Published in: Kaiser M. I. (2018): ENCODE and the Parts of the Human Genome. *Studies in History and Philosophy of Biological and Biomedical Sciences* 72: 28-37. DOI: 10.1016/j.shpsc.2018.10.008

ENCODE and the Parts of the Human Genome

Marie I. Kaiser

Bielefeld University, kaiser.m@uni-bielefeld.de

Abstract

This paper examines a specific kind of part-whole relations that exist in the molecular genetic domain. The central question is under which conditions a particular molecule, such as a DNA sequence, is a biological part of the human genome. I address this question by analyzing how biologists in fact partition the human genome into parts. This paper thus presents a case study in the metaphysics of biological practice. I develop a metaphysical account of genomic parthood by analyzing the investigative and reasoning practices in the ENCODE (ENCyclopedia Of DNA Elements) project. My account reveals two conditions that determine whether a molecule is a part of the human genome (i.e., a genomic part). First, genomic parts must possess a causal role function in the genome as a whole, that is, their functions must contribute to the genome directing the overall functioning of the cell. Second, genomic parts must have a specific chemical structure and be actual segments of the DNA sequence of the genome.

Keywords

part-whole relation, ENCODE, genome, causal role function, structure

1 Introduction

This paper is concerned with a specific kind of part-whole relations in the biological domain – with the relations that exist between specific DNA sequences and the human genome of which they are parts. My central goal is to analyze genomic part-whole relations by explicating the conditions under which some biological objects, such as specific DNA sequences, are parts of the human genome (i.e., genomic parts), whereas others are not. The relations between parts and wholes and the conditions under which they exist have fascinated philosophers for a long time. Van Inwagen calls the question 'Under which conditions is *X* a part of *Y*?' the "special composition question" (1990, 21). In this paper, I address a special version of van Inwagen's question. I am not concerned with parthood in general, which is the object of mereology (e.g., Simons 1987), but only with part-whole relations in the biological domain. Even more specifically, I focus on those part-whole relations that are studied in molecular genetics, namely the relation between specific DNA sequences and the human genome as a whole.

My philosophical analysis is guided by the assumption that to be *a biological part* of the human genome that is relevant to biological processes and phenomena is different from merely being *part* of the human genome, for instance, in the sense of being contained within or being a mereological part of the human genome (Jansen and Schulz 2014, Kaiser forthc.). Biological parts of the human genome will differ from mereological parts in at least two

ways. First, only a subset of mereological parts will also be biological parts because the latter must satisfy additional criteria. To specify these criteria for genomic parts is the goal of this paper. Second, some biological parts of the human genome will not be mereological parts because they conflict with central principles of extensional mereology, for example, if genomic parts overlap (e.g., overlapping genes). For reasons of simplicity, in what follows, I will speak about the parts of the human genome or about genomic parts without always indicating that I focus on *biological* parts, rather than on mereological parts or physical parts.

The central question of this paper is a metaphysical question. It asks what in the world makes something to a part of the human genome. I approach this metaphysical question in a naturalistic way, that is, by drawing metaphysical conclusions from the knowledge that biologists gain about genomic part-whole relations. More precisely, I analyze the metaphysical assumptions that underlie the reasoning strategies of biologists and their actual practices of decomposing the human genome into parts. The account of genomic parthood that I develop in this paper is thus an instance of what I call 'metaphysics of biological practice' (more on this in Section 3). The aim of my analysis is to explicate the conditions under which genomic part-whole relations exist by articulating two criteria that must be satisfied for something to be a part of the human genome – a causal-functional criterion and a chemical-structural criterion. My metaphysical analysis focuses on the reasoning and investigative practices of the ENCODE (ENCyclopedia Of DNA Elements) project because this project aims at providing a complete list of the parts of the human genome (ENCODE 2004, 636).

I proceed as follows. After introducing the ENCODE project (Section 2), I specify the goals and methodology of a metaphysics of biological practice and I point out how it deviates from other approaches (Section 3). Section 4 gives an overview about the methods that ENCODE researchers use to identify the functional parts of the human genome. I explain why these different methods are exemplifications of what I call the General Strategy (GS). In Section 5, I argue that GS suggests a biochemical activity criterion for genomic parthood, and I show why this criterion is implausible. Section 6 points out how biochemical activities differ from causal role functions and how the latter can be used to develop a convincing first criterion for genomic parthood. In Section 7, I propose a second criterion, which accounts for the fact that also chemical-structural features determine genomic part-whole relations.

2 The ENCODE Project

At the beginning of the 21th century, in April 2003, scientists declared that they had completed the Human Genome Project whose goal was to determine the exact nucleotide sequence of the human genome and to identify all of its genes.[1] It became quickly clear that knowing the "blueprint of human biology" is not enough. We also need an "instruction

[1] This date was due to the Golden Anniversary of Watson and Crick's discovery of the structure of DNA (1953). Strictly speaking, we still do not fully know the exact nucleotide sequence of the human genome. Until the end of 2017, scientists had discovered 95-97% of it.

manual for reading the blueprint" (Maher 2012, 46). To understand the functioning of the genome and to use this knowledge to understand and cure human diseases we need to know how to interpret the sequence of the human genome. The ENCyclopedia Of DNA Elements (ENCODE) Project, whose pilot project started in 2003, promises to provide such an interpretation. Its goal is to annotate the human genome[2] by partitioning it into its functional elements or parts. ENCODE seeks to provide a "complete catalog, or 'parts list'" (2004, 636) identifying not only protein-coding genes but also and in particular non-coding genes, transcriptional regulatory elements, and sequences that mediate chromosome structure and dynamics. The underlying idea is that "you've got to put all the parts down on the table before putting it together" (Ewan Birney, quoted in Maher 2012, 48). This means that you first need to individuate those DNA sequences that play a functional role – for instance in transcription or gene regulation – before you can then integrate this knowledge into a general picture of how the genome affects human traits and diseases.

Among the most important findings of the ENCODE project was that functional DNA elements or parts make up the vast majority (80,4%) of the human genome because each of the identified functional elements "participates in at least one biochemical RNA- and/or chromatin-associated event in at least one cell type" (ENCODE 2012, 57). This is a surprising result given that protein-coding genes constitute only slightly more than 1% of the human genome. It seems as if biologists are increasingly forced to pay attention to non-gene DNA sequences in the human genome (Greally 2007, 782). ENCODE's claim that they were able to "assign biochemical functions for 80% of the genome" (2012, 57) has provoked much criticism (e.g., Doolittle 2013, Graur et al. 2013). It appears to be at odds with the fact that only 3-8% of the genome has been undergoing purifying selection (given the controversial assumption that natural selection determines functions) and it seems to question the common view that most of the human genome is junk DNA, which can be dismissed as irrelevant (Kolata 2012). Critics accuse ENCODE of making use of the wrong concept of function – namely a causal role concept of function rather than a selected-effect concept (Graur et al. 2013) –, which results in a multiplication of spurious functions and thus overestimates functionality. From a philosophical perspective, it is very interesting to discuss whether ENCODE in fact employs the wrong concept of function (e.g., Elliott et al. 2014; Doolittle et al. 2014) or whether ENCODE's understanding of biological function is rather appropriate to that specific type of research (i.e., Big Data applied to the biomedical context; Germain et al. 2014).

In this paper, I will only touch on this discussion because I focus on another aspect of the ENCODE project that deserves philosophical scrutiny. It is ENCODE's self-declared goal to provide an "encyclopedia of DNA elements" (2012, 57) or a "parts list" (2004, 636) that identifies all and only those DNA sequences that are functional parts of the human genome. Providing such a parts list seem to be a precondition for understanding how the different

[2] The ENCODE project started with annotating the human genome but it extended its research to the genomes of other model organisms. My account of genomic parthood applies also to non-human genomes.

genomic parts work together and affect human biology and disease. From a philosophical standpoint, the ENCODE project constitutes a telling case study because it is among the very few cases, in which biologists are explicitly concerned with identifying part-whole relations.[3] ENCODE researchers are, however, not concerned with individuating the mereological parts (or the physical parts) of the human genome. It is an unchallenged fact that nucleotides (or atoms) are building blocks of the genome und, as such, parts of it. But there is a difference between an object being *part of* the human genome (i.e., being a mereological part) and an object being *a* part of the human genome, and ENCODE is concerned only with the latter. The goal of the ENCODE project is not to decompose the human genome into mereological parts, such as nucleotides, but to individuate such parts that are relevant to other biological processes and phenomena and thus can be generally called *biological* parts (Kaiser forthc.). In this paper, I specify the grounds on which ENCODE treats parts as relevant to other biological processes and phenomena. It will turn out that ENCODE identifies an object as a part of the human genome if it possesses a function of a special kind. The biological parts that ENCODE individuates are thus functional parts. They clearly differ from (recall Section 1) and should not be lumped together with mereological parts, which are not the objects of study of ENCODE.

3 Metaphysics of Biological Practice

The term 'metaphysics of biological practice' does not denote a novel overarching metaphysical system that, for instance, ascribes particular importance to biological entities or practices. Instead, metaphysics of biological practice is a novel approach that brings together philosophical interest in metaphysical issues and a philosophy of biology that pays special attention to how biological research is actually practiced. The philosophical accounts that can be developed in the framework of metaphysics of biological practice are diverse but they share some basic features. First, they are metaphysical accounts, that is, they make claims about how the world is like and which kinds of entities exist. Second, they are naturalistic accounts, that is, they regard scientific knowledge as central to developing and justifying metaphysical claims. Third, they are practice-oriented accounts, that is, they do not focus on our best scientific theories but draw metaphysical conclusions from the actual practice of the biological sciences. I explicate these features in turn.

First, accounts in the metaphysics of biological practice are metaphysical in character. They make claims about which kinds of entities exist in the world, what is their nature, and how are they related to each other. According to a standard view, metaphysics explicates the general structure of reality, rather than explicating features of our knowledge or representations of this reality, which is the goal of epistemology (van Inwagen and Sullivan

[3] If biologists decompose systems into parts or provide reductive explanations of biological phenomena they are also concerned with part-whole relations but more implicitly (Kaiser 2015).

2017).[4] My account of genomic parthood, for instance, makes claims about a certain kind of part-whole relations that exist out there in the living world. It provides an understanding of the general characteristics of genomic part-whole relations and of the conditions under which they hold. Its subject matter is a certain aspect of the world rather than our knowledge about or representations of this aspect of the world.

Second, metaphysicians of biological practice agree with naturalistic or scientific metaphysicians that we should let science be our guide to metaphysics (e.g., Ladyman and Ross 2007; Callender 2011; Ross et al. 2013; Chakravartty 2013). As Maudlin puts it, "when choosing the fundamental posits of one's ontology, one must look to scientific practice rather than to philosophical prejudice" (2007, 1). Metaphysics of biological practice stands in the tradition of naturalistic metaphysics in so far as it looks into the biological sciences to develop metaphysical claims about the living world. For example, I assume that a metaphysical account of genomic parthood must be developed and justified on the basis of biological knowledge about the human genome and its parts.

Third, in two other respects metaphysics of biological practice clearly deviates from traditional naturalistic metaphysics. On the one hand, metaphysics of biological practice abandons the fundamentalist aspirations that many metaphysicians of science share. Its aim is not to identify the fundamental building blocks of the universe, and it rejects the assumption that only the physical sciences deliver valuable metaphysical insights. Metaphysicians of biological practice emphasize that there exist urgent metaphysical questions that require analyzing biological rather than physical knowledge. On the other hand, metaphysics of biological practice has joined philosophy of biology in its "practice turn" (Soler et al. 2014) insofar as it shares the methodological conviction that a philosophical account should emerge from a detailed analysis of scientific practice.[5] Metaphysics of biological practice abandons the exclusive focus on our best scientific theories, it takes into account the entire variety of biological practice, and it draws philosophical attention to the roles that metaphysical assumptions play in achieving certain aims in biological practice. In developing my account of genomic parthood, for instance, I analyze the methods that ENCODE researchers use to individuate genomic parts, how they draw conclusions from empirical data, and how they spell out the aims and importance of their research. My goal is to specify the conditions under which genomic part-whole relations exist by way of analyzing ENCODE's reasoning strategies and composition practices.

One might object that the naturalistic character of my account makes it dependent on a substantial version of scientific realism that cannot be sustained. The argument goes as

[4] An alternative notion of metaphysics is the neo-Kantian view, according to which metaphysics makes claims about the general features of our conceptual structure, rather than about a mind-independent reality (Strawson 1959, 9).

[5] I understand the term 'scientific practice' in a broad sense. It encompasses not only material aspects of science, such as experimental apparatus and instruments, but also scientific activities and investigative strategies, such as modeling or manipulating (e.g., Chang 2011, Waters 2014), and more long-term results of science, such as theories and explanations.

follows: Drawing metaphysical claims about the world from scientists' reasoning and investigative practices presupposes that the claims made in these practices are true and that the theoretical terms that they include refer to entities that exist in the world independently of scientific investigation. Scientific practices that consist of false claims and non-referring theoretical terms do not provide us with epistemic access to the world and thus lead to metaphysical claims that might be false. Decade-long disputes about scientific realism, however, have shown that a convincing defense of scientific realism is not available. Hence, metaphysics of biological practice is unfeasible because it presupposes an unwarranted realism concerning the biological sciences. My response to this objection is twofold.

On the one hand, metaphysics of biological practice is based only on a selective scientific realism. Not all elements of biological practice are suitable for drawing metaphysical inferences from them. Metaphysical claims are claims about how the world is like (e.g., which kinds of entities exist in the world). If we want these claims to be true we should draw them from such elements of biological practice that consist of (or imply) factual statements about the world that are most likely to be true. The traditional strategy of naturalistic metaphysics to ensure epistemic access to the world is to draw metaphysical claims from our best scientific theories only. Shifting philosophical attention from theories to practice, however, requires a different strategy, such as to focus on those elements of biological practice that have proven to be successful because they have contributed to achieving central goals of science.[6]

On the other hand, metaphysicians of biological practice do not need to ultimately defend selective scientific realism before they can start their work. The metaphysical claims that are drawn from biological practice can be interpreted as having a *provisional* character. Provisional metaphysical claims describe general features of the world, given that the biological practices they are drawn from can be interpreted realistically. If we conceive of metaphysics of biological practice as making provisional claims we can stick to the view that metaphysics makes claims about reality, while allowing for the possibility that the analyzed biological practices contain false statements and terms that do not refer and thus may give rise to false metaphysical claims. For example, it might turn out in the future that the critics of ENCODE are right in that ENCODE identifies too many functions which in fact do not exist. In that case some of my metaphysical claims would turn out as false and I would have to revise them. The provisional character of metaphysical claims, however, does not render metaphysics of biological practice uninteresting. It merely avoids unwarranted foot stamping.

4 ENCODE's General Strategy

ENCODE researchers use many different methods and techniques to identify the parts of the human genome (for an overview see, e.g., ENDODE 2011, 5, Table 1). RNA-seq, for instance,

[6] Of course, this strategy raises the question of how to spell out the notion of scientific success in a way that ensures a realistic interpretation of successful practices.

is a method for mapping and quantifying the complete set of transcripts in a cell for a specific developmental stage or physiological condition. Another method is 'ChIP-seq', which is the abbreviation for 'Chromatin ImmunoPrecipitation followed by sequencing'. ChIP-seq uses antibodies to identify the location of specific proteins (e.g., transcription factors or histones) that are directly or indirectly bound to DNA. The identified DNA fragments are then sequenced. All these methods produce a huge bulk of data of different kinds, which is why the ENCODE project is assigned to the so-called big-data sciences (Birney 2012, 49). Moreover, in its "technology development phase" (ENCODE 2004, 636; 2011, 1), it has been a major goal of ENCODE to develop new and advance existing technologies for annotating the human genome.

Although the methods that ENCODE researchers apply are divers, from a general perspective, they are all exemplifications of the same strategy of individuating genomic parts. I refer to this strategy as the *General Strategy* (GS); others have called it the "biochemical signature strategy" (Stamatoyannopoulos 2012, 1602).[7] All methods used in the ENCODE project share a common feature: they pick out certain DNA sequences as parts of the human genome by virtue of the biochemical activities that these sequences engage in. The notion of an activity has been explicated in the philosophical debate about biological mechanisms (e.g., Machamer et al. 2000; Illari and Williamson 2013; Kaiser 2018). Activities are what concrete material objects do. Proteins, for example, bind to DNA sequences and organisms reproduce. Activities are said to be "active doings" (Illari and Williamson 2013, 71), they are temporally extended, and they are the things that produce changes (Machamer et al. 2000, 3). One activity can involve one or many objects, at least one of which must be actively involved in the activity (e.g., the protein binding to DNA; Kaiser 2018).

[7] In my analysis, I prefer the notion of an activity to the notion of a signature for two main reasons. First, most ENCODE researchers prefer to speak about biochemical activities. Second, 'signature' is an epistemic concept, which is ontologically ambiguous. Signatures are signs or marks; they provide evidence for or serve as a trace for something else. For example, the biochemical signature "not being formaldehyde cross-linked" (detected by the method FAIRE-seq) is a mark of open chromatin regions. From an ontological perspective, however, it is unclear to which ontological category signatures belong.

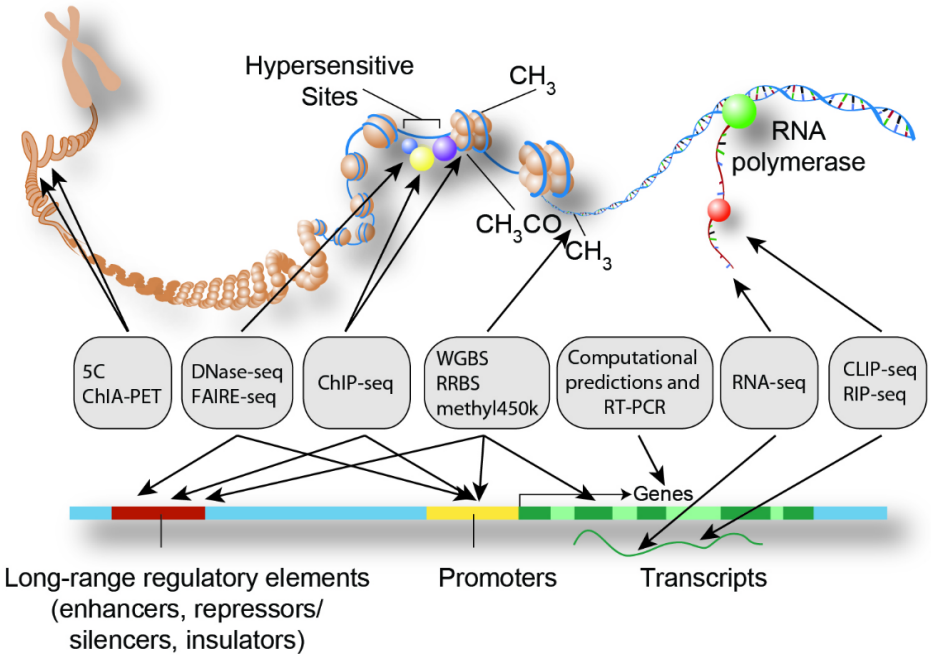

Fig. 1: Schematic representation of the major methods that ENCODE uses to detect genomic parts. Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI) (https://www.encodeproject.org/)

Figure 1 exemplifies ENCODE's General Strategy. The grey boxes list the different methods that ENCODE researchers apply. The arrows leading from the boxes to the upper representation of the chromatin indicate the biochemical activities that each method tracks. For example, DNase-seq is a method for identifying open chromatin regions, which are DNA regions that are nucleosome-depleted and thus loosely packed to allow access to RNA and enzymes, such as DNA polymerases or nucleases. The method DNase-seq utilizes DNase I enzymes to cut live chromatin preparations at sites where nearby there are specific non-histone proteins. That way, open chromatin regions which are hypersensitive to DNase I are identified and distinguished from DNA regions that are more resistant to DNase I because they are tightly wrapped around histone proteins forming nucleosomes and higher-order structures. In sum, the method DNase-seq picks out particular DNA sequences because they display the biochemical activity "being cut by DNase I". This is why one arrow in Figure 1 leads from the DNase-seq box up to the blue circle (representing DNase I), which is attached to the blue line (representing the DNA strand).

Another example is ChIP-seq, which is a method that uses antibodies to select certain kinds of proteins (e.g., transcription factors or chemically modified histone proteins) that are bound to DNA and form with it regions of cross-linked chromatin. Depending on the kind of protein at which the antibody is directed (e.g., at certain kinds of transcription factors or at histones modified in a certain way, e.g., lysine 4 of histone protein H3 being trimethylated (H3K4me3)) the targets of this method vary. ChIP-seq can be used to select DNA sequences that display, for instance, the biochemical activity "binding a specific kind of transcription factors" or the biochemical activity "associating with histone modification H3K4me3". Some

arrows in Figure 1 thus lead from the ChIP-seq box up to the yellow circle (representing a transcription factor) and to the acetyl group attached to the histone DNA complex.

It must be noted that the various methods that ENCODE utilizes do not work in isolation but are interrelated in several, interesting ways. First, some methods are used to identify the same kinds of parts or regions of parts. The results of applying one method may thus confirm or contest the results of another. For example, DNase-seq and FAIRE-seq are both methods for identifying open chromatin regions, in which different types of regulatory elements, such as promoters and enhancers, are located. These methods are independent because DNase-seq measures the biochemical activity "DNAse I cleaving", whereas FAIRE-seq detects "formaldehyde cross-linking". But since both methods identify DNA regions free from nucleosomes (i.e., open chromatin regions) their results complement or confirm each other.[8] Second, some methods individuate genomic parts more fine- or course-grained than others and thus are applied in succession. For example, ChIP-seq identifies quite fine-grained where in open chromatin regions specific proteins (directly or indirectly) bind to the DNA. Usually, ChIP-seq is thus employed after DNase I or FAIRE experiments identify open chromatin regions.

In conclusion, all different methods and techniques used in the ENCODE project share a General Strategy for identifying genomic parts: Biologists pick out particular DNA segments with a specific base sequence as genomic parts because these DNA segments engage in specific kinds of biochemical activities, such as "being cut by DNase I" or "associating with histone modification H3K4me3".

5 Why a Biochemical Activity Criterion Fails

This section presents a first attempt to formulate a criterion for genomic parthood, which specifies the conditions under which something is a biologically relevant part of the human genome, rather than being a mereological part or merely being spatially contained in the genome. This first criterion is suggested by ENCODE's General Strategy introduced in the previous section. The goal of this section is to first motivate this criterion and then to point out its shortcomings.

5.1 The Biochemical Activity Criterion

ENCODE's General Strategy is to select those DNA sequences as genomic parts that engage in specific kinds of biochemical activities that are detected by methods, such as RNA-seq and ChIP-seq. This General Strategy suggests that the biochemical activity that a particular DNA sequence displays is central to its status as a part of the human genome. In other words, the fact that ENCODE's methods identify genomic parts by virtue of their biochemical activities suggests that biochemical activities determine genomic parthood. Accordingly, the following criterion for genomic part-whole relations can be formulated:

[8] In contrast to DNase-seq, FAIRE-seq does not require the permeabilization of cells or the isolation of nuclei and thus can be used to analyze any cell type.

Biochemical Activity Criterion (BAC)
A DNA sequence *X* is a part of the human genome *G* if and only if
X engages in a certain kind of biochemical activity *A* that is tracked by one (or more) of ENCODE's methods.

Prima facie, BAC is a plausible criterion for genomic parthood. It is quite flexible because it allows for biochemical activities of different kinds to be relevant to genomic parthood. BAC corresponds well to the methods that ENCODE researchers use and to the empirical data they collect because the empirical data primarily record which DNA sequences display which biochemical activities. For example, RNA-seq picks out DNA sequences that produce RNAs and ChIP-seq can be used to gather empirical data about which DNA sequences bind to transcription factors.[9] Finally, BAC seems to be in line with how ENCODE researchers explicate the notion of a functional element: "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)." (ENCODE 2012, 57)

5.2 Divergent Epistemic Roles of Activities and Functions

Despite its initial plausibility, BAC is deficient for two major reasons. First, though BAC nicely accounts for the fact that ENCODE's methods and techniques are targeted at biochemical activities it emerges from a too narrow focus. Methods are only one element of ENCODE's research practice that philosophers should pay attention to. A convincing account of genomic parthood must capture also how ENCODE researchers interpret empirical data, discuss their empirical findings, specify their goals, and place the ENCODE project in a broader context. Second, if philosophers widen their perspective and take into account more than just the methods and techniques that ENCODE apply they will recognize that ascribing functions to DNA sequences is crucial to identifying them as genomic parts. To begin with, functions of DNA sequences are essential to ENCODE's goals and major findings. The ENCODE project seeks to identify the *functional* parts of the human genome, and one of ENCODE's major results is that functions can be assigned to 80% of the genome (e.g., ENCODE 2012, 57).

Even more importantly, functions play an important epistemic role in the ENCODE project. As shown in Section 4, ENDODE's General Strategy is to identify those DNA sequences as genomic parts, which engage in one of the kinds of biochemical activities detected by ENCODE's methods. This strategy raises the question of why certain methods are conceived of as appropriate means to detect genomic parts, whereas others are ignored. Likewise, one might wonder why certain kinds of biochemical activities are seen as indicators

[9] Most biochemical activities are tied to single DNA sequences, that is, only one DNA sequence is engaged in one biochemical activity.

of genomic parthood, whereas others are treated as irrelevant. What distinguishes relevant from irrelevant activities and legitimate methods from illegitimate ones? The answer is: It is the assumption that certain biochemical activities of DNA sequences are correlated with functions that these DNA sequences have in the genome as a whole. In other words, methods are legitimate if and only if they track such kinds of biochemical activities that are indicators of functions.[10] For example, early on it has been discovered that active promoters were correlated with alterations in chromatin structure that gave rise to nuclease hypersensitivity of the underlying DNA (e.g., Wu 1980). In the ENCODE project, the method DNase-seq is used to detect DNA sequences, which are hypersensitive to the nuclease DNase I, that is, which engage in the biochemical activity "being frequently cut by DNase I". This biochemical activity is regarded as relevant to genomic parthood because it is correlated with the function of DNA sequences to act as promoters (i.e., to initiate the start of gene transcription). This is why the arrows in Figure 1 do not only lead from the grey boxes to the top, indicating the biochemical activity that a method is targeted at, but also lead from the grey boxes down, indicating the functional element that a method individuates. To conclude, it is the function that renders a biochemical activity relevant and that legitimates a method as an adequate mean to detect genomic parts. DNase-seq, for example, is a legitimate method for individuating genomic parts because it tracks a biochemical activity (DNase I cleaving) that indicates a function of the individuated DNA sequences (the capacity to initiate gene transcription).

These considerations reveal that biochemical activities and functions of DNA sequences play very different epistemic roles in ENCODE's practices of individuating the parts of the human genome. Biochemical activities are the *targets of measurement*. They are the entities that are actually being measured and that ENCODE's methods are directed at. ENCODE's empirical data primarily record which DNA sequences display which biochemical activities, and particular DNA sequences are singled out as genomic parts by virtue of the biochemical activities they engage in. On the contrary, functions are the *targets of annotation*. They are the entities that justify whether or not a particular DNA sequence is a part of the human genome. Functions determine how to partition the human genome into parts (i.e., how to annotate the genome). ENCODE researchers employ certain methods that track certain biochemical activities because they assume that these biochemical activities are reliably correlated with specific functions of these DNA sequences. That is, specific biochemical activities serve as indicators for specific functions. For instance, if the antibodies used in ChIP-seq bind histones that are methylated at lysine 4 of histone protein H3 (H3K4me1), this method annotates DNA sequences that probably function as enhancers, that is, which bind specific transcription factors and thereby increase the likelihood that the transcription of a certain gene will occur (ENCODE 2011, 7; 2012, 59). Gaining knowledge

[10] Accordingly, Stamatoyannopoulos states that the "biochemical signature strategy… was motivated by the recognition of common biochemical or biophysical events that invariably attended certain types of noncoding functional elements" (2012, 1602).

about biochemical activities is only a mean to the end of gaining knowledge about the DNA's functions.[11] In conclusion, it is the function of a particular DNA sequence that justifies individuating it as a genomic part. The biochemical activity of the DNA sequence serves as an indicator of this function and thus can be used as a mean to single out genomic parts.

The different epistemic roles of biochemical activities and of functions have interesting consequences for our metaphysical picture of genomic part-whole relations. It should have become clear that if we want to specify the conditions under which something is a part of the human genome we should focus on functions, not on biochemical activities. The reason is that biochemical activities serve only as means to discover functions, whereas having a function is what makes a DNA sequence to a part of the human genome. Biochemical activities are crucial to understanding how scientists discover genomic part-whole relations but only functions determine which DNA segments in fact are parts of the human genome. BAC traces genomic parthood back to biochemical activities and thus fails to explain the importance of functions to genomic parthood. We must thus reject BAC and develop an alternative functional criterion.

6 A Functional Criterion for Genomic Parthood

This section develops a functional criterion for genomic parthood (as an alternative to BAC), which specifies the functional relationship that must hold if a biological object is a part of the human genome. In 2012, the ENCODE Consortium reported that they are able to "assign biochemical functions for 80% of the [human] genome" (2012, 57). This is a surprising result because only a very small fraction of the genome (approximately 1-2%) consists of protein-coding genes and only 3-8% of the genome has been undergoing purifying selection. In the media, ENCODE's findings were interpreted as debunking the concept of junk DNA because much of the DNA that used to be dismissed as junk in fact turns out to play "critical roles" and to be "active and needed" (Kolata 2012). ENCODE's claim that 80% of the human genome is functional initiated a heated debate about the right concept of function, its correct application, and other methodological and conceptual mistakes that the ENCODE project putatively involves (e.g., Graur et al. 2013; Doolittle et al. 2014; Elliott et al. 2014; Germain et al. 2014). At the center of the criticism lies the claim that ENCODE adopts a misleading concept of function – namely a causal role (CR) concept rather than a selected effect (SE) concept of function – and that this concept of function "divorces genomic analysis from its evolutionary context" (Graur et al. 2013, 579), results in a proliferation of spurious functions, and thus overestimates functionality.

The aim of this section is not to get involved in this dispute and to ultimately defend the CR concept of function against its alternatives. This would require an extensive argumentation than lies beyond the scope of this paper. Furthermore, my analysis does not

[11] It must be emphasized that we are only beginning to understand the genome and the functional roles that different genomic parts play in it. Even though the ENCODE project ultimately aims at ascribing functions to DNA sequences this functional knowledge is, at present, only fragmentary (cf. Germain et al. 2014).

need an ultimate defense of the CR account. This paper focuses on part-whole relations in the molecular-genetic domain and it offers an account of the conditions under which genomic part-whole relations exist by analyzing the reasoning and decomposition practices in the ENCODE project. For the purposes of my analysis, it is sufficient to assume that ENCODE researchers in fact use the notion of a biological function according to the CR approach[12] and that this is adequate because a CR concept of function suits ENCODE's biomedical research goals best (Germain et al. 2014, 815f.). Rather than defending a CR account of function, I focus on the question of how it can be used to develop a functional criterion for genomic parthood and in what respects this requires revising and supplementing Cummins' CR account (1975, 1983).

6.1 Cummins' CR Account of Function

The CR account of function traces back to Cummins' (1975) idea of functional analysis. According to Cummins, "to ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some capacity of a containing system" (1975, 765). This understanding of functions neatly applies to part-whole relations because it can be read as specifying the function of a part relative to a capacity of a whole (Cummins 1983, 29).[13] As Figure 2 illustrates, a causal role function is ascribed to a part *X* in an overall system *S*, where *X* has the capacity φ and *S* has the capacity ψ.

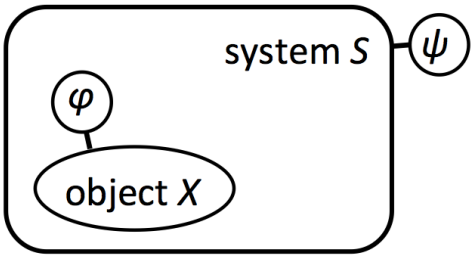

Fig. 2: Causal-role Functions and Part-whole Relations

In Cummins' view, φ is a causal role function of *X* in *S* if it plays a causal role in *S*, which means that φ contributes to *S*'s capacity to ψ. We can functionally explain the capacity ψ of *S* by describing the causal role function that *X* has in *S*, that is, how *X*'s φ-ing (i.e., *X* manifesting its capacity φ) contributes to *S*'s ψ-ing. For example, we can explain the capacity of the protozoan (*S*) to survive despite constantly absorbing water (ψ) by describing how the function of the contractile vacuole (*X*) to eliminate excess water (φ) contributes to the

[12] This assumption is supported by several claims that ENCODE researchers make. For example, they state that "our present state of knowledge may not yet permit definitive declaration of the ultimate biological role(s), function(s) or mechanism(s) of action of any given genomic element" (2011, 1).

[13] This does not render my account of genomic parthood circular because CR functions do not presuppose that the object (*X*) having the function is a biologically significant part of the system (*S*), in which the function plays a causal role. *X* can, for instance, be merely contained in *S* without being biological part of it (Jansen and Schulz 2014; Kaiser forthc.). Hence, using CR functions to specify a criterion for genomic parthood is not circular.

capacity of the protozoan to survive. Even though Cummins is more interested in explaining how functional analysis works than in explicating the concept of function and its metaphysical underpinnings (1983, 28, fn. 2), we can reconstruct the following definition of causal role functions from his writings.[14]

Cummins' CR Account of Function
An object *X* has a function *φ* in a system *S* with the capacity *ψ* if and only if *φ* has a causal role in *S*, that is, *X*'s *φ*-ing contributes to *S*'s *ψ*-ing.

6.2 Specifying the Variables

In Section 5, I argued that although ENCODE researchers use biochemical activities to single out particular DNA sequences as genomic parts it is the *function* of a DNA sequence that justifies its status as a part of the human genome. This section explicates the basic idea that genomic parts are essentially functional parts by using Cummins' account of CR functions to develop a functional criterion for genomic parthood. This strategy is legitimate because ENCODE researchers actually employ the CR concept of function (e.g., ENCODE 2011, 1). Even more, they have been accused for relying on the CR concept of function because it is said to be too permissive and to ignore the evolutionary context (e.g., Graur et al. 2013; Doolittle et al. 2014; Elliott et al. 2014). When applying Cummins' account to the case of genomic parthood, two questions must be addressed: First, what is the ontological nature of CR functions of genomic parts; are they dispositions or manifestations of dispositions? Second, what are the overall system *S* and its capacity *ψ*, in regarding to which CR functions of genomic parts are individuated?

I start with the first question. In the case of genomic parthood, the biological objects (*X*) having the CR functions are molecules, such as DNA segments, proteins (e.g., histones, DNA polymerases, transcription factors), or RNAs. Examples of the CR functions (*φ*) that these objects possess include, for instance, to act as a promoter (i.e., to initiate gene transcription) or to be a gene (i.e., to code for proteins). From a metaphysical perspective, the question arises whether CR functions are dispositions of objects (*φ*) or manifestations of dispositions (*X*'s actual *φ*-ing). A disposition is a property that an object has but that is only manifested under specific conditions (e.g., a muscle fiber manifests its disposition to contract if a neuronal stimulus arrives). By contrast, manifestations of biological dispositions typically are temporally extended processes (Kaiser and Krickel 2017). Hence, we might, for example, ask whether the CR function of a promoter is the disposition to initiate gene transcription or the manifestation of this disposition, that is, the actual initiation of gene transcription.

[14] The two most important passages read as follows: "*x* functions as a *φ* in *s* (or: the function of *x* in *s* is to *φ*) relative to an analytical account *A* of *s*'s capacity to *ψ* just in case *x* is capable of *φ*-ing in *s* and *A* appropriately and adequately accounts for *s*'s capacity to *ψ* by, in part, appealing to the capacity of *x* to *φ* in *s*." (Cummins 1975, 762) "To ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some capacity of a containing system." (Cummins 1975, 765)

Cummins' own approach seems ambivalent regarding this metaphysical question. On the one hand, he clearly states that CR functions are dispositions: "to ascribe a function to something is to ascribe a capacity to it" (1975, 765). On the other hand, we also find claims, such as "the function of *x* in *s* is *φ*-ing" (1975, 762) and in his examples he uses activity-language rather than disposition-language.[15] More importantly, ENCODE's functional reasoning and decomposition practices are metaphysically flexible at this point. One the one hand, ENCODE researchers use activity-language rather than disposition-language to speak about genomic functions (e.g., they refer to the "action of promoters" (2012, 61) and discuss the function of genes to be transcribed and to code for proteins) and a DNA sequence must do something, it must engage in a biochemical activity (e.g., binding to transcription factors or producing RNA; ENCODE 2011, 1; 2012, 57) to be identified as a genomic part. This suggests that CR functions of genomic parts are manifestations of dispositions (*φ*-ing), rather than dispositions themselves. On the other hand, the CR functions that mark genomic parts are distinct from the biochemical activities by which they are picked out (recall Section 5) and they need not belong to the same ontological category. Furthermore, ENCODE researchers ascribe CR functions to DNA sequences even if they do not manifest the relevant disposition at all times. For example, promoters do not permanently bind RNA-polymerases and genes do not produce RNAs at all times but they do not do not cease to be genomic parts if they stop manifesting their dispositions. This suggests that CR functions of genomic parts are dispositions (*φ*), rather than their manifestations. On my view, a functional criterion of genomic parthood must capture this metaphysical flexibility and allow for CR functions to be dispositions and manifestations of dispositions.

Let us turn to the second question and specify the other two variables in Cummins' account. In the case of genomic parthood, what are the system *S* and its capacity *ψ* with respect to which genomic parts are individuated? The relevant system seems to be clear since we are seeking a criterion for identifying the parts of the human genome. Hence, *S* is the human genome as a whole (the specific nature of the genome will be discussed in the next section). The relevant capacity *ψ* that guides ENCODE's individuation of genomic parts seems to be the capacity of the human genome to direct the overall functioning of the cell, which in turn influences human traits and diseases. Specifying the relevant capacity *ψ* of the genome as the capacity to direct the overall functioning of the cell accounts for the fact that the cellular context is central to DNA sequences performing their causal role functions in the genome. For example, in eukaryotes a DNA sequence can only act as a promoter and initiate the start of gene transcription if, for example, the protein RNA polymerase II is present in the cell and binds to the promoter. The CR function of the genomic part to initiate gene transcription contributes to the genome directing the overall functioning of the cell because the regulation of gene transcription affects the overall functioning of the cell.

[15] For example, "the function of the contractile vacuole in fresh-water protozoans is to eliminate excess water from the organism" (Cummins 1975, 758).

The preceding considerations give rise to the following functional criterion for genomic parthood.

Causal-role Functional Criterion (CFC)
A molecule *X* is a part of the human genome *G* only if
X has a causal role function *F* in *G*, that is, *F* contributes to *G* directing the overall functioning of the cell.

CFC states that a molecule, such as a DNA sequence, must possess a CR function in order to be a part of the human genome. Having a CR function requires that the putative genomic part has or manifests a disposition and thereby contributes to the genome directing the overall functioning of the cell. CFC deliberately leaves it open whether the contributes-to relation exists between the disposition of a part and the disposition of a whole or whether it exists between the manifestations of these dispositions (e.g., whether the capacity of a promoter to initiate gene transcription contributes to the capacity of the cell to direct its overall functioning or whether the actual gene initiation contributes to the cell actually directing its overall functioning). Furthermore, CFC shows in what way the parts of the human genome are unified. All genomic parts must contribute to the human genome's behavior to direct the overall functioning of the cell. The thesis that all CR functions are individuated with regard to the same behavior or capacity of the human genome also solves ENCODE's putative "permissiveness problem" (Elliott et al. 2014, 16) because it undermines the claim that the CR account gives rise to an extremely liberal, almost arbitrary criterion for ascribing functions to genetic elements.

I close this section by addressing three objections that might be raised against CFC. First, one might worry that CFC is circular because it specifies what it means for a putative genomic part to have a CF function by reference to the functioning of the cell. The circularity reproach can be refuted on the following grounds. CFC does not refer to CR functions in general, but rather specifies CR functions of molecules (e.g., of DNA sequences) by referring to the overall functioning of the cell. The functioning of the cell need not be spelled out in terms of CR functions. But even if it were it would not render CFC circular because CF functions of molecules would be defined in terms of different kinds of CR functions, namely in terms of CR functions of the cell.

Second, one might argue that I have failed to identify the correct behavior or capacity (ψ) of the whole that determines the CF functions of its parts. Instead of *G*'s capacity to direct the overall functioning of the cell one might, for instance, suggest to characterize ψ as *G*'s capacity to determine the linear sequence of amino acids or as *G*'s capacity to provide information for protein and RNA synthesis. Both alternative suggestions, however, ignore the context of the cell and they are capacities that rather pertain to particular DNA sequences than to the genome as a whole. In addition, the first suggestion focuses on genomic parts that function as protein-coding genes and neglects non-protein-coding genes,

transcriptional regulatory elements, and sequences that mediate chromosome structure and dynamics (ENCODE 2004, 636).

Third, critics might claim that it is not the genome and its capacity to direct the overall functioning of the cell that determines the functions of genomic parts, but rather the organism and its relative capacity to survive and reproduce. This claim is appealing because it allows for a neat integration of causal-role and etiological accounts of function (cf. Griffiths 1993). However, the claim conflicts with ENCODE's practice of individuating genomic parts because ENCODE researchers do not refer to evolutionary contexts to ascribe functions. Proponents of an etiological theory of function might object that a metaphysics of biological practice must be stronger normative at this point and replace ENCODE's CR concept of function by a more appropriate, etiological one. If the philosophical debate about functions were settled because an indisputable argument for why the level of organisms and their fitness has priority in ascribing functions is available, I would agree. But until it has been shown that "biological sense can only be derived from evolutionary context" (Graur et al. 2013, 579) I will stay out of this dispute and work with ENCODE's CR notion of function.

7 A Structural Criterion for Genomic Parthood

My main claim in this section is that genomic part-whole relations are not only functionally but also structurally constrained. Whether a particular object is a part of the human genome depends not only on whether it has a causal role function in the genome but also on whether it has a specific chemical nature. This section thus supplements the functional criterion for genomic parthood with a structural criterion.

ENCODE's General Strategy of individuating genomic parts (recall Section 4) is characterized as a "departure from the widely accepted reductionist approach to genome function" (Stamatoyannopoulos 2012, 1602). The non-reductionist character of ENCODE's approach is typically traced back to the fact that ENCODE studies the genome in its cellular context, rather than as an isolated DNA molecule, and that ENCODE draws attention to the many different molecules in that context and to how they affect genome function rather than focusing on the nucleotide sequence only. The notion of methodological reduction that underlies this characterization is twofold. Methodological reduction can either mean to focus on a system (i.e., the genome) and its parts and to ignore or simplify the system's context (i.e., the cellular context) or methodological reduction can mean to focus on the structure of a system or object and to ignore or simplify its function (Kaiser 2015). Accordingly, ENCODE's approach is non-reductionist because it avoids methodological reductions of both kinds. It studies the genome in its cellular context (including many different proteins, RNAs and other cell organelles) and it investigates how specific DNA elements function in the genome as a whole and how this affects genome functioning.

The non-reductionist character of the ENCODE project emphasizes that one cannot gain knowledge about genomic part-whole relations by studying isolated DNA molecules and their nucleotide sequence. Instead, one must take into account many other molecules, such

as RNAs and proteins (e.g., transcription factors or histones), and investigate how they affect the functioning of particular DNA sequences in the human genome. Without these various proteins and RNAs a genome could never exhibit its characteristic behaviors (e.g., gene regulation and transcription) and direct the overall functioning of the cell. The vast importance of these non-DNA molecules suggests that not only DNA sequences but also these molecules (i.e., proteins and RNAs) can be parts of the human genome.[16] ENCODE's methods would not work without proteins and RNAs and the genome would not be able to direct the overall functioning of the cell without them.

Surprisingly, however, ENCODE researchers exclude proteins and RNAs from being candidate parts of the human genome. The ENCODE project lists exclusively DNA sequences as genomic parts. Proteins and RNAs are regarded as essential to applying ENCODE's methods (e.g., ChIP-seq and RNA-seq) and to DNA sequences exhibiting its causal-role functions but they are not considered to be parts of the human genome. In other words, there seem to be an inconsistency in the ENCODE project between emphasizing the importance of proteins and RNAs (e.g., by highlighting the anti-reductionist character of ENCODE's approach) on the one hand and excluding proteins and RNAs from being genomic parts on the other. This apparent inconsistency can be resolved by pointing out that, in addition to the functional criterion for genomic parthood, there is a second necessary criterion implicitly at work in ENCODE's composition practice. This second criterion is a structural criterion that excludes various molecules, such as proteins and RNAs, from being genomic parts.

The structural criterion arises from the metaphysical view that the nature of the human genome as a whole is essentially structural. The structural view of the genome implicitly shapes the ENCODE project and constrains its composition practices. The structural view says that, as a carrier of heritable, genetic information, the genome must have a certain chemical structure. It must be a deoxyribonucleic acid that consists of nucleotides with one of four different bases (i.e., adenine, cytosine, guanine, and thymine). The nucleotides form two strands with sugar-phosphate backbones, which are bound together by hydrogen bonds between the bases (A pairs with T, and C with G). The structural view of the human genome traces back to Watson and Crick's double helix model of the DNA (1953). Accordingly, Stamatoyannopoulos states that "[a]lthough the ENCODE project formally originated in the post-genome era, its intellectual origins lie some 40 years earlier" (2012, 1602). In the ENCODE project, Watson and Crick's double helix model of the DNA plays the role of a background theory that assists the identification of the genome (as a structural entity) and of its parts. The structural nature of the human genome constrains the individuation of its parts in two ways. It follows from it, first, that genomic parts must consist of nucleotides

[16] Another support for this assumption can be found at the official homepage of the ENCODE project: "The goal of ENCODE is to build a comprehensive parts list of 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". (https://www.encodeproject.org/; my emphasis)

and, second, that genomic parts must be actual segments of the nucleotide sequence of the human genome. The following criterion captures the chemical-structural constraint on genomic parthood.

Chemical-Structural Criterion (CSC)
A molecule *X* is a part of the human genome *G* only if
X is an actual segment of *G*'s DNA sequence.

Adding CSC as a second criterion for genomic parthood resolves the apparent inconsistency in ENCODE's composition practices. Proteins and RNAs are excluded from being parts of the human genome because they have the wrong kind of chemical structure. Proteins consist of amino acids and RNAs are ribonucleic acids. Hence, they cannot be actual segments of DNA and fail to satisfy CSC.

8 The Account of Genomic Parthood and the Limits of the ENCODE Project

The central goal of this paper was to analyze ENCODE's investigative and reasoning practices to reveal the conditions under which some biological objects are parts of the human genome whereas others are not. In the preceding sections, I have developed two criteria, one functional and the other structural, which together constitute my metaphysical account of genomic parthood.[17]

Account of Genomic Parthood
A molecule *X* is a part of the human genome *G* if and only if
(1) *Causal-role Functional Criterion (CFC)*: *X* has a causal role function *F* in *G*, that is, *F* contributes to *G* directing the overall functioning of the cell and
(2) *Chemical-Structural Criterion (CSC)*: *X* is an actual segment of *G*'s DNA sequence)

Among the virtues of this account is that it enables us also to recognize the limitations of the ENCODE project. The goal of this last section is to address a possible objection to the first criterion of my account of genomic parthood and thereby to clarify what the ENCODE project can do and what it cannot do and how this affects my metaphysical claims about genomic part-whole relations.

For somebody who wants to pin down which DNA sequence, exactly, is a genomic part and which is not, CFC might be disappointing. The term 'contributes to' is quite vague and does not tell us how to count individual CR functions and assess the sameness of CR functions. Assume that ENCODE discovers that a particular DNA segment D_1, which consists

[17] As I have emphasized in the introduction, these two criteria determine what the biological parts of the human genome are (i.e., those parts that are relevant to biological processes and phenomena and that are thus studied by biologists); they do not concern the mereological parts or the physical parts of the human genome.

of a specific sequence of 92 nucleotides, functions as an enhancer that binds transcription factors and thereby contributes to the regulation of gene transcription. Does a DNA sequence D_2 that consists of the nucleotide sequence of D_1 plus one nucleotide at the end, which is a free rider and does not affect the binding of transcription factors, have the same function? Can one DNA sequence simultaneously have more than one function? Can there be different degrees of function? How can we specify the term 'contributes to' in a way that excludes that any combination of genomic parts is a genomic part as well?

I agree that these are interesting and urgent philosophical questions. However, the metaphysical claims that one makes when answering these questions cannot be drawn from ENCODE's practices of individuating genomic parts and thus lie beyond the scope of this paper. ENCODE identifies particular DNA sequences as parts of the genome because they display a biochemical activity of a certain kind (e.g., being cut by DNAse I or associating with histones that are chemically modified such that lysine 3 of histone protein H3 is methylated (H3K4me1)). The biochemical activities are treated as indicators for certain kinds of CR functions (e.g., to act as promoters, that is, to initiate gene transcription, or to act as enhancers, that is, to bind transcription factors and thus to increase the likelihood that gene transcription will occur) because it is assumed that certain kinds of biochemical activities are correlated with certain kinds of functions.

Two points must be noted here: First, even if ENCODE individuates very specific DNA sequences as genomic parts they do so by ascribing CR functions of quite general kinds, such as acting as a promoter. ENCODE's composition practice is thus quite *coarse-grained* and is not concerned with identifying fine-grained token functions. For instance, when specifying the putative functions that are regularly associated with specific histone modifications they refer to CR functions of coarse-grained kinds, such as "regulatory elements in dynamic chromatin", "enhancers and other distal elements", "promoters/transcription starts", "genes", "transcribed portions of genes", and "repressive domains" (ENCODE 2012, 59, Table 2). This practice of ascribing coarse-grained CR functions to genomic parts does not allow for any inference that would help us to answer the above questions.

Second, ENCODE's practices of ascribing CR functions to genomic parts are not more fine-grained for the same reason as for why ENCODE speaks about "putative functions" (2012, 59) and "candidate functional elements" (2011, 1). The empirical data that ENCODE gathers is data about which DNA sequence engages in which kind of biochemical activity. Ascribing CR functions to DNA sequences, however, requires much more than knowing biochemical activities. It requires telling a complex story about how a particular DNA sequence interacts with other molecules in the cell, how this relates to the biochemical activity that ENCODE's methods measure, and how the DNA sequence thereby contributes to the genome as a whole directing the overall functioning of the cell. At present, this

complex knowledge about CR functions is only fragmentary.[18] Accordingly, ENCODE can only be said to have specified *potential* CR functions of the genomic parts it individuates (Germain et al. 2014, 819). ENCODE's claim to have assigned "functions for 80% of the genome" (2012, 57), strictly speaking, is false. ENCODE's findings merely show that 80% of the human genome potentially or likely is functional, not that it definitely is functional.

Recognizing these limits of the ENCODE project does not present a problem for my account of genomic parthood because the fact that ENCODE provides only fragmentary information about CR functions does not challenge my claim that CR functions determine what the parts of the human genome are. Moreover, seeing the limits of the ENCODE project explains why my functional criterion is not more fine-grained and does not allow for counting individual CR functions and assessing the sameness of particular CR functions.

9 Conclusion

This paper presents a case study in the metaphysics of biological practice. It analyzes the investigative and reasoning practices in the ENCODE project to reveal the conditions under which some molecules, such as DNA sequences, are biological parts of the human genome and others are not. The result of my analysis is that genomic part-whole relations are determined functionally as well as structurally. The two criteria that together constitute my metaphysical account of genomic parthood specify these functional and structural constraints. On the one hand, genomic parts must possess a causal role function in the genome as a whole, that is, their functions must contribute to the genome directing the overall functioning of the cell. This is why not any arbitrary DNA sequence actually is a part of the human genome. On the other hand, the structural nature of the genome implies that genomic parts must have a specific chemical structure and be actual segments of the DNA sequence of the genome. This is why RNAs and proteins (e.g., histones or transcription factors) cannot be parts of the human genome.

References

Birney E (2012) The making of ENCODE: Lessons for big-data projects. Nature 489:49-51. doi: 10.1038/489049a.

Callender C (2011) Philosophy of Science and Metaphysics. In: French S, Saatsi J (eds) Continuum Companion to the Philosophy of Science, Continuum, pp 33–54.

Chakravartty, A (2013) On the Prospects of Naturalized Metaphysics. In: Ross D, Ladyman J, Kincaid H (eds) Scientific Metaphysics. Oxford University Press, Oxford, pp 27-50.

Chang H (2011) The Philosophical Grammar of Scientific Practice, Int Stud Philos Sci 25:205-221.

Cummins RC (1975) Functional analysis, J Phil 72:741-765.

[18] In some passages, ENCODE admits this knowledge gap: "our present state of knowledge may not yet permit definitive declaration of the ultimate biological role(s), function(s), or mechanism(s) of action of any given genomic element" (2011, 1).

Cummins RC (1983) The Nature of Psychological Explanation. MIT Press, Cambridge.

Doolittle WF (2013) Is junk DNA bunk? A critique of ENCODE. Proc Natl Acad Sci U S A 110(14):5294-5300.

Doolittle WF, Brunet TDP, Linquist S, Gregory TR (2014) Distinguishing between “function” and “effect” in genome biology. Genome Biol Evol 6:1158-1165.

Elliott TA, Linquist S, Gregory TR (2014) Conceptual an Empirical Challenges of Ascribing Functions to Transposable Elements. Am Nat 184(1):14-24.

The ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. Science 306:636-640.

The ENCODE Project Consortium (2011) A user’s guide to the encyclopedia of DNA elements (ENCODE). PLOS Biol 9(4):e1001046.

The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57-74.

Germain P-L, Ratti E, Boem F (2014) Junk or functional DNA? ENCODE and the function controversy. Biol Philos 29:807-831.

Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E (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.

Greally JM (2007) Encyclopedia of humble DNA. Nature 447:784-785.

Griffiths PE (1993) Functional Analysis and proper Functions. Brit J Phil Sci 44:409-422.

Illari P, Williamson J (2013) In Defense of Activities. J Gen Philos Sci 44:69-83.

Jansen L, Schulz S (2014) Crisp Islands in Vague Seas: Cases of Determinate Parthood Relations in Biological Objects. In: Calosi C, Graziani P (eds) Mereology in the Sciences. Parts and Wholes in Contemporary Scientific Contexts. Springer, Cham, pp 163-188.

Kaiser MI (2015) Reductive Explanation in the Biological Sciences. Springer, Cham.

Kaiser MI (2018) The Components and Boundaries of Mechanisms. In: Glennan S, Illari P (eds) The Routledge Handbook of Mechanisms and Mechanical Philosophy. Routledge, New York, pp 116-130.

Kaiser MI (2018) Individuating Part-whole Relations in the Biological World. In: Bueno O, Chen RL, Fagan MB (eds) Individuation, Process and Scientific Practices. Oxford University Press, Oxford, 63-87.

Kaiser MI, Krickel B (2017) The Metaphysics of Constitutive Mechanistic Phenomena. Brit J Phil Sci 68:745-779.

Kolata G (2012) Bits of mystery DNA, far from ‘Junk,’ play crucial role. The New York Times, pp 5-7. http://www.nytimes.com/2012/09/06/science/far-from-junk-dna-dark-matter-proves-crucial-to-health.html. Accessed July 19, 2017.

Ladyman J, Ross D (eds) (2007) Every Thing Must Go. Metaphysics Naturalized. Oxford University Press, Oxford.

Machamer P, Darden L, Craver CF (2000) Thinking About Mechanisms. Philos Sci 67:1-25.

Maher B (2012) The human encyclopedia. Nature 486:46-48.

Maudlin T (2007) The Metaphysics within Physics. Oxford University Press, Oxford.

Ross D, Ladyman J, Kincaid H (eds) (2013) Scientific Metaphysics. Oxford University Press, Oxford.

Simons P (1987) Parts. A Study in Ontology. Clarendon, Oxford.

Soler L, Zwart S, Lynch M, Israel-Jost V (eds) (2014) Science after the Practice Turn in the Philosophy, History, and Social Studies of Science. Routledge, New York.

Stamatoyannopoulos J (2012) What does our genome encode? Genome Res 22:1602-1611.

Strawson P (1959) Individuals. An Essay in Descriptive Metaphysics. Routledge, London.

van Inwagen P (1990) Material Beings. Cornell University Press, New York.

van Inwagen P, Sullivan M (2017) Metaphysics. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy (Spring 2017 Edition), https://plato.stanford.edu/archives/spr2017/entries/metaphysics/

Waters CK (2014) Shifting Attention from Theory to Practice in Philosophy of Biology. In: Galavotti MC, Dieks D, Gonzalez WJ, Hartmann S, Uebel T, Weber M (eds) New Directions in the Philosophy of Science. Springer, Cham, pp 121-139.

Watson JD, Crick FHC (1953) Molecular Structure of Nucleic Acids. Nature 171:737-738.

Wu C (1980) The 59 ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature 286:854-860.