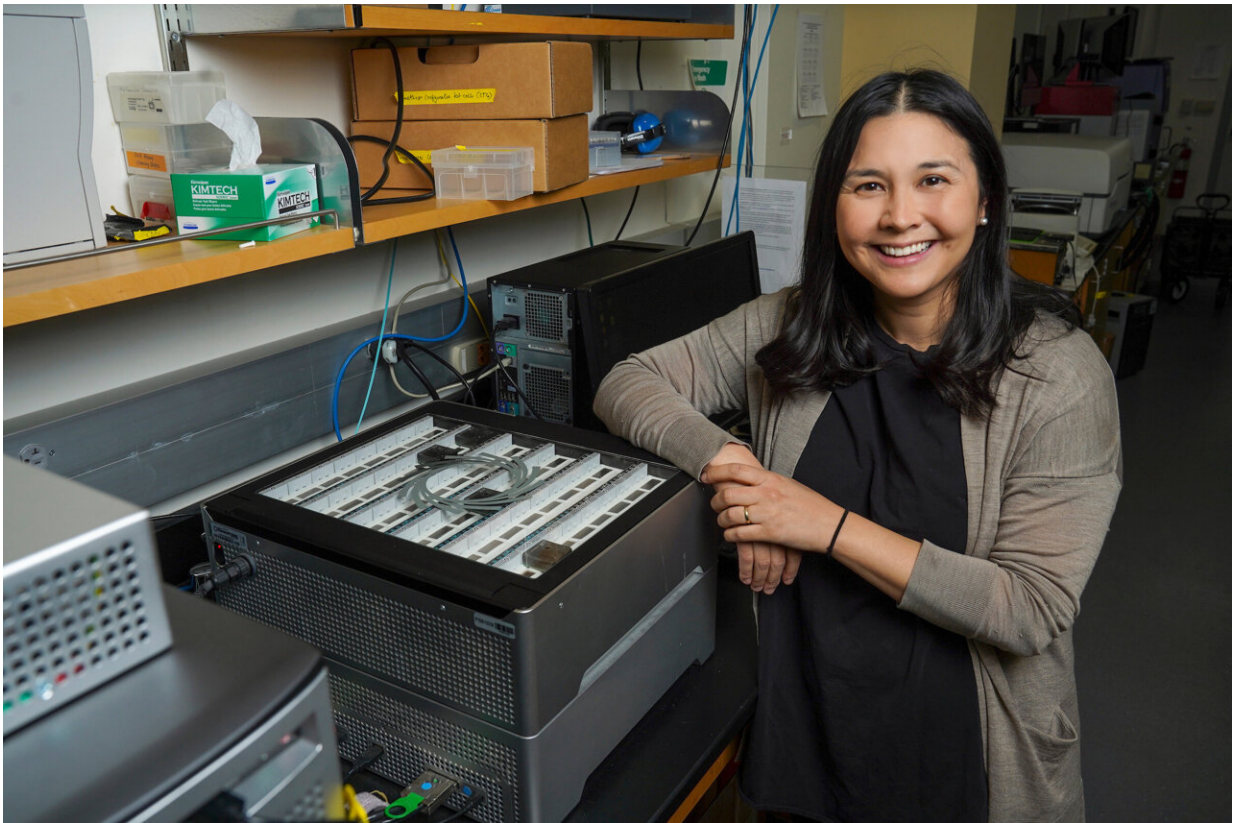


# New human reference genome opens unexplored regions

March 31 2022

---



Megan Dennis, assistant professor of biochemistry and molecular medicine at the UC Davis School of Medicine and MIND Institute, collaborated on studies completing sequencing of the human genome. Credit: Karin Higgins/UC Davis

A complete sequence of the human genome has finally been published

by an international consortium of scientists. The new reference genome fills in many gaps that were left by earlier drafts, especially in the centromeres at the middle of chromosomes and large-scale duplications.

The work is described in a series of papers published April 1 in *Science* by the Telomere-to-Telomere (T2T) Consortium. A number of University of California, Davis investigators contributed to the studies, including Megan Dennis, assistant professor of biochemistry and [molecular medicine](#) at the UC Davis School of Medicine and MIND Institute, with Integrative Genetics and Genomics graduate students Daniela Soto and Colin Shew, as well as Charles Langley, distinguished professor of evolution and ecology at the UC Davis College of Biological Sciences along with his daughter Sasha Langley, a project scientist at UC Berkeley.

The original human [genome](#) sequence, published in 2001, left out about eight percent of the DNA, Dennis said. The areas left out included nearly-identical duplications containing functional genes as well as centromeres and telomeres in the middle and at the tip of chromosomes respectively. These areas contain long runs of repeated sequences.

"These are important regions but difficult to sequence," Dennis said.

Sequencing a genome is rather like slicing up a book into snippets of text then trying to reconstruct the book by piecing them together again. Stretches of text that contain a lot of common or repeated words and phrases would be harder to put in their correct place than more unique pieces of text.

Earlier DNA sequencing technology could only read relatively short runs of sequence.

"A major leap in technology has been long-read sequencing," Dennis

said. Newer generation sequencers can decode much longer pieces, as much as a million base-pairs or "letters" of DNA. That means the chunks are much larger and easier to assemble back into the original sequence.

"It's a game changer," Dennis said.

UC Davis researchers contributed to the project by carrying out some of the long-read sequencing with machines at the Genome Center, and by analyzing variants and duplicated sequences.

The new reference genome comes from a single human sample, although not exactly a person. The DNA came from a cell line derived from a bundle of cells called a hydatidiform mole. These form when an egg in the uterus loses its own genome but gets fertilized by a sperm. The resulting cell ends up with two identical copies of each chromosome, unlike most human cells, which carry two slightly different copies. Despite its odd origin, there's nothing to suggest anything out of the ordinary with the cell line's genome, Dennis said.

The sperm came from a person of European descent. In contrast, the original human reference genome was stitched together from several people, creating some errors and artifacts.

## **Exploring the centromere**

About 90 percent of the new sequence actually comes from the centromeres of chromosomes, Langley said. Structurally distinct and containing long stretches of repetitive DNA, these regions are notoriously difficult to study.

"We used to say that you would warn young geneticists not to venture into the centromere because you'll never get out," Langley said.

But these days centromeres are a hot topic in biology. This is where the machinery that separates paired chromosomes during meiosis—formation of sperm and eggs—attaches, a fundamental step in inheritance. It contains large amounts of heterochromatin, or areas where DNA and proteins seem to be more condensed and compact.

Geneticists have known about heterochromatin, seen as dark spots in chromosomes, for decades. Recent thinking suggests that heterochromatin plays an important role in how genes are turned on and off by shifting parts of the DNA into a different phase from the rest of the chromosome, like blobs of oil in water. This would effectively create compartments in the nucleus where specific genes could be turned on or off.

Another mystery of centromeres is how and why they consistently form in the same place, because there is no specific genetic code for them to do so. They are determined "epigenetically," or outside the genome. Basically, your centromeres are where they are because that's where they were in the sperm and egg from which you were conceived.

The Langleys and their coauthors were able to compare the [centromere](#) sequences from the new reference genome with other published sequences, providing evidence that human centromeres can in fact move around a bit. This has been found in other animal species.

"Now we will be better able to understand how these things happen," Langley said.

## **Applications**

Having the original human genome sequence has been a powerful tool for discovery in biomedical sciences over the past 20 years. The new reference will help researchers better understand variation, especially in

those areas that were not well covered before or with mistakes and artifacts, Dennis said.

"It's already being used to reanalyze genomes collected by the 1000 Genomes Project, discovering and verifying thousands of new variants," she said. The 1000 Genomes Project is an international collaboration to create a catalogue of human genetic variation.

Those new, confirmed genetic variants can then, for example, be associated with disease states and clinical outcomes using sequencing data from patients, such as autistic individuals, Dennis said.

The T2T Consortium includes 114 scientists at 33 institutions and is co-chaired by Adam Phillippy, NHGRI and Karen Miga, UC Santa Cruz.

**More information:** Sergey Aganezov et al, A complete reference genome improves analysis of human genetic variation, *Science* (2022).  
[DOI: 10.1126/science.abl3533](https://doi.org/10.1126/science.abl3533).  
[www.science.org/doi/10.1126/science.abl3533](https://www.science.org/doi/10.1126/science.abl3533)

Provided by UC Davis

Citation: New human reference genome opens unexplored regions (2022, March 31) retrieved 8 May 2024 from <https://phys.org/news/2022-03-human-genome-unexplored-regions.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.