The Unseen Genome: Gems among the Junk

Sent to Scientific American in 2003 but not printed by Scientific American

The publication of the article “The Unseen Genome: Gems among the Junk”, by W.Wayt Gibbs (October 2003), may turn out to be a critical step in facing up to the possibility that almost the whole of every DNA strand in a cell is potentially-executable “code”. The original “junk” hypothesis – that only the sections of code that form a template for protein construction represent “valid” code – was hastily arrived at, and led to researchers treating these sections of code as uninteresting. Now the hypothesis is being tested and found to be, at least in part, wrong. Perhaps it is entirely wrong.

A useful analog for DNA might be a computer program. Suppose that we knew absolutely nothing about a programming language, but had studied samples of code and noticed that it contained strings of characters enclosed in quotation marks, and that these strings could be matched with words in the program output, like “RESULTS” and “ ACHIEVED” in:

```
1000 PRINT "RESULTS ";
1100 PRINT X1,X2,X3;
1200 PRINT " ACHIEVED"
1300 GOTO 700
```

Looking at this code, would we say that everything other than the printable strings is “junk” that does nothing? Surely not. By analogy, we should think of protein-defining sections of DNA as being like “the stuff in quotes”, and we should expect the rest of the DNA to be just as important – possibly more important. This is a much more logical conclusion. If the “junk” were really junk, and cells were able to recognize and ignore junk when “executing” the DNA, then cells that edit out the junk while copying the DNA would have evolved and prospered, being more energy-efficient. Yet the “junk” is still there; so it is probably not junk.

Of course, not all of the code in the DNA may be executed in a particular cell at a particular point in its life. As in a computer program, we should expect the code to contain conditional statements that cause branching, such that only certain sections are executed in a given situation.

If researchers were to approach the task of understanding how DNA as a whole operates, with the assumption that the majority of DNA is “potentially executable code” that plays a vital role in cell construction (and multicellular lifeform assembly), then they would be more likely to find out how it all works.

Malcolm Hamer
New York