

Team develops benchmark for detecting large genetic mutations linked to major diseases

June 15 2020



The 'book analogy' is a popular way of explaining genome changes. If the human genome is considered a book, a new NIST-developed benchmark will help scientists better detect large chapters that are missing (deleted chapters) or not in the original (inserted chapters). Credit: N. Hanacek/NIST

Many serious diseases, including autism, schizophrenia and numerous cardiac disorders, are believed to result from mutation of an individual's DNA. But some large mutations, which still make up only a small fraction of the total human genome, have been surprisingly challenging to detect.

Now, researchers at the National Institute of Standards and Technology (NIST) have developed a way for laboratories to determine how accurately they can detect these mutations, which take the form of large insertions and deletions in the human [genome](#). The new method and the benchmark material enable researchers, clinical labs and commercial technology developers to better identify large genome changes they now miss and will help them reduce false detections of genome changes.

The researchers present their new benchmark in *Nature Biotechnology*.

Scientists in the Human Genome Project generated the first reference genome in the late 1990s, pieced together from a collection of genome sequences from different individuals. When scientists sequence DNA, they are essentially randomly chopping up the DNA into smaller pieces, which then need to be pieced back together like a puzzle.

The building blocks of DNA include four types of bases: adenine (A), cytosine (C), guanine (G) and thymine (T), strung together to form 23 chromosomes in human cells. These genetic codes contain all the information of life. To understand the genetic basis for a given disease, scientists sequence a person's DNA and compare it against a reference genome. Differences between the individual's DNA sequence and the reference genome are called variants. Some of these variants, which can range from insertions and deletions of 50 to tens of thousands of letters (or bases) of the roughly 6.4 billion bases that make up the human genome, are found to be linked to a disease.

Over the last eight years, the NIST-led Genome in a Bottle consortium (GIAB), which includes members from the federal government, academia and industry, developed whole human genome benchmarks for small variants for seven individuals. For this new paper, NIST worked with GIAB to develop a new benchmark for large insertions and deletions. To form this benchmark, NIST integrated results from 19 different analysis approaches by GIAB members, using GIAB's public data from a well-characterized set of human DNA from a family of Eastern European Ashkenazi Jewish ancestry (NIST Reference Material 8392).

The NIST Genome in a Bottle Consortium is a public-private-academic consortium hosted by NIST to develop the technical infrastructure (reference standards, reference methods, and reference data) to enable translation of whole human genome sequencing to clinical practice. In this animation, learn more about the genome sequencing process and why standards are such an important part of this process.

"Just like a company making rulers could compare their ruler to a standard measuring stick to make sure it is measuring the correct distance, clinical laboratories doing DNA sequencing can measure NIST reference material DNA and compare their answer to this new benchmark to help make sure they measure large insertions and deletions well," said NIST biomedical engineer Justin Zook.

Laboratories have accurately detected many small insertions and deletions in the genome for years. One would think detecting larger insertions and deletions would be easier, but it's actually harder because "the most widely used sequencing technologies output relatively short strings of genetic code, making it hard to reconstruct what's happening," says Zook. With new DNA sequencing technologies, it is now possible to detect many more large insertions and deletions.

Imagine the genome as a book. The benchmark helps scientists detect large chapters that are missing (deleted chapters) or not in the original (inserted chapters).

"DNA sequencing is like shredding the book into smaller pieces and then trying to find any differences between the book that was shredded and a similar book, perhaps the same book before it went through editorial revisions," said Zook. Even though the DNA is broken into [smaller pieces](#), the new DNA sequencing technologies make it possible to read the larger pieces, making it easier to find these larger insertions and deletions.

The NIST Genome in a Bottle Consortium is a public-private-academic consortium hosted by NIST to develop the technical infrastructure (reference standards, reference methods, and reference data) to enable translation of whole [human genome](#) sequencing to clinical practice. In this animation, learn more about why developing these reference materials is so important.

This benchmark for large insertions and deletions will improve the accuracy of DNA sequencing technologies and analysis methods, reducing the likelihood of errors such as false positives and negatives. A false positive means detecting an [insertion](#) or deletion in the genome that's not real, while a false negative means not detecting a change in the genome when it's actually there.

Reducing false positive and negative numbers is critical, especially in clinical settings where many diseases such as autism, schizophrenia and cardiovascular disease have been linked to structural variants. For example, if a clinical laboratory is sequencing a patient's DNA, a false negative can result in missing the change in the genome that is causing the disease, leading to incorrect treatments.

Down the road, applications of the benchmark will help labs detect disease-associated structural variants by validating their methods.

For NIST researchers, next steps include characterizing difficult regions of the genome that contain repetitive sequences. DNA sequence technologies and methods continue to improve, enabling researchers to push into more challenging regions of the genome and identify structural variants that are harder to detect.

But according to Zook, this is precisely why this area is fun to work in, as technologies have changed and improved in the past 30 years. He credits the collaboration with GIAB as being key to these efforts: "All of this work wouldn't be possible if we weren't able to collaborate with a group of diverse people with different areas of expertise."

More information: Justin M. Zook et al, A robust benchmark for detection of germline large deletions and insertions, *Nature Biotechnology* (2020). [DOI: 10.1038/s41587-020-0538-8](https://doi.org/10.1038/s41587-020-0538-8)

Provided by National Institute of Standards and Technology

Citation: Team develops benchmark for detecting large genetic mutations linked to major diseases (2020, June 15) retrieved 23 April 2024 from <https://phys.org/news/2020-06-team-benchmark-large-genetic-mutations.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.