FUNCTION WAR: AN EVALUATION OF ENCODE PROJECT AND JUNK DNA IN THE LIGHT OF PHILOSOPHY OF BIOLOGY

Atikur Rahman

ABSTRACT: For a large period in biology, it was thought that only 1-2% of the entire human DNA carries protein-making signals, and the remaining 98% of DNA does not carry protein-making signals. They were called junk DNA or non-coding DNA. The Encode Project began in 2003 to explore the function of the remaining 98% of human DNA in the context of human genome research. From the ENCODE project of 2007 and 2012, we know that about 80% of the DNA in the human body is not junk. The subsequent criticisms of the Encode project have been numerous. But no matter the criticism, the results of the Encode Project have always proved true. In this paper, I will evaluate the result of the Encode project for the case of junk DNA in the light of the philosophy of biology.

KEYWORDS: Philosophy of Biology; Encode Project; Junk DNA

1. Introduction: Every living cell contains a molecule called Deoxyribonucleic acid or DNA, and these molecules contain all the necessary information for the formation and maintenance of a living cell. Four nucleotide bases named adenine (A), cytosine (C), guanine (G), and thymine (T) are arranged linearly in a different order to form this DNA. [1] These different sequences are called DNA sequences, and this sequence determines all the information needed by the organism. However, the information that exists throughout the DNA is not the same, the parts that contain information are called genes. These genes are carried from one generation to the next. [2] There are two main parts in a cell, one is the nucleus

and the other is the cytoplasm. The cytoplasm contains organelles surrounded by various membranes. And the nucleus contains DNA, the container and carrier of heredity. However, not all organisms are the same. Many unicellular organisms do not have a well-formed nucleus (that is, the nuclear material is not surrounded by a nuclear membrane) and other organelles. They are called prokaryotes or primitive cells. And those who have a well-structured nucleus are called eukaryotes. DNA, the carrier and carrier of heredity within a eukaryotic cell, is organised into chromosomes within a membrane-enclosed organelle nucleus. DNA and some nuclear proteins (histone) are twisted together to form a well-organised dense structure, called Chromatin. Chromatins are further condensed to form chromosomes. As the cell divides, this patchy shape continues to simplify, making the chromatin look like a string of beads under the microscope. Proteins are signalled from euchromatin or less condensed chromatin but not from heterochromatin or highly condensed chromatin. All the information stored in an organism is collectively called the genome. This information is stored in the DNA contained within the genome of the cell's chromosomes. The small parts of DNA from which the code of RNA and the signal for making proteins necessary for organisms are stored, those parts are called genes. [3]

1.1 Junk DNA: A non-coding DNA (ncdna) sequence is a component of an organism's DNA that does not encode a protein sequence. Some non-coding regions appear to be mostly non-functional. Like intron, pseudogene, intergenic DNA, etc. "Junk DNA" broadly refers to "all DNA sequences that do not play a functional role in growth, physiology, or some other organism-level ability." [4] The term "junk DNA" was coined in the 1960s. However, it only became widely known in 1972 in a paper by Susumu Ohno. [5] The rate at which mutations occur in humans, Compared to that of the human genome, is much larger. And therefore the number of deleterious mutations per generation is very low. If essential information were conserved throughout the genome, many more deleterious mutations would occur per generation. That is, not all of the human genome is functional, but due to mutations at different times, this DNA remains as garbage. [6] Phylogenetics, the branch of molecular evolution, is based on this theory. Their task is to find molecular homology in the DNA of different species and create an evolutionary sequence based on it.

2. Meaning of "Function" In Philosophy of Biology: Function in physiology is an activity or process performed by a system in an organism, such as sensation or movement in an organism. [7] This definition was central to the biological interpretation of classical antiquity. [8] There are three concepts of the term "function" in the philosophy of biology. For example- 1. Causal role 2. Selected Effects 3. Goal contribution.

Causal Role- The origin of causal role theories of biological activity can be traced back to a 1975 paper by Robert Cummins. [9] Cummins defines the functional role of a component of a system as the causal role of the component that has a causal effect on the larger containing system. For example, the heart has the actual causal role of pumping blood into the circulatory system; So the function of the heart is to pump blood. Again, a biologist might say that kidneys play a role in removing waste from the bloodstream, so this is a function of the kidneys. This idea has been criticised, because it is too narrow concept of function. For example, the heart also has the causal effect of making a sound, But we cannot think of producing sound as the function of the heart. Robert Cummins replied to such objections by saying, there is no objective way to distinguish between true function and other effects. A component's effect may be relevant in explaining different overall abilities. The limits of what abilities should be explained depend on the researchers' particular explanatory interests. The heart is related to the blood circulation area of the body. The heart can be said to work as a pumping mechanism so the function of the heart (biological function) is pumping as it plays a role in blood circulation Now blood circulation is very important for living organisms. But the heart here acts as a sound maker rather than the heart being able to do anything in producing sound. In that case, there is no problem even if this word is not called the biological function of the heart. But anything that has a good or bad effect on any system can be called a biological function. Several academics have responded to other criticisms of the causal role. [10] [11] [12]

Selected effect- According to the selective effects theory of biological function, the function of a biological trait is the function for which the trait was selected, as argued by Ruth Millikan. [13] For example, the function of the heart is to pump blood, because that is the function for which the heart was selected through

evolution. In other words, pumping blood is what causes the heart to grow. This concept of function has also been criticised for being too restrictive. It is not always clear which behaviour has contributed to the selection of a trait, because biological traits may have a function, even if they are not selected for. As such beneficial mutations are not primarily selected for, but they have a function.[14]

Target contribution– Target contribution theory seeks to create a middle ground between causal role and selective effect theory, as [15] Boorse defines the function of a biological trait as such, target contribution is that which statistically contributes to the survival and reproduction of that trait. So for example, zebra stripes sometimes serve to confuse predators. This role of the zebra stripe will contribute to the survival and reproduction of zebras and hence the function of the zebra stripe is to confuse predators. According to this concept, whether a particular causal role of a trait is its function depends on whether that causal role contributes to the survival and reproduction of the organism. [16]

2.1 C-Value paradox: C-value is the amount of DNA in a haploid cell. The word "C-Value" is used from 'constant' or 'characteristic' because the value of this C-Value is the same in different types of cells of the same organism. [17] It might seem natural that the more complex an organism is, the larger the amount of information stored within its cells, and the larger its genome should be. However, the ideal cell or eukaryotic cell does not follow this rule. That is, in reality, the genome of an organism that is as complex as its structure is not larger than that of a relatively simple organism. Rather, there is a huge difference. An example is an amoeba. Amoeba's genome is about a hundred times larger than that of humans. Initially, researchers expected that the amount of DNA would correlate with an organism's biological complexity. Yet studies have shown that there is no such relationship. Some relatively simple organisms have a larger C value than more complex organisms. To resolve this paradox, Molecular biologists proposed that the majority of an organism's genome consists of DNA that does not code for proteins or regulate gene expression. The researchers concluded that non-coding DNA serves no real purpose. They declared it to be an artefact or debris from the evolutionary process. However, the C value paradox is not a problem for the Encode project. As such, organisms with larger genomes than humans must have more functional elements if the encode is true. This makes no sense from an

evolutionary point of view. Yet it is possible to explain the presence of the largest genomes in organisms less complex than humans. It may be that the extra DNA plays a role other than coding for proteins and regulating gene expression. Such as, some researchers have suggested that non-coding DNA may have a functional role. They developed a model in which non-coding DNA determines the volume of the cell's nucleus. As the overall cell volume increases, the nuclear volume increases and so the DNA material must also increase for the nuclear material of the cell to effectively communicate with the cell's cytoplasm. These researchers' DNA' model also provided a solution to this C value paradox. [18]

2.2 Mutational Load Argument: Compared to the rate at which mutations occur in humans, the human genome is much larger. And therefore the number of deleterious mutations per generation is very low. If essential information were conserved throughout the genome, many more deleterious mutations would occur per generation. For that, not all of the human genome is functional, but due to mutations at various times, this DNA remains as garbage. Many researchers claim that the results of the Encode Project may not be true for the mutational load. Encode critic Dan Graur has argued that Encode's empirical conclusions may not be correct because "consideration of mutational load leads to the conclusion that the functional fraction in the human genome cannot exceed 15%. [19] Therefore, the rest of the DNA remains junk. Even if Dan Graur's arguments were not true,for sake of the argument, if most genomes were biochemically functional, the number of deleterious mutations would be enormous. But since DNA has no biochemical function, if the purpose of non-coding DNA is to reduce the chance of harmful mutations in the organism, it still sounds like a wonderful, important function right? Mutational load is not such a problem for the Encode project. According to mutational load, the number of deleterious mutations would be high if most of the genome were functional. Recently, three researchers have studied this in the journal Genome Biology and Evolution. They noted that these arguments wrongly assume that there could potentially exist a person with no deleterious mutations in their genome: "Our approach is different from previous work that compared mean fitness at mutation-selection equilibrium with the fitness of an individual who has no deleterious mutations; we show that such an individual is exceedingly unlikely to exist. We find that the functional fraction is not very likely to be limited substantially by mutational load and that any such limit, if it exists, depends

strongly on the selection coefficients of new deleterious mutations.

…By comparing the population mean fitness at mutation-selection equilibrium to that of an individual who possesses no deleterious mutations, Graur (2017) concluded that, for likely values of the human per-base deleterious mutation rate, the functional fraction must be small.

In this article, we present a different approach to analyzing mutational load and the human functional fraction. We do not take the fitness of an individual with zero deleterious mutations to be a meaningful value, because in a finite population of a realistic size such an individual will never exist. Instead, we consider the fitness of the fittest individual likely to exist in a finite population. We conclude — while making no claims about the actual functional fraction as determined by comparative studies—that a mutational load argument is unlikely to set a low limit on the functional fraction of the human genome, and that any attempt to set such a limit must take into account the fitness effects of new deleterious mutations. [Emphasis added.]"

They end their paper as follows:

"Our conclusion is simply that an argument from mutational load does not appear to be particularly limiting on (function) f."

3. Encode Project and Function of non-coding DNA: The Encyclopedia of DNA Elements (ENCODE) is a public research project aimed at identifying the functional elements of the human genome. 2007 and 2012 Encode research has shown that 80% of the human genome has biochemical functions. [21] Noncoding DNA contains sequences that act as regulatory elements and Determine when and where genes are turned on and off. This is important. Because if the gene expression is disrupted, the cell will not be able to function according to the environment. Even disruption of gene expression can lead to cancer. Another point needs to be clarified. We have previously discussed the definition of function in the light of philosophy of biology. The definition of function depends on the particular explanatory interest of the researcher. Moreover, any good/bad effect on a system must be called a function. So what is the function of biochemical activity determined by the Encode project? Of course, we can see that if the non-coding DNA does not express the gene, then the possibility of disease increases in the human body, for this, gene expression is important for survival. Now we will discuss some more functions of non-coding

DNA and RNA. Transposons are important for genome complexity. Non-coding DNA acts as transposable elements in plant development. [22] Recent studies have shown that non-coding DNA plays an important role in ensuring the proper bundling of chromosomes inside the cell nucleus, which is essential for cell survival. This function appears to be conserved across many species. [23] Transposons play an important role in viability in mice and possibly all mammals. When researchers knocked out a specific transposon in mice, half of their mice died before birth. This is an important function of "junk DNA" for the survival of mammals. [24] Other research shows that long non-coding RNAs (lncRNAs) are widely expressed and have key roles in gene regulation. Recent studies have begun to unravel how the biogenesis of lncRNAs is distinct from that of mRNAs and is linked with their specific subcellular localizations and functions. Depending on their localization and their specific interactions with DNA, RNA, and proteins, lncRNAs can modulate chromatin function, regulate the assembly and function of membrane-less nuclear bodies, alter the stability and translation of cytoplasmic mRNAs and interfere with signalling pathways. Many of these functions ultimately affect gene expression in diverse biological and physiopathological contexts [25]

Discussion: The definition of "function" in the philosophy of biology has been hotly debated over the past centuries. However, it is logical to assert that non-coding DNA has a function as they have beneficial or harmful effects on an organism. The concepts of causal role and selected effect are quite successful in determining the meaning of function. Also, the argument raised as a critique of the causal role "that the heart also produces sound but we do not attribute this sound production to a function of the heart" this critic cannot be compared to the results of the Encode project. Because the function of non-coding DNA, be it biochemical activity or biological function, holds the highest concept of causal role. Although the mutational load argument is quite fruitful, it does not create any problem for the "biological function" of non-coding DNA. Moreover, the C value paradox does not create any problems with the results determined by the Encode project. So junk DNA can no longer be called junk.

Conclusion: Through the Encode project, we have come to know about the function of so-called junk DNA. Although the Encode project has been criticized

for different meanings of "function", we can see that Encode's results are correct, that Encode's defined functions are also correct and that junk DNA is not junk. Future research is expected to help get a clearer understanding of this issue.

Email: ar3670700@gmail.com

REFERENCES:

1. Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738. doi: https://doi.org/10.1038/171737a0

2.Bateson, W., & Mendel, G. (2013). *Mendel's principles of heredity*. Courier Corporation.

3. "What is a gene?: MedlinePlus Genetics". *MedlinePlus*. 17 September 2020. Retrieved 4 January 2021.

4. Palazzo, A. F., & Gregory, T. R. (2014). The case for junk DNA. *PLoS genetics*, 10(5), e1004351. doi:10.1371/journal.pgen.1004351

5. Ohno, S. (1972). So much junk DNA in our genome. In Evolution of Genetic Systems, *Brookhaven Symp. Biol.* (pp. 366-370). PMID: 5065367

6. Kimura, M., & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics,* 54(6), 1337. doi: 10.1093/genetics/54.6.1337

7. Fletcher, John (1837). On the functions of organized beings, and their arrangement. Rudiments of physiology, Part 2. *On life, as manifested in irritation*. John Carfrae & Son. pp. 1–15.

8.Tipton, J. A. (2014). *Philosophical biology in Aristotle's parts of animals*. Heidelberg: Springer.

9. Robert Cummins. (1975) Functional Analysis. *The Journal of Philosophy*, Vol. 72, No. 20 pp. 741-765

10. Function without purpose: the uses of causal role function in evolutionary biology." *Biology and Philosophy* 9: 443-469 https://doi.org/10.1007/BF00850375.

11.Ariew, A., Cummins, R., & Perlman, M. (eds.) (2002). *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press

12. Griffiths, P. E. (1993). "Functional analysis and proper functions." *British Journal for the Philosophy of Science* 44: 409-422.

13.Millikan, Ruth (1989). "In Defense of Proper Functions". *Philosophy of Science*. 56 (2): 288–302. https://www.jstor.org/stable/187875

14.Neander, Karen (1991). "Functions as Selected Effects: The Conceptual Analyst's Defense". *Philosophy of Science*. 58 (2): 168–184 doi:

https://doi.org/10.1086/289610

15.] Boorse, Christopher (1977). "Health as a Theoretical Concept". *Philosophy of Science.* 44 (4): 542–573. doi: https://doi.org/10.1086/288768

16.Bigelow, John; Pargetter, Robert (1987). "Functions". *The Journal of Philosophy*. 84 (4): 181–196.

17. Greilhuber J, Doležel J, Lysák M, Bennett MD (2005). "The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents". *Annals of Botany*. 95 (1): 255–60. doi: 10.1093/aob/mci019

18.T. Cavalier-Smith, "Nuclear Volume Control by Nucleoskeletal DNA, Selection for Cell Volume and Cell Growth Rate, and the Solution of the DNA C-Value Paradox," *Journal of Cell Science* 34, (1978): 247-278. doi: 10.1242/jcs.34.1.247

19. Graur, D. (2017). An upper limit on the functional fraction of the human genome. *Genome biology and evolution*, 9(7), 1880-1885 doi: https://doi.org/10.1093/gbe/evx121.

20.Galeota-Sprung, B., Sniegowski, P., & Ewens, W. (2020). Mutational load and the functional fraction of the human genome. *Genome biology and evolution*, 12(4), 273-281. doi: https://doi.org/10.1093/gbe/evaa040

21. ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489(7414), 57. doi: https://doi.org/10.1038/nature11247

22.Ariel, F. D., & Manavella, P. A. (2021). When junk DNA turns functional: Transposon-derived non-coding RNAs in plants. *Journal of Experimental Botany*, 72(11), 4132-4143. doi: https://doi.org/10.1093/jxb/erab073

23. Jagannathan, M., Cummings, R., & Yamashita, Y. M. (2018). A conserved function for pericentromeric satellite DNA. *Elife,* 7, e34122.doi: https://doi.org/10.7554/eLife.34122

24.Modzelewski, A. J., Shao, W., Chen, J., Lee, A., Qi, X., Noon, M., ... & He, L. (2021). A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. *Cell,* 184(22), 5541-5558. doi: https://doi.org/10.1016/j.cell.2021.09.021

25.Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature reviews: Molecular cell biology*, 22(2), 96-118. doi:https://doi.org/10.1038/s41580-020-00315-9