

MAR 31 2022

Research

Complete Human Genome Deciphered for the First Time

Summary

Twenty years after scientists first reported reading the human genome, the Telomere-to-Telomere consortium has now unveiled a complete version. Their work uncovered hundreds of new genes and opens the door to new insights into evolution and disease.



Successfully completing the entire human genome has revealed a whole chromosome's worth of previously missing DNA. Credit: Irving Geis/HHMI

Evan Eichler has always been drawn to the most complex regions of humanity's genome – those with bizarrely long stretches of repeated DNA or with extra copies of genes. He suspected these regions might play crucial roles in evolution and disease. That's why, more than 20 years ago, he became part of the Human Genome Project, the \$3 billion effort to read every letter of a person's DNA.

But after the project claimed victory in 2003, Eichler was only a little closer to his scientific goal. The sequencing effort had failed to read many big chunks of DNA – more than eight percent of the genome. Scientists knew these missing chunks contained highly repetitive sequences, and largely dismissed them as junk. Not so, says Eichler, a Howard Hughes Medical Institute (HHMI) Investigator at the University of Washington. "It turned out that many of the regions I was interested in were in the gaps." He became committed to finishing the job – reading the entire genome, tricky bits and all.

Now he and a team of about 100 scientists, led by Adam Phillippy of the National Human Genome Research Institute (NHGRI) and Karen Miga of the University of California, Santa Cruz, (UCSC) have finally gotten it right. In new work first posted as a preprint on bioRxiv.org and now published March 31, 2022, in the journal *Science*, they describe the first ever sequencing of an entire human genome, adding a whole chromosome's worth of previously hidden DNA – the missing eight percent. In the genetic manuscript for life, "we are seeing chapters that were never read before," says Eichler.

Or as University of Washington geneticist Robert Waterston puts it: "There are no longer any hidden or unknown bits."

"I think that is psychologically a big thing," adds Waterston, a leader in the original Human Genome Project who was not involved in the new effort. "I just admire these scientists for sticking with it."

An intricate puzzle

The human genome is made up of just over six billion individual letters of DNA – about the same number as other primates like chimps – spread among 23 pairs of chromosomes. To read a genome, scientists first chop up all that DNA into pieces hundreds to thousands of letters long. Sequencing machines then read the individual letters in each piece, and scientists try to assemble the pieces in the right order, like putting together an intricate puzzle.



HHMI Investigator Evan Eichler sees the new, complete sequence as a “Rosetta stone” for understanding complex genetic variation underlying disease and evolution and is a principal investigator of a pan-genome effort to sequence the complete genomes of many humans. Credit: Ron Wurzer/AP Images for HHMI



HHMI Investigator Erich Jarvis helped assemble the DNA pieces to finish the human genome and is now co-leading the pan-genome project. Credit: Peter Ross

One challenge is that some regions of the genome repeat the same letters over and over again. Repetitive regions include the centromeres, the parts that hold the two strands of chromosomes together and that play crucial roles in cell division, and ribosomal DNA, which provides instructions for the cell's protein factories. Still other repetitive parts include new genes that may help species adapt. In the past, all that repetition made it impossible to assemble some chopped-up pieces in the correct order. It's like having identical puzzle pieces – scientists didn't know which went where, leaving big gaps in the genomic picture.

Another snag: most cells contain two genomes – one from the father and one from the mother. When researchers try to assemble all the pieces, sequences from each parent can mix together, obscuring the actual variation within each individual genome.

In the mid-2000s, as scientists tried to figure out how to overcome the barriers, “we came up with the idea of getting a complete genome by sequencing just one of the genomes instead of solving two at the same time,” recalls Eichler. He knew just where to find it – from a set of cell lines being studied by

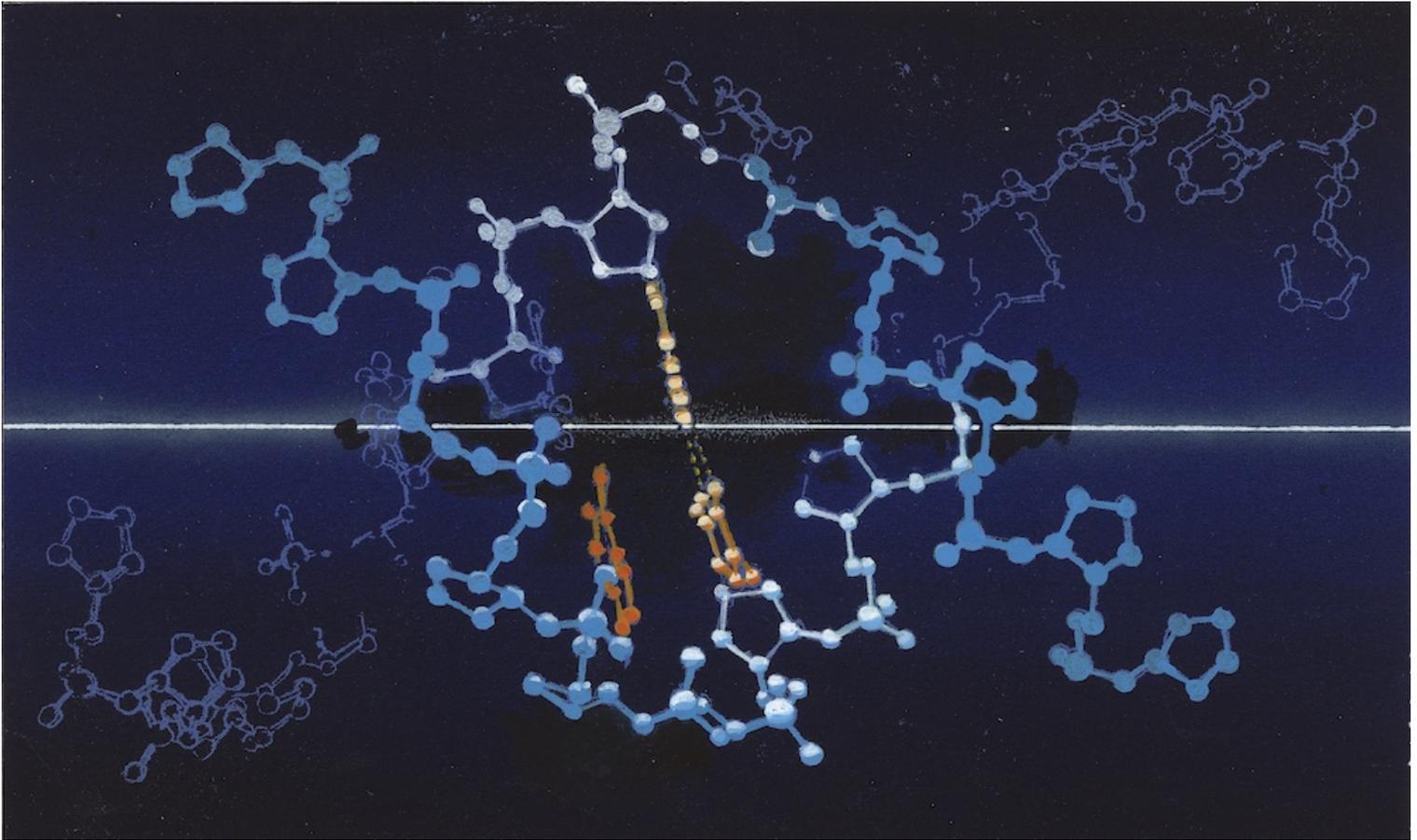
University of Pittsburgh reproductive geneticist Urvashi Surti. Because of a rare glitch in normal development, the cells end up with two copies of the father's DNA and none of the mother's.

Such a cell line, with only one genome, "is what made this genome assembly possible," says HHMI Investigator Erich Jarvis, a Rockefeller University neurogeneticist who collaborated on the new work.

Fired up

Other key advances included rapid improvements in the gene sequencing machines made by Oxford Nanopore Technologies and Pacific Biosciences. By 2017, NHGRI's Phillippy and UCSC's Miga realized that a new Nanopore machine's ability to accurately read a million letters of DNA at a time had opened the door to finally tackling the genome's hard bits. They created the Telomere-to-Telomere (T2T) consortium to sequence each chromosome from one end, or telomere, to the other. Around the same time Eichler's team had shown the value of using Pacific Biosciences technology to resolve more complex forms of genetic variation.

There was no guarantee of success. But "we had the benefit of youthful optimism and we were fired up by the promise of these new technologies," recalls Phillippy. The team ran their Nanopore machines nonstop for six months and brought in scores of scientists to assemble the pieces and analyze the results. At the same time, sequencing data were being generated by other team members and Pacific Biosciences using their long-read sequencing platform. In particular, the project got a boost when Pacific Biosciences introduced a new sequencing machine which generated long-read sequencing reads that were greater than 99 percent accurate. "It was the last piece of the puzzle – like putting on a new pair of glasses," says Phillippy. The Pacific Biosciences technology couldn't cover all parts of the genome equally well, but the scientists realized that by combining the long-read sequencing with the Oxford Nanopore data, they could fill all the gaps.



The human genome is made up of just over six billion individual letters of DNA, spread among 23 pairs of chromosomes. Now, scientists describe the first ever sequencing of an entire human genome, uncovering hundreds of new genes and opening the door to new insights into evolution and disease. Credit: Irving Geis/HHMI

By summer 2020, the consortium had assembled two chromosomes and planned what Phillippy calls a hackathon to get the other 21, working remotely over Zoom and Slack during the pandemic. One key aha moment came when the team tried to assemble the most difficult regions of the genome – the highly repetitive DNA in the centromeres. The researchers realized that the algorithms for assembling the pieces couldn't handle the repetition, but the human eye could. On the computer screen, the scientists saw where the different repetitive sequences had become tangled together. Then, they untangled it manually, “like untangling a string in your yo-yo,” Jarvis says. By summer's end, the team had sequenced every chromosome.

Earthquake of genetic changes

As each new chapter in our genetic book of life emerged, researchers dove in to look for biological meaning. Their results appear in six papers in *Science* and more than a dozen papers elsewhere. For example, the team discovered unexpectedly high levels of genetic variation in centromeres and other regions – “a whole new treasure chest of variants that we can study to see if they have functional significance,” says Phillippy.



A key to reading every letter of the human genome was using a rare cell line in which the copies of each chromosome are identical, rather than having one chromosome from each parent. Credit: NHGRI

The data offer “the foundation for a new era” in studying centromeres, says Miga, who co-led the T2T centromere satellite working group. Scientists will now be able to explore how this newly discovered variation contributes to disease, and how centromere DNA changes over time, she says.

The T2T results also point to more complex patterns of variation in genes that may have helped create the human species – and could explain our rapid evolution. The full genome sequence reveals that some genes associated with bigger brains are highly variable, Eichler explains. One person might have 10 copies of a particular gene, while others might have only one or two. This variation can spell trouble during fertilization, when chromosomes from mom and dad line up and swap pieces. The mismatched

genes can lead to “an earthquake” of gene alterations, Eichler explains. As a result, “these regions become a crucible for both rapid evolutionary changes and disease susceptibility, both within and between species,” he says.

The successful completion of a single genome is hardly the last word. Consortium members are already working to sequence a genome with different chromosomes inherited from each parent. They’re also beginning a pan-genome effort to read the entire DNA sequences of hundreds of people from around the world. “The goal is to create as complete a human genome as possible, representing much more of human diversity,” explains Jarvis, co-leader of the pan-genome effort.

But the new sequence is the indispensable first step, says Eichler. “Now we have a Rosetta stone for looking at complete variation in hundreds of thousands of other genomes going forward.”

###

Citations

Sergey Nurk et al. “The complete sequence of a human genome.” Posted on bioRxiv.org on May 27, 2021. doi: 10.1101/2021.05.26.445798. Published in *Science* on March 31, 2022. doi: 10.1126/science.abj6987.

Mitchell Vollger et al. “Segmental duplications and their variation in a complete human genome.” Posted on bioRxiv.org on May 26, 2021. doi: 10.1101/2021.05.26.445678. Published in *Science* on March 31, 2022. doi: 10.1126/science.abj6965

Scientist Profiles



Evan E. Eichler

University of Washington

Genetics Computational Biology



Erich D. Jarvis

The Rockefeller University

Neuroscience Molecular Biology

FOR MORE INFORMATION

Meghan Rosen

301-215-8859

rosenm2@hhmi.org

GET HHMI NEWS BY EMAIL

Sign up

 [Subscribe to RSS](#)

Related Links

[The Eichler Lab](#) 

[The Jarvis Lab](#) 

[Erich Jarvis | HHMI Profile](#) 

[Project to Read Genomes of All 70,000 Vertebrate Species Reports First Discoveries](#) 

[Whole Genome Sequencing Identifies New Genetic Signature for Autism](#) 

hhmi

[Contact Us / Careers](#)

© 2022 Howard Hughes Medical Institute / [Privacy Policy & Cookie Notice](#) / [Terms of Use](#) / [Log In](#)