Junk DNA Isn't Junk After All

Letter sent to Scientific American in 2004 but not printed by Scientific American

The article “The Hidden Genetic Program”, by John S. Mattick (October 2004), is the latest in a series of articles which reveal the growing recognition that introns (also known as “junk DNA” – the large sections of DNA in between the protein-defining strings) is not, after all, “junk”. It is becoming clear that intronic DNA, through a number of mechanisms which we are only starting to observe and understand, plays a role in assembling proteins into operating structures.

Viewing DNA as something analogous to a computer program, rather than a series of protein-defining strings separated by junk, is much more likely to lead to a full understanding of how DNA works: it implies that we accept that at least some, and possibly most, of the intronic DNA consists of instructions about how to build a cell, or at least how to split one cell into two cells. Of course, not all of the intronic DNA code may be “executed” in a particular cell at a particular point in its life. As in a computer program, we should expect the intronic DNA “program” to contain conditional statements that cause branching, such that only certain sections are executed in a given situation.

One point that has not yet been covered in the articles on this topic is the need to search for two distinct classes of “program” within the intronic DNA of eukaryotes. The first class consists of programs that play a vital role in *intracellular processes,* particularly during cell division. The second class consists of programs that play a role in *multicellular lifeform assembly*. Overall, we should expect at least half of intronic DNA to be concerned with intracellular processes, rather than multicellular lifeform assembly.

Multicellular lifeforms, such as human beings, are clearly seen to be complex when studied with an optical microscope. However, the information required to assemble roughly 100 trillion cells into a human being (given a means of generating cells) is no more than, and is possibly less than, the information required to assemble about 100 trillion molecules of various compounds into a cell. (We are reminded of the remarkable complexity of cells only occasionally by molecular biologists.)

Congenital diseases, and susceptibility to other diseases, can arise from three distinct characteristics of a person’s DNA: (A) errors in protein-defining strings, (B) errors in the intracellular-process-defining intronic DNA, or (C) errors in the multicellular-lifeform-assembly intronic DNA. Diseases related to Type (A) errors have been easy to identify. However, in order to understand the relationships between Type (B) and Type (C) errors and the diseases that they may cause, or make an individual susceptible to, it is first necessary to identify which parts of intronic DNA are related to intracellular processes, and which parts are related to multicellular lifeform assembly.

Malcolm Hamer
New York