amount of junk DNA. Thus, these values favor the view that he is arguing against. If even for these values that favor the competing view, the predicted amount of junk DNA has to be large, the implication that the genome is mostly functional is undermined.

**The theoretical framework**: The evolutionary models used to derive the equations in the model are mutation/selection balance models. Selection and mutation are the "forces" or causes of changes in frequencies in a population and equilibrium or stable states are identified. Some things are worth mentioning about the application of such models to the question at hand (the following is far from a thorough discussion of this field; the theoretical assumptions based on which these models have been developed are discussed extensively in the literature (19) (20) (21)).

Equilibrium analysis ignores the details of the historical process of evolution and instead focuses on equilibrium states (22) (23). An example of how this manifests itself is that Graur uses an average mutation rate, which presumably averages over multiple time periods and sources of variation in the mutation rate. It is difficult to use analyses of this kind to study the functionality of some aspect of the genome that may have originated in a specific constellation of selection/mutation, e.g., early in the evolution of the genetic systems or in a context with more stress-induced mutations, or through mutations in specific genomic hotspots and has become entrenched since. It is interesting that ENCODE also ignores history, but in a different way, by only addressing current biochemical activity in the genome.

The empirical values that are used carry assumptions about these issues. In other words, the choice of values reflects assumptions about the significance of these historical factors. Formal modeling allows you to ignore details irrelevant to a specific question. Yet choosing the empirical values, once such models are applied to a concrete empirical question, may involve further assumptions. Do you take measurements of mutation rate (for example) at a single point in time or over long evolutionary periods, for example? How do you come up with the estimate? This concern is not about the population genetics models per se and what can be learned from them regarding the results of selection. It relates to issues that should be addressed when the models are applied to concrete evolutionary scenarios. Justifying a formal model as a furtuiful way to understand how selection operates is different and possibly in tension with assumptions justifying the use of particular empirical values for the parameters when applying the model to a specific question.

To summarize, Graur uses values that disfavor his conclusion the most, thus showing that even in that case, his conclusion stands, and a huge fertility value is required. However, this claim depends on the mathematical relations in the model, which is based on a specific selection/mutation regime, and not the actual history of genomes.

**Substantive assumptions:**

1. The model assumes a fixed genome size ignoring the pre-history of the human genome, from single-celled organisms, archaea, and bacteria (or before), and upward, including the role that may have been played by horizontal gene transfer and endosymbiosis (24) (25) (26). This history shaped many regions of the genome that are operational, indeed SE functional, such as centromeres (27). ENCODE and Graur (15) ignore this "pre-history of the human genome" in different ways.
2. The population genetics models from which the model is derived are models of standard diploid Mendelian genetics. Some regions of the genome differ from standard diploid Mendelian genetics in various ways and have specific hereditary and mutational processes. These include the telomeres, the centromeres, and histone regions that may require different models (28) (29).
3. The models assume that mutations are non-directional (regarding selection or adaptive value).
4. The models assume that to understand the magnitude of the mutational load, one only needs to determine the deleterious mutation rate, not the distribution of fitness effects.
5. The probability of a mutation happening in the model is independent of whether that particular region is functional or operational. Another assumption is related: all mutations occurring in the nonfunctional fraction of the genome are believed to be neutral. Mutations in functional fractions are believed to be either neutral or deleterious (given that advantageous mutations are believed to be too rare to affect the modeling).
6. Epistasis is ignored. That is, the fitness contributions of each locus are believed to be independent of each other. The model further assumes that mating patterns are random, not considering processes such as assortative mating and niche construction that affect them.

While Graur takes his calculations as evidence tha ENCODE's claims are wrong, given the upper bound he derives, it is important to emphasize that Graur's and ENCODE's approaches are in many ways different and not easily comparable. Graur uses population genetic assumptions and mathematical modeling to come to his conclusions. ENCODE, on the other hand, primarily made use of biochemical assays. The context of the questions addressed by ENCODE is neither populational nor historical. Also, Graur employs a definition of function that limits functioning to an evolutionary discourse. In contrast, the ENCODE project is situated within a context that aims at identifying functions in terms of those properties that potentially affect the phenotype – selected or not. Indeed, in the context of ENCODE, there is no distinction between a chemical interaction having a positive or a negative effect, an essential distinction in the population genetic models. Although many commentators tried to resolve the debate by pointing out these differences, the clarifications did not clear the air. In the next section, we aim at a closer look into the context and assumptions of the ENCODE project.

**2. The ENCODE Project**

The results of the human genome project (HGP) that provided the sequence of the human genome were puzzling. It became clear that, on the one hand, human and chimpanzee DNA align almost 99% and, on the other hand, that the human genome is full of non-coding elements, many of them active and inactive transposable elements. Even though there were known exceptions, it was generally believed that most transposable elements do not contribute to the organism's fitness (30) (31) (32). From an anthropocentric perspective, these results were somewhat disappointing, given that they did not provide a biological basis for the hard-to-shake belief in human superiority and exceptional complexity compared to other species.

The idea that the DNA sequence coding for proteins that supposedly determines an 'organism's traits sets *Homo sapiens* apart from all other phyla, long viewed with some suspicion, could not be maintained. But researchers were endorse an alternative hypothesis, also with a long history: it is not the genes themselves that explain the relative complexity of an organism, but the regulation of that genetic material, i.e., the regulatory mechanisms that underlie gene expression. The ENCODE project set out to map the gene regulatory landscape and thus to determine all the active – that is, transcribed – parts of the human genome and, in this way, to provide a guide, or interpretation, to the map constructed by the HGP. It reflected a change in attention from the genome to the transcriptome.

Now in its fourth iteration, the ENCODE project was initiated and funded by the U.S. National Human Genome Research Institute (NHGRI) in 2003 to catalog the functional elements in the human genome, "including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active."[[2]](#footnote-2) A paradigmatic example of Big Science, the ENCODE project established publication embargoes on the publication of analyses based on the data collected and, more generally, created a situation in which some scientists felt that there were insiders and outsiders (33).[[3]](#footnote-3)

ENCODE's primary methodological setup was an experimental, to be precise, biochemical one. By studying which parts of the genome can be shown to interact with effector molecules necessary for transcribing DNA into RNAs – transcription factors – the ENCODE consortium aimed to identify the active parts of the human genome. Furthermore, they assayed which parts of the genome show certain histone modifications that are generally associated with transcription, as well as measured the sensitivity to DNase I activity – assuming that those stretches of DNA that are currently transcribed are more susceptible to DNase I cleavage given that they are temporarily disentangled and single-stranded. In addition, they assayed RNA transcription directly. By conducting these assays, they aimed to – in a sense – catch gene expression "in the act."

Patterns of gene expression are, of course, not identical in all cells of the organism. The different specializations of individual cell types define which genes will be expressed. Also, different conditions, such as stress or disease, alter gene expression. One of the best-known metaphors for the differences in gene expression across the cells of the organism is Waddington's epigenetic landscape: as the cell specializes, it goes down an irreversible path of gene regulatory changes in order to fit its specific niche (34). As ENCODE aimed at identifying transcriptionally active regions in all possible conditions, they worked on three cultured cell lines extensively and with 144 other cell types less extensively (33).

ENCODE's most cited, most notorious, and also most debated result was the following: according to their biochemical screens, 80% of the human genome is functional. That is, from about 80% of the genome, it was possible to assay reproducible biochemical activity. Given this result, the ENCODE consortium declared all textbooks wrong regarding their estimates of functional and nonfunctional parts of the genome: "*The human genome is packed with at least four million gene switches that reside in bits of DNA that once were dismissed as "junk" but that turn out to play critical roles in controlling how cells, organs and other tissues behave.... At least 80 percent of this DNA is active and needed*."[[4]](#footnote-4) ENCODE thus declared the end of junk DNA.

This claim about the scarcity of junk in the human genome stirred waves of criticism. One of the most straightforward critiques was that the ENCODE consortium used the wrong notion of function. Restricted by their experimental methodology, ENCODE can only detect biochemical affinity, not evolutionary function. In other words, just because a sequence is being transcribed does not mean that its transcription is "needed" for the organism to survive or that it has any evolutionarily relevant impact whatsoever on the phenotype. And, even if certain biochemical functions play a role that affects fitness in a particular way, this does not mean that they were selected for that functioning but merely that they came to function in this way. One illustrative way of explaining this differentiation is through Sidney Brenner's distinction between junk and garbage (35). While garbage is thrown out immediately, it does not mean that junk not thrown out remains functional. In other words, if junk sits around in the garage and might keep a cupboard from getting dusty, this does not mean it is kept to keep that cupboard from becoming dusty. Thus, opponents criticized that ENCODE proponents labeled everything that is not garbage as functional. Conversely, their opponents labeled anything not produced or maintained by selection as junk. As we will see, this leads to very different perspectives not only on the evolutionary history of genomes but also on what genomes are.

So far, commentators on the ENCODE debate have mostly focused on a narrow analysis of function to argue why or why not the ENCODE consortium has used the right concept to present its findings. Most authors seem to accord with both the ENCODE consortium and its critics that there is a strong distinction to be made between causal role (CR) and selected effect (SE) functioning. What is, however, striking is that even though the debate is so focused on function and, particularly, a debate on what is function in an evolutionary sense, it is not entirely straightforward to see how the ENCODE debate connects to other recent debates such as the one focusing on what is the correct evolutionary framework. In what way does ENCODE fit with or relate to the agenda of the traditional Neo-Darwinist Modern Synthesis, given that ENCODE has been attacked primarily by evolutionary biologists? And how does ENCODE sit with alternative proposals, such as the extended evolutionary synthesis (EES)?

As is true for all debates (in science, on the rims of science, and beyond), they are never only about a particular disagreement, a particular data point, or a particular built-in assumption. What is at stake is a tightly-meshed net of a broad range of social, political, and epistemic vectors. In the case of the ENCODE debate, much disentangling is yet to be done. We argue that the debate will not and cannot be settled by further clarification on the use of the term "function" since, as we shall argue, the ENCODE consortium and its critics did not really disagree about the definition of function. As we shall show, there is more that unites them regarding their perspective on function than what separates them.

**3. An anatomy of the encode debate: function, evidence, methods**

Probably the most straightforward but also most substantial area of disagreement within the ENCODE debate is that surrounding the appropriate notion of function. As pointed out in the introduction, members of the ENCODE committee claimed that by reporting biochemical activity from over 80% of all regions of the genome, they could declare these regions functional. This sparked immediate critiques, focusing on the notion of function. The debate was primarily presented as one regards two non-coextensive definitions of biological function: causal role (cr) functioning and selected effect (se) functioning. It is important to note that this distinction was not "native" to biological discourse but was imported by the critics of the ENCODE project from the philosophy of science.

Selected effect (se) function is often defined as *A biological function is a selected effect if a trait T has a function F iff Ts performing F is the reason why T has been selected and maintained through evolution* (36) (37) (38). Another version of the se definition of functioning is to say that *a trait T has a function F iff it is currently/will be under selection because it does F* (39)*.* Strictly speaking, saying that something is functional in an SE sense is relative to a specific function.

The causal role (cr) notion of function is defined relative to a specific system and its capacities rather than evolutionary history. It is defined as a relation between the capacity of a system and the activities of its parts*: An object X has a function φ in a system S with the capacity ψ iff φ has a causal role in S, that is, X's φ-ing contributes to S's ψ-ing* (40).

It is important to note that both these definitions may lead to questions about "levels" of selection or biological organization. We will see concerns about levels of selection and levels of organization reappear again and again and see that these concerns are related to how the genome is conceptualized.

Even though ENCODE's strategies were methodologically diverse, they all individuated parts of the genome similarly (41) – by their biochemical signature (42). Segments thus identified are "parts of the human genome by virtue of the biochemical activities that these sequences engage in" (41). Genomic parts are those particular DNA segments that engage in particular biochemical activities ("being transcribed," "being cut," "associating with," etc.). ENCODE researchers justified the biochemical signature strategy as a means of avoiding reductionism – in other words, of only looking at the sequence and not at the context and therefore privileging easily identified protein-coding genes – and as a way to accommodate those regions that are needed but are not themselves transcribed (42). This is, however, only part of the answer. Together with this definition comes the assumption that particular types of activities correlate with functions of these parts in the genome as a whole (41).

The notion of causal role need not be equated with biochemical activity, and there are a variety of kinds of causes in biology. Accordingly, ENCODE critics argued that by using a version of the cr definition of function that equates cr with biochemical activity, the ENCODE project had engorged the meaning of function to a level where its extension is massive and thus, as a result, almost meaningless (43). In other words, this is both a limited view of biological causality and an overly permissive experimental setup. This "loose" concept of functioning moreover led to ENCODE's claim about the absence of junk DNA in the human genome. If it is sufficient for a segment to be functional that it is transcribed or engages in any biochemical activity whatsoever, then the cr concept leads to the conclusion that the genome is predominantly functional. The question is, however, whether ENCODE really got rid of junk DNA or whether it simply relied on a different notion of function, namely a biochemically-operationalized cr function, and is thus irrelevant for this question.

Why the ENCODE debate triggered so many heated responses needs some unpacking. Both sides seem to agree about the distinction between cr and se functioning; each simply uses a different one of the two notions of function. So why the heated disagreement? The ENCODE authors described 80% of the human genome as functional in a cr kind of way, whereas the ENCODE critics opined that 80% of the genome could not be functional in a se kind of way. Using the term junk DNA loosely, as has become prevalent, these notions translate into the claim that from the ENCODE perspective, 80% of the genome can be assigned cr functionality and therefore cannot be regarded as junk. From the perspective of ENCODE critics, the claim is that only those sections of the genome that have se function are not junk. ENCODE opponents primarily informed their critique by pointing to population genetics' estimates that only 3-8 percent of the genome has been undergoing purifying selection. As ENCODE critic Ford Doolittle summarized:

"Those of us who speak of excess DNA as informationally junk mean that its presence is not to be explained by past and/or current selection at the level of organisms—that it has no informational function construable historically as an SE [selected effect]. Those who say that almost the whole of the human genome is functional informationally do so on the basis of an operational diagnosis embracing a non-historical CR [causal role] definition of function." (43).

It is, however, important to keep in mind that given its methodological setup, the ENCODE project did not identify 80% of the human genome to be functional but only as potentially functional. ENCODE's claim to have assigned "functions for 80% of the genome" (33) is, in the end, inaccurate (41). The ENCODE consortium always went with function as:

"Operationally, we define a functional element as a discrete genome segment that encodes a defined product (for example, protein or non-coding RNA) or displays a reproducible biochemical signature (for example, protein binding, or a specific chromatin structure)." (33).

It would seem this should have ended the debate. However, the ENCODE consortium parted with the mainstream notion of cr function in how ENCODE proposed to identify function. Elliot et al. list four common means of identifying (cr) function: 1) context-specific transcription, 2) positional information, 3) sequence conservation, 4) experimental manipulation (44). ENCODE approach was to apply the following five criteria for (biochemical) functioning: A genomic sequence functions if (1) it is transcribed into RNA (but not necessarily translated into a protein), or (2) it contains or is adjacent to a transcription binding factor, or (3) it is a methylated CpG dinucleotide, or (4) it is located in an area of open chromatin, or (5) it is found organized in nucleosomes containing certain histone modifications (33). These criteria of functioning lead to the worries of evolutionary biologists with the cr concept of function. By operating with this concept, one accepts many "irrelevant" functions (45) (43).

Thus, the opponents' issue with ENCODE's claim was that it was taken as implying that junk DNA, such as transposons, are functional with regard to the organism level (44). By assaying for transcription, it is, however, impossible, given ENCODE's methodological setup to delineate between what contributes to the organism level and what contributes to the transposon as such (assuming the two can be separated). Particularly for transposons, it is exactly their role as scaffolded Darwinian individuals – evolutionary individuals within individuals that make use of the organism they lodge in to replicate themselves. With ENCODE's methodological setup, there is no way to distinguish between transposons functioning in contributing to the phenotype and replication events of transposons with little or no impact on the organism's fitness. In conclusion, TEs are peculiar in at least four ways: 1) evolutionary processes operate on them in two ways: on the organism level and on the level of the genome or that of the individual TE. 2) the origin of the transposon sequence and the reason for its persistence are often decoupled, particularly if the respective TE attains organism-level functioning. 3) it is hard to distinguish between a TE playing a causal role in a particular phenotype or a TE having merely beneficial side effects (44). 4) TE have been shown to potentially provide evolvability advantages to lineages, specifically because of their activation in times of stress (though debates remain on how best to understand the evolutionary origin and maintenance of this activation).

In addition to TE-specific problems also come broader problems regarding how ENCODE assigns function. For instance, genomic elements may have critical functional roles without being specific in their activity. Spacers are a good example. While it is highly likely that a particular length of a stretch of nucleotides has been under selection, the actual sequence of As, Cs, Ts, and Gs does not seem to be of any importance. By their approach, ENCODE would miss such sequences. Problems with associating sequence with function continue: If a particular sequence S is determined as functional by ENCODE's standards and if we would add one neutral nucleotide to that sequence: would we still talk about the same function? Are there degrees of functioning? Do we believe that one segment can have more than one function (9)? Questions such as these highlight the significance of the qualification that excess DNA is "informationally junk" in the quotation from Ford Doolittle given above. Structural as opposed to informational roles need not align with the distinction between functional and nonfunctional or with the distinction between SE and CR functions. Genomic regions that have structural roles (the centromeres and telomeres in particular) do not sit well with the junk DNA paradigm, as we indicated in section 2.

An additional issue is that the dynamics of interactions of the genome, its transcripts, and its effectors, depend on an economy of resources. For example, a particular process P taking place is highly likely limited by the number of processes P that can take place at the same time and the number of effectors E whose activities engage in the process. If there are more Ps than Es, Ps will compete for associating with E. If there is a new P and its propensity for interacting with E is higher than zero, the introduction of a new P will shift interaction dynamics and, in a major or minor way, titrate E away. These shifts will be in almost all cases entirely quantitative. It is a question of more or less process P (e.g., transcription of a sequence). How do we thus describe the function of a sequence S that can engage a process P? What is its functioning? There is, on the one hand, its biochemical, ENCODE-functioning of being transcribed at a particular rate and thus possibly having some effect on the phenotype but also for a systemic contribution to a whole net of interactions that are modified, which, in turn, might also have a phenotypic result.

In conclusion, there are several issues pertaining to ENCODE's concept of functioning, some general and some specifically related to how the genome is understood. Those discrepancies between ENCODE and its critics, however, cannot be explained fully by reference to a disagreement on what the notion of function is. Both seem to agree on a strict separation of the cr and the se concept, the proximate and the ultimate, the physiological and the evolutionary. In a sense, the disagreement does not seem very thick since the root of the debate seems to be easy to identifty by considering the disciplinary context of the participants. Rather than problems with the distinction of cr and se functioning, we noted that there are other problems with ENCODE's method of assigning cr function. In the next section, we discuss the broader context of the ENCODE debate. Particularly we will ask how ENCODE sits with proposals that are in tension with classic population genetics, such as the Extended Evolutionary Synthesis. We will then ask what all these views mean for understanding genomes.

**4. ENCODE and the Extended Evolutionary Synthesis**

In the previous section, we tried to show that ENCODE and its critics do not have a major disagreement about function. Both agree at least to a degree on two definitions of function, on their applicability, and also that these two concepts are somewhat distinct. We argued that while there is no disagreement with regards to a distinction of se and cr function – even though ENCODE might be guilty of publicly insisting too little on it – other aspects of function and functioning explain the dispute: the question of purpose/context relativity; the disciplinary situatedness of what is the adequate way to define functioning; and the organizational level the function claim refers to. What is, however, shared amongst both sides is the emphasis on a distinction of se and cr functioning, treating both notions as somewhat exclusive and non-gradual. It is the aim of this and the following sections to situate the particularities of the ENCODE debate with broader debates and, particularly, try to explore why the ENCODE debate seems to be set apart from recent debates on what is the correct framework to account for evolution, the Extended Evolutionary Synthesis (EES) in particular. Furthermore, we will question whether insisting on a clear distinction between se and cr functioning is necessary.

The bearing of the HGP as well as ENCODE on human biology explains why these debates became that public. Both projects situate the human body after the century of the gene in the century of postgenomics (46) (47). One important issue critics of ENCODE accused the project of is its alleged soothing of the creationist agenda by increasing the estimates of how much of the human genome is functional. This issue needs disentangling, as it incorporates two not entirely equivalent strands of reasoning. ENCODE’s have implications for both speciesism and creationism. On the one hand, the quest for the map of the human genome is entrenched with speciesism. This started with the G-value paradox, the finding that the size of genomes does not seem to align with any perceived complexity of a respective species. The big-science sequencing projects of the end of the 20th century promised to shed light on that paradox by sequencing genomes of several model species. The G-value paradox remained, however, now in the form of a coding-gene paradox: the number of coding genes does not seem to align with any perceived complexity of a respective species.

This finding was hard to align with the belief that there are major transitions in complexity and that jawed vertebrates, and even more so, rational and sensitive individuals are more complex than others. In addition, the HGP revealed that the human genome contained massive quantities of junk DNA, confirming theoretical predictions, thus triggering the question of why the human genome is so "wasteful." Thus, ENCODE, on the one hand, investigated the status and activities of junk DNA to identify a particular level of complexity in gene regulation particular to humans. On the other hand, the claim that there is no junk DNA could also soothe another agenda besides implicit speciesism. That the human genome contains much junk was also a result that was hard to accommodate for creationists: given an omniscient and omnipotent creator, why would there be anything wasteful in the genome? In other words: why is the genome not perfect? The creationist issue with junk DNA, thus, primarily plays out in the selected effect perspective on function and, consequently, the criticism of ENCODE opponents also rests on their problem with ENCODE's interpretation of the term functioning. Though, as Germain et al. point out, even if ENCODE would have claimed that the genome is functional in a se sense, it would not be evidence for but merely compatible with intelligent design (48).

The ENCODE proclamations were thus accused by some of opening a crack in the door for creationists by implying that the human genome is perfectly designed. This is a serious allegation in the U.S. in particular, given the many public debates and court cases on whether evolution or creation should be taught in schools (49). The ENCODE debate is not the only current debate accused of opening a crack in the door for creationists. A similar debate concerns the Extended Evolutionary Synthesis – a scientific intellectual movement founded in 2008, making a case for a more inclusive account of evolution that incorporates, amongst others, evo-devo, niche construction, epigenetics, and cultural evolution as relevant causal factors (50) (51).

EES and ENCODE have an interesting relationship. There is barely any scholarship investigating how both perspectives might go together (one exception being (9)). We shall now take it upon us to explore how both relate and how they differ. Both EES and ENCODE are situated in the aftermath of the HGP but provide different types of answers to the end of the era of (coding-)gene centrism. Both ENCODE and critics like Graur are on the same neo-Darwinian side. The EES, in contrast, can broadly be said to belong to a tradition, with many precursors throughout the 20th century, that emphasizes environmental influences, development, and agency of organisms, as factors that are internal to the evolutionary process and not only external constraints (i.e., selection pressures). EES goes in a different direction than ENCODE as it tries to contextualize the results of the HGP. Both share critics within the same disciplinary background, as both projects have received extensive criticism from population geneticists and mathematical modelers. For both, this criticism has something to do with their disciplinary situatedness. ENCODE is primarily situated within the biomedical disciplines, and even though many EES proponents are evolutionary biologists, they come from different, more marginal traditions within evolutionary biologists (but also, for instance, molecular epigenetics) using different types of hypothesizing, methodological setups, and experimental techniques. In short, they differ in thought style.

One critical area in which an EES perspective differs from ENCODE and its critics concerns function. The EES seeks to incorporate how certain physiological and developmental processes that affect the genome impact heritable variation (genetic, epigenetic, or other). This comes down to several issues. First, whether the genome is responsive to environmental conditions and, if so, in what way? Are there changes to the genome that are not random? Second, whether such changes (often epigenetic and regulatory) are heritable and to what extent (e.g., whether they affect the germline), and whether they can lead to changes in DNA sequence. In several fields pertaining to epigenetics, quantitative trait loci (QTL) approaches are often seen as an opposition to explaining variation in certain traits through other gene-regulatory mechanisms such as histone modification, DNA methylation, or small regulatory RNAs. The dominant view - not contested by the ENCODE consortium - is that gene activity in itself does and cannot change the regulatory logic of the wiring of the genome (9). An EES perspective challenges that view. To a degree, certain environmental changes are believed to be possibly canalized into genetic changes. Gene regulatory changes are believed to be potentially heritable and might even lead to genomic changes (52). A related but separate issue is whether plasticity might alleviate certain selection pressures and whether this explains why certain parts of the genome were not subjected to purifying selection. These ideas challenge the separation of concerns between proximate and ultimate explanations (53). This separation is shared by ENCODE (implicitly) and its critics.

An important case is that of small regulatory RNAs, molecules involved with many aspects of gene regulation. In mammals, they are particularly associated with development and disease, as well as transposon control. Some small RNAs locate in the nucleus and were shown to cause DNA methylation and histone modifications. In turn, such epigenetic marks and small RNA activity on genetic loci have been shown to alter the susceptibility to DNA mutations (54). The EES seeks to incorporate and study the various ways by which the organism's interactions with the environment may find their way back into the genome. Through its interest in the gene regulatory landscape, ENCODE is not entirely separated from these endeavors since it, for example, provides information on the transcription of non-coding RNAs and the histone/DNA modification landscape as well as context sensitive responses of the genome to environmental stresses. Questions regarding the responsive genome were, however, not asked in the context of the ENCODE debate. The ENCODE project, therefore – and surprisingly – remains within a conceptual framework that is quite compatible with that of the century of the gene, as well as the Modern Synthesis of evolution.

**5. Disciplinary Context and Public Reception**

The ENCODE debate is a special one, particularly when it comes to its intensity and the public arenas it has appeared in. Germain et al. provide two reasons for why the debate got so heated. On the one hand, many scholars felt unease with ENCODE's relations with public media outlets and the types of marketing strategies proponents pushed. ENCODE was blamed for promoting their results in a particular way, not necessarily caring whether their results would be communicated adequately. On the other hand, critique was perceived as too critical and too rhetorical in its criticisms. It was argued that the way opponents launched their criticisms did not accord with a proper scientific tone. Both sides of the debate were thus concerned with the scientificity of the debate, or rather, whether individuals lived up to the ideal of scientificity, as epitomized by the Mertonian norms (55) as well as particular stereotypes of scientists such as, e.g., the "modest witness" (56). Employing this kind of credibility strategy (57) or boundary work (58) are hallmarks of scientists negotiating a debate within a public arena (58) (59).

What is the "right" notion of function is essentially relative to a particular research purpose (48). The initiation of the ENCODE project is situated in the realization that sequencing the human genome alone would not bring the much-promised advancement in biomedicine; i.e., having a map of the human genome does not cure cancer by itself. An important aim of ENCODE was for biomedicine to harness biological knowledge about the genome. ENCODE, in a sense, served as the interpretation device of the HGP. Another important motivation for the ENCODE project was the rise of demand for personalized drugs, given the hypothesis that drug inefficiency is likely linked with genetic variations within a population. Thus, there is another aspect to ENCODE's use of the term functioning – it is a methodological, pragmatic one: they declared those parts functional that might be deemed relevant for biomedical research and are subjectable to biomedical inquiry and intervention.

Two other aspects of ENCODE's biochemical setup were particularly important for the biomedical context: On the one hand, it promised to identify QTLs that could contribute to disease phenotypes. On the other hand, given that molecular epigenetics proposed that disease phenotypes were dependent on particular epigenetic signatures of genes, ENCODE, with its definition of function that included the histone and DNA modification landscape, promised to provide informative data for biomedical endeavors (48). These considerations run orthogonal to the argumentations of ENCODE critics on what is a "relevant" function. Those functioning bits of the human genome that make, e.g., a human breath and will thus have been selected on the grounds of enabling humans to breathe might not vary so much between populations and thus, be of very little interest to biomedical research, particularly in the personalized medicine context: "making people healthier most often has very little to do with granting them more offspring" (48). Thus, what is a "relevant" function is almost the opposite of what evolutionary biologists would consider what it means to be functional.

The ENCODE findings got taken up in a particular way by media outlets and several commentators accused the ENCODE consortium of making fairly little attempt to clarify the misunderstanding (48). Their claims were transformed into the pronouncement that 80% of the genome is "critical" and "needed," which is imprecise and untrue. It, however, points towards explanations why the ENCODE debate was triggered, given that terms such as "needed" do not steer clear of evolutionary, se connotations. If something is "needed" or "critical," it means that its loss or a particular mutation would change the organism's phenotype and fitness, respectively. That something is "needed" also implies that it is under selection currently. But by merely contributing to a particular phenotype, it does not mean that a particular sequence is crucial per se. Given the cr definition of function, a DNA segment can be functional but non-essential. In conclusion, the differentiation between the technical notions of junk DNA and SE-functional genome segments got conflated with the distinction of being currently essential or non-essential at the level of individuals and hypothesized medical interventions.

**6. What is the genome?**

The Human Genome Project, together with numerous discoveries about the organization and regulation of genomes, put an end to the hope that genomes could be conceptualized simply as collections of possibly interacting genes. Rather, the physical organization and regulatory mechanisms remain central topics of research and fundamental to how biologists think of genomes and genetic systems. This new work continues and revisits inventive speculations about how best to conceptualize genomes going back for a century (9).

In the scientific discourse, the term genome has come increasingly to refer strictly to the nucleotide sequence, while the term chromatin, which existed for a long time as well and refers to the structure of DNA and proteins in eukaryotic cells, has become a shorthand for referring to work on the physical aspects of genomes. This terminological divide-and-conquer (as well as various other creative terms such as transcriptomics and epigenomics) should be understood as indicating that the two faces of the hereditary material are of active concern for scientists. The ENCODE project focused primarily on transcription. It thus endeavored to map transcription factor (TF) binding sites but also histone modifications since the latter affect the physical organization of chromatin (i.e., the nucleosomes) with implications for transcription as well other chromatin processes, which affect, among other things, mutations and recombination.

This broader perspective helps contextualize the debate over what percentage of the DNA sequence is functional. Many processes beyond gene expression constantly take place in and around the genome, and how best to conceptualize genomes as dynamic, three-dimensional entities is and will likely remain for some time an open-ended question. To put it starkly, whether informally or even in scholarly publication, the ENCODE debate was prompted by the whiff of optimality, or even of intelligent design, coming from the claim that the genome is almost entirely functional. The opponents of ENCODE proclaimed that junk DNA was a manifestation of a Darwinian worldview, which predicts imperfections. Yet there are clearly Darwinian views that attempt to go beyond the Neo-Darwinian population genetics view of the role of genes in evolution, namely the EES and Third Way approaches. Scientists working within these broad perspectives have suggested various ways of conceptualizing genomes beyond the static nucleotide sequence model (Caporale, 2006; Fedoroff and Botstein, 1992; Fontdevila, 2011; Shapiro, 1992, 2011).

 How best to incorporate the variety of ideas and speculations these and other scientists suggested with the empirical findings of ENCODE and theoretical analysis of its critics and other work on the population genetics of the genome is very much an open question. What does seem clear is that the image of eukaryotic genomes at the moment is not one of static, well-designed sequences of nucleotides or, alternatively, long, mostly junk sequences, with coding sequences and some regulatory sequences strewn in. Rather, genomes are complex dynamic systems within other interacting levels of biological organization and levels of selection with a variety of mechanisms that affect, among other things, both mutation and recombination, and that ameliorate to an extent many kinds of imperfections, noise, and ever-changing environmental challenges.

**7. Concluding Remarks**

In this article, we tried to provide a broad perspective on the ENCODE debate. We argued that the disagreements in the ENCODE debate, that is the dispute between ENCODE proponents and evolutionary biologists, is quite resolvable. In particular, we agreed with others who already noted that ENCODE at most indicates which regions of the genome are potentially functional, not that they are functional. We further argued that even when it comes to some issues that explain the dispute, such as the concept of function used being relative to ENCODE's situatedness in molecular biology and particularly biomedicine, ENCODE is quite in line with canonical understandings of the genome, conceptualizing it as essentially rigid, with epigenetic and gene regulatory marks being solely enablers of a particular function yet that cannot shape the hardwired functioning in the genome.

We argued that the ENCODE project continues the tradition of thought that dominated the century of the gene and also the HGP. We proceeded by trying to uncover the assumptions guiding ENCODE. This brings us back to the start of this paper. We discussed Graur's elaborations on why ENCODE's estimate that 80% of the human genome is functional (in a se sense) is impossible, based on the mathematics of population genetics. We tried to highlight the many built-in assumptions of Graur's calculations. Some of the assumptions reflect the perspective that the genome is unresponsive to the environment and others relate to the equilibrium style of analysis paradigmatic of classic population genetics. We tried to show that even though ENCODE uses a different concept of function, namely cr and not se. However, the distinction between these two notions of function is shared by the participants in the debat such, as Graur and ENCODE proponents, as are built-in assumptions regarding the very nature of the genome . This is thus, in essence, Kuhnian "normal science" going about its business.

We contrasted the ENCODE debate as it unfolded with the EES and Third Way approaches and tried to show that while at first glance they seem related, their outlooks on genomes and gene regulation are different. It is exactly with the EES that one could locate a "real" dispute about functioning. While both ENCODE proponents and critics work with a strict differentiation of cr and se functioning, which when applied to the genome manifests in its nonresponsiveness, perspectives on genomes such as those suggested by the EES move towards a dissolution of a strict differentiation (65) (9) (64). When the responsiveness of genomes to environments and organism agency are considered, the causal analysis of evolutionary change is extended. If genomic change is not only caused by random mutations but is also susceptible to more directed processes that are correlated with gene regulatory changes, also those segments of the genome that would traditionally be labeled functioning in a cr kind of way could readily be subjected to changes that selection could act upon. These processes are not part of the modeling tradition Graur uses, and it is far from straightforward how these perspectives are best reconciled. This challenge goes beyond the day-to-day hustle and bustle of "normal science," which partly explains why the debate about the EES is fraught with mutual misunderstanding (50). We also emphasized that while the critics of ENCODE are correct in saying that its analysis is ahistorical, a related claim about ahistoricity can be made about mutation/selection models.

The worry of evolutionary biologists of identifying too many areas of the genome as potentially functional is less of a concern in the molecular biological and biomedical sciences than within evolutionary biology. For a biomedical intervention, it has little importance whether a function F is currently present because it was selected in ancestral genomes. Similarly, whether a particular function F might be selected currently or in the future might be of interest but is not within the realms in which molecular biological and biomedical interventions usually occur. This difference in focus is related to the contentious proximate/ultimate distinction (53). Biomedical sciences are primarily interested in the "how?" questions and whether it is possible to interfere with the "how?". The "why?" question, however, is not immediately relevant. Tied to this is the question of what, exactly, a relevant phenotypic difference is – another definition ENCODE proponents and their critics might disagree about: A very slight phenotypic difference might be sufficient in the medical context to aim for a particular intervention, whereas it might seem quite irrelevant from an evolutionary perspective.

The se/cr distinction was not "native" to the ENCODE discourse but got imported from an ongoing discourse in the philosophy of science. Within the philosophy of science, there is somewhat of a consensus that se and cr definitions of function are complementary, given that several instances within biology cannot be approached by a se concept alone. They are not easily separable, however (40). In fact, some disciplines within biology might find the SE concept entirely useless. Examples are, for instance, oncology or functional morphology (44). Also, it has recently been argued that the cr concept of functioning is more basic than the se concept, as se functions are logically dependent on cr functions (66). The philosophical consensus on the plurality of the use of functioning is deeply practice-based meaning that philosophers take the prevalence of a particular concept in a particular branch of biology as a good indication of its epistemic benefits (44).

Is there a way to make ENCODE fit for the age of the post-genome? We agree with what many others have argued before, ENCODE mainly is to be looked at as a reservoir of data. Thus, in principle, it should be easy to use it in a way that fits other perspectives on the human genome. It is, of course, important to note, again pointing to ENCODE's disciplinary situatedness, that ENCODE is not a project that was designed to answer questions pertaining to the field of evolution. ENCODE-generated data cannot settle debates currently going on within evolutionary biology. But, and this connects to a point we tried to make throughout this article, it contributes to our perspective on the organism, and how much the physiological and the evolutionary should be separated. By the insistence of ENCODE critics on a strict separation of cr and se functioning, this separation is, obviously, maintained.

Conceptually, ENCODE is not tied to any concept of function and is primarily disposed to a particular concept given its methodological setup and purpose. The question comes down to what it means for a transcript to function biochemically. If we regard the genome as responsive to environmental cues, then being transcriptionally active means more than just potentially contributing to a particular property of a system. It also means being a sequence currently being shaped by interactions with the environment and thus more vulnerable to directed changes to the genomic sequence. In that scenario, ENCODE did not only catch gene expression "in the act" but does the same for epigenetic canalization. However, can this really be done without studying environmental conditions, levels of biological organization, changes throughout the life-cycle, and the pre-history of genome archictectures? The proponents of new approaches to evolution would think not.

1. Eddy SR, The C-value paradox, junk DNA and ENCODE. *Current biology*. 2012;*22*(21): R898-R899.
2. Ohno S, So much'junk'DNA in our genome. In *Evolution of Genetic Systems, Brookhaven Symp. Biol.* 1972: 366-370
3. Palazzo AF, Gregory TR. The Case for Junk DNA. PLoS Genet 2014;*10*: e1004351.
4. Gregory TR, The evolution of the genome. London: Academic Press; 2005
5. Zuckerkandl E. Revisiting junk DNA. *Journal of molecular evolution* 1992, *34*(3): 259-271.
6. McClintock B. Mechanisms that rapidly reorganize the genome. Stadler Genetic Symposia 1978, 10: 25-48
7. McClintock B. The significance of responses of the genome to challenge. Science 1984, *226*: 792–801.
8. Comfort N. The Tangled Field. Cambridge, MA: Harvard University Press, Cambridge; 2001
9. Lamm E. The genome as a developmental organ. The Journal of Physiology 2014, *592*: 2283–2293.
10. Dawkins R. *The selfish gene*. Oxford: Oxford University Press; 1978
11. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature 1980*, *284*(5757): 601-603.
12. Orgel LE, Crick, FH. Selfish DNA: the ultimate parasite. *Nature 1980*, *284*(5757): 604-607.
13. Burt A , Trivers R. Genes in conflict Cambridge, MA: Harvard University Press; 2006
14. Lamm E, Jablonka E. The Nurture of Nature: Hereditary Plasticity in Evolution. Philosophical Psychology 2008 *21*: 305–319.
15. Graur D. Rubbish DNA: The Functionless Fraction of the Human Genome. In: Saitou N, editor. Evolution of the Human Genome I: The Genome and Genes. Tokyo: Springer Japan; 2017. p. 19–60.
16. Muller HJ. Our load of mutations. American Journal of Human Genetics 1950; *2*: 111- 176.
17. Paul DB. “Our load of mutations” revisited. J Hist Biol 1987; *20*: 321–335.
18. Lamm E. Systems Thinking Versus Population Thinking: Genotype Integration and Chromosomal Organization 1930s–1950s. J Hist Biol 2015; 48(4): 1–37.
19. Bulmer MG. Maintenance of genetic variability by mutation–selection balance: a child's guide through the jungle. *Genome 1989*, *31*(2): 761-767.
20. Wagner GP. Multivariate mutation-selection balance with constrained pleiotropic effects. *Genetics 1989*, *122*(1): 223-234.
21. Bürger, R. Mathematical properties of mutation-selection models. *Genetica 1998*, *102*: 279-298.
22. Beatty J, Desjardins, EC. Natural selection and history. Biol Philos 2009, *24*: 231–246.
23. Lewontin RC. The Bases of Conflict in Biological Explanation. Journal of the History of Biology 1969,*2*: 35–45.
24. Brown JR. Ancient horizontal gene transfer. *Nature Reviews Genetics 2003*, *4*(2): 121-132.
25. Timmis JN, Ayliffe MA, Huang CY, Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature reviews genetics 2004*, *5*(2): 123-135.
26. Hotopp JCD. Horizontal gene transfer between bacteria and animals. *Trends in genetics 2011*, *27*(4): 157-163.
27. Dawe RK. RNA interference, transposons, and the centromere. *The Plant Cell 2003*, *15*(2): 297-301.
28. Grewal SI and Elgin SC Heterochromatin: new possibilities for the inheritance of structure. *Current opinion in genetics & development 2002*, *12*(2): 178-187.
29. Campos, E. I., Stafford, J. M., & Reinberg, D. (2014). Epigenetic inheritance: histone bookmarks across generations. *Trends in cell biology*, *24*(11), 664-674.
30. Erwin DH, Valentine JW. "Hopeful monsters," transposons, and Metazoan radiation. *Proceedings of the National Academy of Sciences 1984*, *81*(17): 5482-5483.
31. Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences 1997*, *94*(15): 7704-7711.
32. Capy P, Gasperi G, Biémont C, Bazin C. Stress and transposable elements: co-evolution or useful parasites?. *Heredity 2000*, *85*(2): 101-106.
33. Consortium, T.E.P. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLOS Biology 2011, *9*, e1001046.
34. Jablonka E, Lamm E. Commentary: The epigenotype—a dynamic network view of development. Int J Epidemiol 2012, *41*: 16–20.
35. Brenner S. Refuge of spandrels. *Current Biology 1998*, *8*(19): 669.
36. Wright L. Functions, The Philosophical Review 1973, 82: 139–168.
37. Millikan RG. In Defense of Proper Functions, Philosophy of Science 1989, 56: 288–302
38. Neander K. The Teleological Notion of ‘Function‘ Australasian Journal of Philosophy 1991. *69*(4): 454-468
39. Godfrey-Smith P. A modern history theory of functions. Nous 1994, *28*: 344–362.
40. Millikan, RG. Biofunctions: Two Paradigms. In Ariew A, Cummins R, Perlman M, editors. Functions: New Essays in the Philosophy of Psychology and Biology, Oxford: Clarendon Press 2002, p. 113-143
41. Kaiser MI. ENCODE and the parts of the human genome. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 2018*, *72*: 28-37.
42. Stamatoyannopoulos JA. What does our genome encode?. *Genome research 2012*, *22*(9): 1602-1611.
43. Doolittle WF. Is junk DNA bunk? A critique of ENCODE. *Proceedings of the National Academy of Sciences 2013*, *110*(14): 5294-5300.
44. Elliott TA, Linquist S, Gregory TR. Conceptual and empirical challenges of ascribing functions to transposable elements. *The American Naturalist 2014*, *184*(1): 14-24.
45. Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, Elhaik E. On the immortality of television sets:“function” in the human genome according to the evolution-free gospel of ENCODE. *Genome biology and evolution 2013*, *5*(3): 578-590.
46. Fox Keller E. The century of the gene. Cambridge, Ma: Harvard University Press, 2000.
47. Rheinberger HJ, Müller-Wille S. *The gene: From genetics to postgenomics*. Chicago: University of Chicago Press, 2018.
48. Germain PL, Ratti E, Boem F. Junk or functional DNA? ENCODE and the function controversy. *Biology & Philosophy 2014*, *29*(6): 807-831.
49. Nelkin D. Creation versus evolution at the millennium. *Science as Culture 2000*, *9*(4): 535-542.
50. Laland K, Uller T, Feldman M, Sterelny K, Müller GB, Moczek A, Jablonka E, Odling-Smee J, Wray GA, Hoekstra HE. Does evolutionary theory need a rethink? Nature 2014, *514*: 161–164.
51. Pigliucci M, Müller GB. Evolution: the extended synthesis. Cambridge, MA: MIT Press, 2010.
52. Jablonka E. Genes as Followers in Evolution – A Post-synthesis Synthesis? Biology and Philosophy 2016, *21*: 143–154.
53. Laland KN, Sterelny K, Odling-Smee J, Hoppitt W, Uller T. Cause and Effect in Biology Revisited: Is Mayr's Proximate-Ultimate Dichotomy Still Useful? Science 2011, *334*: 1512–1516.
54. Auboeuf D. Genome evolution is driven by gene expression‐generated biophysical constraints through RNA‐directed genetic variation: A hypothesis. *Bioessays 2017*, *39*(10): 1700069.
55. Merton, RK. A note on science and democracy. *J. Legal & Pol. Soc. 1942*, *1*: 115.
56. Haraway, DJ. Modest\_witness@ second\_millennium. In *Modest\_Witness@ Second\_Millennium. FemaleMan© \_Meets\_OncoMouseTM*, London:Routledge, 1997
57. Epstein S. *Impure science: AIDS, activism, and the politics of knowledge*. Los Angeles, CA: Univ of California Press, 1996
58. Gieryn TF. Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American sociological review 1983*: 781-795.
59. Mullins NC. The development of a scientific specialty: The phage group and the origins of molecular biology. *Minerva 1972:* 51-82.
60. Caporale LH. The Implicit Genome. Oxford: Oxford University Press, 2006.
61. Fedoroff N, Botstein D. The Dynamic genome. Plainview, NY: Cold Spring Harbor Laboratory Press, 1992
62. Fontdevila A. The Dynamic Genome: A Darwinian Approach. Oxford: Oxford University Press, 2011.
63. Shapiro JA. Natural genetic engineering in evolution. Genetica 1992, *86*: 99–111.
64. Shapiro JA. Evolution: a view from the 21st century. London: Pearson Education, 2011
65. Lamm E. The Metastable Genome: A Lamarckian Organ in a Darwinian World? In Jablonka E, Gissis S, editors, Transformations of Lamarckism: From Subtle Fluids to Molecular Biology, Cambridge, MA: MIT Press. 2011
66. Griffiths PE. Function, homology, and character individuation. *Philosophy of science 2006*, *73*(1): 1-25.
1. Graur D (2013) The Origin of Junk DNA: A Historical Whodunnit. Judge Starling. Available: <http://judgestarling.tumblr.com/post/64504735261/the-origin-of-junk-dna-a-historical-whodunnit>. (Access 1/6/2022). [↑](#footnote-ref-1)
2. <https://www.encodeproject.org/help/project-overview/> (Accessed 1/6/2022). [↑](#footnote-ref-2)
3. See, for example, “The Cost to Science of the ENCODE Publication Embargo”. <https://caseybergman.wordpress.com/2012/09/05/the-cost-to-science-of-the-encode-publication-embargo/> (Access 1/6/2022). [↑](#footnote-ref-3)
4. https://www.nytimes.com/2012/09/06/science/far-from-junk-dna-dark-matter-proves-crucial-to-health.html retrieved 13.6.2022. [↑](#footnote-ref-4)