Biology & Philosophy (2022) 37: 28
https://doi.org/10.1007/s10539-022-09853-2

Causal-role myopia and the functional investigation of junk DNA

Stefan Linquist[1]

Received: 3 August 2021 / Accepted: 4 January 2022 / Published online: 29 June 2022

Abstract

The distinction between causal role (CR) and selected effect (SE) functions is typically framed in terms of their respective explanatory roles. However, much of the controversy over functions in genomics takes place in an investigative, not an explanatory context. Specifically, the process of component-driven functional investigation begins with the designation of some genetic or epigenetic element as functional —i.e. not junk— because it possesses properties that, arguably, suggest some biologically interesting organismal effect. The investigative process then proceeds, in a bottom-up fashion, to search for those effects. I argue that this process encounters a problem reminiscent of one that Gould and Lewontin (1979) associated with the adaptationist program. Just as their stereotypical adaptationst became trapped in the myopic pursuit of one selectionist hypothesis after another, so can the investigation of CR functions in genomics lead to an unending series putative organism-level CR functions for junk DNA. This is an acute problem for genomics, because (1) eukaryotic genomes are littered with transposable elements (TEs) and their deactivated descendants which (2) often masquerade as interesting CR-functional components and (3) it is experimentally onerous to determine whether they lack such a function. I further argue that selectionist reasoning about TE-host coevolutionary dynamics can greatly streamline the investigative process. Importantly, selectionist hypotheses need not be well confirmed to be illuminating in this context. Informed selectionist reasoning about the strategic roles of TEs in the genome offers a corrective to the idea that most of our DNA is somehow CR (and possibly SE) functional for the organism.

Keywords Function concepts · Genomics · Causal-role · Selected-effect · Transposable elements · Junk DNA

✉ Stefan Linquist
linquist@uoguelph.ca

[1] University of Guelph, OntarioGuelph, Canada

"Obviously, the philosophically correct thing to be is a pluralist, regardless of the position one holds." - David Hull, 1998.

Introduction

The pluralist consensus in the philosophy of biology claims that there are at least two distinct conceptions of function –selected effect (SE) and causal role (CR) – which perform complementary explanatory duties. Put simply, SE functions are said to explain "why a thing is there" while CR functions explain "how it works." Much of the ensuing debate has involved a boundary dispute over their respective scientific applications. Justin Garson (Garson 2016) provides a helpful distinction for navigating this terrain. Between-discipline pluralism holds that scientific disciplines can be classified according to which of the two function concepts they employ. Evolutionary biology and ecology are thought to employ SE functions. Meanwhile, such disciplines as functional morphology (Amundson and Lauder 1994), neuroscience (Craver 2001, 2013), and molecular biology (Griffiths 2006) have been identified as "experimental" sciences (Weber 2004) that employ CR function concepts. As Paul Griffiths describes it, "unless anatomy, physiology, molecular biology, developmental biology, and so forth turn their attention to specifically evolutionary questions, they investigate function in the causal sense" (2006, 3). Garson's second category, within-discipline-pluralism, sees at least some disciplines engaging in both tasks. Perhaps the discipline of animal physiology is an example where both CR and SE functional explanations are developed simultaneously. Even in this case, however, it should be possible to identify precisely which concept a researcher is using at a given point in time. This, I gather, is a common assumption for both forms of pluralism: that any scientific tokening of the word "function" can be disambiguated as either the CR or SE sense.

But how exactly does one disambiguate controversial or unclear cases? If the dispute over within- vs. between-discipline pluralism is to be resolved, there must be some way of answering this question. There needs to be some decision procedure for establishing exactly what a given researcher means on a given occasion when functional terms are employed.

The most obvious method, the one that I suspect philosophers typically rely on, is to look at the surrounding explanatory context. As a rule, if a biologist is explaining the evolutionary origins or the selective maintenance of a trait, then they must mean "function" in the SE sense. Alternatively, if a researcher is presenting a mechanistic explanation of how a trait contributes to the capacities of some containing system, then "function" is used in the CR sense.

This is a reasonable heuristic. The two modes of explanation are easily distinguishable because they describe events at different scales and involve distinct sorts of causal process. More important perhaps, very different kinds of evidence are required to establish the truth of a CR as opposed to an SE functional explanation. CR explanations are typically supported with direct experimental evidence showing that the modification or removal of some component affects the relevant system-level capacities. SE functional explanations, in contrast, require evidence about historical pro-

cesses that are often not amenable to direct experiment and involve more inferential arguments.

However, this reasonable picture can easily lead to a more misguided doctrine, that selectionist reasoning has no place in the scientific *investigation* (as opposed to the explanation) of CR functions (e.g. Cummins 1976; Amundson and Lauder 1994; Griffiths 2006; Brandon 2013). Because reasoning about selective history is often associated with SE functional explanation, other epistemic benefits of this practice, especially pertaining to the investigation of proximate mechanisms, are easily overlooked. The aim of this paper is to show that even when a researcher is not interested in explaining the evolutionary origin or selective maintenance of some trait, informed speculation about its selective history is sometimes useful (perhaps essential) for investigating its biologically interesting causal roles.

The discipline of genomics offers an instructive case. A central objective of genomics is to determine the potential functions of so called "noncoding" genetic elements. This category includes not only the DNA falling outside of known protein-coding regions, but also various epigenetic marks such as methylation patterns and histone modifications, as well as certain DNA products that do not encode protein (e.g. long noncoding RNA). Genomics researchers sometimes attempt to assign functions to these entities by engaging in *component-driven* functional research (discussed below). This begins with some type of entity thought to be functionally salient. For example, a strand of DNA that attracts transcription factors to its binding site. The component-driven method proceeds to search for biologically interesting system-level effects. For example, manipulating the binding site and looking for developmental effects in a certain tissue. The reason that this procedure benefits from (or maybe even requires) selectionist reasoning stems from an important fact about all plant and animal genomes: they are largely structured by transposable elements (TEs), also known as "jumping genes." These entities behave like parasites, self-replicating within the germline and reinserting daughter copies into new chromosomal locations. I will say more about TEs and their evolutionary dynamics in what follows. The important thing to note at this point is that TE insertions are often functionally ambiguous. It can be very difficult to determine whether the local properties of a TE insertion are related to some interesting CR function for the organism or, alternatively, whether they are associated with some selfish strategy of the TE and perhaps phenotypically neutral for the organism. This ambiguity can lead a researcher down blind alleys, fruitlessly entertaining one biologically interesting organism-level capacity after another when, in fact, none exist. This pursuit can be avoided, I argue, by thinking carefully about the coevolutionary dynamics between TEs and the host genome.

Before proceeding, a few points of clarification will help situate my position. First, this is not an argument for methodological adaptationism (Godfrey-Smith 2001). That method is for investigating SE, not CR functional hypotheses. Specifically, the methodological adaptationist asks, "what ecological role(s), if any, was this trait selected to perform?" They pursue this question by entertaining and testing various selectionist hypotheses, up to a point where the list of plausible candidates runs dry. In contrast, the mode of CR functional investigation that I am highlighting asks, "given some conspicuous structure or trait, what causal contributions does it make,

if any, to some containing system?" Selectionist hypotheses help guide and constrain the process, pointing to likely experimental hypotheses and by suggesting more (or less) plausible interpretations of data. Such applications of selectionist reasoning in CR functional investigation will be described in more detail momentarily.

Notably, however, my central argument mirrors a popular *criticism* of methodological adaptationism. In their famous "Spandrals" paper, Steven Gould and Richard Lewontin (1979) argued that the adaptationist program leads to a kind of myopic focus on selectionist hypotheses while systematically ignoring non-adaptationist alternatives: "If one adaptive argument fails, try another." (1979, 152). Lloyd (2015) has developed this criticism, arguing that choices made early on in a research program tend to guide and constrain the hypotheses that follow. If a researcher begins, even provisionally, with the assumption that some trait is an adaptation, then, according to Lloyd, the ensuing research project will take a very different shape than if it had started with the assumption that it is a developmental byproduct or the result of drift. This is not the place to evaluate whether methodological adaptationism is threatened by this argument. Rather, my contention is that a similar myopia can arise in the pursuit of CR functions. Indeed, the discipline of genomics offers many examples where some non-coding genetic element is presumed to have some biologically significant effect or other on the organism (Nowak 1994; Ponicsan et al. 2010; Kim et al. 2012; Mattick and Dinger 2013; Biscotti et al. 2015; Larsen 2018; Parenteau and Abou Elela 2019). When one CR functional hypothesis fails, the protocol, often, is to try another. This process can go on indefinitely, or so I argue below, leading to the systematic neglect of alternative hypotheses.

Some readers might already be anticipating an objection that stems from the permissiveness of CR functional ascription. I have just suggested that some genetic elements might lack an "interesting" CR function at the organism level, and that the failure to consider selectionist hypotheses leads some researchers to overlook this possibility. Technically speaking, it is impossible for any entity contained within an organism to lack CR functions altogether. This is because CR functions are defined in relation to the interests of an investigator. Suppose that some strand of DNA is truly junk in the sense that it has no biologically significant phenotypic effect on the organism. The sequence is still part of multiple containing systems. It is located on a chromosome, inside a nucleus, within a cell, and so on. It also engages in various causal interactions with its surroundings. For instance, in germ cells the "non-functional" sequence engages with transcription machinery. So, what could it possibly mean for a DNA sequence to lack an organismal CR function? What, for that matter, are debates in genomics about the existence of "junk" or non-functional DNA about?

This touches on a problem that the functional pluralist must somehow contend with, the fact that any entity whatsoever can be construed as CR functional so long as we imagine the right sorts of investigative interests (Amundson and Lauder 1994). Hence, an investigator interested in the sheer volume of DNA within the cell could view junk DNA as "having the function" of contributing to this property. More outlandishly, one of the CR functions of junk DNA is to contribute to disagreements in the field of genomics. What exactly might be wrong with such claims? Within genomics, as with any biological science, there are implicit criteria governing functional ascriptions. It isn't kosher to ascribe functions to just any DNA sequence.

The recent controversy generated by ENCODE's claim that over 80% of the human genome has some "biochemical function" is an example where conventions surrounding functional ascription were violated (Eddy 2012; Graur et al. 2013; Doolittle 2013; Elliott et al. 2014). Presumably some of these conventions are non-arbitrary –they reflect ontological facts about the scientific subject matter. To avoid trivializing the ENCODE debate, and discussions over junk DNA more broadly, we must identify some reasonable constraints on the kinds of systems or capacities to which an element contributes in order to be legitimately regarded, in genomics, as functional. What exactly are those constraints?

One possible solution to the permissiveness problem appeals to implicit selectionist considerations that might be doing important work behind the scenes. Ruth Millikan (Millikan 2002) argues that considerations about ecological function are necessary for individuating the containing systems to which an entity functionally contributes. Along similar lines, Linquist, Doolittle and Palazzo (Linquist et al. 2020) argue that genomics researchers operate with two distinct conceptions of SE function: origin functions explain how a sequence or trait originated in a population, maintenance functions explain how it has persisted in some system, for instance, by contributing to organismal survival and reproduction. It is implausible that genomics researchers are always talking about origin functions when identifying the functional contributions of genetic elements. However, Linquist et al. (2020) propose that they are quite reasonably operating with the conception of a maintenance function –an under-appreciated species of SE function. One advantage of this proposal is that maintenance functions do not encounter the permissiveness problem. This is because they are not interest relative. Either some element is maintained by selection at some level or it is not, regardless of how one chooses to carve up the system and its capacities. I also suspect that this reading of the junk DNA debate better accords with scientific practice than functional pluralism (e.g. Graur et al. 2013; Graur et al. 2015). However, my argument in this paper does not presuppose this interpretation.

An alternative way to avoid the permissiveness problem assumes that there are other, non-selectionist criteria in the background when genomics researchers ask whether a sequence or trait has a non-trivial CR function. For instance, Germain, Ratti and Boem (Germain et al. 2014) propose that biomedical interests constrain whether certain elements are considered functional in genomics. Hence, it makes sense in some contexts to say that a sequence has the function of contributing to a cancerous tissue. Arguably, this is not a form of SE functional attribution (Garson 2017), and could thereby constrain the practice of CR functional attribution without appealing to some species of SE function. Alternatively, Griffiths (2006, 2009) proposes that genomics researchers might be interested in "future directed" functions which, allegedly, are not implicitly about selective history (see Sect. 5). My aim in this paper is to show that regardless of how one resolves the permissiveness problem, selectionist reasoning should not be abandoned in genomics. Obviously, if we pursue the path of Linquist et al. (2020) and substitute CR functions with SE maintenance functions, selectionist considerations are relevant. What remains less obvious is that selectionist reasoning remains unavoidable even for those who pursue a different path, arguing genomics largely involves the pursuit of CR functions. The one caveat I insist upon is that, in order to do justice to scientific debates over junk DNA, propo-

nents of CR function must identify some constraint on either the kinds of systems or the sorts of capacities that could be considered relevant to a CR-functional investigation. Aside of selective origin or maintenance, I don't know what those criteria are. I therefore use the phrase "biologically interesting" as a placeholder for the relevant notion, whatever it might be. Hence, if genomics researchers are to traffic in CR functions and avoid the permissiveness problem, then they must restrict their focus to biologically interesting system-level capacities. My contention is that selectionist reasoning plays an important epistemic role in the investigation of which components contribute to such capacities, and how they do so.

The argument is structured as follows. Section 3 develops the distinction between functional explanation and functional investigation. This allows for a fairly precise account of component-driven CR functional investigation. Section 4 explains the problem of functional ambiguity in genomics. In particular, I discuss the phenomenon of causal role myopia: the positing of one CR functional hypothesis after another without considering the possibility that an entity lacks any biologically interesting effect on the organism. The final section addresses a challenge from Griffiths (2009), who argues that although selectionist thinking might be instrumental for individuating systems, it should be "forward looking" and not historical. Before proceeding, however, it is perhaps helpful to sketch my thesis using a true story that has nothing to do with either genomics or philosophy.

A toy example

A few years ago, my daughter acquired a musical toy. On the surface are buttons that play different notes and on the underside is a circular cavity that amplifies the sound generated within. Inserted tightly into the cavity is a circular plastic stopper –effectively a cap– that can be removed and reinserted with a moderate amount of force. What, we might ask, is the function of this cap?

Philosophers will be quick to point out that the question is ambiguous. Am I asking how the cap got there in the first place? Or is my question about how the cap contributes to certain capacities of the system? Let us set aside questions of origin and focus on the second question about CR function. There are all sorts of boring system-level capacities to which the cap contributes that no one in their right mind would entertain. Obviously, the cap contributes to the overall mass of the toy. Who cares? What we are interested in are the *interesting* capacities of the system (following my earlier caveat) to which this component functionally contributes.

One hypothesis is that the cap functions as a noise dampener. We could test this suggestion by taking decibel readings with the cap removed versus in its home. Another hypothesis is that the cap encloses a storage compartment for other, smaller toys. This function could be tested by loading the compartment and shaking the toy around to see how well it holds. A third possibility is that the cap functions to keep small fingers out of the inner workings. Or maybe it functions to keep stray liquids out of the electronics. The list of potential CR functions is sizable. I had my preferred ranking about their relative probabilities. Though I refrained from engaging in any serious experimentation.

Then one day we received an important piece of information. It turns out that the cap did not originally come with the musical toy. A child who owned the toy before us ripped it off some other plaything and inserted it into a suitable hole. And there it stayed. This important piece of information caused me to reorder my rankings. For instance, there is no reason to think that the cap would be effective in keeping fingers out of the internal component. After all, if a child's fingers put it in place they could just as easily remove it. Nor would we expect the cap to do a particularly good job at storing items or dampening sound. After all, it had been inserted into the hole essentially by accident. It suddenly become much more plausible that the cap might not have any interesting CR functions to speak of. This became a live hypothesis only after making the discovery about the cap's "selective" history. The analogy between this system and the roles of TEs in the genome will hopefully become obvious as we proceed.

CR functions in philosophy and genomics

As I mentioned earlier, a common way of disambiguating token instances of "function" is by looking at explanatory context. Accordingly, pluralist philosophers who defend the importance of CR functions in biology point to textbook examples of functional explanation –cases where the investigative process is finished. For instance, Amundson and Lauder (1994) cite the centrality of CR functions for the discipline of functional anatomy. Craver (2013) points to the textbook explanation of a neurotransmitter to show that mechanistic explanation in neuroscience is a species of CR, not SE functional explanation. Arno Wouters (Wouters 2005; Weber 2004) likewise cites examples from "experimental" biology –which includes such fields as molecular biology and physiology—to suggest that CR functions are the dominant explanatory framework therein.

For all of this emphasis on functional explanation, relatively little has been said about the various non-explanatory uses of function concepts in biology. Non-explanatory uses include not only prediction, but also such informal practices as the formulation of research questions, the ranking of competing hypotheses, or even the procurement of research funding. To be clear, I use "explanation" as a success term. It is the goal toward which much (but not all) scientific effort is ultimately directed. Although explanation is obviously an important part of science, to focus on it exclusively is to overlook a considerable amount of scientific practice. Nowhere is this better illustrated than by the science of genomics. This discipline was born out of the Human Genome Project (HGP) with its achievement of generating a comprehensive mapping of human DNA sequences. Much subsequent work in genomics has likewise involved high-resolution mapping of various species' genomes. Such mappings provide a valuable resource for comparative and experimental research. For example, they are useful in formulating and testing hypotheses about the specific phenotypic effects of certain genetic elements. However, as early critics of the HGP argued, a genomic mapping cannot itself explain the mechanistic functions of genes (Tauber and Sarkar 1992; Rosenberg 1994). This point extends to mapping practices we find in more recent extensions of the HGP. For instance, the aim of the ENCODE project

was "to delineate all functional elements encoded in the human genome" by cataloging "regions of transcription, transcription factor association, chromatin structure and histone modification" (The ENCODE Project Consortium 2012, p. 57). Despite their emphasis on function, ENCODE researchers did not attempt to identify the phenotypic effects of the elements under investigation, let alone explain the causal processes by which those effects are generated. Instead, they focused on a handful of very localized biochemical properties associated with nuclear DNA. These were regarded as reliable proxies for more interesting functional roles. Like many research projects in functional genomics, the aim of ENCODE was annotation, not explanation. It is therefore interesting that the ENCODE project nevertheless generated so much controversy. Their highly publicized finding that roughly 80% of the human genome has some "biochemical function" was criticized for its non-standard use of the term function (Doolittle et al. 2014; Brunet and Doolittle 2014) and for its reliance on overly-permissive functional proxies (Eddy 2012; Niu and Jiang 2013; Graur et al. 2013; Doolittle 2013; Elliott et al. 2014). I will have more to say about these objections in what follows. My current point is simply that the recent debate in genomics over how to distinguish functional from junk DNA cannot be about the explanatory roles of function concepts, because ENCODE ascribed functions without proffering explanations.

What then was the ECNODE controversy about? My sense is that many critics objected to ENCODE's working definition of "biochemical function" because, in the critics' view, it is an unsuitable category for functional *investigation.* I use "functional investigation" to refer to the scientific process leading up to a functional explanation. Described in philosophical terms, functional investigation in genomics can either be component-driven or capacity-driven. Component-driven functional investigation starts with a given category of genetic element that is presumed to contribute to some system-level capacity of biological interest. It proceeds to generate and test more refined hypotheses about how such components contribute to specific system-level capacities. For example, this process might start by classifying lncRNA as a type of element that is of potential functional significance for the organism, simply because it is so abundant in cell nuclei. The investigation proceeds to determine whether, for example, a specific RNA transcript is involved in regulating the expression of a certain gene. Capacity-driven functional investigation moves in the opposite direction. It starts with some system-level capacity of biological interest and then "works downwards" to identify the genetic and epigenetic components involved in generating that capacity. For example, which genetic mutations are involved in breast cancer? Which DNA repair mechanisms suppress mutation? And so on.

Before exploring the problem of causal-role myopia in genomics, it is important to highlight a second trend in the philosophy of CR functions. It is often argued CR functional explanations make no reference to the selective history of an organism. This suggestion can be traced to Robert Cummins' (1976) pioneering work on CR functions. Famously, Cummins was interested in a form of explanation that he dubbed *functional analysis*, where the CR functional components of a system explain how some system-level capacity is generated. Importantly for Cummins,

> Functional analysis can properly be carried out in biology quite independently of evolutionary considerations: a complex capacity of an organism… may be explained by appeal to a functional analysis regardless of how it relates to the organism's capacity to maintain the species. (1976, p. 756).

Notice that in this passage Cummins slides between functional investigation and functional explanation. He begins by alluding to the investigative conditions under which functional analysis is "properly carried out" but then alludes to the a-historical nature of CR functional explanation itself. This overlooks the possibility that the role of evolutionary considerations might differ across these two contexts. Setting this quibble aside, I take his central point to be that functional explanation need make no refence to natural selection. This follows from Cummins' contention that it is up to the scientist or "investigator" to choose which system-level capacities to functionally analyse. According to Cummins, evolutionary considerations do not constrain an investigator's decision on this matter. An example from genomics illustrates Cummins' point.

Suppose that we are interested in developing a strain of rice that can withstand dramatic temperature fluctuations. We proceed to grow various strains under a range of temperature conditions, singling out those which are most resilient. The next step might be a genomic comparison looking for allelic differences among the strains. Several candidate alleles are identified that potentially contribute to temperature resilience. Further experimentation on those alleles reveals precisely how they generate this system-level capacity. Following Cummins, it seems obvious that selective history is nowhere to be found at the final step of this process, where we have an explanation of how certain alleles contribute to temperature resilience. As far as our explanation is concerned, it wouldn't matter if God had created our rice strains de novo or if they had arrived on a spaceship from another planet. A slightly less obvious question is whether the investigation leading up to this explanation could have proceeded without taking selective history into account. Many proponents of CR function consider selective history irrelevant even to functional investigation (Amundson and Lauder 1994; Weber 2004; Wouters 2005; Griffiths 2006, 2009). However, this remains a controversial issue (Garson 2011, 2017).

As it was touched upon earlier, Millikan (2002) presents a compelling reason why evolutionary history plays an essential but often implicit role in CR functional investigation. Millikan's focus is on the system-level capacities that are chosen for functional analysis. How does a biologist determine which objects to regard as "systems" in the first place? As Millikan puts it, "[l]iving chunks of matter do not come, just as such, with instructions about what are allowable conditions of operation and what is to count as allowable input" (Millikan 2002, p. 21). CR functional investigation, to get off the ground, presupposes a considerable amount of biological information. Hence, according to Millikan, if corn plants had suddenly appeared on planet Earth five minutes before our investigation, we might not know how to individuate them as systems. At the very least, we wouldn't have expectations about their temperature thresholds. It is only by viewing them as familiar a species of plant (*Zea mays*), which has been selected to survive and reproduce under familiar terrestrial conditions, that we are able identify them as systems with certain capacities. It is in pro-

viding this essential background information that, according to Millikan, selective history informs CR functional investigation.

There are two sorts of objection that one might raise against this argument. The stronger objection simply denies that evolutionary considerations are necessary for individuating systems and their capacities. It is perhaps conceivable that biologists could hit upon a reasonable way to individuate alien corn plants as plants. Likewise, they would eventually hit upon their temperature thresholds through trial and error. Hence there is room for skepticism about Millikan's claim that evolutionary considerations *necessarily* inform CR functional investigation[1]. A somewhat milder objection relies on a distinction between evolutionary history and selective history. As Amundson and Lauder (1994) point out, many biological systems and their capacities are standardly individuated in terms of homology (common ancestry). However, they add, the relation of homology does not depend on selective history. Hence, while the individuation and functional investigation of a system might rely on evolutionary history, qua phylogeny, it does not presuppose selective history. This applies to our corn example. Our ability to successfully individuate (Earthly) corn plants and zero-in on reasonable temperature thresholds (arguably) depends not on background knowledge about the conditions under which they were selected. Rather, it is an extrapolation from other plants that we have encountered plus the knowledge of their common ancestry[2].

In practice, however, it difficult to tease these alternatives apart. We know that most traits are shaped by a combination of selective history and common ancestry. What does it mean to systematically ignore one factor and focus on the other? It is a curious feature of the philosophical debate over CR functions that phylogeny and selection are being artificially separated, as if one factor is more basic than the other. Surely scientists should draw on both types of consideration when functionally investigating biological systems.

My sense is that Amundson and Lauder's resistance to selective history is motivated by an epistemic worry. As they note, hypotheses about the kinds of selective pressures that might have shaped a system in the past can be difficult to confirm. This is reminiscent of a sentiment that one sometimes encounters in the experimental sciences, that evolutionary reasoning is just so "wishy washy." On this view, the hard-headed business of CR functional investigation can and should limit investigations to experimentally demonstrable effects. Amundson and Lauder explain,

> Even within modern populations, studies designed to show selection on a given trait are difficult, and are subject to numerous alternative interpretations and confounding effects… Functional morphologists do not have the luxury of simply asserting that the SE function of a structure X is F (as philosophers so

[1] I find it difficult to evaluate this objection because my intuitions about how CR functional investigation might proceed in a non-Darwinian world are so unreliable. To once again sympathetically quote David Hull: "I remember when I had intuitions about what counts as a functional system and what does not; but after studying counter-example after counter-example my intuitions have become so battered that they are no longer of any use whatsoever" (1998 p. 224).

[2] See Griffiths (2006) for a more detailed account of the roles of homology in functional investigation and functional analysis.

> regularly do with the heart): there must be direct evidence selection acted on structure X for effect F. (Amundson and Lauder 1994, p. 248)

At the same time, most experimental biologists often aren't equipped to develop and test evolutionary hypotheses. Evolutionary considerations are unfamiliar territory for many genomics researchers and, the thinking goes, this shouldn't stand in the way of their making progress in such fields as cellular biology or biomedicine.

The point is well taken that CR functional investigation certainly can and often should proceed without a *well confirmed* explanation of how the components in a system evolved. An important question is whether merely *plausible* selectionist hypotheses can do the job. This question will be addressed in the following two sections.

A second response to Amundson and Lauder's epistemic worry is that not all biological systems have the same causal transparency. It might be possible to CR functionally investigate a morphological trait such as vertebrate jaw without having to consider its selective history. It plays an unambiguous role in a containing system which has capacities –biting and chewing– that no one could reasonably overlook. We do not find jaw-like structures repetitively distributed throughout the body, playing a number of potentially different causal roles at different locations. Nor are jaws capable of self-replicating independently of other organs. Hence, when it comes to identifying the relevant system, discerning its capacities, and investigating the contributions of specific jaw features to those capacities, a scientist's job involves no difficult matters of individuation.

Matters are quite different when it comes to genetic elements within eukaryotic genomes. It is important to recognize that the genome is much more than just a blueprint for coding an organism. It is estimated that less than 2% of the nuclear DNA in the human genome is comprised of "normal" protein-coding genes (Hatje et al. 2019). At least half of the human genome is the result of TE activity (Bannert and Kurth 2004). Some families of TE are more active than others. And within a family, some TE insertions are more causally significant than others. It is therefore a mistake to generalize about TEs as a class or even as a family (Doolittle, 2022) . However, it is a safe bet that most of the token TE insertions in the genome are no longer capable of transposition. Most of our DNA is comprised of the slowly mutating "corpses" of previously active TE insertions. As I explain in the following section, some of this junk DNA retains a level of biochemical activity that can give the misleading impression that it is a promising CR-functional component in the organism. Likewise for active TEs, which exhibit certain switch-like behaviours, it can mistakenly seem as if they perform interesting organism-level causal roles when this is not the case (Linquist and Fullerton 2021). It can be extremely difficult to confirm these CR functional hypotheses experimentally. This places the experimentalist in a similar epistemic predicament to the methodological adaptationist, with indefinitely many functional hypotheses on the table and no simple way to eliminate them experimentally. Somewhat ironically, reasoning about selective history can help the experimentalist to rank the likelihood of those CR functional hypotheses.

A third response to the epistemic objection raised by Amundson and Lauder (1994) is that just because something is difficult doesn't mean that it is optional. It would be nice if genetic elements wore their CR-functional significance, as it were, on their

sleeves. Then it might be possible to simply read off from a genetic sequence the level (if any) at which it makes a biologically interesting contribution. Unfortunately, functional genomics is not so simple. Genetic elements are functionally ambiguous and an indispensable tool for identifying likely system-level contributions is to consider their selective history.

It is perhaps important to be clear that this view is not widely accepted within the discipline of functional genomics (Brunet and Doolittle 2015). While I doubt that most genomics researchers would come out and say that selective history is irrelevant for their CR-functional investigations (Kellis et al. 2014), it is clear from the behaviour of organizations like ENCODE that selective history does not inform their choice of CR-functional proxies. According to ENCODE, a sufficient condition for classifying a DNA sequence as "biochemically functional" is if it exhibits at least one of the following properties, at least once, in at least 1 of the147 cell types analyzed: (1) it is transcribed into RNA (but not necessarily translated into a protein), (2) it contains or is adjacent to a transcription binding factor, (3) it is a methylated CpG dinucleotide, (4) it is located in an area of open chromatin, or (5) it is found organized in nucleosomes containing certain histone modifications (The ENCODE Project Consortium 2012). This is an extremely permissive set of criteria. In the following section we shall consider why the first two conditions –transcription and proximity to regulatory regions– are unreliable proxies for identifying biologically interesting components. This might not be so obvious when the genome is viewed from an a-historical perspective, but it becomes immediately apparent once we consider the kinds of selective processes that structure it.

Not all genomics researchers adopt such permissive criteria as ENCODE. Some TEs remain dormant in the genome until the organism encounters physiological stress, at which point they jump into action. Many genomics researchers regard such "switch-like" behaviour as a strong indication that the element in question is a functional component of the organism. Again, in the next section we consider how this inference could lead a researcher down blind alleys which are potentially avoided by taking selective history into account.

An interesting question is whether causal-role myopia is characteristic of only component-driven CR functional investigation or, in addition, whether this phenomenon also arises in capacity-driven research. Indeed, there is an asymmetry. In the case of component-driven research it is possible for a salient component to have no biologically interesting capacity at the organism level. This opens the door to an open ended, but ultimately fruitless investigation. By contrast, in the case of capacity-driven CR investigation there will always be some lower-level components responsible for the capacity in question. Once a biologically interesting capacity has been identified, it is just a matter of drilling down. Hence, capacity-driven research, however circuitous, has a logical endpoint whereas component-driven research has no stopping rule if, indeed, the focal component lacks any biologically interesting system-level effect. This is why I suggest that selectionist reasoning might not only be beneficial to, but necessary for component-driven CR investigation. It might be the only way to develop an informed ranking of likely CR functional hypotheses.

Transposable elements: parts or parasites?

When it comes to "normal" protein-coding genes, transcription into RNA is often associated with translation into protein and thus serves as a reliable indicator of some organism-level CR functional role. After all, proteins are the building blocks of phenotypic traits. When it comes to the expanses of DNA outside protein-coding regions, much of this also codes for RNA (Johnson et al. 2005). However, most of this "noncoding" RNA is not further transported outside the nucleus and is therefore not translated into protein. These RNA molecules persist in the nucleus for a short period before decaying. An increasingly popular hypothesis in genomics is that much of this noncoding RNA is likely to be involved in the regulation of genes responsible for normal organismal development (Mattick et al. 2010; Mattick and Dinger 2013). In support of this idea, there is at least some evidence that specific noncoding RNAs can sometimes take on such organism-level functions (Palazzo and Lee 2015). The key question is whether it is reasonable to assume, when engaged in CR functional investigation, that most RNA transcripts are functional components in this sense. A popular sentiment in genomics is that since non-coding RNA is so abundant in the nucleus, it surely must be doing something functionally important for the organism (e.g. Mattick 2001; Willingham and Gingeras 2006).

This sentiment has given rise to an explosion of research, much of it in the field of biomedical genomics, attempting to isolate the regulatory functions of this or that strand of noncoding RNA (Hüttenhofer et al. 2005). These experiments can be difficult to perform and results are often inconclusive (Ponting and Hardison 2011). The usual way of testing the functionality of an RNA transcript is by silencing it. One then looks for an effect on the organismal phenotype. If such an effect is found, this is indicative of an organism-level CR function, but not necessarily an SE function (Linquist et al. 2020). Importantly, however, failure to identify a phenotypic effect is not conclusive evidence against CR functionality. One possibility is that the phenotypic effects were too cryptic to be detected. Even from one lab to the next, subtle differences can affect gene expression (Crabbe et al. 1999; Francis et al. 2003). Organisms raised in the relatively sterile conditions of a genomics lab do not encounter many of the environmental factors that they do in the wild. Presumably, certain regulatory functions are triggered only under specific conditions in nature. Unless one controls for every possible environmental factor that the organism is likely to encounter in the wild, there remains a chance that an RNA transcript might be interestingly functional. For a researcher dedicated to the idea that noncoding RNA must have some regulatory function, this process could go on indefinitely, or, at least until the funding runs out.

A second epistemic obstacle surrounds the looming possibility of functional redundancy (Ahituv et al. 2007; McLean and Bejerano 2008). Suppose that an RNA transcript does perform some important regulatory function. It seems likely that the organism would have evolved some redundancy in response to cases where the transcript is naturally silenced by mutation. Perhaps there are two, or three, or more kinds of RNA transcript that vary slightly in their sequence but perform the same regulatory function. Silencing just one of these transcripts would have no detectable phenotypic effect. Again, a researcher dedicated to the idea that noncoding RNA must have some

organismal function ("why else is it so abundant?") could pursue this inconclusive tack for a very long time.

A selectionist approach to the genome provides a corrective to this line of thinking. Active transposons impose a net fitness cost on the organism by inserting into "normal" genes, typically disrupting their expression, and occasionally causing massive chromosomal rearrangements (Kidwell and Lisch 2001). This imposes a selection pressure on the organism to remove these elements or otherwise silence TE activity. One mechanism for silencing TEs is through RNA interference (Roessler et al. 2018). The cell identifies RNA transcripts that descend from TEs and prevents them from exiting the nucleus. This prohibits the TE from producing the enzymes required to copy and reinsert into new locations in the genome. It has recently come to light that the nucleus serves as a filter for all sorts of RNA transcripts that would otherwise cause cellular damage (Martin and Koonin 2006; Palazzo and Lee 2018). The existence of this nuclear barrier means that a considerable amount of RNA is screened off from having any downstream functional significance. Thinking in selectionist terms, the evolutionary emergence of the nucleus allowed eukaryotes to relax the amount of "policing" taking place on the front line of the DNA strand, because rogue transcripts were now prevented from escaping the outer wall of the nucleus.

Other TEs are silenced "in their tracks" by being bound up by methylation and prohibited from transcription (Lisch and Bennetzen 2011). Methylation marks need to be re-inserted into the correct location each cell cycle and it is a difficult achievement to continue generating them over successive generations. Hence, we should expect methylation marks to quickly become relaxed after they cease to have their inhibitory functions. During the period when an otherwise active TE is bound up in a methyl group it gradually accumulates point mutations. Eventually it loses the ability to generate a readable RNA. Even if the RNA manages to breach the "outer wall" it won't be capable of generating a transposition-friendly enzyme. At this point the selection pressure on the organism to continue binding the TE is relaxed. It can be permitted to continue producing its decrepit transcripts because they no longer pose a mutational threat.

Given these considerations about TE/host evolutionary dynamics, one would expect the preponderance of noncoding RNA contained in the nucleus to lack any downstream causal significance for the organism. An evolutionary perspective suggests that only on very rare occasions will noncoding RNA acquire some secondary, regulatory function that is beneficial for the organism. This has implications for component-driven CR functional investigation in genomics. Unless a noncoding RNA transcript has some distinctive properties (e.g. its sequence is highly conserved over evolutionary time), it is unreasonable to regard that element as a likely component in some organism-level functional system (Palazzo and Lee 2015).

Therefore, it is quite surprising that ENCODE chose transcription as a proxy for CR functionality. A TE can continue to be transcribed long after its capcity for replication has faded. What were they thinking? Some apologists for ENCODE have suggested that the aim of this consortium was not merely to identify elements that contribute to organismal functions. In addition, ENCODE was interested in identifying elements that the contribute to disease (Germain et al. 2014). These authors make an important point, that at least some functional investigations in genomics

are interested in genetic sequences that contribute to things like tumors, which can be viewed as functional subsystems in their own right. But this is no excuse for ENCODE. If most RNA transcripts are produced by decaying TEs and enjoy only an ephemeral existence trapped behind the nuclear wall, they probably don't have the ability to impact the cell even in a negative way. This is why genomics researchers who think in terms of coevolutionary dynamics regard these structures as "transcriptional noise" (Palazzo and Gregory 2014).

As for ENCODE, one possible explanation for their oversight is that these biomedical genomicists were simply unacquainted with the theory behind TE evolutionary dynamics (Brunet and Doolittle 2015). Trained to think in terms of "normal" protein-coding genes, they applied conventional criteria for functionality to an unconventional domain. Elsewhere, it has been reported that the popularity of transposon research has declined over the past two decades withing biomedical science, despite increasing in almost every other branch of biology that was considered (Linquist and Fullerton 2021). Understanding why some genomics researchers overlook TE evolutionary dynamics is an ongoing area of research.

A slightly more credible indicator that a genetic element has some organism-level CR function is if it is located within the regulatory region of a "normal" gene. These regions contain transcription factor binding sites that effectively switch genes on and off. Interestingly, active TEs also contain binding sites that are typically used for their own selfish replication. However, if a TE were to land in the regulatory region of a gene and become stuck there, one could imagine that it could open up new regulatory opportunities (Bannert and Kurth 2004). The TE might draw in transcription factors under some novel conditions that now cause the gene to become active at a different point in the organism's development. It is conceivable that this will occasionally have beneficial effects on the organism. Hence, some genomics researchers describe TEs as "genomic engineers" because of their capacity to "rewire" regulatory gene networks (Shapiro 2000).

ENCODE researchers inferred that any TE located near a regulatory region potentially has a biologically interesting CR function and, in this sense, should not be written off as junk. Apparently, anything that resides in a regulatory region is considered, on this view, to be a linkely candidate for regulating the adjacent gene. ("Why else would it be there?") Once again, this inference ignores coevolutionary dynamics. If we think strategically from the perspective of a TE, it is advantageous to insert into locations that the host cell is incapable of silencing. When the cell silences a TE through methylation, adjacent sections of the chromosomes can get wrapped up along with it. Methylation is a blunt instrument for TE suppression. By inserting within or near regulatory regions, TEs make it potentially very costly for the cell to silence them through methylation without also wrapping up important regulatory regions. It might be less costly to simply "allow" the TE to persist unmolested within a regulatory region than to shut down the entire site at an elevated cost to the cell. This "safe-haven" hypothesis (Kidwell and Lisch 1998) provides an explanation for the accumulation of TEs within regulatory regions of the genome, but it does not

imply that these insertions will have biologically interesting effects on the organismal phenotype[3].

What does this mean for CR functional investigation? One lesson is that we can easily imagine how, in the absence of co-evolutionary reasoning, a researcher might become locked into the myopic search for organism-beneficial effects. The assumption that any insertion located in a regulatory region must be involved (somehow!) in gene regulation provides the fuel for this open-ended expedition. A selectionist perspective suggests a more thoughtful approach. We do not assume that TE insertions into regulatory regions must be organism-level CR-functional. Before heading down the experimental path we might first attempt to rule out the safe-haven hypothesis using comparative information. For instance, it is sometimes possible to estimate the "age" of an insertion. If an insertion is relatively recent, then it is less likely that it has some functional significance for the organism than if it has persisted at this location for a long time. Another important question is the diversity of TEs that inhabit a specific location. Suppose that within a sample of organisms there is considerable variation in the families of TE that are located at the same regulatory region. This suggests that we are looking at a safe-haven which any TE (regardless of its capacity to attract transcription factors) is happy to occupy. I do not suggest that such evidence is conclusive. However, it provides a way of estimating the likelihood that a given TE insertion is CR-functional for the organism. If the likelihood is low, then a researcher shouldn't waste her time laboring under the conviction that everything located in a regulatory region must have some regulatory function.

Moving away from the criteria that ENCODE proposed for identifying functionally salient genetic elements, another interesting property of certain TEs is their tendency to become active when the organism is exposed to physiological stress, such as a high dose or radiation or heat shock (Wessler 1996). Barbara McClintock proposed in her early work that the movement of TEs in and out of protein-coding regions was a normal mechanism for gene regulation (Comfort 1999). Later in her career she proposed the "genome shock" hypothesis to account for the elevated activity of TEs under stressful conditions (Fedoroff 2012). I interpret this (perhaps charitably) not as an SE functional hypothesis that aims to explain why conditionally activated TEs are maintained in the genome, but merely as a CR functional hypothesis about their occasional beneficial effects on organismal fitness. The basic idea is that TEs sit

[3] An objection to the safe-haven hypothesis, thanks to Ford Doolittle (pers. comm.) claims that it is unable to account for the fixation of a TE at a specific location. Suppose that a single TE inserts into a safe location in one token organism only. By hypothesis, this causes a non-lethal reduction in the fitness of its offspring. For the insertion to persist at this site, the effective population size must be such as to allow for a certain amount of genetic drift. Notably, some argue that this is a reasonable expectation for most eucaryotes (Lynch 2007).That aside, Doolittle's worry is that unless drift is exceptionally strong, it would be exceedingly unlikely for the deleterious insertion to reach fixation. Therefore, if a researcher identifies a TE that is fixed at some location, it is reasonable to assume that it has been favorably selected at the organism level. My response is that this argument assumes that the insertion into a safe-haven is a one-off event, or at least very infrequent occurrence. On the contrary, I take the safe-haven hypothesis to be that TEs are regularly reinserting into these favorable locations with a certain frequency. Effectively, this constitutes a lower level "ecological" process (Linquist et al. 2015) that counteracts the deleterious effects on the organism. Of course, determining the relative strengths of these counteracting forces would require modelling and careful experimentation.

dormant in the genome until the organism encounters stress, at which point there is a burst of TE insertion into new chromosomal locations. When this burst occurs in the germ line it elevates the mutation rate within gametes and ultimately increases the phenotypic diversity of viable offspring. If stressful conditions are correlated with changes in the selective environment, we can imagine that TE bursts will occasionally benefit organismal fitness (Lu et al. 2017). This assumes that an organism with more diverse offspring will have a better chance of adapting to the novel environment than one with less offspring diversity.

Some genomics researchers regard such "switch like" behaviour as strong evidence for precisely this type of CR-function (Shapiro 2000; von Sternberg 2002; Shapiro and von Sternberg 2005). This inference appears to be informed by familiar examples of conditionally activated genes. For instance, the genes that activate the neck spine in *Daphnia pulex* are switched on when this organism senses the presence of a predator. Such familiar examples set the expectation that any conditionally activated genetic response is likely to function as a switch that has beneficial effects for the organism. Once again, we can imagine how researchers who embrace this expectation might persist in the search for some biologically interesting phenotypic effect associated with stress-induced TE activation. By now, you won't be surprised to hear that evolutionary considerations call this assumption into question. As we have discussed, it is generally in the interest of the cell to suppress TE activity. Such cellular "immunity" mechanisms as RNA interference or TE methylation are themselves corrigible systems. Under stressful conditions methylation marks can become dislodged. Cell nuclei sometimes break down. It is no surprise that TE replication would become elevated under these conditions. What appears from the organism-level to be a finely tuned switch might simply be the cell's loss of control over TE suppression.

In this section I have considered two of the functional proxies employed by ENCODE for (I am assuming) the identification of DNA that is likely to have interesting CR-functions. Reasoning about selective history reveals that these are poor choices. Any TE insertion will exhibit these properties at some point in its lifecycle and we know that most of the DNA in our genome was generated by TE activity. Conditional activation under stress is another property commonly interpreted as indicative if CR-functionality. Again, considerations about TE-host evolutionary dynamics suggest a more qualified interpretation. I have further suggested that the myopic pursuit of CR-functions for these types of element is understandable in the absence of selectionist reasoning, thus explaining (if not excusing) ENCODE's follies.

Griffiths' paradox

Paul Griffiths (2009) has developed an account of "forward looking" function that is very similar to the line I have been defending. His view, like mine, is motivated in part by Millikan's worry about how to individuate the containing systems to which CR functions are legitimately ascribed. However, Griffiths argues that selective history cannot in principle resolve this issue. He presents his argument as a logical paradox:

P1) If considerations about selective history are necessary for investigating the CR functions of components in contemporary organisms, then it must be possible to accurately identify the selective functions of those components in their ancestors.

P2) However, "ascriptions of selected function are generated by causal analysis of the capacities of ancestral organisms to survive and reproduce in ancestral environments" (Griffiths 2009, p18). In other words, to determine whether some component has a SE function, one must first establish that it has a certain CR function.

P3) But according to P1, establishing that some component has a CR function requires accurately identifying the selective functions of those components in some (further back) ancestor. This gives rise to an infinite regress.

C1) "Therefore, a purely causal [non-historical] analysis of how the parts of ancestors were adaptive must be possible without knowing what those parts were adaptations for" (ibid.).

C2) "If this is possible for ancestors, it is possible for living organisms."

I suspect that Griffiths would agree with the suggestion that considerations about TE/host coevolutionary dynamics are instructive, if not essential for determining whether a given genetic element is likely to be a part or a parasite. Where Griffiths draws the line is in regarding these as hypotheses about selective *history*. Instead, he wants to think of them as "forward looking." I am unclear as to why such functional analysis should have any temporal direction, forward or back. Perhaps the kind of selectionist reasoning that Griffiths condones is more accurately described as a-historical. He mentions optimality analysis and evolutionary game theory as examples of "forward looking" reasoning. Such models provide insight into what Griffiths calls the "survival value" of a trait. As he explains,

> Tinbergen's concept of 'survival value' opens up the possibly of a genuine evolutionary perspective that is not an historical perspective, and thus not prone to the vicious regress identified [earlier]. Rather than focusing on causal capacities that featured in past episodes of selection, we should focus on causal capacities that contribute to survival and reproduction (survival value)…the vicious regress is avoided by adopting this forward-looking evolutionary perspective on the organism (p 14).

I am sympathetic with the attempt to identify a kind of functional analysis that focuses on survival value, but which makes no historical commitments. Elsewhere, I defend a methodological distinction between "purely ecological" and "purely evolutionary" modes of idealization ((Linquist 2016, 2019). The former abstracts away from the history of an entity and regards it as a static type that stands in certain functional relations to the environment. A purely evolutionary approach ignores specific functional relationships to the environment and considers how properties internal to the entity explain its changes over time. The kind of analysis that Griffiths is calling "forward looking" sounds exactly like what I have described as a purely ecological form of functional analysis. It is interesting to also compare Wimsatt's (2013) observation that

> causal role-functional analyses are common in areas like functional morphology, where one is dealing with idealized archetypes rather than intraspecific variation, and the features under discussion *have no variation in the relevant respects*. (original italics, 23)

.The common thread between these three views, as I see it, is that there exists a certain mode of "ecological" functional reasoning that necessarily takes place in the abstract, either (according to Wimsatt) because we lack epistemic access to actual selective history, or (following Griffiths) in order to avoid a logical regress, or (as I argue) because it an efficient idealization strategy when ecological causes outweigh historical causes.

However, I do not think that this accurately describes the kind of evolutionary reasoning that is essential for CR functional investigation in genomics. In the examples described in the previous section, it is necessary to regard the relevant selection pressures to have occurred in the past. It is by entertaining likely hypotheses about how certain elements came to be in the genome in the first place (an essentially historical hypothesis) that we are able to generate informed hypotheses about the likely CR effects of those entities.

How then to deal with Griffiths' paradox? An obvious place to put pressure is on the first premise (P1), which suggests that an *accurate understanding* (i.e. knowledge) of historical selection pressures is necessary for formulating reasonable CR functional hypotheses. I see no reason to accept such a strong claim. As we saw in the previous section, some fairly basic evolutionary reasoning reveals that certain (TE-derived) sequences are expected to acquire such properties as methylation, proximity to genes, and conditional activation even without having interesting organismal effects. This is enough to exclude such properties as reliable indicators of organismal-level CR functions. Griffiths anticipates an objection along these lines.

> A common response when I have presented the paradox… is that while biologists cannot *know* the selected function of a part before they describe it, they can *hypothesize* a selected function and this hypothesis helps them describe its form and (causal) function. If we examine this suggestion in more detail, however, it morphs into the alternative, forward-looking heuristic that I have advocated in this paper. Suppose that a biologist examines a stretch of genome or a body part of a little-studied organism. They can draw on no prior understanding of the role this part plays in the life of the organism, or even whether it really is part of the organism, as opposed to a parasite… Can it seriously be suggested that the first thing they should do is to hypothesise that the part evolved because of a particular set of selection pressures? What reason could there be to choose one selection pressure rather than another, given that nothing is known about the form and function of the part?

Griffiths is setting up a scenario in which the proponent of selectionist reasoning has no information about the entity in question. Under such a state of ignorance, it would indeed seem arbitrary to hypothesize a particular selection pressure. Of course, this is not the situation in which genomics researchers find themselves when conducting

component-driven research. Any such investigation must begin at least with some minimal annotation of the DNA sequence under investigation. The question is where to go from there. At least in the case of genomics, it is indeed a "serious" suggestion that the first thing to ask is whether a given sequence derives from TE activity. This automatically invokes a plausible picture of its coevolutionary history with the host genome. Knowing this information enables us to adjust our prior probabilities with respect to the element. For instance, if the element does in fact descend from a TE then we should not take its proximity to a gene, the fact that it is methylated, or its conditional activation as evidence of some interesting CR function. More important perhaps, knowing the history of an element can help us avoid the kind of myopic reasoning that can arise in genomics. I see no vicious regress here.

Conclusions

I have argued that in the discipline of genomics, component-driven functional investigation runs the risk of causal-role myopia. The tendency to posit one organism-level capacity after another as the putative CR function of some genetic element can proceed indefinitely because (1) the genome is littered with TEs and their partially deactivated descendants which (2) masquerade as components with interesting CR-functions and (3) it is experimentally onerous to determine whether a given element lacks any such function. The fact that ENCODE appears to have fallen victim to this kind of reasoning suggests that CR myopia is not a hypothetical concern.

Granted, reasoning about selective history cannot decisively rule out whether a given element lacks some biologically interesting CR-function. However, this type of reasoning allows a researcher to assign likelihoods to specific CR-functional hypotheses. For instance, if one knows that a given genetic elements descends from a TE, then, even without some further organismal function, the default expectations are that it will probably be methylated, it most likely possesses binding sites, and it could become conditionally activated under stress. Hence, none of these properties should be taken on their own to suggest that the element is likely performing some interesting CR function for the organism. Selectionist reasoning is also a valuable source of hypotheses about potential CR functions. For instance, Doolittle (2022) proposes that TE activity could contribute to the evolvability of an entire species. Whether this effect is likely to become the product of natural selection, and thus become an SE function is a more complicated question. It seems foolish for genomics researchers to discount Darwinian reasoning, given the promise of streamlining their search for CR functions. The precise reasons why so many genomics researchers are inclined to interpret the entire genome in organism-functional terms (Lynch 2007) is a topic for future research.

The popularity of functional pluralism in philosophical circles is more easily diagnosed. I take the core assumptions of pluralism to be that selectionist reasoning is associated with the investigation of SE origin functions only, and that CR functions are best investigated without the burden of confirming selectionist hypotheses about the distant past. I have argued that this apprehension stems partly from a failure to distinguish informed selectionist reasoning from the more onerous confirmation of

an adaptationist hypothesis. Some philosophers appear to have developed an allergy to adaptationist reasoning in most biological contexts (Lloyd 2015). It is therefore important to note that selectionist reasoning about TE/host coevolutionary dynamics leads to the conclusion (perhaps refreshing, for some) that most of our DNA is probably not CR functional, and hence not SE functional – at least, not at the level of the organism.

Acknowledgements An early version of this paper was presented at a 2018 workshop in Halifax on Evolutionary Roles of Transposable Elements and Non-coding DNA: the Science and Philosophy. A later version was presented to Ford Doolittle's postdoctoral research group. Thanks to participants at both events for their helpful comments and criticisms. This work was funded by the Social Science and Humanities Research Council of Canada.

References

Ahituv N, Zhu Y, Visel A et al (2007) Deletion of Ultraconserved Elements Yields Viable Mice. PLOS Biol 5:e234. https://doi.org/10.1371/journal.pbio.0050234

Amundson R, Lauder GV (1994) Function without purpose. Biol Philos 9:443–469. https://doi.org/10.1007/BF00850375

Bannert N, Kurth R (2004) Retroelements and the human genome: New perspectives on an old relation. Proc Natl Acad Sci 101:14572–14579. https://doi.org/10.1073/pnas.0404838101

Biscotti MA, Canapa A, Forconi M et al (2015) Transcription of tandemly repetitive DNA: functional roles. Chromosome Res 23:463–477. https://doi.org/10.1007/s10577-015-9494-4

Brandon RN (2013) A General Case for Functional Pluralism. In: Huneman P (ed) Functions: selection and mechanisms. Springer Netherlands, Dordrecht, pp 97–104

Brunet TDP, Doolittle WF (2014) Getting "function" right. Proc Natl Acad Sci 111:E3365–E3365. https://doi.org/10.1073/pnas.1409762111

Brunet TDP, Doolittle WF (2015) Multilevel Selection Theory and the Evolutionary Functions of Transposable Elements. Genome Biol Evol 7:2445–2457. https://doi.org/10.1093/gbe/evv152

Comfort NC (1999) "The Real Point is Control": The Reception of Barbara McClintock's Controlling Elements. J Hist Biol 32:133-162. https://doi.org/10.1023/A:1004468625863

Crabbe JC, Wahlsten D, Dudek BC (1999) Genetics of Mouse Behavior: Interactions with Laboratory Environment. Science 284:1670–1672. https://doi.org/10.1126/science.284.5420.1670

Craver CF (2001) Role Functions, Mechanisms, and Hierarchy. Philos Sci 68:53–74. https://doi.org/10.1086/392866

Craver CF (2013) Functions and Mechanisms: A Perspectivalist View. In: Huneman P (ed) Functions: Selection and Mechanisms. Springer, pp 133–158

Cummins R (1976) Functional Analysis. J Philos 72:741–765. https://doi.org/10.2307/2024640

Doolittle WF (2013) Is junk DNA bunk? A critique of ENCODE. Proc Natl Acad Sci 110:5294–5300. https://doi.org/10.1073/pnas.1221376110

Doolittle WF, Brunet TDP, Linquist S, Gregory TR (2014) Distinguishing between "Function" and "Effect" in Genome Biology. Genome Biol Evol 6:1234–1237. https://doi.org/10.1093/gbe/evu098

Doolittle WF (2022) All about levels: transposable elements as selfish DNAs and drivers of evolution. Biology and Philosophy (forthcoming)

Eddy SR (2012) The C-value paradox, junk DNA and ENCODE. Curr Biol 22:R898–R899. https://doi.org/10.1016/j.cub.2012.10.002

Elliott TA, Linquist S, Gregory TR (2014) Conceptual and Empirical Challenges of Ascribing Functions to Transposable Elements. Am Nat 184:14–24. https://doi.org/10.1086/676588

Fedoroff NV (2012) McClintock's challenge in the 21st century. Proc Natl Acad Sci 109:20200–20203. https://doi.org/10.1073/pnas.1215482109

Francis DD, Szegda K, Campbell G et al (2003) Epigenetic sources of behavioral differences in mice. Nat Neurosci 6:445–446. https://doi.org/10.1038/nn1038

Garson J (2016) A Critical Overview of Biological Functions. Springer International Publishing, Cham

Garson J (2017) A Generalized Selected Effects Theory of Function. Philos Sci 84:523–543. https://doi.org/10.1086/692146
Garson J (2011) Selected effects and causal role functions in the brain: the case for an etiological approach to neuroscience. Biol Philos 26:547–565. https://doi.org/10.1007/s10539-011-9262-6
Germain P-L, Ratti E, Boem F (2014) Junk or functional DNA? ENCODE and the function controversy. Biol Philos 29:807–831. https://doi.org/10.1007/s10539-014-9441-3
Godfrey-Smith P (2001) Three Kinds of Adaptationism. In: Orzack SH, Sober E (eds) Adaptationism and Optimality, 1st edn. Cambridge University Press, pp 335–357
Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B Biol Sci 205:581–598. https://doi.org/10.1098/rspb.1979.0086
Graur D, Zheng Y, Azevedo RBR (2015) An Evolutionary Classification of Genomic Function. Genome Biol Evol 7:642–645. https://doi.org/10.1093/gbe/evv021
Graur D, Zheng Y, Price N et al (2013) On the Immortality of Television Sets: "Function" in the Human Genome According to the Evolution-Free Gospel of ENCODE. Genome Biol Evol 5:578–590. https://doi.org/10.1093/gbe/evt028
Griffiths PE (2009) In What Sense Does "Nothing Make Sense Except in the Light of Evolution"? Acta Biotheor 57:11–32. http://dx.doi.org.subzero.lib.uoguelph.ca/https://doi.org/10.1007/s10441-008-9054-9
Griffiths PE (2006) Function, Homology, and Character Individuation. Philos Sci 73:1–25. https://doi.org/10.1086/510172
Hatje K, Mühlhausen S, Simm D, Kollmar M (2019) The Protein-Coding Human Genome: Annotating High-Hanging Fruits. BioEssays News Rev Mol Cell Dev Biol 41:e1900066. https://doi.org/10.1002/bies.201900066
Hüttenhofer A, Schattner P, Polacek N (2005) Non-coding RNAs: hope or hype? Trends Genet 21:289–297. https://doi.org/10.1016/j.tig.2005.03.007
Johnson JM, Edwards S, Shoemaker D, Schadt EE (2005) Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet 21:93–102. https://doi.org/10.1016/j.tig.2004.12.009
Kellis M, Wold B, Snyder MP et al (2014) Reply to Brunet and Doolittle: Both selected effect and causal role elements can influence human biology and disease. Proc Natl Acad Sci 111:E3366–E3366. https://doi.org/10.1073/pnas.1410434111
Kidwell MG, Lisch DR (2001) Perspective: Transposable Elements, Parasitic Dna, and Genome Evolution. Evolution 55:1–24. https://doi.org/10.1111/j.0014-3820.2001.tb01268.x
Kidwell MG, Lisch DR (1998) Transposons unbound. Nature 393:22–23. https://doi.org/10.1038/29889
Kim Y-J, Lee J, Han K (2012) Transposable Elements: No More "Junk DNA." Genomics Inform 10:226–233. https://doi.org/10.5808/GI.2012.10.4.226
Larsen PA (2018) Transposable elements and the multidimensional genome. Chromosome Res 26:1–3. https://doi.org/10.1007/s10577-018-9575-2
Linquist S (2016) Against Lawton's Contingency Thesis; or, Why the Reported Demise of Community Ecology Is Greatly Exaggerated. Philos Sci. https://doi.org/10.1086/684024
Linquist S (2019) Why Ecology and Evolution Occupy Distinct Epistemic Niches. Philos Top 47:143–165. https://doi.org/10.5840/philtopics20194718
Linquist S, Cottenie K, Elliott TA et al (2015) Applying ecological models to communities of genetic elements: the case of neutral theory. Mol Ecol 24:3232–3242. https://doi.org/10.1111/mec.13219
Linquist S, Doolittle WF, Palazzo AF (2020) Getting clear about the F-word in genomics. PLoS Genet 16:e1008702. https://doi.org/10.1371/journal.pgen.1008702
Linquist S, Fullerton B (2021) Transposon dynamics and the epigenetic switch hypothesis. Theor Med Bioeth 42:137–154. https://doi.org/10.1007/s11017-021-09548-x
Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. Curr Opin Plant Biol 14:156–161. https://doi.org/10.1016/j.pbi.2011.01.003
Lloyd EA (2015) Adaptationism and the Logic of Research Questions: How to Think Clearly About Evolutionary Causes. Biol Theory 10:343–362. https://doi.org/10.1007/s13752-015-0214-2
Lu L, Chen J, Robb SMC et al (2017) Tracking the genome-wide outcomes of a transposable element burst over decades of amplification. Proc Natl Acad Sci 114:E10550–E10559. https://doi.org/10.1073/pnas.1716459114
Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci 104:8597–8604. https://doi.org/10.1073/pnas.0702207104

Mattick JS (2001) Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep 2:986–991. https://doi.org/10.1093/embo-reports/kve230
Mattick JS, Dinger ME (2013) The extent of functionality in the human genome. HUGO J 7:2. https://doi.org/10.1186/1877-6566-7-2
Mattick JS, Taft RJ, Faulkner GJ (2010) A global view of genomic information – moving beyond the gene and the master regulator. Trends Genet 26:21–28. https://doi.org/10.1016/j.tig.2009.11.002
McLean C, Bejerano G (2008) Dispensability of mammalian DNA. Genome Res 18:1743–1751. https://doi.org/10.1101/gr.080184.108
Millikan R (2002) Biofunctions: Two Paradigms. In: Ariew A, Cummins R, Perlman M (eds) Functions: New Essays in the Philosophy of Psychology and Biology. Clarendon Press
Niu D-K, Jiang L (2013) Can ENCODE tell us how much junk DNA we carry in our genome? Biochem Biophys Res Commun 430:1340–1343. https://doi.org/10.1016/j.bbrc.2012.12.074
Nowak R (1994) Mining Treasures from "Junk DNA. Science. https://doi.org/10.1126/science.7508142
Palazzo AF, Lee ES (2015) Non-coding RNA: what is functional and what is junk? Front Genet 0. https://doi.org/10.3389/fgene.2015.00002
Parenteau J, Abou Elela S (2019) Introns: Good Day Junk Is Bad Day Treasure. Trends Genet 35:923–934. https://doi.org/10.1016/j.tig.2019.09.010
Ponicsan SL, Kugel JF, Goodrich JA (2010) Genomic gems: SINE RNAs regulate mRNA production. Curr Opin Genet Dev 20:149–155. https://doi.org/10.1016/j.gde.2010.01.004
Ponting CP, Hardison RC (2011) What fraction of the human genome is functional? Genome Res 21:1769–1776. https://doi.org/10.1101/gr.116814.110
Roessler K, Bousios A, Meca E, Gaut BS (2018) Modeling Interactions between Transposable Elements and the Plant Epigenetic Response: A Surprising Reliance on Element Retention. Genome Biol Evol 10:803–815. https://doi.org/10.1093/gbe/evy043
Rosenberg A (1994) Subversive Reflections on the Human Genome Project. PSA Proc Bienn Meet Philos Sci Assoc 1994:329–335. https://doi.org/10.1086/psaprocbienmeetp.1994.2.192943
Shapiro JA (2000) Transposable elements as the key to a 21st century view of evolution. In: McDonald JF (ed) Transposable Elements and Genome Evolution. Springer Netherlands, Dordrecht, pp 171–179
Shapiro JA, von Sternberg R (2005) Why repetitive DNA is essential to genome function. Biol Rev 80:227–250. https://doi.org/10.1017/S1464793104006657
Tauber AI, Sarkar S (1992) The Human Genome Project: Has Blind Reductionism Gone too Far? Perspect Biol Med 35:220–235. https://doi.org/10.1353/pbm.1992.0015
von Sternberg R (2002) On the roles of repetitive DNA elements in the context of a unified genomic-epigenetic system. Ann N Y Acad Sci 981:154–188. https://doi.org/10.1111/j.1749-6632.2002.tb04917.x
Weber M (2004) Philosophy of Experimental Biology. Cambridge University Press
Wessler SR (1996) Plant retrotransposons: Turned on by stress. Curr Biol 6:959–961. https://doi.org/10.1016/S0960-9822(02)00638-3
Willingham AT, Gingeras TR (2006) TUF Love for "Junk". DNA Cell 125:1215–1220. https://doi.org/10.1016/j.cell.2006.06.009
Wimsatt WC (2013) Evolution and the Stability of Functional Architectures. In: Huneman P (ed) Functions: Selection and Mechanisms. Springer, pp 19–41
Wouters A (2005) The Function Debate in Philosophy. Acta Biotheor 53:123–151. https://doi.org/10.1007/s10441-005-5353-6