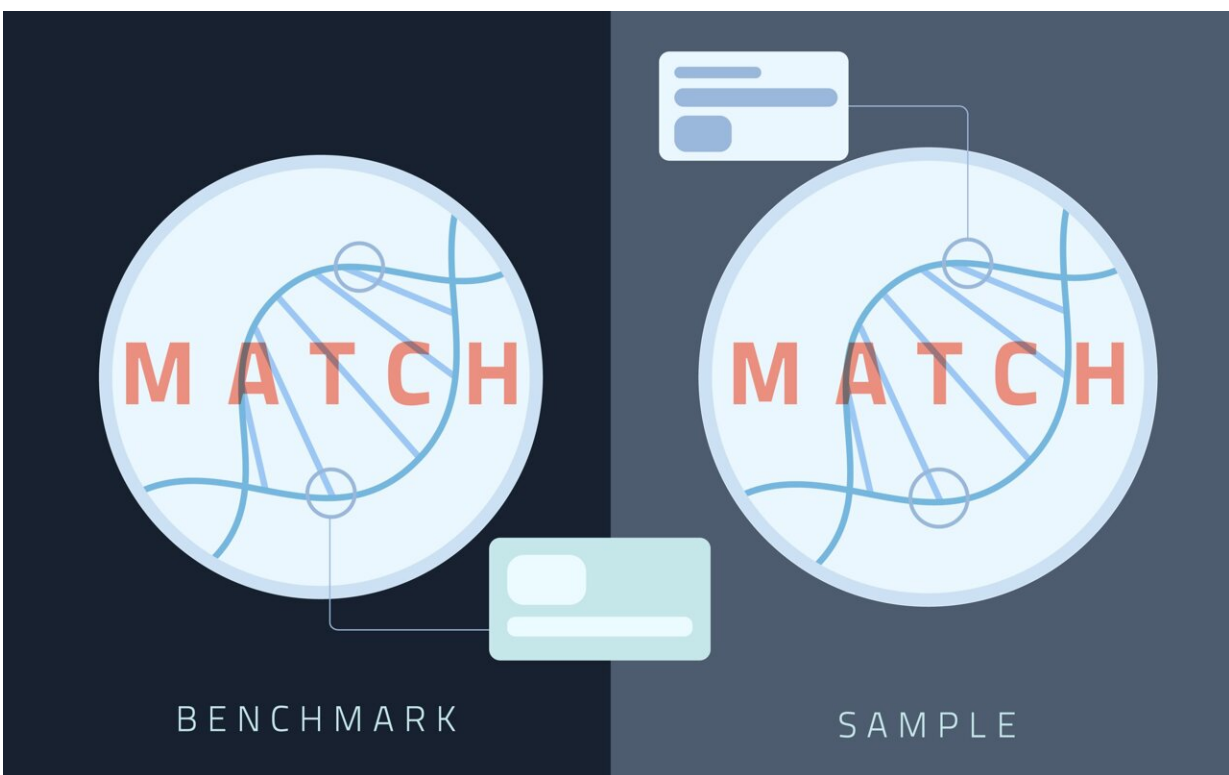


# New benchmark could improve detection of genetic variants linked to spinal muscular atrophy, other diseases

February 7 2022

---



NIST's genome sequencing benchmarks are highly accurate sequences of DNA that clinics and research labs can use as a kind of answer key when testing their own sequencing methods. Credit: B. Hayes/NIST

The stretches of DNA that differ from person to person, called variants,

are a major part of what makes us unique, but they can also put us at greater risk of disease. Although we can currently spell out between 80% and 90% of the millions that are in the human genome, the remaining variants may hold clues for treating an array of diseases. Today the list of variants yet to be decoded has shrunk sizably.

A team led by researchers at the National Institute of Standards and Technology (NIST), Baylor College of Medicine and DNAnexus has characterized over 20,000 variants in 273 genes of medical importance. In a study published in the journal *Nature Biotechnology*, the researchers applied both cutting-edge and long-standing DNA sequencing methods to decipher the genetic codes of the variants with a high degree of certainty. Using their results, they formulated benchmarks that will help labs and clinics sequence the genes more accurately, which is critical for gaining a better understanding of a host of diseases and eventually developing treatments.

"Some of these genes, which have previously been very difficult to access, are suspected to have some connection to disease. Others have very clear clinical importance," said NIST biomedical engineer Justin Zook, a co-author of the study. "SMN1, for example, is a gene we characterized that is directly associated with spinal muscular atrophy, a rare but severe condition."

The new benchmark is the latest produced by the Genome in a Bottle (GIAB) consortium, a NIST-hosted collaborative effort aimed at improving DNA sequencing technologies and making them practical for clinical application.

These benchmarks are highly accurate sequences of DNA that clinics and research labs can use as a kind of answer key when testing their own sequencing methods. By sequencing the same genome used to develop a benchmark and then comparing their result to the benchmark itself, they

could learn how well they can detect certain variants.

Over the years, producing benchmarks for some regions of the genome has proved much more difficult than others. There are several reasons, many of which are tied to the general approach people use to sequence DNA.

Rather than sequencing entire genomes in one go, DNA sequencing technologies read out sequences of small fractions of DNA first, and then attempt to place them together correctly, similar to a puzzle set. Reference genomes, the first of which was completed by the [Human Genome Project](#), are nearly full genomes, stitched together from several people's DNA, that serve as guides for where to place the puzzle pieces.

Since we share close to [99.9% of our genetic makeup](#) as a species, any human genome will have mostly the same code as the reference genome. This means putting together a genome is a matter of laying out the pieces based on where they match up with the reference. Most variants fall in line using this process. Certain types throw a wrench into it.

In particular, a type called a structural [variant](#) can create large differences between a genome and a reference genome. They range from 50 up to thousands of letters, or bases, and take many forms, including inserted, deleted or rearranged code. The more distinct a genome is from the reference, the harder it is to use the reference as a guide, Zook said.

Structural variants could cause labs to unintentionally misplace chunks of DNA, and, in a clinical setting, that sort of error may cause a disease-linked variant to evade detection or a harmless variant to create alarm. On top of the human costs, treatments prescribed needlessly or too late due to these mismeasurements could establish the need for more expensive or invasive treatments for patients down the road, driving up health care costs drastically.

However, recent advances in sequencing technology have cleared some of these obstacles. In the new study, the GIAB consortium applied the latest technology to decode some of the most elusive regions of the [human genome](#) with either a known or suspected connection to diseases.

A key player in the effort was high fidelity, or HiFi, sequencing, which can sequence longer stretches of DNA. Common DNA sequencing methods can read about a hundred bases, but with HiFi sequencing, you can accurately read tens of thousands at a time, Zook said.

"Instead of having a thousand-piece puzzle, where you have these little, tiny pieces that you have to put together, it's more like having a hundred-piece puzzle where you have bigger pieces that you can put together," Zook said.

The team specifically employed HiFi with hifiiasm, a state-of-the-art software tool that simultaneously solves another issue that has hampered DNA sequencing.

Rather than reading both copies of an individual's chromosomes (one from mother, the other from father), previous methods sequenced an amalgamation of both, causing them to create errors and miss important details unique to each copy.

With hifiiasm, the researchers could independently spell out the separate copies of a person's genome. In the case of this study, the [genome](#) was from a single person, designated HG002, who had consented to publicizing their genetic code through the [Personal Genome Project](#).

The authors used these technologies in addition to previously established methods, leveraging the strengths of each at once. In the end, their approach allowed them to unearth the sequences of more than 20,000 variants—including dozens of the difficult-to-assess structural

variants—across 273 genes, and did so with higher accuracy than could be achieved just using a single method.

In addition to [spinal muscular atrophy](#), the researchers characterized variants in genes connected to heart disease, diabetes, celiac disease and many other conditions.

The team also unexpectedly encountered errors in the two reference genomes they were using. Some could cause sequencing methods to misread genes that cause serious conditions, including homocystinuria, which is associated with skeletal, cardiovascular and nervous system disorders and is usually detected through newborn screening, Zook said. With their newly benchmarked variants, the authors proposed corrections to the reference genomes they used.

The benchmarks themselves are [now publicly available](#) for labs to put to good use. To do so, interested researchers or clinicians would first need to sequence HG002 samples, which can be accessed through the NIST Office of Reference Materials, and then check their results against the benchmarks.

The study marks a significant step in the GIAB consortium's ongoing journey to improve the accuracy of DNA sequencing. But with thousands of important genes left to characterize containing variants that are difficult to pin down, the researchers aim to trudge on, applying the latest and greatest technologies as they become available.

**More information:** Chen-Shan Chin, Curated variation benchmarks for challenging medically relevant autosomal genes, *Nature Biotechnology* (2022). [DOI: 10.1038/s41587-021-01158-1](https://doi.org/10.1038/s41587-021-01158-1). [www.nature.com/articles/s41587-021-01158-1](https://www.nature.com/articles/s41587-021-01158-1)

Provided by National Institute of Standards and Technology

Citation: New benchmark could improve detection of genetic variants linked to spinal muscular atrophy, other diseases (2022, February 7) retrieved 25 April 2024 from

<https://medicalxpress.com/news/2022-02-benchmark-genetic-variants-linked-spinal.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.