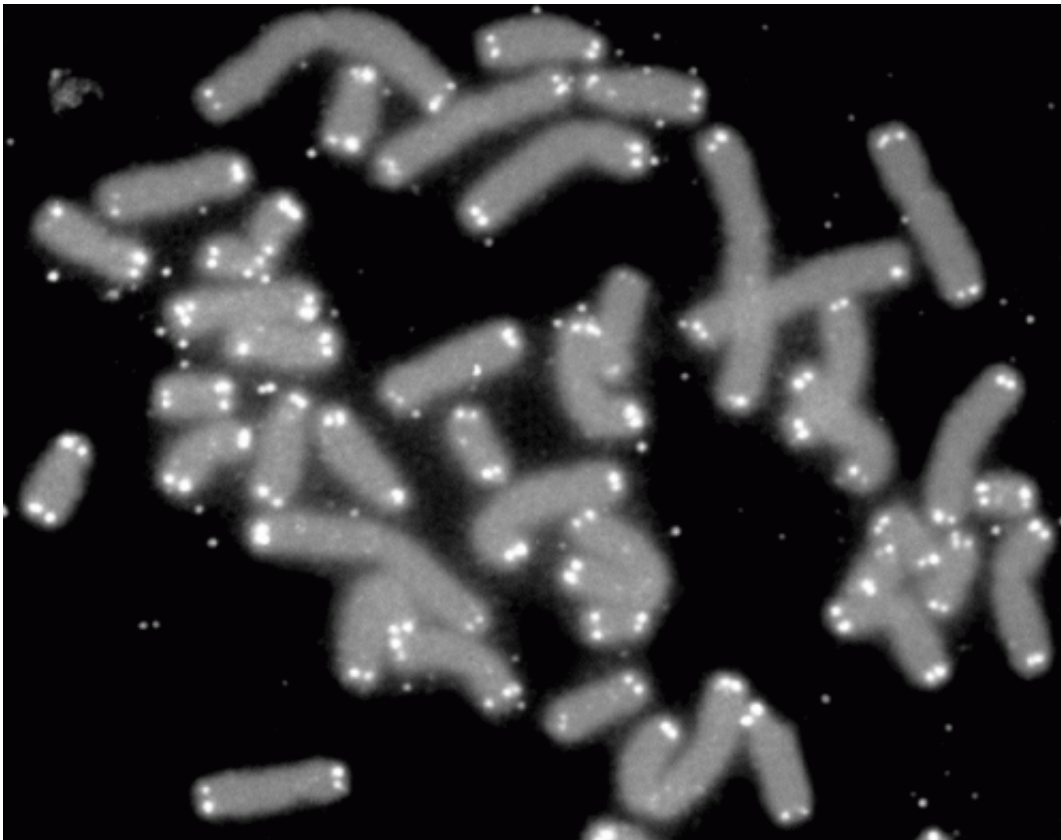


Repeats are key to understanding humanity's genome

March 31 2022



Human chromosomes (grey) capped by telomeres (white). Credit: PD-NASA; PD-USGOV-NASA

It was like a map of New York missing all of Manhattan. The human reference genome finally has all its blank spots filled in, and seeing everything we missed the first time around is both repetitive—and

enlightening.

"We're realizing there's a lot of human variation out there," says geneticist Rachel O'Neill, director of UConn's Institute for Systems Genomics. And in a contra-intuitive twist of fate, the variation comes in large part from the repeats.

A significant amount of human genetic material turns out to be long, repetitive sections that occur over and over. Although every human has some repeats, not everyone has the same number of them. And the difference in the number of repeats is where most of [human genetic variation](#) is found.

This insight—that the repeats are important—is one of many significant findings from the Telomere-2-Telomere (T2T) project, a globe spanning collaboration of institutions that filled in the missing sections of the original human genome assembly. O'Neill is a principal investigator on that project, and an author on four of the six T2T papers being published in *Science* on March 31.

"It took the invention of new methods of DNA sequencing and computational analysis, and the dedication of a remarkable team of scientists, to complete the reading of the 8% of the human genome that was too complex and repetitive in its structure to be resolved 20 years ago. It was worth the wait—a rich array of surprising architectural features is revealed, with major consequences for understanding [human evolution](#), variation, and [biological function](#)," says Francis Collins, White House Science Advisor and former director of the National Institutes of Health.

As amazing as it was at the time, the original Human Genome Project left about 8% of the genome blank.

"That's the equivalent of an entire chromosome in human DNA," O'Neill says. That last 8% includes numerous genes and repetitive areas. Most of the newly added DNA sequences were near the repetitive telomeres (long, trailing ends of each chromosome) and centromeres (dense middle sections of each chromosome).

The blanks were a result of the "short-read" technology the Human Genome Project used. It was the only technology for genome mapping available 20 years ago, and it could only read the equivalent of a few words of the genetic code at a time. For example, imagine a section of the genome consisted of the sentence "All work and no play makes Jack a dull boy," repeated nine times in a row. The short read technology would reveal only pieces of it such as "All work", "Jack a", "makes Jack", et cetera. The researchers pieced those short sections together to make the sentence "All work and no play makes Jack a dull boy," but they had no way to know it was repeated nine times.

The T2T project, however, had better tools. The new long-read DNA technology can read entire sentences, even paragraphs, at a time. So the researchers were able to see large chunks, or even full sections, of repeats.

"Generating a truly gapless sequence of the [human genome](#) is a major milestone. We would have loved to have done this 20 years ago, but the technology had to advance. This new reference is a truly solid foundation, without cracks, on which to understand human biology. There are no missing pieces!" says Bob Waterston, a biologist at the University of Washington who worked on the original Human Genome Project.

Many early-career researchers and trainees played pivotal roles in the T2T project. At UConn, Savannah Hoyt, Gabrielle Hartley, and Patrick Grady in Rachel O'Neill's lab, and Luke Wojenski in Leighton Core's

lab, were deeply involved in the work. One of their major contributions was developing a compendium of the repeats in the genome. They found the repetitive sections contained mobile elements, which are sections capable of jumping from one part of the genome to the other (the classic example is jumping genes that lead to color changes in corn kernels, such as from red to white); viruses; and new repeats no one had identified before, including some that carry genes. Some of the giant repeats with 10, 20, or 30 copies repeat back-to-back and contain genes that could account for much of human diversity. In the example sentence from earlier, imagine "Jack" is a gene. One person might have 5 copies of it. Another might have 25.

The T2T team got the first look at complete sequences for the central parts of every human chromosome. Called centromeres, they join the different arms of each X-shaped chromosome together. O'Neill's team found that centromeres contain known mobile elements as well as new repeats. Much of the DNA in the centromere seems to be important for maintaining a cell's genetic information through the generations. Centromeres are already known to play a role in DNA replication when cells reproduce, and if they shift their position in the chromosome significantly they can give rise to entirely new species. The complete, gapless centromere sequences constructed by the T2T project will allow a much more nuanced understanding of human centromeres and what they do. The next steps in human genetics research can use the complete T2T [genome](#) assembly as a jumping off point to identify interesting areas of our DNA.

"The next phase of research will sequence the genomes of many different people to fully grasp human diversity, diseases and our relationship to our closest relatives, the other primates," O'Neill says.

More information: Savannah J. Hoyt et al, From telomere to telomere: the transcriptional and epigenetic state of human repeat

elements, *Science* (2022). [DOI: 10.1126/science.abk3112](https://doi.org/10.1126/science.abk3112).
www.science.org/doi/10.1126/science.abk3112

Provided by University of Connecticut

Citation: Repeats are key to understanding humanity's genome (2022, March 31) retrieved 25 April 2024 from <https://phys.org/news/2022-03-key-humanity-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.