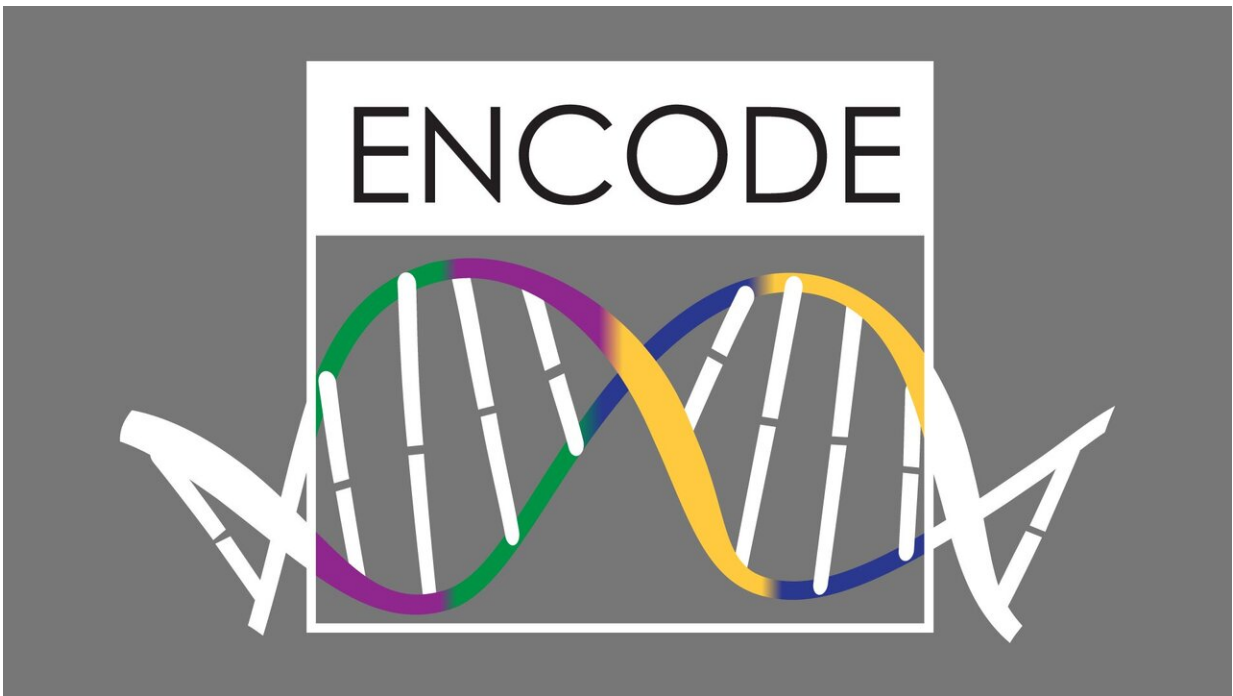


ENCODE3: Interpreting the human and mouse genomes

July 29 2020



Logo for the ongoing ENCODE project (Encyclopedia of DNA Elements).
Credit: ENCODE

Scientists around the world have access to a rich trove of information through the Encyclopedia of DNA Elements (ENCODE)—annotated versions of the human and mouse genomes that are vital for interpreting their genetic codes. In the July 29, 2020 issue of the journal *Nature*, an international consortium of approximately 500 scientists reports on the

completion of Phase 3 of an ongoing project, an achievement 20 years in the making that will help reveal how genetic variation shapes human health and disease.

Funded by the National Human Genome Research Institute, ENCODE launched in 2003, soon after the [human genome](#) was first sequenced. Its researchers are developing a comprehensive catalog of the human and mouse genomes' functional elements—dense arrays of protein-coding genes, non-coding genes, and [regulatory elements](#). Thousands of researchers worldwide have taken advantage of ENCODE data, using it to shed light on cancer biology, cardiovascular disease, human genetics, and other topics.

"When the first draft of the human genome was completed... it became immediately clear that while we had the primary sequence of the genome, or we had a draft of it... we needed to have an annotation for the genome," says Cold Spring Harbor Laboratory Professor Thomas Gingeras, whose team has been contributing to the ENCODE project since its inception. "We knew where the genes were located. Where the regulatory mechanisms and loci were located was significantly underdeveloped."

In Phase 3, researchers took advantage of the latest genetic technologies to glean data from biological specimens and deeply investigate the regulatory regions outside of [genes](#), where most of the genome's person-to-person variation lies. Their data identifies some 900,000 candidate regulatory elements from the human genome and more than 300,000 from the mouse, which can be explored through ENCODE's new online browser.

Gingeras's team is investigating [genome](#) elements that instruct cells about how and when to transcribe DNA sequences into RNA. In a companion publication to the ENCODE report, a team led by Gingeras

and collaborator Roderic Guigó at the Centre for Genomic Regulation detail work identifying molecular fingerprints that can be used to identify five groups of human cells. "Our work redefines, based on gene expression, the basic histological types in which tissues have been traditionally classified," Guigó says.

Those findings are now available through the ENCODE database. Meanwhile, the project has begun its fourth phase, employing new technologies and investigating additional cell types. Gingeras notes:

"This encyclopedia is a living resource. It has a beginning but really no end. It will continue to be improved, and grown, as time goes on."

More information: Expanded encyclopaedias of DNA elements in the human and mouse genomes, *Nature* (2020). [DOI: 10.1038/s41586-020-2493-4](https://doi.org/10.1038/s41586-020-2493-4) , www.nature.com/articles/s41586-020-2493-4

Provided by Cold Spring Harbor Laboratory

Citation: ENCODE3: Interpreting the human and mouse genomes (2020, July 29) retrieved 24 June 2024 from <https://phys.org/news/2020-07-encode3-human-mouse-genomes.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.