

# Uncharted genetic territory offers insight into human-specific proteins

July 13 2022

---



Uncharted territories in the human genome. Credit: Karen Arnott/EMBL-EBI

When researchers working on the Human Genome Project completely mapped the genetic blueprint of humans in 2001, they were surprised to find only around 20,000 genes that produce proteins. Could it be that humans have only about twice as many genes as a common fly? Scientists had expected considerably more.

Now, researchers from 20 institutions worldwide bring together more than 7,200 unrecognized gene segments that potentially code for new proteins. For the first time, the study makes use of a new technology to find possible proteins in humans—looking in detail at the protein-producing machinery in cells. The new study suggests the gene discovery efforts of the Human Genome Project were just the beginning, and the [research consortium](#) aims to encourage the scientific community to integrate the data into the major human genome databases.

The study recently published in *Nature Biotechnology*, was co-led by Dr. Jorge Ruiz-Orera from Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC) in Germany, Dr. Sebastiaan van Heesch from the Princess Máxima Center for [pediatric oncology](#) in the Netherlands, Dr. Jonathan Mudge from the European Molecular Biology Laboratory—European Bioinformatics Institute (EMBL-EBI) in the United Kingdom, and Dr. John Prensner from the Broad Institute of MIT and Harvard in the United States.

## **New gene sequences remained out of reach**

In the past few years, thousands of frequently very small open reading frames (ORFs) have been discovered in the human genome. These are spans of DNA sequence that may contain instructions for building proteins. Several authors of the current study have previously found ORFs and described them in [scientific journals](#): Van Heesch, together with MDC-Professors Norbert Hübner and Uwe Ohler described [new mini-proteins](#) in the [human heart](#) and reported on them in *Cell* in 2019; Prensner also published on ORFs in *Nature Biotechnology* in 2021. Yet none of these previously virtually unexplored segments were included afterwards in reference databases. Other sequences were reported in journals such as *Science* or *Nature Chemical Biology*, but remained largely out of reach for most members of the scientific community—despite evidence that they produce RNA molecules that

subsequently bind to ribosomes, the cell's protein factories.

Traditionally, protein-coding regions in genes have been identified by comparing DNA sequences from multiple species: the most important coding regions have been preserved during animal evolution. But this method has a drawback: coding regions that are relatively young, i.e., that arose during the evolution of primates, fall through the cracks and are therefore missing from the databases.

So now the task is to integrate the largely ignored ORFs into the largest reference databases, because researchers have so far had to specifically search for them in the literature if they wanted to study them.

As a first step, the international research team collected information on sequences that had been discovered using ribosome profiling—a technique that determines which part of the messenger RNA (mRNA) the ribosome interacts with. They then assembled the data into a standardized catalog. This was no small feat, as data obtained in a wide variety of ways from different laboratories cannot simply be combined.

Once this was accomplished, the international consortium labored over central questions that define our very notion of the human genome: What is a gene? What is a protein? Do we need flexible notions of whether ribosomes always produce a [protein](#) or rather some other cellular output?

The group now calls for the human genome databases used by scientists worldwide to be revised. Ensembl-GENCODE are configuring this ORF catalog as a component of their reference annotation database. The approach will be supported by many others like UniProt, HGNC, PeptideAtlas and HUPO.

## **ORFs likely play a role in common diseases**

Dr. Sebastiaan van Heesch, group leader at the Princess Máxima Center for pediatric oncology, says that their "research marks a huge step forward in understanding the genetic make-up and complete number of proteins in humans. It's tremendously exciting to enable the research community with our new catalog. It's too soon to say whether all of the unexplored sections of DNA truly represent proteins, but we can clearly see that something unexplored is happening across the human genome and that the world should be paying attention."

"For too long, the [scientific community](#) has been mostly left in the dark about these ORFs," says Jonathan Mudge of the EMBL-EBI. "We're very proud that our work will be able to let researchers across the world start to study them. This is the point at which they enter the mainstream of genomic and medical science—an effort which we expect to have wide-ranging ripple effects."

"It is especially remarkable that most of these 7,200 ORFs are exclusive to primates and might represent evolutionary innovations unique to our species," reports Jorge Ruiz-Orera, an [evolutionary biologist](#) working in Hübner's lab at the MDC. "This shows how these elements can provide important hints of what makes us humans."

So, what's next? John Prensner, Broad Institute of MIT and Harvard, says that "these ORFs almost certainly will be contributing factors to many human traits and diseases, both rare diseases and common ones such as cancer. The challenge is now to figure out which ones have which roles in which diseases."

**More information:** A community-driven roadmap to advance research on translated open reading frames, *Nature Biotechnology* (2022). [DOI: 10.1038/s41587-022-01369-0](https://doi.org/10.1038/s41587-022-01369-0)

Provided by Max Delbrück Center for Molecular Medicine

Citation: Uncharted genetic territory offers insight into human-specific proteins (2022, July 13)  
retrieved 23 June 2024 from <https://phys.org/news/2022-07-uncharted-genetic-territory-insight-human-specific.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.