How Much of the Human Genome is Functional?

Nielsen, Henrik

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

How Much of the Human Genome is Functional?

Henrik Nielsen
Department of Bio and Health Informatics,
Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

In his 2002 article "The chicken and the Orphean egg: On the function of meaning and the meaning of function", Claus Emmeche mentions two crucial characteristics of living systems that make them radically different and irreducible to physics and chemistry:
(1) biosystems (organisms) contain genetic information;
(2) biosystems (organisms) have functions.
The "genetic information" of point (1) could, from a naïve viewpoint, be equated with the genome of the organism, but that would ignore point (2). Not all DNA of an organism is necessarily functional, and therefore, not all DNA necessarily carries information in any non-trivial sense of the word.

The human genome comprises roughly 3.1 billion base pairs of DNA, but only slightly more than 20,000 protein-coding genes (estimates vary a bit). This means that only around 1% of the DNA is directly protein-coding. What is the rest doing? If we take into account known RNA-coding genes and regulatory regions, we end up with only 2-3% of the DNA having a function we can account for. The rest may have functions we don't know yet, or it may be "junk"—DNA that is just there without actually doing anything good for us.

So how much of the human genome is functional? The answer depends on what you mean by "function" in biology. Using a "selected effect" concept of function, various groups have estimated the fraction to be 5-15%. However, using a "causal role" concept of function, the ENCODE consortium in 2012 reported that they had found function for 80% of the genome, prompting science writers to talk about the "eulogy for junk DNA". Of course, this sparked a heated debate.

From a biosemiotic perspective, both the "selected effect" and the "causal role" definitions seem to miss the point. The "selected effect" definition is diachronic, making it impossible to talk about function without taking history into account and thereby turning a concept such as "a new function" into a contradiction in terms. The "causal role" definition is synchronic, but completely misses the crucial characteristics of living systems. According to Emmeche, however, "Any biofunction is something (a process or a structure) that has meaning for the organism as an interpretant system".

I will argue that Emmeche's biosemiotic concept of function, when applied to the genome, is intricately linked to Bateson's definition of information as "a difference that makes a difference", and that this will result in an estimate of functional DNA that is closer to the "selected effect" than to the "causal role" estimate. In other words, the rumours of the death of "junk DNA" are exaggerated.