

“Junk” DNA—Past, Present, and Future, Part 1

Georgia Purdom, Ph.D., Answers in Genesis

Keywords

junk DNA, genetics, ENCODE, Susumu Ohno, pseudogenes, genetic fossil, mutations

While the month of June is commonly known as a time to survey other peoples' junk at yard sales across America, this year it featured a survey of a different kind of junk—“junk” DNA. The ENCODE (Encyclopedia of DNA Elements) Pilot Project which analyzed 1% of the human genome reported the culmination of 3½ years of work by hundreds of scientists in 11 countries. Their findings were published in 29 papers in the journals *Nature* and *Genome Research*.

Unfortunately, for many years this notion that non-coding DNA was not functional (“junk”) actually inhibited science.

History of “junk” DNA

A geneticist, Susumu Ohno, was the first to coin the term “junk” DNA in 1972.¹ He used the term to refer to pseudogenes (commonly thought of as defunct relatives of known genes that do not code for proteins), but with time its meaning broadened to include all non-coding DNA (DNA that does not contain genes and does not produce proteins).² Ohno stated, “The earth is strewn with fossil remains of extinct species; is it a wonder that our genome too is filled with the remains of extinct genes?”³ Due to his evolutionary presupposition, he assumed that non-coding DNA was merely a “genetic fossil” that may have been useful somewhere in our evolutionary past but had been discarded as we evolved into more complex, higher organisms. Since this “junk” DNA was no longer needed, it would not be under selective pressure, and mutations could accumulate without any harm to the organism.

Unfortunately, for many years this notion that non-coding DNA was not functional (“junk”) actually inhibited science. Many scientists didn't spend their time studying it because of their evolutionary presuppositions that it was worthless DNA. It is rather an ironic situation, since many times creationists are accused of inhibiting science because of their presuppositions (the God-did-it-so-there's-nothing-more-to-study idea; but in reality, it's just the opposite—“God did it” so we have every reason to study it!). Creationists for many years have argued that non-coding DNA is not junk (see “*Junk*” DNA: evolutionary discards or God's tools? and “*Junk*” DNA is not junk) because of their presupposition that God intelligently designed all DNA and it does have a purpose (granting that DNA has been negatively affected by the curse and some of its original functions may have been lost).

ENCODE Pilot Project Findings

The ENCODE Pilot Project studied 1% of the human genome or approximately 30 million bases spread throughout the genome.⁴ To their surprise, the scientists discovered that much of the “junk” or non-coding DNA was actually transcribed into RNA and was functional.

Most biology textbooks relay the genetic dogma that “one gene=one protein”. Proteins consist of a coding region and regulatory regions in the DNA. The coding region of the DNA (commonly called a gene) is transcribed into RNA (messenger RNA) and “read” to make a protein which carries out many of the functions of the cell. The regulatory regions determine many aspects of the protein production such as when the protein is made, how much is made, how long it's made, etc. Many regulatory regions have already been identified in the DNA (typically close to the gene), but it was known that many more had to exist for the complex level of protein regulation that exists in the cells. Was it possible that this information might exist in the “junk” or non-coding DNA?

“Junk” or non-coding DNA was not thought to be transcribed (this is typically reserved for genes); however, the ENCODE Pilot Project discovered that almost all DNA is transcribed into RNA.⁵ Ewan Birney, a coordinator of ENCODE at the European Bioinformatics Institute in England, states, “The genome looks like it is far more of a network of RNA transcripts that are all collaborating together. Some go off and make proteins; [and] quite a few, although we know they are there, we really do not have a good understanding of what they do.”⁶ It appears

that many of the regulatory regions of proteins may be transcribed into RNA and serve their function as RNA (other examples exist such as microRNAs and siRNAs). Thus, rather than being linear and one dimensional the genome should be viewed as a three dimensional network (does this sound familiar—evolutionary tree vs. creationist orchard?). ENCODE discovered some of the functions of the RNA transcripts including the regulation of transcription, replication of DNA, and the structure of chromosomes (which likely plays a role in the regulation of protein production).⁷ This may help account for the fact that even though we only have 20,000–25,000 genes in our DNA, we produce 100,000–300,000 proteins.

Even the definition of the word gene will need to be rewritten. A common definition of the word gene is “a section of the DNA that encodes a specific functional product (protein)”.⁸ One group working on the ENCODE Pilot Project has suggested, “A gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products”.⁹ Due to the complexity of the genome, it seems a bit like trying to define the undefinable.

What does the future hold?

Here are just a few of the many quotes from scientists commenting on the ENCODE Pilot Project results:

- Tim Hubbard, Wellcome Trust Sanger Institute, “We are now seeing the majority of the rest of the genome is active to some extent”.¹⁰
- Francis Collins, director of the National Human Genome Research Institute (NHGRI), “There’s a lot more going on than we thought”.¹¹
- Eric Green, scientific director of NHGRI, “The take-home message [from the ENCODE results] is, ‘Oh, my gosh, this is really complicated’”.¹²

I couldn’t agree more! As a molecular geneticist by training, I can honestly say that much of what I learned in college and graduate school will need to be completely rewritten.

Scientists are making plans to study the remaining 99% of the genome in a similar fashion. It is hoped that studying the non-coding sequences will lead to a greater understanding of disease processes. The likelihood of developing Type 2 diabetes has already been linked to mutations in non-coding sequences.¹³

I read the ENCODE Pilot Project results with great enthusiasm as it spoke volumes for the complexity that the Creator God has designed and His intelligence in doing so. It also reinforced the fact that beginning with biblical presuppositions in any type of science (historical or operational) is supreme. It is interesting to ponder how much sooner this type of research might have been employed had many scientists recognized the hand of God in creating the genome rather than the cruel effects of time and chance.

Part 2 will focus on the ENCODE Pilot Project findings as related to conservation/non-conservation of non-coding DNA across mammalian species.

Footnotes

1. Kuska, B., 1998. Should scientists scrap the notion of junk DNA? *Journal of the National Cancer Institute* **90**:1032–1033.
2. Kuska, Ref. 1.
3. Kuska, Ref. 1.
4. The Encode Project Consortium, 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**:799–816.
5. The Encode Project Consortium, Ref. 4.
6. BBC News, 2007. Human genome further unravelled.
7. The Encode Project Consortium, Ref. 4.
8. Genelabs Technologies, Inc. glossary.
9. Gerstein, M. B., et al., 2007. What is a gene, post-ENCODE? History and updated definition. *Genome Research* **17**:669–681.
10. BBC News, Ref. 6.
11. Washington Post, 2007. Intricate toiling found in nooks of DNA once believed to stand idle.
12. Washington Post, Ref. 11.
13. Washington Post, Ref. 11.