

Researchers generate complete human X chromosome sequence

July 14 2020



Image depicting the puzzle pieces of DNA sequences coming together. Credit: Ernesto del Aguila III (NHGRI)

Researchers at the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health (NIH), have produced the first end-to-end DNA sequence of a human chromosome. The results, published today in the journal *Nature*, show that generating a precise, base-by-base sequence of a human chromosome is now possible, and will enable researchers to produce a complete sequence of the human genome.

"This accomplishment begins a new era in genomics research," said Eric Green, M.D., Ph.D., NHGRI director. "The ability to generate truly complete sequences of chromosomes and genomes is a technical feat that will help us gain a comprehensive understanding of genome function and inform the use of genomic information in medical care."

After nearly two decades of improvements, the reference sequence of the human genome is the most accurate and complete vertebrate genome sequence ever produced. However, there are hundreds of gaps or missing DNA sequences that are unknown.

These gaps most often contain repetitive DNA segments that are exceptionally difficult to sequence. Yet, these repetitive segments include genes and other functional elements that may be relevant to human health and disease.

Because a human genome is incredibly long, consisting of about 6 billion bases, DNA sequencing machines cannot read all the bases at once. Instead, researchers chop the genome into smaller pieces, then analyze each piece to yield sequences of a few hundred bases at a time. Those shorter DNA sequences must then be put back together.

Senior author Adam Phillippy, Ph.D., at National Human Genome Research Institute (NHGRI) compared this issue to solving a puzzle.

"Imagine having to reconstruct a jigsaw puzzle. If you are working with smaller pieces, each contains less context for figuring out where it came from, especially in parts of the puzzle without any unique clues, like a blue sky," he said. "The same is true for sequencing the human genome. Until now, the pieces were too small, and there was no way to put the hardest parts of the genome puzzle together."

Of the 24 human chromosomes (including X and Y), study authors

Phillippy and Karen Miga, Ph.D., at the University of California, Santa Cruz, chose to complete the X chromosome sequence first, due to its [link with a myriad of diseases](#), including hemophilia, chronic granulomatous disease and Duchenne muscular dystrophy.

Humans have two sets of chromosomes, one set from each parent. For example, biologically female humans inherit two X chromosomes, one from their mother and one from their father. However, those two X chromosomes are not identical and will contain many differences in their DNA sequences.

In this study, researchers did not sequence the X chromosome from a normal human cell. Instead, they used a special cell type—one that has two identical X chromosomes. Such a cell provides more DNA for sequencing than a male cell, which has only a single copy of an X chromosome. It also avoids sequence differences encountered when analyzing two X chromosomes of a typical female cell.

The authors and their colleagues capitalized on new technologies that can sequence long segments of DNA. Instead of preparing and analyzing small pieces of DNA, they used a method that leaves DNA molecules largely intact. These large DNA molecules were then analyzed by two different instruments. Each of them generates very long DNA sequences—something previous instruments could not accomplish.

After analyzing the human X chromosome in this fashion, Phillippy and his team used their newly developed computer program to assemble the many segments of generated sequences. Miga's group led the effort to close the largest remaining sequence gap on the X chromosome, the roughly 3 million bases of repetitive DNA found at the middle portion of the chromosome, called the centromere.

There is no "gold standard" for researchers to critically evaluate the

accuracy of assembling such highly repetitive DNA sequences. To help confirm the validity of the generated sequence, Miga and her collaborators performed several validation steps.

"We have never actually seen these sequences before in our genome, and do not have many tools to test if the predictions we are making are correct. This is why it is important to have specialists in the genomics community weigh in and ensure the final product is high-quality," Miga said.

The effort is part of a broader initiative by the Telomere-to-Telomere (T2T) consortium, partially funded by NHGRI. The consortium aims to generate a complete reference sequence of the human genome.

The T2T consortium is continuing its efforts with the remaining human chromosomes, aiming to generate a complete human genome sequence in 2020.

"We don't yet know what we'll find in the newly uncovered sequences. It is the exciting unknown of discovery. This is the era of complete genome sequences, and we are embracing it wholeheartedly," Phillippy said.

Potential challenges remain. Chromosomes 1 and 9, for example, have repetitive DNA segments that are much larger than the ones encountered on the X chromosome.

"We know these previously uncharted sites in our genome are very different among individuals, but it is important to start figuring out how these differences contribute to human biology and disease," Miga said. Both Phillippy and Miga agree that enhancing sequencing methods will continue to create new opportunities in human genetics and genomics.

Provided by NIH/National Human Genome Research Institute

Citation: Researchers generate complete human X chromosome sequence (2020, July 14)
retrieved 15 May 2024 from <https://medicalxpress.com/news/2020-07-scientists-closer-entire-human-genome.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.