

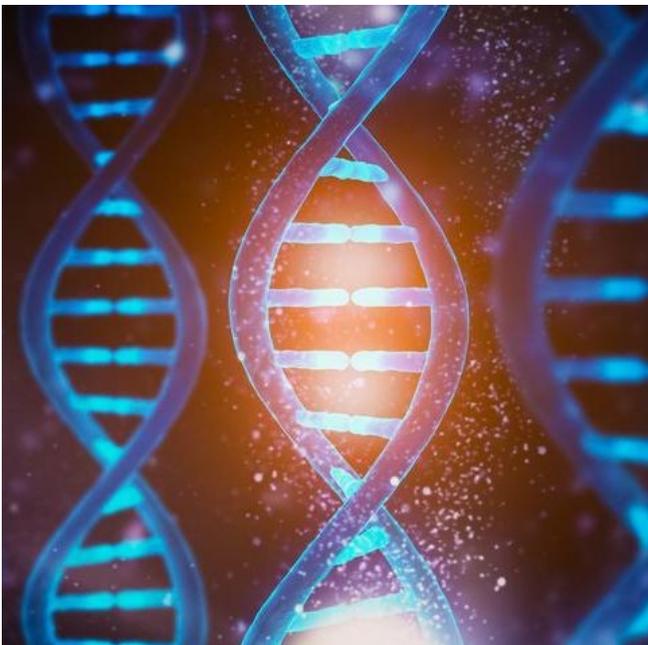
Scientists Say They've Finally Sequenced the Entire Human Genome. Yes, All of It.

This is a technological triumph.



By [Caroline Delbert](#)

Aug 30, 2021



Getty Images

In June 2000, scientists announced the first draft of the human genome sequence had been completed. Generating the sequence was a technical feat—one that allowed scientists to begin reading humanity's genetic blueprint—but it was still missing about 8 percent of the genome.

Now, an international consortium of about 100 scientists, dubbed the Telomere-to-Telomere (T2T) Consortium, say they've finally assembled [the human genome](#) in its entirety. If their work, which [was published](#) May 27 to the pre-print website *bioRxiv*, holds up to peer review, it could change the

future of medicine. As researchers become more familiar with humanity’s genetic code, they can, for example, make more precise and effective medicines—including the kind of gene-focused treatment that powered [the first effective COVID-19 vaccines](#).

The first draft led to a boom in technology, like CRISPR, and therapeutics, but scientists didn’t realize how incomplete the draft was at the time, says Karen Miga, Ph.D., a genomics researcher at the University of California, Santa Cruz, and a member of the T2T Consortium.

The 2000 sequence was the product of the Human Genome Project (HGP) and private company Celera Genomics. HGP claimed it had mapped the whole human genome at the time, but was careful in how it defined its success: “‘Finished sequence’ is a technical term meaning that the sequence is highly accurate (with fewer than one error per 10,000 letters) and highly contiguous (with the only remaining gaps corresponding to regions whose sequence cannot be reliably resolved with current technology).”



Alan Phillips Getty Images

“Current technology” is doing a lot of heavy lifting here. At the time, HGP used a technology called bacterial artificial chromosome (BAC), where scientists used a bacterium to clone each piece of the genome, and then studied them in smaller groups. A complete “BAC library” is about 20,000 carefully prepared bacteria with cloned genes inside. But that BAC process inherently misses some portions of the whole genome.

Humans have 46 chromosomes, in 23 pairs, that represent tens of thousands of individual genes. Each gene consists of base pairs made of adenine (A), thymine (T), guanine (G), and cytosine (C). There are billions of these base pairs in the human genome. The base pairs in the untouched 8 percent of the 2000 genome draft, it turns out, are made of many, many repeated patterns that were too difficult to study using BAC or similar methods.

For the latest sequence, T2T turned to the California-based Pacific Biosciences (PacBio) and the U.K.-based Oxford Nanopore Technologies. PacBio uses a system called HiFi, where base pairs are circularized (made into a circle) and repeatedly read to ensure accuracy. The system is just a few years old and represents a big step forward in both length and accuracy for those longer sequences. Oxford Nanopore, meanwhile, presses strands of base pairs through a microscopic nanopore—just one molecule at a time—where an electrical current zaps them in order to observe what kind of molecule they are. By zapping each molecule, scientists can identify the full strand.

You Have to Read This

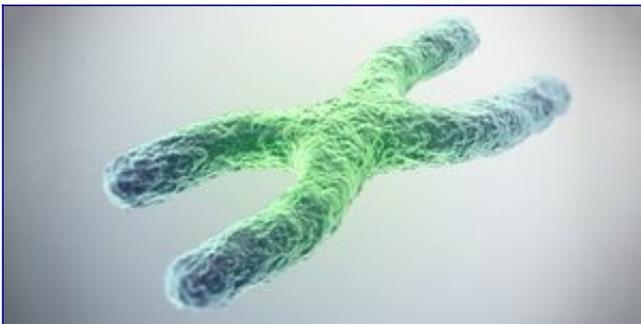


Image of a chromosome

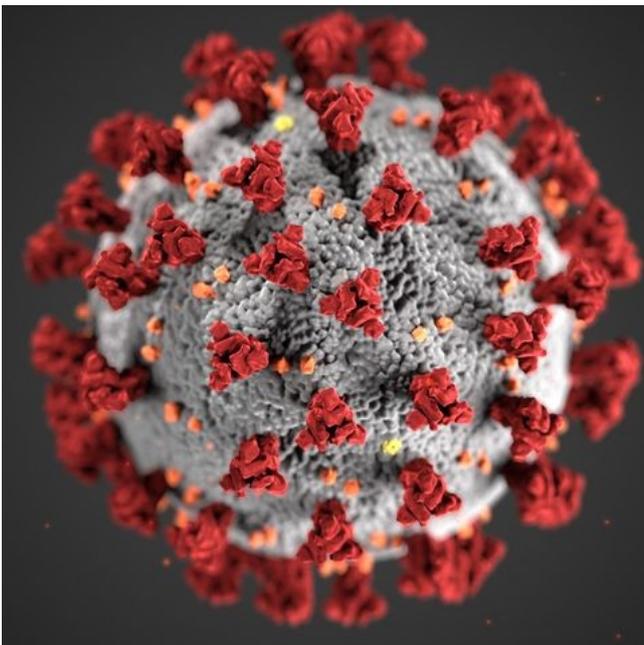
[Did Scientists Just Find a Way to Reverse Aging?](#)

The amount of ground T2T covered is staggering. “Over the last 20 years we’ve had a genomic revolution where we’ve started to put function, base by base, across the genome,” Miga says. “Now we’re going to present to the community 200 million bases, which have not been looked at before, to start to assign function and start to understand how our genome works.”

There's still more work to do. One snag is that both projects studied cells that had just 23 chromosomes instead of the full 46. That's because they use cells derived from the reproductive system, where eggs and sperm each carry half of a full chromosomal load. The cell used in the latest research is from a hydatidiform mole, a kind of reproductive growth that represents an extremely early, nonviable union between a sperm and an egg cell that has no nucleus.

Choosing this kind of cell, which has been kept and cultured as a cell line used for research purposes, cuts the huge sequencing job in half, but, in this case, it only carries [genetic information](#) from the father's chromosome. After the study passes through the peer-review process, both PacBio and Oxford will attempt to sequence a genome that includes genetic information from both parents.

Another snag is that the material from which this genome was sequenced represents only the genetic information of a single person. The consortium has tapped another team, the Human Pangenome Reference Consortium, and is planning to sequence the genomes of people from different regions around the globe, and build an ethnically diverse array of genetic material to study.



COURTESY CDC/ALISSA ECKERT, MSMI; DAN HIGGINS, MAMS

Since the beginning of the pandemic, scientists have raced to figure out why some people contract severe cases of Covid from the SARS-CoV-2 virus, while others experience few to no symptoms at

all. In [a recent study](#) published to the journal *Nature*, an international team of roughly 3,000 scientists combed through genetic data from 46 studies and pinpointed 13 locations in the human genome that influence the susceptibility and severity of the disease. This information could help scientists identify high-risk populations and develop better treatments.

To see this research in action, check out “[Can America’s Fastest Supercomputer Defeat COVID for Good?](#)”

[Source](#)