



**Embargoed until Wednesday, 13 June 2007, 18:00 BST**

## New findings challenge established views about human genome

### ENCODE research consortium uncovers surprises related to organization and function of human genetic blueprint

**Hinxton, 14 June 2007** – The ENCyclopedia Of DNA Elements (ENCODE), an international research consortium organised by the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health (NIH), today published the results of its exhaustive, four-year effort to build a “parts list” of all biologically functional elements in 1 percent of the human genome in the journal *Nature*. The analysis was led by the European Molecular Biology Laboratory’s European Bioinformatics Institute (EMBL-EBI), drawing on expertise from 35 groups from 80 organizations around the world. The project served as a pilot to test the feasibility of a full-scale initiative to produce a comprehensive catalog of all components of the human genome crucial for biological function.

The findings promise to reshape our understanding of how the human genome functions. They challenge the traditional view of our genetic blueprint as a tidy collection of independent genes, pointing instead to a network in which genes, regulatory elements and other types of DNA sequences interact in complex, overlapping ways.

“By integrating 200 datasets generated by various high-throughput methods we now have a very good idea what 1 percent of our DNA might be doing. Our results reveal important principles about the organization of functional elements in the human genome, providing new perspectives on everything from DNA transcription to mammalian evolution. In particular, we gained significant insight into DNA sequences that do not encode proteins, which we knew very little about before,” said Ewan Birney, Ph.D., head of genome annotation at EMBL-EBI, who led ENCODE’s massive data integration and analysis effort.

The ENCODE consortium’s major findings include the discovery that the majority of human DNA is transcribed into RNA and that these transcripts extensively overlap one another. This broad pattern of transcription challenges the long-standing view that the human genome consists of a small set of discrete genes, along with a vast amount of “junk” DNA that is not biologically active.

The new data indicate that the genome contains little unused sequences; genes are just one of many types of DNA sequences that have a functional impact. The consortium identified many previously unrecognized start sites for transcription and regulatory sequences that contrary to traditional views are located not only upstream but also downstream of transcription start sites.

Other surprises in the ENCODE data have major implications for our understanding of the evolution of genomes. Until recently, researchers had thought that most DNA sequences with important biological function would be constrained by evolution making them likely to be conserved as species evolve. But about half of the functional elements in the human genome do not appear to have been constrained during evolution, suggesting that many species’ genomes contain a pool of functional elements that provide no specific benefits in terms of survival or reproduction.

“This impressive effort has uncovered many exciting surprises and blazed the way for future efforts to explore the functional landscape of the entire human genome,” said NHGRI Director Francis S. Collins, M.D., Ph.D. “Because of the hard work and keen insights of the ENCODE consortium, the scientific community will need to rethink some long-held views about what genes are and what they do, as well as how the genome’s functional elements have evolved. This could have significant implications for efforts to identify the DNA sequences involved in many human diseases.”

In addition to coordinating the analysis and integration of the ENCODE data, EMBL-EBI researchers in collaboration with the BioSapiens Network of Excellence (NoE) have investigated as part of the ENCODE effort how RNA transcripts are processed in human cells. In the 27 March issue of *PNAS* they reported that alternative splicing, the phenomenon that the same RNA transcript can be cut at two or more different positions to make different products, is very common in humans. It is unlikely, however, that alternative splicing adds substantially to the variety of functions and structures among proteins.

#### Source Article

The ENCODE project consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 14 June 2007

Tress, M.L. et al. The implications of alternative splicing in the ENCODE protein complement. *PNAS*, 27 March 2007

#### Contact:

Louisa Wright PhD, EMBL-EBI Scientific Outreach Officer, Hinxton, UK, Tel: +44 1223 494665, [www.ebi.ac.uk](http://www.ebi.ac.uk), [louisa@ebi.ac.uk](mailto:louisa@ebi.ac.uk)  
Anna-Lynn Wegener, EMBL Press Officer, Heidelberg, Germany, Tel: +49 6221 387 452, [www.embl.org](http://www.embl.org), [wegener@embl.de](mailto:wegener@embl.de)

Over the next couple of years the ENCODE project will be scaled up to the entire genome. The Ensembl project, a joint EMBL-EBI and Sanger Institute project, jointly headed by Ewan Birney, has already generated some initial genome wide datasets with early full scale datasets. This integration has led to the identification of just over 110,000 regulatory elements across the human genome. In parallel the BioSapiens NoE is creating a pipeline for the systematic annotation of the proteins potentially produced by alter-

native splicing throughout the human genome. "The collaboration with the ENCODE project holds great potential for new discoveries by the Biosapiens network" said Professor Janet Thornton, BioSapiens coordinator.

"The goal for the next five years is delivering a more complete understanding across our genome" said Birney, "the ENCODE pilot project is the first step towards this goal." ●

#### **About EMBL:**

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 19 member states (Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom). Research at EMBL is conducted by approximately 80 independent groups covering the spectrum of molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg, and Outstations in Hinxton (the European Bioinformatics Institute), Grenoble, Hamburg, and Monterotondo near Rome. The cornerstones of EMBL's mission are: to perform basic research in molecular biology; to train scientists, students and visitors at all levels; to offer vital services to scientists in the member states; to develop new instruments and methods in the life sciences and to actively engage in technology transfer activities. EMBL's International PhD Programme has a student body of about 170. The Laboratory also sponsors an active Science and Society programme. Visitors from the press and public are welcome.

#### **About EBI:**

The European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL) and is located on the Wellcome Trust Genome Campus in Hinxton near Cambridge (UK). The EBI grew out of EMBL's pioneering work in providing public biological databases to the research community. It hosts some of the world's most important collections of biological data, including DNA sequences (EMBL-Bank), protein sequences (UniProt), animal genomes (Ensembl), three-dimensional structures (the Macromolecular Structure Database), data from microarray experiments (ArrayExpress), protein-protein interactions (IntAct) and pathway information (Reactome). The EBI hosts several research groups and its scientists continually develop new tools for the bio-computing community.

#### **About the NHGRI:**

NHGRI is one of 27 institutes and centers at NIH, an agency of the Department of Health and Human Services. For more about NHGRI, visit [www.genome.gov](http://www.genome.gov). The National Institutes of Health – "The Nation's Medical Research Agency" – includes 27 institutes and centers, and is a component of the U.S. Department of Health and Human Services. It is the primary federal agency for conducting and supporting basic, clinical and translational medical research, and it investigates the causes, treatments, and cures for both common and rare diseases. For more, visit [www.nih.gov](http://www.nih.gov).

#### **About the ENCODE consortium:**

The ENCODE project, which stands for the Encyclopedia Of DNA Elements, was launched in September 2003 by the National Human Genome Research Institute (NHGRI). The project involves a public research consortium comprising of several research teams from the United States, Canada, Singapore, Spain and the United Kingdom. The aim of the project is to identify all functional elements in the human genome sequence. For more information on the ENCODE project, including a complete list of participants and the consortium's data release and accessibility policies, go to: [www.genome.gov/ENCODE](http://www.genome.gov/ENCODE).

#### **About the BioSapiens Network of Excellence:**

BioSapiens is a Network of Excellence, funded by the European Union's 6th Framework Programme, and made up of bioinformatics researchers from 25 institutions based in 14 countries throughout Europe. The objective of the BioSapiens is to provide a large scale, concerted effort to annotate genome data by laboratories distributed around Europe, using both informatics tools and input from experimentalists. For more information on BioSapiens, details of training opportunities and access to the annotation data, please visit [www.biosapiens.info](http://www.biosapiens.info).

#### **Policy regarding use**

*EMBL press releases may be freely reprinted and distributed via print and electronic media. Text, photographs & graphics are copyrighted by EMBL. They may be freely reprinted and distributed in conjunction with this news story, provided that proper attribution to authors, photographers and designers is made. High-resolution copies of the images can be downloaded from the EMBL web site: [www.embl.org](http://www.embl.org)*