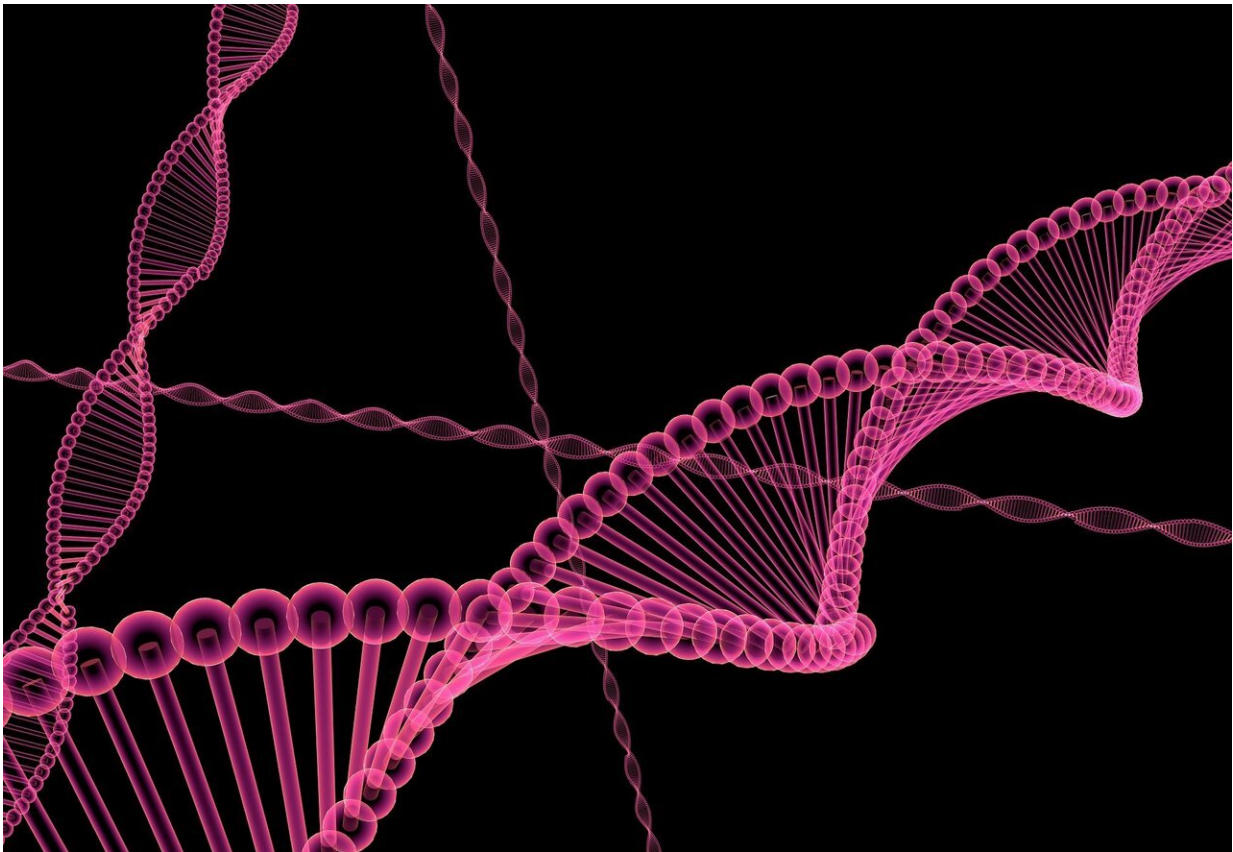


Researchers discover one million new components of the human genome

February 9 2024, by Anika Hazra



Credit: CC0 Public Domain

Researchers at the University of Toronto's Donnelly Center for Cellular and Biomolecular Research have found nearly one million new exons—stretches of DNA that are expressed in mature RNA—in the

human genome.

[The findings](#) were published in the journal *Genome Research*.

There are around 20,000 [protein-coding genes](#) in humans that contain approximately 180,000 known internal exons. These protein-coding regions account for only one percent of the entire [human genome](#). The vast majority of what remains is a mystery—aptly referred to as the "dark genome."

"We've started to chip away at the dark genome by finding nearly one million previously unknown exons through a method called exon trapping," said Timothy Hughes, principal investigator on the study and professor and chair of the department of molecular genetics in U of T's Temerty Faculty of Medicine.

"The technique involves an assay with plasmids to find exons in DNA fragments of unknown composition," said Hughes, who holds the Canada Research Chair in decoding [gene regulation](#) and the John W. Billes Chair of Medical Research at U of T. "While exon trapping is not widely used anymore, it proved to be effective when used in combination with high-throughput sequencing to scan the entire human genome."

Exons are segments of the genome that can encode proteins to direct tissue development and biological processes within the body. They are considered to be autonomous if they don't require external assistance to splice into a mature RNA transcript, which is then translated into a protein.

The team behind the study was driven to test the exon definition model that guides research in molecular genetics after questioning one of its assumptions—that the accurate removal of non-protein-coding intron

regions of the genome is aided by clear and consistent indicators of where exons begin and end. This assumption does not seem to hold in all cases as the splicing of exons does not always go smoothly, sometimes resulting in mature RNA transcripts that contain non-functional components.

"Almost none of the newly discovered exons are found consistently across genomes of different species," said Hughes. "They seem to appear in the human genome mainly due to random mutation and are unlikely to play a significant role in our biology. This is evidence that evolution in humans involves a lot of trial and error—most likely enabled by the vast size of our genome."

It is helpful to document randomly mutated exons within the human genome as their translation could potentially be harmful. Long non-coding RNA exons, which are autonomous but often have no known function, have been connected to the development of cancer. Of the roughly 1.25 million known and unknown exons the team found through exon trapping, almost four percent were long non-coding RNA exons.

In addition, the exons residing within non-coding introns, called pseudoexons, can mutate to make a weak splice site stronger. This results in the exon being included in a mature RNA transcript, potentially leading to disease.

"This is an interesting study that broadens our knowledge of sequences across the human genome that have the potential to be recognized as exons in transcribed RNA," said Benjamin Blencowe, professor of [molecular genetics](#) in U of T's Temerty Faculty of Medicine, who was not involved in the study.

"While the significance of the majority of the newly detected exons is unclear, some of them may be activated in certain contexts—for

example, by disease mutations—and therefore cataloging them is important. This study will further serve as a valuable resource facilitating ongoing efforts directed at deciphering the splicing code."

A stronger understanding of the factors impacting exon inclusion in mature RNA can help improve programs like SpliceAI, a widely used tool for predicting splice sites and aberrant splicing. SpliceAI can be trained on new data such as that produced through this study to refine its prediction capabilities.

"SpliceAI often doesn't provide details on the characteristics of exons and has a poor ability to predict splicing in exons that aren't already catalogued," said Hughes.

"Our [exon](#) trapping data contains biologically meaningful information that can be fed into SpliceAI and other splicing predictors to open up new paths for exploring the dark genome."

More information: Nicholas Stepankiw et al, The human genome contains over a million autonomous exons, *Genome Research* (2023). [DOI: 10.1101/gr.277792.123](https://doi.org/10.1101/gr.277792.123)

Provided by University of Toronto

Citation: Researchers discover one million new components of the human genome (2024, February 9) retrieved 27 April 2024 from <https://medicalxpress.com/news/2024-02-million-components-human-genome.html>

<p>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.</p>
--