Researchers discover new class of objects encoded within the genome

October 6 2010

Despite progress in decoding the genome, scientists estimate that fully 95 percent of our DNA represents dark, unknown territory. In the October 1 issue of the journal *Cell* researchers at The Wistar Institute shed new light on the genetic unknown with the discovery of the ability of long non-coding RNA (ncRNA) to promote gene expression. The researchers believe these long ncRNA molecules may represent so-called gene enhancer elements-short regions of DNA that can increase gene transcription. While scientists have known about gene enhancers for decades, there has been no consensus about how these enhancers work.

These findings join a growing body of evidence that the classic "central dogma" of genetics is incomplete. In the central dogma, chromosomal DNA is transcribed into RNA, which is then translated by the cell into proteins. In recent years, however, scientists have found that not all transcribed RNA molecules become translated into proteins. In fact, studies have shown that whole swathes of the genome are transcribed for unknown reasons.

In the present study, the Wistar researchers pinpoint 3,000 long ncRNAs and estimate that there could be a total of between 10,000 to 12,000 long ncRNA sequences within our DNA. This number is comparable to the 20,000 genes that are known to encode proteins. Most long ncRNAs are encoded in DNA near genes known to be important to both stem cells and cancer. This observation also suggests that targeting ncRNAs may represent a new strategy in slowing cancer growth.

"We are excited, first of all, because this is a new discovery about the very nature of human DNA; a new class of genetic object and a new layer of genetic regulation," said Ramin Shiekhattar, Ph.D., Wistar's Herbert Kean, M.D., Family Professor and senior author of the study.

"Secondly, we may have solved, in part, a great mystery in modern genetics. These long non-coding RNA sequences may account for the activity of enhancer elements, which have been well-studied but never quite characterized," Shiekhattar said.

Almost three years ago, while at the Centre for Genomic Regulation in Barcelona, Spain, Shiekhattar began a prospective hunt for non-coding RNA sequences using GENCODE, a database that annotates the human genome with currently available scientific evidence. After filtering out protein-coding transcripts and non-coding RNAs that might overlap known protein-coding genes, they found approximately 3,000 long ncRNA sequences. At the time, GENCODE only accounted for a third of the genome, so Shiekhattar estimates that there are likely more.

The researchers mapped the ncRNA sites within the genome, and found that ncRNAs tended to be located near genes that influence how stem cells change into other cell types. Shiekhattar and his colleagues then developed new assays to screen cell cultures for these ncRNA sequences, and discovered that ncRNAs were found extensively in a variety of cell types.

The idea that molecules of RNA can have a DNA-regulating effect is well established. More than 1,000 so-called microRNAs are known to science, for example, and their effect on silencing genes has been well described. According to Shiekhattar, he assumed that long ncRNAs would also silence genes, not promote their activation. To his surprise, the researchers found that depleting a cell of ncRNAs actually decreased the degree of overall gene expression of neighboring genes, revealing a

role for ncRNAs in potentiating gene expression.

In fact, when Shiekhattar and his colleagues depleted adult stem cells of a specific long ncRNA, known as ncRNA-activating 7 (ncRNA-a7), it had the same effect as depleting the protein product of a nearby gene, Snai1, which regulates how the cells migrate. Their studies further showed that inserting an ncRNA next to a gene for luciferase-the enzyme responsible for a firefly's glow-increased the amount of protein produced by that gene in cells grown in culture. While not all long ncRNAs may act like enhancers, the majority of the ones the team studied do, Shiekhattar says.

"We know long non-coding RNAs can promote gene expression, but what we need to know now is how they do it," Shiekhattar said, "which is precisely the object of our ongoing research plan."

Provided by The Wistar Institute