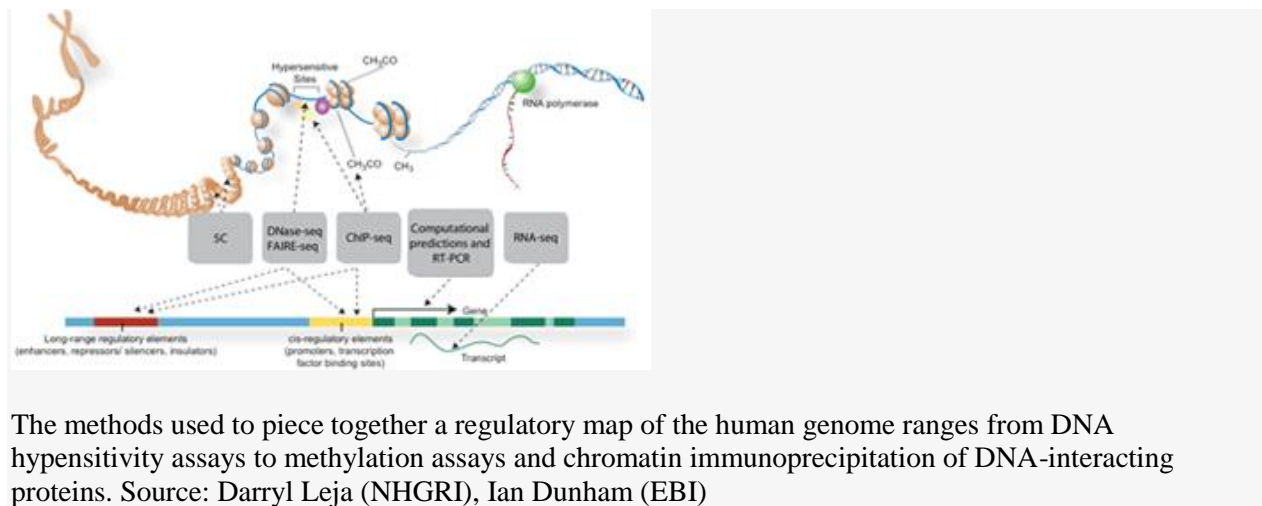


09/06/2012 (In *BioTechniques* – by Sarah C.P. Williams)

## ENCODE Finds 80% of the Genome is Functional

In 30 papers published simultaneously, the five-year ENCODE project reports the mapping of more than four million regulatory sites across the human genome.

In an effort that rivals the original human genome project in scale and scope, researchers from around the world have been collaborating for the past five years to understand the non-coding regions of the human genome—the more than 95% of the genome that’s been dubbed “junk DNA” in the past. Now, with the simultaneous publication of 30 papers describing their findings, the team has reported that more than 80% of the human genome does indeed have a function.



The project, called the Encyclopedia of DNA Elements (ENCODE), involved 440 scientists from 32 labs in the United States, United Kingdom, Spain, Singapore, and Japan. Since 2007, they have collected more than 15 terabytes of raw data that describes places in the genome that contain regulatory binding sites, areas of frequent DNA modification, or roles in managing the larger chromatin structure of DNA.

### SPECIAL EVENT ALERT



ENCODE researcher Piero Carninci from the RICKEN Omics Science Center will speak and answer your questions about his transcription analysis techniques at our free virtual symposium.

[Register today.](#)

“For basic researchers, the ENCODE data represent a powerful resource for exploring fundamental questions about how life is encoded in our genome,” said Eric Green of the National Human Genome Research Institute. “For more clinically oriented researchers, the ENCODE data provide key information about which genome sequences are functionally important.”

While determining the placement and function of regulatory sites in DNA has been done in individual regions before, the new map is the most complete picture, and provides a launching-off spot for future studies in almost every avenue of genetic research. The maps were created using a variety of techniques, including chromatin immunoprecipitation (ChIP) to locate binding sites for 119 transcription factors and histones as well as [chromatin conformation capture](#), methylation analysis assays, and RNAseq. These experiments were performed in 150 cell types from different organs and developmental stages to create a full picture of functionality.

“Because of the millions of these switches, only a small percentage of them are on in any given type of cell, and the pattern of switches is different for each kind of cell,” said John Stamatoyannopoulos of the University of Washington. “One has to survey a lot of different cells to gain a complete picture that can then be compared with a disease landscape.”

As a first step toward applying the new functional genome data to clinical relevance, a multi-institutional team led by Stamatoyannopoulos analyzed gene variants that have already been discovered in genome-wide association studies (GWAS) for their overlap with the newly mapped out functional areas. “Much of the outcome of these studies has not been taken advantage of, and their usefulness has not been fully realized,” said Stamatoyannopoulos.

To see how much of previously identified disease-associated variation is located within DNA regulatory elements, his team treated hundreds of cell types with the nuclease DNase1. Sites with high levels of cleavage by DNase1—called DNase I hypersensitive regions (DHS)—are known to contain DNA regulatory elements. From this data, the researcher determined the placement of these DHSs and then aligned them with more than 5,000 gene variants associated with 207 diseases and 447 traits identified in GWAS.

In a paper published September 5 in *Science*, the team reported that 76% of these disease-associated gene variants fell within DHSs. The next step is to home in on each variant and determine the exact function of the regulatory region it affects and how it may cause disease.

“We now know that a majority of these changes that are associated with common diseases and traits that don’t fall within genes actually occur within the gene controlling switches,” says Stamatoyannopoulos. “And this phenomenon is not confined to a particular type of disease. It seems to be present across the board for a very wide variety of different diseases and traits.”

## References

1. Maurano, MT, R. Humbert, and E. Rynes. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* Vol. 337: 1190-1196.