

Studies Show Benefits of Long-Read Sequencing in Toughest Rare Disease Cases

Aug 27, 2021 | [Andrew P. Han](#)

Premium

NEW YORK – Long-read sequencing can successfully provide diagnoses in previously intractable hereditary disease cases, according to recent studies, suggesting that long-read, whole-genome sequencing is on track to become the first-line genetic test of choice.

Last month, researchers led by Danny Miller and Evan Eichler of the University of Washington described the use of adaptive sampling on the Oxford Nanopore Technologies Gridlon platform to identify previously undetected, disease-causing genomic variation in individuals who remained undiagnosed after a complete clinical genetics work-up. In some cases, this included exome sequencing or short-read whole-genome sequencing.



The team used a targeted approach to find pathogenic or "likely pathogenic" variants in six out of 10 individuals with suspected Mendelian conditions and variants of unknown significance in two more. They also were able to identify previously known single-nucleotide variants, copy number changes, repeat expansions, and even methylation differences seen in prior testing of 40 patients, providing more detail on structural changes in about 20 cases.

"There's a cost in genetics of running multiple tests and bringing the patient back," Miller said. "This tech can simplify that process and make the analysis a lot more straightforward. It's beneficial for patients and families and for the healthcare system."

Their results, published in the [American Journal of Human Genetics](#), are congruous with similar studies at the HudsonAlpha Institute for Biotechnology and Children's Mercy Hospital, each using long-read technology from Pacific Biosciences.

"We do find that as presented here, too, that there is clinically significant genetic variation that is undetectable by short-read sequencing," said Tomi Pastinen, director of the Genomic Medicine Center at Children's Mercy. His team is working on a study of more than 200 cases and presented data on 100 patients at this year's American College of Medical Genetics and Genomics virtual conference. The data out so far are evidence that current methods of genetic testing for rare diseases, whether by short-read sequencing or microarrays, are incomplete, he said.

Pastinen suggested that the targeted approach used in the UW-led paper could be used as "a follow-up tool" for other tests or strong clinical hypotheses. "It's not a quantum leap," he said. "But they tested a number of positive controls, which is a nice feature."

Miller said the targeted method used by his team was more proof of concept and that long-read WGS "will eventually be the only clinical genetic test we do," he said. "Because it's a single dataset, you can query multiple times."

Yet many challenges remain, including developing bioinformatics tools and reference datasets and building a case for reimbursement.

As the name implies, long-read sequencing provides the ability to analyze longer stretches of the genome without having to piece them together from smaller parts, including regions where short reads crap out, such as repetitive regions. Those regions often contain so-called structural variants and many studies have shown that they are associated with disease, including cancers.

Long reads have been effective in detecting SVs but they have their drawbacks. Generally speaking, they offer lower throughput than short-read platforms and, until recently, had significantly lower single-read accuracy.

And they're not without blind spots. A [recent study](#) from the Human Genome Structural Variation Consortium published in the *American Journal for Human Genetics* analyzed SVs in samples from the 1000 Genomes Project. The study found that assembly-based methods of sequencing, which often are based on long reads, missed some large copy number variants that are detected with other methods.

Still, many researchers, like Miller, believe long read WGS is the future of clinical genetic testing, and the companies that make the technologies, namely PacBio and Oxford Nanopore, are driving proof-of-concept studies. In addition to its [collaboration](#) with Children's Mercy, PacBio is [working with Rady Children's Hospital](#) in San Diego on a similar study. It has also [partnered with Invitae](#) to build an instrument for clinical long-read WGS. Oxford Nanopore, over in the UK, is developing its "Q Line" of instruments intended for clinical use.

Miller, a resident physician in the UW division of medical genetics with a doctorate in physiology, said he developed his chops on the Oxford Nanopore platform sequencing fruit fly genomes but "always had an eye on what I could do with humans, eventually."

The low barrier to entry made his study an "easier pitch, from the perspective of me, at my training level," he said. While Oxford Nanopore also offers ultra-long reads of up to 4 Mb, Miller said he was fine working with 50 kb to 60 kb reads. "They're very useful and I think they will be useful clinically."

For the study, Miller and his team used a special feature of the nanopore platform: the ability to preselect genomic targets and have the device spit out any reads that don't match. This "read-until" feature was introduced in 2014 but [unlocked last year for targeted sequencing](#).

Miller said it's a fast and "straightforward" way of targeting sequences without using hybridization or amplification chemistries. "You just go to a genome browser, type in a gene, get the coordinates, and put it in a BED file," he said. "Then you're done." Moreover, it's pretty cheap. When purchasing reagents at scale, nanopore sequencing costs about \$650 per sample, he said, compared to about \$1,000 for short-read sequencing.

Pastinen said he wished the authors had compared the "real-life benefits of targeting reads versus doing Oxford Nanopore whole-genome sequencing."

One advantage of WGS is that it's unbiased, said Susan Hiatt, first author of the HudsonAlpha study, published in April in [Human Genetics and Genomics Advances](#). "Whole-genome long-read sequencing is the best way to go for sure. You can do this targeted sequencing, too, if you know where to look."

In the study, her team found disease-causing structural variants in two out of six cases. "We had a guess there was some sort of structural variant, but we weren't looking for a particular gene or set of genes," she said.

"It's only six cases, so we don't know if that's the real diagnostic rate," Hiatt noted. Establishing a diagnostic yield will be a critical next step for the technology. Her team will be looking at another 200 probands over the course of the year using PacBio's platform. Already, they're seeing results. "It's showing

us that we are going to find a significant number of variants," she said.

Finding variants is one thing, but putting them in context will be its own challenge. In the UW study, Miller said they were allowed to tell patients and their providers the study results if they wanted to receive them. "We clinically validated some of these findings, but in all cases, the institutional review board did not allow us to interpret the results," he said. "It's going to get even more challenging when we do whole-genome sequencing and find novel structural variants. How do you explain a complex expansion of a repeat that alters methylation? There will be a lot of interesting genetic counseling with long-read sequencing."

Improving the amount and availability of reference data will be key to making calls about clinical significance, Pastinen said. "In principle, you need reference data on similarly targeted but nonaffected samples," he said. "The data resources are not there yet for a robust rollout of all variants. Most of our reference data is our own data of a very small number of individuals. It's insufficient for high production level analysis."

But WGS had the benefit of generating reference data across the genome, which can be used for future cases. The All of Us project will be generating some long-read data from a "normal" population, but Pastinen said there needs to be more data sharing.

Bioinformatics tools are yet another resource the field still needs to develop. In addition to providing more data, HudsonAlpha's additional cases are giving the researchers a chance to experiment with different bioinformatics pipelines.

"There's a lot of different options out there. This will get us comfortable enough to say, 'This is the way we're going to go, so now we can scale it up,'" Hiatt said. So far, a lot of their pipeline comes directly from PacBio, including their SV caller.

And, of course, the field needs to do the ultimate head-to-head comparison with short reads, which would be a start to solving the problem of reimbursement.

"There hasn't, to my knowledge, been a systematic study comparing the incremental diagnostic rate of long reads over short reads," Miller said. "That's probably what we really need to get payors to reimburse for the test."

"I don't know how insurance will respond," he added. "Just showing we can solve rare disease cases, I don't know if that's enough."

Filed Under [Sequencing](#) [Molecular Diagnostics](#) [Clinical Sequencing](#) [North America](#)

[University of Washington](#) [HudsonAlpha](#) [Children's Mercy](#) [targeted sequencing](#)

[whole-genome sequencing](#) [Advances in Clinical Genomics Profiling](#)

[Genomics: Clinical Implementation](#) [Advances in Clinical Genomics Profiling](#)

[Genomics: Clinical Implementation](#)