

## DOUBLE HELIX

# 20 YEARS AFTER THE HUMAN GENOME PROJECT, RESEARCHERS DECIPHER THE MISSING 8 PERCENT OF HUMAN DNA

A decade ago, researchers sequenced 92 percent of the human genome. They just cracked the last 8 percent.

**A TEAM OF** international researchers has finally managed to sequence a complete human genome. The breakthrough comes almost more than three decades after the Human Genome Project, which began in 1990, set out to accomplish this very goal.

Now, researchers have the full set of instructions in human DNA that is inside every cell in our body and tells our cells how to develop, survive and reproduce. In a series of papers published Thursday in the journals Science and Nature Methods, the researchers outline how they managed to go where no scientists had gone before. The findings could revolutionize how we understand human evolution as well as fundamentally alter how we treat myriad diseases.



If you thought this already happened, you're 92 percent right. In 2003, [The Human Genome Project](#) announced they'd sequenced the vast majority of the genome — about 92 percent. But there were some undecipherable gaps in the genome, amounting to about 8 percent of the total biological blueprint.

Deciphering the remaining 8 percent has been a process akin to [getting to Mars](#) — experts had a general idea of what is there, but actually developing the technology to see it at a granular level was an overwhelming challenge.

Some researchers wrote off the remaining 8 percent as largely inconsequential. Other researchers, some of whom worked on the initial project, couldn't get those pesky gaps out of their heads.

[Evan Eichler](#), a Howard Hughes Medical Institute Investigator at the University of Washington and one of the researchers who worked on the initial Human Genome Project tells *Inverse* that the researchers interested in this particular mystery are “the happy misfits of the Human Genome [Project]...the original human genome has driven so much discovery and research, getting 92 percent was good enough for a lot of people. It just wasn't good enough for us.”

**THE BACKGROUND** — DNA has four building blocks (nucleotides) which researchers designate with the letters A, C, G, and T. Sequencing is basically reading the order of those letters within DNA. It's a simple concept that can be extremely complicated to execute well.

The genome they fully sequenced is a genome originally studied by a reproductive geneticist in the mid-2000s. When two sets of DNA, from a biological mother and father, combine, it can be challenging to distinguish the variation within each individual genome. The genome used for The Human Genome Project as well as this latest development had an anomaly in which cells ended up with two copies of the father's DNA and none of the mother's. This anomaly allows researchers to study a single genome.

[Eric Green](#) is the Director of the National Human Genome Research Institute and involved in the “Telomere-to-Telomere Coalition,” as well as the previous Human Genome Project. He tells *Inverse* that after the Human Genome Project wrapped in

2003, “we knew what we were missing.”

“We knew we were missing the highly repetitive stuff. Those tend to be enriched near the ends of chromosomes, at the telomeres, and also at the middle of the chromosomes, at the centromeres.”

A genome with missing pieces is, “like a paragraph with sentences missing,” Green adds.

**HOW THEY DID IT** — The researchers were able to sequence those hard-to-read repeating sequences thanks to massive improvements in technology.

Eichler likens the technological developments to using a powerful magnifying glass to examine extremely tiny puzzle pieces.

ADVERTISEMENT

“Anyone who has done a puzzle knows that if you have big puzzle pieces, it’s a piece of cake. It’s when you get those tiny ones where there’s not a whole lot of difference.”

“In terms of repeating sections, you have a lot of pieces of the puzzle that look the same and you spend hours and hours trying to get those pieces in the right place,” he says.

Better technology meant the researchers could look at those tiny, repeating sequences more closely than ever before, and, eventually, put the right pieces into place. Unlike a puzzle, there’s no picture on a box for reference — the scientists put together that picture in real-time.

**WHAT THEY FOUND** — It turns out those repeating sequences that many researchers wrote off as “junk, or inconsequential” are actually extremely important, according to Eichler.

“These aren’t lightweight regions of our genome,” he says. “If you took away these regions, we would not be human. Our cells wouldn’t divide, and we wouldn’t produce proteins. Pretty much dead, right?”

Some of these sequences are also responsible for our big, human brains, he says.

“It turns out, several of the genes that are important for making a bigger brain are mapping precisely in these repeat regions that we have characterized.”

While the technology allowed them to see the sections more clearly, they still needed the human eye.

— DNA has four building blocks (nucleotides) which researchers abbreviate with the letters A, C, G, and T. Sequencing is the reading of the order of those letters within DNA. [Getty/MR.Cole\\_Photographer](#)

Green likens that process, called “sequence finishing” to a copy-editing a piece of writing. “I’d like to think of it as copy editing where, in some cases, they’re finding overt typos. But in other cases, they’re finding just awkward grammar,” he says.

Sequence finishing, he explains, is very labor-intensive. Skilled scientists will look through what the software reports. Sometimes the software will flag something as suspicious — this happens frequently with repeating sentences — and the scientists have to check and see if whatever the computer is flagging as potentially inaccurate actually is inaccurate. Much like spellcheck and grammar software doesn’t always understand the reality of phrasing or spelling, the sequencing software doesn’t always understand anomalies either. In both cases, a skilled human eye is required to check the computer’s work.

ADVERTISEMENT

**WHAT IT MEANS FOR HEALTH** — In addition to telling us vital information about the evolution of our species and our distinctive brains, these newly sequenced sections may have huge implications for conditions with a clear genetic link, like Down Syndrome and some cancers.

Centromeres, where many of these repeating sections are found, are vital for the segregation of chromosomes. Diseases like cancer and conditions like cancer and Down Syndrome are “essentially dysfunction of segregation,” Green says. Having a complete human genome could give researchers the tools to better understand how to treat these diseases.

This is just one human genome, but the breakthrough could speed up other advances in genetics. In much the same way that mRNA vaccine technology was in the works for over a decade before we got the first FDA-approved mRNA vaccine, completing one puzzle puts researchers in an excellent position to sequence others.

**WHAT'S NEXT** — Some of the researchers involved in the consortium are now turning their focus to sequencing a (more complicated) diploid human genome, one containing two sets of DNA, one from each parent. They also want to analyze a wide variety of human genomes from people with different ethnic and geographic backgrounds to determine what genetic variations may be specific to certain regions or ethnicities.

A completed human genome is just the beginning of how we understand and use the human genome — in research and clinical practice.

“This is not the end,” Eichler says. “This is the beginning — a stepping stone for all kinds of things.”

It’s also a testament to the value of “consortium science,” Green explains.

“What was required to accomplish this was really a collective toolbox of three different DNA sequencing methods, all sorts of theories, sophisticated computational tools, not just by one person, but by various people from multiple countries,” he says.

“Not unlike the moonshot, the problem was so hard that you needed a lot of creative minds coming together and bringing their own discipline and expertise to the problem, bouncing ideas off each other and incrementally making progress.”

Slow science, perhaps, but a massive and continual payoff.

**RELATED TAGS**[HEALTH](#)[BIOLOGY](#)[MEDICINE](#)

Share: [f](#) [t](#) [link](#) [envelope](#)