

# Human genome far more active than thought

September 6 2012

---

The GENCODE Consortium expects the human genome has twice as many genes than previously thought, many of which might have a role in cellular control and could be important in human disease. This remarkable discovery comes from the GENCODE Consortium, which has done a painstaking and skilled review of available data on gene activity.

Among their discoveries, the team describe more than 10,000 novel genes, identify genes that have 'died' and others that are being resurrected. The GENCODE Consortium reference gene catalogue has been one of the underpinnings of the larger ENCODE Project and will be essential for the full understanding of the role of our genes in disease.

The GENCODE [Consortium](#) is part of the ENCODE Project that, today, publishes 30 research papers describing findings from their nearly decade-long effort to describe comprehensively all the active regions of our [human genome](#). ENCODE was launched in 2003 after the completion of the [Human Genome Project](#), and brought together an international group of scientists tasked with identifying and describing all functional regions of the [human genome sequence](#).

"We have uncovered a staggering array of genes in our genome, simply because we can examine many genomes in a detail that was not possible a decade ago," says Dr Jennifer Harrow, GENCODE principle investigator from the Wellcome Sanger Institute. "As sequencing technology improves, so we have much more data to explore."

"But our work remains a skilled effort to annotate correctly our human genome - or, more precisely, our human genomes, for each of us differ. These vast texts of [genetic information](#) will not give up their secrets easily. GENCODE has made amazing strides to enable immediate access of its reference gene set by other researchers."

The team more accurately described the genes that contain the genetic code to make proteins: they found 20,687 such protein-coding genes, a value that has not changed greatly from previous work. The new set captures far more of the alternative forms of these genes found in different cell types.

More significant are their findings on genes that do not contain [genetic code](#) to make proteins - non-coding genes - and the graveyard of supposedly 'dead' genes from which some are emerging, resurrected from the catalogue of pseudogenes.

They mapped and described 9,277 long non-coding genes, a relatively new type that acts, not through producing a protein, but directly through its RNA messenger. Long non-coding RNAs derived from these genes can play a significant part in human biology and disease, but they remain only poorly understood.

The new map of such genetic components gives researchers more avenues to explore in their quest to understand human biology and human disease. Remarkably, the team think their job is not complete and believe that there may be another 10,000 of these genes yet to be uncovered.

"Our initial work from the Human Genome Project suggested there were around 20,000 protein-coding genes and that value has not changed greatly," says Professor Roderic Guigo, GENCODE principle investigator from Centre for Genomic Regulation, Barcelona. "However

GENCODE has shown that long non-coding RNAs are far more numerous and important than previously thought"

"The limited knowledge we have of the class of long non-coding RNAs suggests they might play a major role in regulating the activity of other genes. If this is generally true of this group, we have much more to explore than we imagined."

As dramatic, GENCODE has catalogued for the first time a set of more than 11,000 pseudogenes by examining the entire human genome. There is some emerging evidence that many of these genes, too, might have some biological activity.

The GENCODE team predict that at least 9% of pseudogenes may be active with some controlling the activity of other genes. Pseudogenes have been implicated in many biological activities, such as the prevention of certain elements known to be involved in the development of cancer.

"At the announcement of the Human Genome Project draft sequence, we emphasized this was the end of the beginning, that 'at present most genes - probably tens of thousands - remain a mystery'", says Dr Tim Hubbard, lead principle investigator of GENCODE from the Wellcome Trust Sanger Institute. "Today, we describe many thousands of [genes](#) for the first time."

"If the Human [Genome Project](#) was the baseline for genetics, ENCODE is the baseline for biology, and GENCODE are the parts that make the human biological machine work. Our list is essential to all those who would fix the human machine."

The GENCODE human reference set will be updated every three months to ensure that models are continually refined and assessed based

on new experimental data deposited in the public databases.

**More information:** Publication details can be found at  
[www.genome.gov/10005107](http://www.genome.gov/10005107)

Provided by Wellcome Trust Sanger Institute

Citation: Human genome far more active than thought (2012, September 6) retrieved 19 April 2024 from <https://medicalxpress.com/news/2012-09-human-genome-thought.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.