

Researchers generate the first complete, gapless sequence of a human genome

Prabarna Ganguly, Ph.D. AND Rachael Zisk | March 31, 2022

Scientists have published the first complete, gapless sequence of a human genome, two decades after the Human Genome Project produced the first draft human genome sequence.

According to researchers, having a complete, gap-free sequence of the roughly 3 billion bases (or “letters”) in our DNA is critical for understanding the full spectrum of human genomic variation and for understanding the genetic contributions to certain diseases. The work was done by the Telomere to Telomere (T2T) consortium, which included leadership from researchers at the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health; University of California, Santa Cruz; and University of Washington, Seattle. NHGRI was the primary funder for the study.

Analyses of the complete genome sequence will significantly add to our knowledge of chromosomes, including more accurate maps for five chromosome arms, which opens new lines of research. This helps answer basic biology questions about how chromosomes properly segregate and divide. The T2T consortium used the now-complete genome sequence as a reference to discover more than 2 million additional variants in the human genome. These studies provide more accurate information about the genomic variants within 622 medically relevant genes.

“Generating a truly complete human genome sequence represents an incredible scientific achievement, providing the first comprehensive view of our DNA blueprint,” said Eric Green, M.D., Ph.D., director of NHGRI. “This foundational information will strengthen the many ongoing efforts to understand all the functional nuances of the human genome, which in turn will empower genetic studies of human disease.”

The now-complete human genome sequence will be particularly valuable for studies that aim to establish comprehensive views of human genomic variation, or how people’s DNA differs. Such insights are vital for understanding the genetic contributions to certain diseases and for using genome sequence as a routine part of clinical care in the future. Many research groups have already started using a pre-release version of the complete human genome sequence for their research.

“Generating a truly complete human genome sequence represents an incredible scientific achievement, providing the first comprehensive view of our DNA blueprint. This foundational information will strengthen the many ongoing efforts to understand all the functional nuances of the human genome, which in turn will empower genetic studies of human disease.”

The full sequencing builds upon the work of the Human Genome Project, which mapped about 92% of the genome, and research undertaken since then. Thousands of researchers have developed better laboratory tools, computational methods and strategic approaches to decipher the complex sequence. Six papers encompassing the completed sequence appear in *Science*, along with companion papers in several other journals.

That last 8% includes numerous genes and repetitive DNA and is comparable in size to an entire chromosome. Researchers generated the complete genome sequence using a human cell line with only one copy of each chromosome, unlike most human cells, which carry two copies of each chromosome. The researchers noted that 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).

“Ever since we had the first draft human genome sequence, determining the exact sequence of complex genomic regions has been challenging,” said Evan Eichler, Ph.D., researcher at the University of Washington and T2T consortium co-chair. “I am thrilled that we got the job done. The complete blueprint is going to revolutionize the way we think about human genomic variation, disease and evolution.”

The cost of sequencing a human genome using “short-read” technologies, which provide several hundred bases of DNA sequence at a time, is only a few hundred dollars, having fallen significantly since the end of the Human Genome Project. However, using these short-read methods alone still leaves some gaps in assembled genome sequences. The massive drop in DNA sequencing costs comes hand-in-hand with increased investments in new DNA sequencing technologies to generate longer DNA sequence reads without compromising the accuracy.

“In the future, when someone has their genome sequenced, we will be able to identify all of the variants in their DNA and use that information to better guide their healthcare. Truly finishing the human genome sequence was like putting on a new pair of glasses. Now that we can clearly see everything, we are one step closer to understanding what it all means.”

Over the past decade, two new DNA sequencing technologies emerged that produced much longer sequence reads. The Oxford Nanopore DNA sequencing method can read up to 1 million DNA letters in a single read with modest accuracy, while the PacBio HiFi DNA sequencing method can read about 20,000 letters with nearly perfect accuracy. Researchers in the T2T consortium used both DNA sequencing methods to generate the complete human genome sequence.

“Using long-read methods, we have made breakthroughs in our understanding of the most difficult, repeat-rich parts of the human genome,” says Karen Miga, Ph.D., a co-chair of the T2T consortium whose research group at the University of California, Santa Cruz is funded by NHGRI. “This complete human genome sequence has already provided new insight into genome biology, and I look forward to the next decade of discoveries about these newly revealed regions.”

According to consortium co-chair Adam Phillippy, Ph.D., whose research group at NHGRI led the finishing effort, sequencing a person’s entire genome should get less expensive and more straightforward in the coming years.

“In the future, when someone has their genome sequenced, we will be able to identify all of the variants in their DNA and use that information to better guide their healthcare,” Phillippy said. “Truly finishing the human genome sequence was like putting on a new pair of glasses. Now that we can clearly see everything, we are one step closer to understanding what it all means.”

Many early-career researchers and trainees played pivotal roles, including researchers from Johns Hopkins University, Baltimore; University of Connecticut, Storrs; University of California, Davis; Howard Hughes Medical Institute, Chevy Chase, Maryland; and the National Institute of Standards and Technology, Gaithersburg, Maryland. The package of six papers reporting this accomplishment appears in today’s issue of Science, along with companion papers in several other journals.

Papers

Nurk et al. The complete sequence of a human genome. *Science* 376. Doi: 10.1126/science.abj6987 (2022)

Gershman et al. Epigenetic patterns in a complete human genome. *Science* 376. Doi: 10.1126/science.abj5089 (2022)

Vollger et al. Segmental duplications and their variation in a complete human genome. *Science* 376. Doi: 10.1126/science.abj6965 (2022)

Hoyt et al. From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science* 376. Doi: 10.1126/science.abk3112 (2022)

Aganezov et al. A complete reference genome improves analysis of human genetic variation. *Science* 376. Doi: 10.1126/science.abl3533 (2022)

Altemose et al. Complete genomic and epigenetic maps of human centromeres. *Science* 376. Doi: 10.1126/science.abl4178 (2022)

About NHGRI and NIH

About the National Human Genome Research Institute (NHGRI): At NHGRI, we are focused on advances in genomics research. Building on our leadership role in the initial sequencing of the human genome, we collaborate with the world's scientific and medical communities to enhance genomic technologies that accelerate breakthroughs and improve lives. By empowering and expanding the field of genomics, we can benefit all of humankind. For more information about NHGRI and its programs, visit www.genome.gov.

About the National Institutes of Health (NIH): NIH, the nation's medical research agency, includes 27 Institutes and Centers and is a component of the U.S. Department of Health and Human Services. NIH is the primary federal agency conducting and supporting basic, clinical, and translational medical research, and is investigating the causes, treatments, and cures for both common and rare diseases. For more information about NIH and its programs, visit www.nih.gov.

Press Contact

NHGRI Press Office

Email: nhgripressoffice@mail.nih.gov

Phone: (301) 402-0911