

## GENOMICS

# The ENCODE project

**The second, genome-wide phase of the Encyclopedia of DNA Elements (ENCODE) project is being reported.**

The function of most of the human genome is unknown. Protein-coding genes account for only a small fraction (about 3%) of the total genome sequence; most functional genomic sequences are likely to have regulatory roles. Understanding human gene organization and regulation and their impact on normal and disease phenotypes requires that functional elements be mapped and annotated across the genome. This is the goal of the ENCODE project.

The initial 5-year pilot phase of the project focused on 1% of the human genome sequence. The second 5-year phase of ENCODE, which began in 2007 and is now coming to fruition, has extended the analysis of functional elements genome wide. A functional element as defined by ENCODE is a genomic sequence that either encodes

a particular product (for instance, a protein or noncoding RNA) or has a consistent biochemical property (for instance, being bound by protein or having a particular biochemical mark).

The laboratories in the ENCODE Project Consortium have developed and applied a huge range of sequencing-based techniques to map functional elements across the genome. To put it succinctly, the ENCODE project has mapped chromatin state and structure, three-dimensional genome organization, DNA methylation, transcription factor binding, RNA transcription and protein expression genome wide. Experiments were conducted in multiple cell types, with the highest priority given to widely studied cell lines but with the list also including a human embryonic stem cell line and, in some cases, primary cells.

It is striking that a large fraction (80%) of the genome overlaps with at least one

ENCODE-defined functional element in at least one examined cell type; an even larger fraction (99%) lies nearby such an element (within 1.7 kilobases). An examination of previously identified disease-associated single-nucleotide polymorphisms shows that they are enriched in ENCODE-annotated regions, suggesting hypotheses for functional consequences of single-nucleotide polymorphisms that can be further tested.

The data generated by ENCODE are vast and can be only very briefly summarized here. The collected ENCODE papers may be examined at <http://www.encodeproject.org/ENCODE/pubs.html> or explored with a dedicated visualization tool at <http://www.nature.com/ENCODE/>.

**Natalie de Souza**

**RESEARCH PAPERS**

The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).