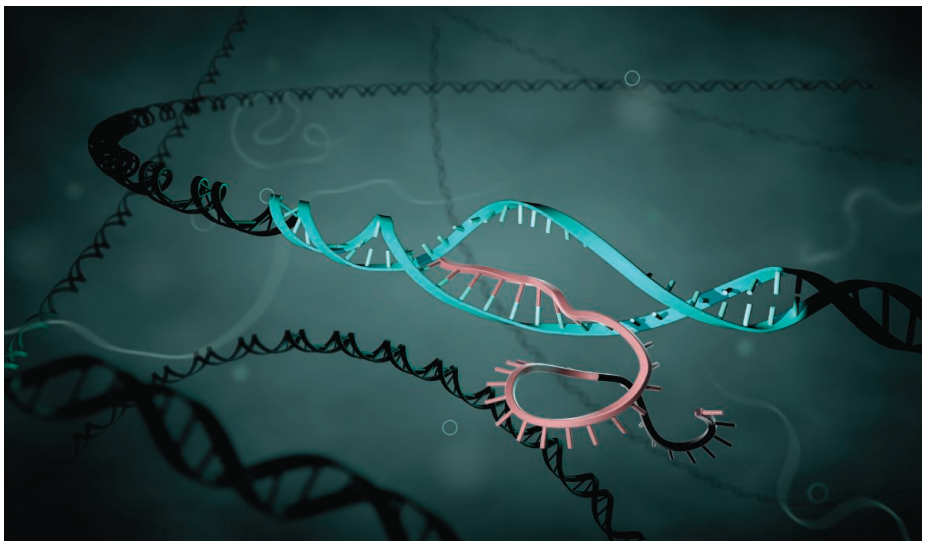

Shining a Light on the Genome's 'Dark Matter'

IT USED TO SEEM SO STRAIGHTFORWARD. DNA told the body how to build proteins. The instructions came in chapters called genes. Strands of DNA's chemical cousin RNA served as molecular messengers, carrying orders to the cells' protein factories and translating them into action. Between the genes lay long stretches of "junk DNA," incoherent, useless, and inert.

That was then. In fact, gene regulation has turned out to be a surprisingly complex process governed by various types of regulatory DNA, which may lie deep in the wilderness of supposed "junk." Far from being humble messengers, RNAs of all shapes and sizes are actually powerful players in how genomes operate. Finally, there's been increasing recognition of the widespread role of chemical alterations called epigenetic factors that can influence the genome across generations without changing the DNA sequence itself.

The scope of this "dark genome" became apparent in 2001, when the human genome was first published. Scientists expected to find as many as 100,000 genes packed into the 3 billion bases of human DNA; they were startled to learn that there were fewer than 35,000. (The current count is 21,000.) Protein-coding regions accounted for just 1.5% of the genome. Could the rest of our DNA really just be junk?

The deciphering of the mouse genome in 2002 showed that there must be more to the story. Mice and people turned out to share not only many genes but also vast stretches of noncoding DNA. To have been "conserved" throughout the 75 million years since the mouse and human lineages diverged, those regions were likely to be crucial to the organisms' survival.

Edward Rubin and Len Pennacchio of the Joint Genome Institute in Walnut Creek, California, and colleagues figured out that some of this conserved DNA helps regulate genes, sometimes from afar, by testing it for function in transgenic mouse embryos. Studies by the group and others suggested that noncoding regions were littered with much more regulatory DNA than expected.

Further evidence that noncoding DNA is vital has come from studies of genetic risk factors for disease. In large-scale searches for single-base differences between diseased and healthy individuals, about 40% of the disease-related differences show up outside of genes.

Genetic dark matter also loomed large when scientists surveyed exactly which DNA was being transcribed, or decoded, into RNA. Scientists thought that most RNA in a cell was messenger RNA generated by protein-coding genes, RNA in ribosomes, or a sprinkling of other RNA elsewhere. But surveys by Thomas Gingeras, now at Cold Spring Harbor Laboratory in New York, and Michael Snyder, now at Stanford University in Palo Alto, California, found a lot more RNA than expected, as did an analysis of mouse RNA by Yoshihide Hayashizaki of the RIKEN Omics Science Center in Japan and colleagues. Other researchers were skeptical, but confirmation soon came from Ewan Birney of the European Bioinformatics Institute and the Encyclopedia of DNA Elements project, which aims to determine the function of every base in the genome. The 2007 pilot results were eye-opening: Chromosomes harbored many previously unsuspected sites where various proteins bound—possible hotbeds of gene regulation or epigenetic effects. Strikingly, about 80% of the cell's DNA showed signs of being transcribed into RNA. What the RNA was doing was unclear.

Other studies revealed that RNA plays a major role in gene regulation and other cellular functions. The story started to unfold in the late 1990s, when plant researchers and nematode biologists learned to use small RNA molecules to shut down genes. Called RNA interference (RNAi), the technique has become a standard way to control gene activity in a variety of species, earning a Nobel Prize in 2006.

To understand RNAi and RNA in general, researchers began isolating and studying RNA molecules just 21 to 30 bases long. It turned out that such "small RNAs" can interfere with messenger RNA, destabilizing it. Four papers in 2002 showed that small RNAs also affect chromatin, the complex of proteins and DNA that makes up chromosomes, in ways that might further control gene activity. In one study, yeast missing certain small RNAs failed to divide properly. Other studies have linked these tiny pieces of RNA to cancer and to development.

The surprises didn't stop at small RNAs. In 2007, a group led by Howard Chang of Stanford and John Rinn, now at Beth Israel Deaconess Medical Center in Boston, pinned down a gene-regulating function by so-called large intervening noncoding RNAs. Rinn and colleagues later determined that the genome contained about 1600 of these lincRNAs. They and other researchers think this type of RNA will prove as important as protein-coding genes in cell function.

Many mysteries about the genome's dark matter are still under investigation. Even so, the overall picture is clear: 10 years ago, genes had the spotlight all to themselves. Now they have to share it with a large, and growing, ensemble.

–ELIZABETH PENNISI

CREDIT: Y. HAMMOND/*SCIENCE*

Published by AAAS