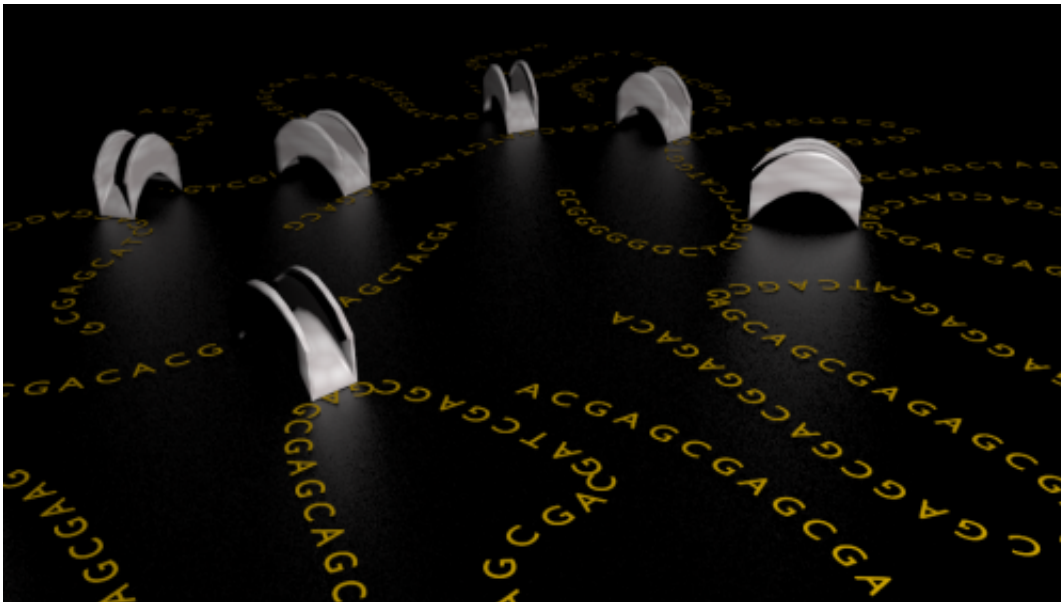


# Scientists re-imagine how genomes are assembled

November 25 2013

---



3D interaction frequencies can be used to bridge gaps in the DNA sequence and assemble genomes. Credit: Noam Kaplan, PhD, UMass Medical School

Scientists at the University of Massachusetts Medical School (UMMS) have developed a new method for piecing together the short DNA reads produced by next-generation sequencing technologies that are the basis for building complete genome sequences. Job Dekker, PhD, and colleagues have shown that entire genomes can be assembled faster and more accurately by measuring the frequency of interactions between DNA segments and by using their three-dimensional shape as a guide.

Employing this technique, they have been able to place 65 previously unaccounted for DNA fragments in incomplete regions of the human genome.

Details of the study appear online in *Nature Biotechnology*.

"The ability of next-generation sequencing technologies to produce hundreds of millions of short reads of DNA sequences has been an incredible boon for biomedical researchers," said Dr. Dekker, co-director of the Program in Systems Biology, professor of biochemistry and molecular pharmacology at UMMS and senior author of the study. "As these DNA sequences have become shorter and shorter, however, assembling complete genomes have become increasingly challenging. After 20 years of intense efforts, even the [human genome](#) still has gaps.

"Using the 3D structure of the [genome](#) as a guide, we have shown that it's possible for these snippets of DNA sequences to be assembled quickly, cheaply and more accurately than current methodologies allow. This elegant and powerful technique will allow us to complete the human genome, assemble the genomes of any other species and facilitate new genetic discoveries more quickly."

In the last decade, as the cost of high-throughput DNA sequencing has come down to as little as a few thousand dollars, sequencing of new genomes has become almost routine. Next-generation sequencing techniques can easily read hundreds of millions DNA sequences at a time. However, these sequences are randomly broken into extremely short pieces and need to be assembled into larger pieces using computer algorithms that can match up overlapping pieces. The end result of this initial assembly is typically a set of as many as 100,000 DNA fragments which then need to be organized with respect to one another in the correct order to create a complete genome.

Hindering this final task is the fact that genomes are full of highly repetitive sequences that appear in a multitude of places. Finding where precisely, among the thousands of possible locations, a particular fragment of DNA resides, is a daunting task. To complete this second step, often referred to as genome scaffolding, scientists rely on labor intensive, low-through put experimental techniques to build reasonably accurate, complete genomes.

"How to assemble these snippets of DNA has become a bottleneck for researchers that can take weeks or months to solve," said Noam Kaplan, PhD, postdoctoral research fellow in the Dekker lab and first author of the *Nature Biotechnology* study.

Tackling this problem, Dekker and Kaplan looked to the three-dimensional structure of the genome as a guide for assembling the linear DNA sequences. Using Hi-C technology, developed by the Dekker lab, they measured how frequently each DNA fragment in the genome interacts with others. DNA sequences that are located near each other in the three dimensional genome tend to interact more frequently, while DNA sequences that are further apart interact less frequently. Computational methods are then used to mathematically determine the linear genomic position of each fragment in the genome based on the 3D interaction frequency data that fits that sequence.

For example, said Kaplan, a sequence may fit into the one-dimensional linear genome in several places. But using the interaction frequency data, it is possible to determine the relationship it has with other sequences and whether it is close to or far away from those sequences. "So while a particular sequence may fit in many places in a linear genome, we can determine if a particular sequence is a better fit, three dimensionally, in one location versus another, based on this interaction data," said Dr. Kaplan.

With this new approach Kaplan and Dekker were able to predict the positions of 65 previously unlocalized fragments.

"We were surprised how well our method worked," said Kaplan. "It is satisfying to see how a simple idea can solve such a difficult and common problem."

Dekker added, "This new approach to genome assembly can help produce higher-quality genome sequences faster and easier than current methods. It will be especially interesting to apply this method to identify chromosomal aberrations, which are a hallmark of cancer."

**More information:** High-throughput genome scaffolding from in vivo DNA interaction frequency, [DOI: 10.1038/nbt.2768](https://doi.org/10.1038/nbt.2768)

Provided by University of Massachusetts Medical School

Citation: Scientists re-imagine how genomes are assembled (2013, November 25) retrieved 10 April 2024 from <https://phys.org/news/2013-11-scientists-re-imagine-genomes.html>

<p>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.</p>
--