

This Week in Genomics  
October 4, 2021

Quanta Magazine

## The Complex Truth About 'Junk DNA'

- *Genomes hold immense quantities of noncoding DNA. Some of it is essential for life, some seems useless, and some has its own agenda.*
  - The 98% of the human genome that does not encode proteins is sometimes called junk DNA, but the reality is more complicated than that name implies.

Jake Buehler

Imagine the human genome as a string stretching out for the length of a football field, with all the genes that encode proteins clustered at the end near your feet. Take two big steps forward; all the protein information is now behind you.

**The human genome has three billion base pairs in its DNA, but only about 2% of them encode proteins. The rest seems like pointless bloat, a profusion of sequence duplications and genomic dead ends often labeled “junk DNA.” This stunningly thriftless allocation of genetic material isn’t limited to humans: Even many bacteria seem to devote 20% of their genome to noncoding filler.**

Many mysteries still surround the issue of **what noncoding DNA is**, and **whether it really is worthless junk or something more**. Portions of it, at least, have turned out to be **vitaly important biologically**. But even beyond the question of its functionality (or lack of it), researchers are beginning to appreciate **how noncoding DNA can be a genetic resource for cells and a nursery where new genes can evolve**.

“Slowly, slowly, slowly, **the terminology of ‘junk DNA’ [has] started to die,**” said Cristina Sisu, a geneticist at Brunel University London.

Scientists casually referred to “junk DNA” as far back as the 1960s, but they took up the term more formally in 1972, when the geneticist and evolutionary biologist **Susumu Ohno** used it to **argue that large genomes would inevitably harbor sequences**, passively accumulated over many millennia, **that did not encode any proteins**. Soon thereafter, researchers acquired hard evidence of **how plentiful this junk is** in genomes, **how varied its origins are**, and **how much of it is transcribed into RNA** despite lacking the blueprints for proteins.

**Technological advances in sequencing, particularly in the past two decades, have done a lot to shift how scientists think about noncoding DNA and RNA**, Sisu said. Although these noncoding sequences don’t carry protein information, they are sometimes shaped by evolution to different ends. As a result, **the functions of the various classes of “junk” — insofar as they have functions — are getting clearer**.

**Cells use some of their noncoding DNA to create a diverse menagerie of RNA molecules that regulate or assist with protein production in various ways**. The catalog of these molecules keeps expanding, with **small nuclear RNAs, microRNAs, small interfering RNAs** and many more. **Some are short segments**, typically less than two dozen base pairs long, while others are an order of magnitude longer. Some exist as double strands or fold back on themselves in hairpin loops. But all of them can **bind selectively to a target, such as a messenger RNA transcript, to either promote or inhibit its translation into protein**.

These RNAs can have substantial effects on an organism's well-being. Experimental **shutdowns of certain microRNAs in mice, for instance, have induced disorders ranging from tremors to liver dysfunction.**

By far the **biggest category of noncoding DNA in the genomes** of humans and many other organisms consists of **transposons, segments of DNA that can change their location within a genome.** These **"jumping genes"** have a propensity to make many copies of themselves — sometimes hundreds of thousands — throughout the genome, says Seth Cheetham, a geneticist at the University of Queensland in Australia. Most prolific are the **retrotransposons**, which spread efficiently by making RNA copies of themselves that convert back into DNA at another place in the genome. **About half of the human genome is made up of transposons; in some maize plants, that figure climbs to about 90%.**

**Noncoding DNA also shows up** within the genes of humans and other eukaryotes (organisms with complex cells) **in the intron sequences** that interrupt the protein-encoding exon sequences. When genes are transcribed, the exon RNA gets spliced together into mRNAs, while much of the intron RNA is discarded. But some of the intron RNA can get turned into small RNAs that are involved in protein production. **Why eukaryotes have introns is an open question, but researchers suspect that introns help accelerate gene evolution by making it easier for exons to be reshuffled into new combinations.**

A **large and variable portion of the noncoding DNA** in genomes consists of **highly repeated sequences of assorted lengths.** The **telomeres** capping the ends of chromosomes, for example, consist largely of these. It seems likely that **the repeats help to maintain the integrity of chromosomes (the shortening of telomeres through the loss of repeats is linked to aging).** But **many of the repeats** in cells serve no known purpose, and they **can be gained and lost during evolution, seemingly without ill effects.**

One category of noncoding DNA that intrigues many scientists these days is the **pseudogenes**, which are usually viewed as the **remnants of working genes that were accidentally duplicated and then degraded through mutation.** As long as one copy of the original gene works, natural selection may exert little pressure to keep the redundant copy intact.

Akin to broken genes, pseudogenes might seem like quintessential genomic junk. But Cheetham warns that some pseudogenes may not be "pseudo" at all. Many of them, he says, were presumed to be defective copies of recognized genes and labeled as pseudogenes without experimental evidence that they weren't functional.

**Pseudogenes** can also evolve new functions. "Sometimes they **can actually control the activity of the gene from which they were copied,**" Cheetham said, if their RNA is similar enough to that of the working gene to interact with it. Sisu notes that the discovery in 2010 that the **PTENP1 pseudogene** had found a second life as an RNA **regulating tumor growth** convinced many researchers to look more closely at pseudogene junk.

Because dynamic **noncoding sequences** can produce so many genomic changes, the sequences **can be both the engine for the evolution of new genes and the raw material for it.** Researchers have found an example of this in the **ERVW-1 gene**, which encodes a **protein essential to the development of the placenta in Old World monkeys, apes and humans.** The gene arose **from a retroviral infection in an ancestral primate about 25 million years ago,** hitching a ride on a retrotransposon into the animal's genome. **The retrotransposon "basically co-opted this element, jumping around the genome, and**

actually turned that into something that's really crucial for the way that humans develop," Cheetham said.

But how much of this DNA therefore qualifies as true "junk" in the sense that it serves no useful purpose for a cell? This is hotly debated. In 2012, the **Encyclopedia of DNA Elements (Encode)** research project announced its findings that about **80% of the human genome seemed to be transcribed or otherwise biochemically active and might therefore be functional**. However, this conclusion was widely disputed by scientists who pointed out that DNA can be transcribed for many reasons that have nothing to do with biological utility.

Alexander Palazzo of the University of Toronto and T. Ryan Gregory of the University of Guelph have described **several lines of evidence** — including evolutionary considerations and genome size — that strongly **suggest "eukaryotic genomes are filled with junk DNA that is transcribed at a low level."** Dan Graur of the University of Houston has argued that because of mutations, less than a quarter of the human genome can have an evolutionarily preserved function. Those ideas are still consistent with the evidence that the "selfish" activities of transposons, for example, can be consequential for the evolution of their hosts.

**Cheetham thinks that dogma about "junk DNA" has weighed down inquiry into the question of how much of it deserves that description.** "It's basically discouraged people from even finding out whether there is a function or not," he said. On the other hand, because of improved sequencing and other methods, "we're in a golden age of understanding noncoding DNA and noncoding RNA," said Zhaolei Zhang, a geneticist at the University of Toronto who studies the role of the sequences in some diseases.

**In the future, researchers may be less and less inclined to describe any of the noncoding sequences as junk because there are so many other more precise ways of labeling them now.** For Sisu, the field's best way forward is to keep an open mind when assessing the eccentricities of noncoding DNA and RNA and their biological importance. **People should "take a step back and realize that one person's trash is another person's treasure," she said.**