https://www.nytimes.com/2021/07/23/science/human-genome-complete.html

**Scientists Finish the Human Genome at Last (edited)** New York Times;Carl Zimmer; July 23, 2021

A century ago, scientists knew that genes were spread across 23 pairs of chromosomes.

By the late 1970s, scientists had gained the ability to pinpoint a few individual human genes and decode their sequence. But their tools were so crude that hunting down a single gene could take up an entire career.

Toward the end of the 20th Century The Human Genome Project was an audacious undertaking, given how much there was to sequence. Scientists knew that the twin strands of DNA in our cells contained roughly three billion pairs of letters — a text long enough to fill hundreds of books.

The White House announced in 2000 that scientists had finished the first draft of the human genome, and details of the project were published the following year. But long stretches of the genome remained unknown,

It turned out that the genome was a very hard puzzle to put together from small pieces. Many of our genes exist as multiple copies that are nearly identical to each other. Sometimes the different copies carry out different jobs.

In 2019, two scientists — Adam Phillippy, a computational biologist at the National Human Genome Research Institute, and Karen Miga, a geneticist at the University of California, Santa Cruz — founded the [**Telomere-to-Telomere Consortium**](https://sites.google.com/ucsc.edu/t2tworkinggroup/home?authuser=0) to complete the genome.

Dr. Phillippy and Dr. Miga put out a call for scientists to join them to finish the puzzle. They ended up with 99 scientists working directly on sequencing the human genome,

They can now say with confidence that the human genome measures **3.05 billion base pairs long**.

Within those new sequences of DNA, the scientists discovered more than 2,000 new genes. Most appear to be disabled by mutations, but **115** of them look as if they **can produce proteins**. The consortium now estimates that the human genome contains **19,969 genes**.

 1. How many total genes did the Telomere to Telomer Consortium find?

 2. How many of those genes can produce proteins?

(full article below)

= = = = =

https://www.nytimes.com/2021/07/23/science/human-genome-complete.html

**Scientists Finish the Human Genome at Last** New York Times;Carl Zimmer; July 23, 2021

The complete genome uncovered more than 100 new genes that are probably functional, and many new variants that may be linked to diseases.Top of Form

Bottom of Form

A century ago, scientists knew that genes were spread across 23 pairs of chromosomes. But pinpointing any single gene and deciphering its sequence was a struggle that could have consumed a career.

Two decades after the draft sequence of the human genome was [unveiled](https://www.nytimes.com/2001/02/11/us/genome-analysis-shows-humans-survive-on-low-number-of-genes.html) to great fanfare, a team of 99 scientists has finally deciphered the entire thing. They have filled in vast gaps and corrected a long list of errors in previous versions, giving us a new view of our DNA.

The consortium has posted [six papers](https://www.biorxiv.org/search/author1%3Amiga%20author2%3Aphillippy%20jcode%3Abiorxiv%20limit_from%3A2021-05-01%20limit_to%3A2021-07-15%20numresults%3A10%20sort%3Arelevance-rank%20format_result%3Astandard) online in recent weeks in which they describe the full genome. These hard-sought data, now under review by scientific journals, will give scientists a deeper understanding of how DNA influences risks of disease, the scientists say, and how cells keep it in neatly organized chromosomes instead of molecular tangles.

For example, the researchers have uncovered more than 100 new genes that may be functional, and have identified millions of genetic variations between people. Some of those differences probably play a role in diseases.

For Nicolas Altemose, a postdoctoral researcher at the University of California, Berkeley, who worked on the team, the view of the complete human genome feels something like the close-up pictures of Pluto from the New Horizons space probe.

“You could see every crater, you could see every color, from something that we only had the blurriest understanding of before,” he said. “This has just been an absolute dream come true.”

Experts who were not involved in the project said it will enable scientists to explore the human genome in much greater detail. Large chunks of the genome that had been simply blank are now deciphered so clearly that scientists can start studying them in earnest.

“The fruit of this sequencing effort is amazing,” said Yukiko Yamashita, a developmental biologist at the Whitehead Institute for Biomedical Research at the Massachusetts Institute of Technology.

While scientists have known for decades that genes were spread across 23 pairs of chromosomes, these strange, wormlike microscopic structures remained largely a mystery.

By the late 1970s, scientists had gained the ability to pinpoint a few individual human genes and decode their sequence. But their tools were so crude that hunting down a single gene could take up an entire career.

Toward the end of the 20th Century, an international network of geneticists decided to try to sequence all the DNA in our chromosomes. The Human Genome Project was an audacious undertaking, given how much there was to sequence. Scientists knew that the twin strands of DNA in our cells contained roughly three billion pairs of letters — a text long enough to fill hundreds of books.

When that team began its work, the best technology the scientists could use sequenced bits of DNA just a few dozen letters, or bases, long. Researchers were left to put them together like the pieces of a vast jigsaw puzzle. To assemble the puzzle, they looked for fragments with identical ends, meaning that they came from overlapping portions of the genome. It took years for them to gradually assemble the sequenced fragments into larger swaths.

The White House announced in 2000 that scientists had finished the first draft of the human genome, and details of the project were published the following year. But long stretches of the genome remained unknown, while scientists struggled to figure out where millions of other bases belonged.

It turned out that the genome was a very hard puzzle to put together from small pieces. Many of our genes exist as multiple copies that are nearly identical to each other. Sometimes the different copies carry out different jobs. Other copies — known as pseudogenes — are disabled by mutations. A short fragment of DNA from one gene might fit just as well into the others.

And genes only make up a small percentage of the genome. The rest of it can be [even more baffling](https://www.nytimes.com/2015/03/08/magazine/is-most-of-our-dna-garbage.html). Much of the genome is made up of virus-like stretches of DNA that exist largely just to make new copies of themselves that get inserted back into the genome.

Loading an electrophoresis gel into a computer system used for DNA sequencing of human chromosomes at the California Institute of Technology in 1989.Credit...Peter Menzel/Science Source

In the early 2000s, scientists got a little better at putting together the genome puzzle from its tiny pieces. They made more fragments, read them more accurately, and developed new computer programs to assemble them into bigger chunks of the genome.

Periodically, researchers would unveil the latest, best draft of the human genome — known as the reference genome. Scientists used the reference genome as a guide for their own sequencing efforts. For example, clinical geneticists would catalog disease-causing mutations by comparing genes from patients to the reference genome.

The newest reference genome came out in 2013. It was a lot better than the first draft, but it was a long way from complete. Eight percent of it was simply blank.

“There’s basically an entire human chromosome that had gone missing,” said Michael Schatz, a computational biologist at Johns Hopkins University.

In 2019, two scientists — Adam Phillippy, a computational biologist at the National Human Genome Research Institute, and Karen Miga, a geneticist at the University of California, Santa Cruz — founded the [Telomere-to-Telomere Consortium](https://sites.google.com/ucsc.edu/t2tworkinggroup/home?authuser=0) to complete the genome.

Dr. Phillippy admitted that part of his motivation for such an audacious project was that the missing gaps annoyed him. “They were just really bugging me,” he said. “You take a beautiful landscape puzzle, pull out a hundred pieces, and look at it — that’s very bothersome to a perfectionist.”

Dr. Phillippy and Dr. Miga put out a call for scientists to join them to finish the puzzle. They ended up with 99 scientists working directly on sequencing the human genome, and dozens more pitching in to make sense of the data. The researchers worked remotely through the pandemic, coordinating their efforts over Slack, a messaging app.

“It was a surprisingly nice ant colony,” Dr. Miga said.

The consortium took advantage of new machines that can read stretches of DNA reaching tens of thousands of bases long. The researchers also invented techniques to figure out where particularly mysterious repeating sequences belonged in a genome.

All told, the scientists added or fixed more than 200 million base pairs in the reference genome. They can now say with confidence that the human genome measures **3.05 billion base pairs long**.

Within those new sequences of DNA, the scientists discovered more than 2,000 new genes. Most appear to be disabled by mutations, but 115 of them look as if they can produce proteins — the function of which scientists may need years to figure out. The consortium now estimates that the human genome contains **19,969 protein-coding genes**.

With a complete genome finally assembled, the researchers could take a better look at the variation in DNA from one person to the next. They discovered more than two million new spots in the genome where people differ. Using the new genome also helped them to avoid identifying disease-linked mutations where none actually exist.

“It’s a great advance for the field,” said Dr. Midhat Farooqui, the director of molecular oncology at Children’s Mercy, a hospital in Kansas City, Mo., who was not involved in the project.

Dr. Farooqi has started using the genome for his research into rare childhood diseases, aligning DNA from his patients against the newly filled gaps to search for mutations.

Switching to the new genome may be a challenge for many clinical labs, however. They’ll have to shift all of their information about the links between genes and diseases to a new map of the genome. “There will be a big effort, but it will take a couple years,” said Dr. Sharon Plon, a medical geneticist at Baylor College of Medicine in Houston.

Dr. Altemose plans on using the complete genome to explore a particularly mysterious region in each chromosome known as the centromere. Instead of storing genes, centromeres anchor proteins that move chromosomes around a cell as it divides. The centromere region contains thousands of repeated segments of DNA.

In their first look, Dr. Altemose and his colleagues were struck by how different centromere regions can be from one person to another. That observation suggests that centromeres have been evolving rapidly, as mutations insert new pieces of repeating DNA into the regions or cut other pieces out.

While some of this repeating DNA may play a role in pulling chromosomes apart, the researchers have also found new segments — some of them millions of bases long — that don’t appear to be involved. “We don’t know what they’re doing,” Dr. Altemose said.

But now that the empty zones of the genome are filled in, Dr. Altemose and his colleagues can study them up close. “I’m really excited moving forward to see all the things we can discover,” he said.

= = = = =

# 7 Journal Articles

 https://www.biorxiv.org/search/author1%3Amiga%20author2%3Aphillippy%20jcode%3Abiorxiv%20limit\_from%3A2021-05-01%20limit\_to%3A2021-07-15%20numresults%3A10%20sort%3Arelevance-rank%20format\_result%3Astandard

for author "miga" and author "phillippy" and posted between "01 May, 2021 and 15 Jul, 2021"

Top of Form

* [A complete reference genome improves analysis of human genetic variation](https://www.biorxiv.org/content/10.1101/2021.07.12.452063v1)

Sergey Aganezov, Stephanie M. Yan, Daniela C. Soto, Melanie Kirsche, Samantha Zarate, Pavel Avdeyev, Dylan J. Taylor, Kishwar Shafin, Alaina Shumate, Chunlin Xiao, Justin Wagner, Jennifer McDaniel, Nathan D. Olson, Michael E.G. Sauria, Mitchell R. Vollger, Arang Rhie, Melissa Meredith, Skylar Martin, Joyce Lee, Sergey Koren, Jeffrey A. Rosenfeld, Benedict Paten, Ryan Layer, Chen-Shan Chin, Fritz J. Sedlazeck, Nancy F. Hansen, Danny E. Miller, Adam M. Phillippy, Karen H. Miga, Rajiv C. McCoy, Megan Y. Dennis, Justin M. Zook, Michael C. Schatz

bioRxiv 2021.07.12.452063; doi: https://doi.org/10.1101/2021.07.12.452063

[Add to Selected Citations](https://www.biorxiv.org/highwire-save-citation/saveapath/%3Bbiorxiv%3Bearly%3B2021%3B07%3B13%3B2021.07.12.452063.atom/nojs/0)

* [From telomere to telomere: the transcriptional and epigenetic state of human repeat elements](https://www.biorxiv.org/content/10.1101/2021.07.12.451456v1)

Savannah J. Hoyt, Jessica M. Storer, Gabrielle A. Hartley, Patrick G. S. Grady, Ariel Gershman, Leonardo G. de Lima, Charles Limouse, Reza Halabian, Luke Wojenski, Matias Rodriguez, Nicolas Altemose, Leighton J. Core, Jennifer L. Gerton, Wojciech Makalowski, Daniel Olson, Jeb Rosen, Arian F. A. Smit, Aaron F. Straight, Mitchell R. Vollger, Travis J. Wheeler, Michael C. Schatz, Evan E. Eichler, Adam M. Phillippy, Winston Timp, Karen H. Miga, Rachel J. O’Neill

bioRxiv 2021.07.12.451456; doi: https://doi.org/10.1101/2021.07.12.451456

[Add to Selected Citations](https://www.biorxiv.org/highwire-save-citation/saveapath/%3Bbiorxiv%3Bearly%3B2021%3B07%3B12%3B2021.07.12.451456.atom/nojs/1)

* [DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome-wide](https://www.biorxiv.org/content/10.1101/2021.07.06.451383v1)

Nicolas Altemose, Annie Maslan, Owen K. Smith, Kousik Sundararajan, Rachel R. Brown, Angela M. Detweiler, Norma Neff, Karen H. Miga, Aaron F. Straight, Aaron Streets

bioRxiv 2021.07.06.451383; doi: https://doi.org/10.1101/2021.07.06.451383

[Add to Selected Citations](https://www.biorxiv.org/highwire-save-citation/saveapath/%3Bbiorxiv%3Bearly%3B2021%3B07%3B07%3B2021.07.06.451383.atom/nojs/2)

* [Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies](https://www.biorxiv.org/content/10.1101/2021.07.02.450803v1)

Ann M. Mc Cartney, Kishwar Shafin, Michael Alonge, Andrey V. Bzikadze, Giulio Formenti, Arkarachai Fungtammasan, Kerstin Howe, Chirag Jain, Sergey Koren, Glennis A. Logsdon, Karen H. Miga, Alla Mikheenko, Benedict Paten, Alaina Shumate, Daniela C. Soto, Ivan Sović, Jonathan MD Wood, Justin M. Zook, Adam M. Phillippy, Arang Rhie

bioRxiv 2021.07.02.450803; doi: https://doi.org/10.1101/2021.07.02.450803

[Add to Selected Citations](https://www.biorxiv.org/highwire-save-citation/saveapath/%3Bbiorxiv%3Bearly%3B2021%3B07%3B02%3B2021.07.02.450803.atom/nojs/3)

* [Epigenetic Patterns in a Complete Human Genome](https://www.biorxiv.org/content/10.1101/2021.05.26.443420v1)

Ariel Gershman, Michael E.G. Sauria, Paul W. Hook, Savannah J. Hoyt, Roham Razaghi, Sergey Koren, Nicolas Altemose, Gina V. Caldas, Mitchell R. Vollger, Glennis A. Logsdon, Arang Rhie, Evan E. Eichler, Michael C. Schatz, Rachel J. O’Neill, Adam M. Phillippy, Karen H. Miga, Winston Timp

bioRxiv 2021.05.26.443420; doi: https://doi.org/10.1101/2021.05.26.443420

[Add to Selected Citations](https://www.biorxiv.org/highwire-save-citation/saveapath/%3Bbiorxiv%3Bearly%3B2021%3B05%3B27%3B2021.05.26.443420.atom/nojs/4)

* [The complete sequence of a human genome](https://www.biorxiv.org/content/10.1101/2021.05.26.445798v1)

Sergey Nurk, Sergey Koren, Arang Rhie, Mikko Rautiainen, Andrey V. Bzikadze, Alla Mikheenko, Mitchell R. Vollger, Nicolas Altemose, Lev Uralsky, Ariel Gershman, Sergey Aganezov, Savannah J. Hoyt, Mark Diekhans, Glennis A. Logsdon, Michael Alonge, Stylianos E. Antonarakis, Matthew Borchers, Gerard G. Bouffard, Shelise Y. Brooks, Gina V. Caldas, Haoyu Cheng, Chen-Shan Chin, William Chow, Leonardo G. de Lima, Philip C. Dishuck, Richard Durbin, Tatiana Dvorkina, Ian T. Fiddes, Giulio Formenti, Robert S. Fulton, Arkarachai Fungtammasan, Erik Garrison, Patrick G.S. Grady, Tina A. Graves-Lindsay, Ira M. Hall, Nancy F. Hansen, Gabrielle A. Hartley, Marina Haukness, Kerstin Howe, Michael W. Hunkapiller, Chirag Jain, Miten Jain, Erich D. Jarvis, Peter Kerpedjiev, Melanie Kirsche, Mikhail Kolmogorov, Jonas Korlach, Milinn Kremitzki, Heng Li, Valerie V. Maduro, Tobias Marschall, Ann M. McCartney, Jennifer McDaniel, Danny E. Miller, James C. Mullikin, Eugene W. Myers, Nathan D. Olson, Benedict Paten, Paul Peluso, Pavel A. Pevzner, David Porubsky, Tamara Potapova, Evgeny I. Rogaev, Jeffrey A. Rosenfeld, Steven L. Salzberg, Valerie A. Schneider, Fritz J. Sedlazeck, Kishwar Shafin, Colin J. Shew, Alaina Shumate, Yumi Sims, Arian F. A. Smit, Daniela C. Soto, Ivan Sović, Jessica M. Storer, Aaron Streets, Beth A. Sullivan, Françoise Thibaud-Nissen, James Torrance, Justin Wagner, Brian P. Walenz, Aaron Wenger, Jonathan M. D. Wood, Chunlin Xiao, Stephanie M. Yan, Alice C. Young, Samantha Zarate, Urvashi Surti, Rajiv C. McCoy, Megan Y. Dennis, Ivan A. Alexandrov, Jennifer L. Gerton, Rachel J. O’Neill, Winston Timp, Justin M. Zook, Michael C. Schatz, Evan E. Eichler, Karen H. Miga, Adam M. Phillippy

bioRxiv 2021.05.26.445798; doi: https://doi.org/10.1101/2021.05.26.445798

[Add to Selected Citations](https://www.biorxiv.org/highwire-save-citation/saveapath/%3Bbiorxiv%3Bearly%3B2021%3B05%3B27%3B2021.05.26.445798.atom/nojs/5)

* [Segmental duplications and their variation in a complete human genome](https://www.biorxiv.org/content/10.1101/2021.05.26.445678v1)

Mitchell R. Vollger, Xavi Guitart, Philip C. Dishuck, Ludovica Mercuri, William T. Harvey, Ariel Gershman, Mark Diekhans, Arvis Sulovari, Katherine M. Munson, Alexandra M. Lewis, Kendra Hoekzema, David Porubsky, Ruiyang Li, Sergey Nurk, Sergey Koren, Karen H. Miga, Adam M. Phillippy, Winston Timp, Mario Ventura, Evan E. Eichler

bioRxiv 2021.05.26.445678; doi: https://doi.org/10.1101/2021.05.26.445678

[Add to Selected Citations](https://www.biorxiv.org/highwire-save-citation/saveapath/%3Bbiorxiv%3Bearly%3B2021%3B05%3B26%3B2021.05.26.445678.atom/nojs/6)

= = = =

Telomere to Teleomere Consortium

https://sites.google.com/ucsc.edu/t2tworkinggroup/home?authuser=0