

## "Junk" DNA Holds Clues to Common Diseases

With the new annotation of the human genome, researchers are finding that **most of the code between genes is controlling crucial functions for life and health**

By Katherine Harmon

When the draft of the human genome was published in 2000, researchers thought that they had obtained the secret decoder ring for the human body. Armed with the code of 3 billion basepairs of As, Ts, Cs and Gs and the 21,000 protein-coding genes, they hoped to be able to find the genetic scaffolds of life—both in sickness and in health.

But in the 12 years since then, very few diseases—almost all of them very rare—have been linked definitively to changes in the genes themselves. And large, genome-wide studies searching for genetic underpinnings for more common diseases, such as lung cancer or autism, have pointed to the nether regions of the genome between the protein-producing genes—**areas that were often thought to contain “junk” DNA that was not part of the pantheon of known genes.**

An international consortium of hundreds of scientists has now deciphered a large portion of the strange language of this junk DNA and **found it to be not junk at all. Rather it contains important signals for regulating our genes, determining disease risk, height and many of the other complex aspects of human biology** that make each one of us different. The findings are described in 30 linked papers published online September 5 in *Nature* and other journals and described at the consortium's Web site. (*Scientific American* is part of Nature Publishing Group.)

Called the Encyclopedia of DNA Elements (ENCODE), the group is focused on understanding not just the elements of the genome but also how they work together. "The complexity of our biology resides not in the number of our genes but in the regulatory switches," Eric Green, director of the National Human Genome Research Institute and collaborator on the ENCODE project, said in a press briefing September 5.

Through more than 1,600 separate experiments, analysis of more than 140 cell types and a massive amount of data analysis, the **group found about 4 million of these so-called switches and can now assign functions to more than 80 percent of the entire genome. Compare that to the roughly 2 percent of the genome that is responsible for the protein-coding genes that researchers have been relying on to look for diseases and traits.** "The genome project was about establishing the set of letters that make up the blueprint," Green said. "When we finally put that blueprint together, we realized we could only really understand very little of it."

These newly catalogued switches not only activate and de-activate genes, but also control how much of each protein gets made and when. They are involved in epigenetic changes, such as DNA methylation, which has been implicated in cardiovascular disease and other conditions. The new data promise to improve our understanding of many common diseases that might have similar genetic underpinnings. Genome-wide association studies (GWAS) have continuously come up short in identifying specific genes for common diseases,

John Stamatoyannopoulos, associate professor of genome sciences at the University of Washington School of Medicine and ENCODE collaborator, said in the briefing.

"Frustratingly, about 95 percent of information from these studies has been pointing to regions of the genome that do not make proteins," he said. But, now with the ENCODE data, they can begin to decipher what genetic switches and functions might be common within and among these diseases. "We're now exploring previously hidden connections between diseases that may explain similar clinical [symptoms]," he noted.

It will most likely be some time before these new findings, which are freely available, are put to use in approved therapies. "The pharmaceutical industry has largely given up on the genome," Stamatoyannopoulos said. "And I think this is going to tremendously reinvigorate the utility of the genome." These additional genetic elements, however, are already in use for screening and testing for diseases such as breast cancer, prostate cancer and autoimmune diseases, Richard Myers, president of HudsonAlpha Institute for Biotechnology in Ala., noted in the briefing.

The group has funding to continue their efforts and does not anticipate a slowdown in discoveries going forward. "Our blueprint is remarkably complicated, and we need to be committed for the long haul to understand it," Green said. Compared with the publication of draft human genome 12 years ago—and with initial findings from the ENCODE project published over the past several years—"the questions that we can now ask are more sophisticated," Green said. And hopefully, those better questions will lead to more satisfying and medically useful answers.