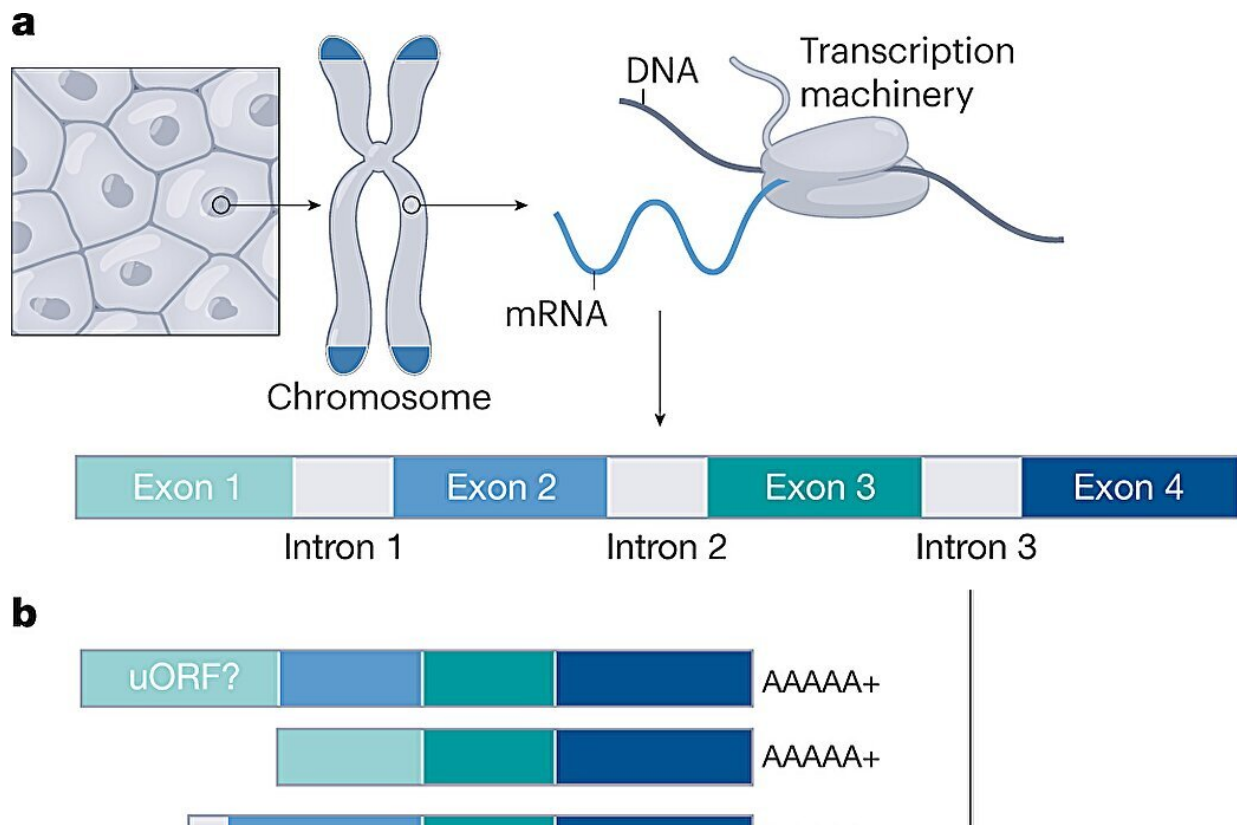


Yes, scientists have sequenced the entire human genome, but they're not done yet

October 13 2023



A major challenge for gene annotation is how to capture the diversity of gene products deriving from each gene locus. Credit: *Nature* (2023). DOI: 10.1038/s41586-023-06490-x

The human genome, from end to end, has been sequenced, meaning scientists worldwide have identified most of the nearly 20,000 protein-

coding genes. However, an international group of scientists notes there's more work to be done. The scientists point out that even though we have nearly converged on the identities of the 20,000 genes, the genes can be cut and spliced to create approximately 100,000 proteins, and gene experts are far from agreement on what those 100,000 proteins are.

The group, which convened last fall at Cold Spring Harbor Laboratory in New York, has now published a guide for prioritizing the next steps in the effort to complete the human gene "catalog."

"Many scientists have been working on efforts to fully understand the [human genome](#), and it's much more difficult and complex than we thought," says Steven Salzberg, Ph.D., Bloomberg Distinguished Professor of Biomedical Engineering, Computer Science, and Biostatistics at The Johns Hopkins University. "We have provided a state of the human gene catalog and a guide on what's needed to complete it."

Salzberg, along with Johns Hopkins [biomedical engineer](#) and associate professor Mihaela Pertea, Ph.D., M.S., M.S.E., postdoctoral researcher Ales Varabyou and 19 other scientists, offered perspectives on the human gene catalog Oct. 4 in the journal [Nature](#).

The scientists say that while the final list of protein coding genes is nearly complete, scientists have not yet fully cataloged the variety of ways that a gene can be cut, or spliced, resulting in "isoforms" of proteins that are slightly different. Some protein isoforms will not affect the protein's function but some may be different enough to result in increased risk for a particular trait, condition or illness.

To complete the catalog, the scientists propose a comprehensive look at how each gene is expressed into functional and nonfunctional proteins and the three-dimensional shape of those proteins.

The scientists also propose a focus on cataloging non-coding RNA genes. RNA is the [genetic material](#) that is transcribed by DNA and follows a molecular path to making proteins. Instead of proteins, non-coding RNA [genes](#) encode other types of molecular material that performs a cellular function.

Finally, the international group emphasizes the importance of enhancing commonly used databases of gene variations that cause illness and disease, improving clinical laboratory standards for annotating DNA sequencing results and developing new technology to enable more effective and precise methods to match the wide array of proteins with their [gene products](#).

More information: Paulo Amaral et al, The status of the human gene catalogue, *Nature* (2023). [DOI: 10.1038/s41586-023-06490-x](https://doi.org/10.1038/s41586-023-06490-x)

Provided by Johns Hopkins University School of Medicine

Citation: Yes, scientists have sequenced the entire human genome, but they're not done yet (2023, October 13) retrieved 28 April 2024 from <https://phys.org/news/2023-10-scientists-sequenced-entire-human-genome.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.