
A User's Guide to the Encyclopedia of DNA Elements

Hinxton, 20 April 2011 - Today the international ENCODE (Encyclopedia Of DNA Elements) Project presented an overview of their ongoing, large-scale efforts to interpret the human genome sequence. Published in the journal *PLoS Biology*, "A User's Guide to the Encyclopedia of DNA Elements (ENCODE)" serves as a guide to help scientists interpret the vast array of data and resources produced by the project. All of the data, tools to study them and the paper itself are freely available through the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) in the UK.

The Human Genome Project and subsequent large-scale genomic efforts have been carried out with the belief that the data they produce should be made freely available to the scientific community to facilitate discovery. The ENCODE database (genome.ucsc.edu/ENCODE) makes its genomic data and related information both available and accessible, offering web-based tools (encodeproject.org) that make it easier for researchers to use the data.

The User's Guide shows how the data can be immediately useful in interpreting associations between single nucleotides and disease. For example, DNA variants upstream of the *c-Myc* proto-oncogene are known to be associated with multiple cancers, but until recently the mechanism behind this association had not been determined. ENCODE data show that the variants can change the binding of transcription factor proteins to an enhancer region, which leads to changes

in expression of the *c-Myc* gene and therefore to the onset of cancer. Similar studies are now possible for the thousands of variants identified in genome-wide association studies. This will go a long way in addressing mechanistic questions of susceptibility to a wide range of human diseases.

Ewan Birney, Senior Team Leader at EMBL-EBI, commented "We knew four years ago, from our publication of ENCODE techniques on 1% of the genome, that we had an unprecedented view of how biology works on those regions. By extending our work to the entire genome, we see the immediate impact on the interpretation of noncoding variants identified in genome-wide association studies. These studies are disease-driven but have not always yielded clear next steps; ENCODE data can open up new paths to follow."

Scientists with the ENCODE Project are applying up to 20 different tests in 108 commonly used cell lines to compile these important data. "Assays that are now fundamental to biology, such as chromatin immunoprecipitation and sequencing or ChIP-seq, were produced by the ENCODE Project," commented John Stamatoyannopoulos of the University of Washington School of Medicine in the US. "Widely used computational tools for processing and interpretation of large-scale functional genomic data have also been developed by the project. The depth, quality, and diversity of the ENCODE data are unprecedented." ●

Source Article

The ENCODE Project Consortium 2011 A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol* 9(4): e1001046. DOI:10.1371/journal.pbio.1001046.

Contact:

Sonia Furtado, EMBL Press Officer, Heidelberg, Germany, Tel: +49 6221 387 8263, www.embl.org, sonia.furtado@embl.de
Mary Todd-Bergman, Outreach Programme Project Leader, Hinxton, UK, Tel: +44 1223 494665, www.ebi.ac.uk, mary@ebi.ac.uk
Katrina Pavelin, EMBL-EBI Scientific Outreach Officer - Hinxton, UK, Tel: +44 1223 494452, www.ebi.ac.uk, katrina@ebi.ac.uk
contactpress@ebi.ac.uk

About EMBL

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 20 member states (Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom) and associate member state Australia. Research at EMBL is conducted by approximately 85 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. Around 190 students are enrolled in EMBL's International PhD programme. Additionally, the Laboratory offers a platform for dialogue with the general public through various science communication activities such as lecture series, visitor programmes and the dissemination of scientific achievements.

About EMBL-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 Protein Databank in Europe), data from gene expression 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 biocomputing community.

Policy regarding use

EMBL press and picture releases including photographs, graphics, movies and videos are copyrighted by EMBL. They may be freely reprinted and distributed for non-commercial use via print, broadcast and electronic media, 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.