Encyclopedia of DNA Elements

ENCODE (**ENC**yclopedia **O**f **D**NA **E**lements) is a multi-year concerted effort by hundreds of researchers across more than 30 scientific institutes in the United States, the United Kingdom, Spain, Singapore and Japan. Its goal is to map functional elements within the human genome to the biological functions they regulate.

Started as a pilot project in 2003 and focusing only on 1 percent of the human genome, the ENCODE project was given its full-scale launch in 2007, when the National Human Genome Research Institute decided that the technology needed by ENCODE was sufficiently advanced to analyze the entirety of the human genome. Finally, in 2012, the project's results were published in more than 30 separate papers on *Nature, Genome Research* and *Genome Biology*. The results are available to the public and the scientific community through a variety of portals.

The ENCODE results, derived from experiments performed on 147 different types of tissues, paint an extremely detailed picture of the human genome and its regulatory elements. A 10-fold higher number of transcriptional start sites in comparison to the number of protein-coding genes; regulatory information distributed in a clustered manner across the genome; and the presence of many distal DNAse I hypersensitive sites, are only a few of the general results. Far from containing large regions of "junk DNA", according to ENCODE the human genome contains over 4 million genomic regions, both coding and non-coding, which interact with proteins to regulate DNA expression, and they span over 80% of the whole genome.

Such a large dataset will require extensive analysis and validation in order to pinpoint the specific biological function of each individual regulatory region, and its role in disease. Still, with an increasing number of disease-inducing mutations being mapped on non-coding regions of the genome, it is clear that the ENCODE results will be of critical importance to researchers seeking to unlock the functionality of the human genome.

Sites:

ENCODE project portal: www.encodeproject.org
University of California, Santa Cruz: http://genome.ucsc.edu/ENCODE/
National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/geo/info/ENCODE.html
European Bioinformatics Institute: http://www.ensembl.org/Homo_sapiens/encode.html4