Long Reads Can Make Short Work of Rare Disease Diagnostics

Recent success stories bring hope that long read sequencing may make diagnosing rare genetic disease more successful

By Julianna LeMieux, Ph.D. (https://www.clinicalomics.com/author/JuliannaLeMieux   )   - August 25, 2021

[Source: jiwasa / Getty Images]

In early June, headlines trumpeted another astonishing feat accomplished by the team at Rady Children's Institute for Genomic Medicine (RCIGM) in San Diego. The study, published in the New England Journal of Medicine, described the sequencing of a newborn's genome and subsequent diagnosis of a rare genetic disease, Thiamine Metabolism Dysfunction Syndrome 2. Within just 13 hours, the Rady team, led by Stephen Kingsmore, sequenced the baby's genome, produced a diagnosis, and started treatment. In 2015, Kingsmore, the RCIGM president and CEO, and colleagues earned the Guinness World Records title for fastest genetic diagnosis. Since then, he has broken his own record more than once.

For many, there is no question that harnessing the power of genomic sequencing for rapid diagnosis of rare genetic diseases is imperative for the medical field to embrace. Several high-profile cases have made headlines, while many others do not make it into the news.

The team at RCIGM accomplished their successes using a rapid whole-genome sequencing (WGS) method known as STATseq. At the center of the

protocol is Illumina sequencing—a sequencing by synthesis (SBS) technology. The short sequences (in the hundreds of bases) are mapped and aligned to the human reference genome. The individual DNA fragments may be short, but the overall sequence is incredibly accurate.

Alex Dickinson, Chromacode

The short read method utilized by RCIGM works in some cases, but not all. Alex Dickinson, Ph.D., chairman of Chromacode, explained that short-read sequencing typically identifies the genetic mutation in a sick newborn about 50% of the time. "That's a long-held number," he stated. Some believe, or hope, that incorporating long read sequencing will allow the other 50% to be found.

Why might longer reads improve genetic disease diagnosis? The answer lies, in part, in the types of mutations that cause disease. Some genetic diseases are caused by single nucleotide variants (SNVs), which require a high level of sequencing accuracy to detect—an area where short reads excel. But other deleterious genomic changes are larger and more complex —including large insertions, deletions, translocations, copy number variants, and structural variants. These larger variations can be missed by short reads but detected in longer sequence reads.

Dickinson uses a jigsaw puzzle analogy. Assembling a genome using short-read sequencing is like putting together a 1,000-piece jigsaw puzzle, where each piece is the same shape such that many pieces end up in the wrong place during assembly—particularly in areas of the picture that are very repetitive. In contrast, long read sequencing is like a puzzle with a few larger pieces—much easier to assemble without errors, even in those repetitive areas.

PacBio and RCIGM team up

Matthew Bainbridge, Ph.D., Rady Children's Institute of Genomic Medicine

In late June, RCIGM announced a joint study with Pacific Biosciences (PacBio) using long-read WGS to diagnose rare diseases. Matthew Bainbridge, Ph.D., associate director of clinical genomics research at RCIGM, told Clinical OMICs that their primary goal is "ultra-rapid sequencing in critically ill infants… to get this technology available throughout the country." After five years, he noted, "we felt that this goal was largely accomplished and our processes hardened enough that we could devote more time to drilling down into our unsolved cases. Further, the technology has greatly improved in accuracy and decreased in price. All of which made now the right time for RCIGM to look at long reads."

Jennifer Stone, Ph.D., vice president of marketing at PacBio said that the study will "explore rare disease cases with the aim to identify numerous variants, both small and structural, that are not readily detectable by short-read sequencing."

PacBio's sequences, which are referred to as HiFi reads, are much longer than Illumina's short reads—up to 25 kilobases (kb). Crucially, PacBio claims these reads are "just as accurate" as short-read sequencing (>99.9%).

PacBio says with HiFi reads, "you no longer need to choose between read accuracy and read length. You can now get highly accurate long reads that can discover all variant types, ranging from CMVs to large structural variants."

Bainbridge explained that "read lengths of 10 kb (or more) is more than suitable to interrogate essentially the entire genome." HiFi, he added, "also has very high single-base accuracy, which means with as little as 20x coverage you can assess both chromosomes and identify single-base

indels with high confidence."

Jenny Ekholm, PacBio

During a virtual event held during rare disease week last April, Jenny Ekholm, Ph.D., staff segment marketing manager at PacBio, gave a talk entitled, "Increasing Solve Rates in Rare and Mendelian Disease Research with Long-Read Sequencing." Ekholm noted that PacBio WGS was first used to diagnose the genetic cause of a Mendelian disease in 2017, in work led by Euan Ashley, M.D., associate dean of the Stanford University School of Medicine. Using HiFi, Ashley's group identified a 2.2-kb heterozygous deletion in a gene associated with the disorder of a 22-year-old male with recurring heart tumors. Ekholm highlighted more recent case studies. One identified a 7-kb insertion in a young girl with intellectual disability, seizures and speech delay; another case found a 12-kb inversion in a five-month-old girl diagnosed with intellectual disability.

What percentage of undiagnosed cases does Bainbridge think they can catch through this long-read project? He answers that their current pipelines, which are fairly sophisticated, push what can be found with short reads to their limit. Thus, he adds, "all that's left are really hard cases." He continued: "If we could solve just an additional 3% to 5% of cases that would be tremendous."

A physician's perspective

Danny Miller, M.D., Ph.D., a resident in the division of medical genetics at the University of Washington, offers a straightforward reason to use long-read sequencing for diagnostics: it helps solve cases. Long-read data is unique, he explained, because it provides a single data source that has the potential to replace nearly all current clinical genetic testing methods. This ability decreases the current number of steps in the process, such as visits to the clinic and blood draws, all of which take time and impact people's

lives.

Danny Miller, University of Washington

In addition, the information held in long-read data surpasses the genetic information currently available to patients. Currently, Miller explained, the typical methods applied to genetic testing are a microarray followed by exome sequencing. After those two methods, the only remaining option is to reanalyze the exome every few years, which is limiting and will never reveal certain mutations, such as intronic variants. It is very rare that a physician obtains a whole-genome sequence, he noted, for the simple reason that insurance companies rarely reimburse for it.

It's difficult, Miller asserted, to have to explain to a parent that while it is suspected their child may have a genetic disorder, there is only about a 50% chance that the cause can be identified. And, to make matters worse, that it may take several years to complete the workup. Long-read sequencing has the potential to reduce the time it takes to complete this workup to days or weeks, while increasing the number of cases solved.

In July, Miller was the lead author on a paper from Evan Eichler's lab at the University of Washington, published in The American Journal of Human Genetics. The work used targeted long-read sequencing on Oxford Nanopore Technologies' (ONT) GridION platform. Of the 40 people tested, 30 were controls while ten lacked a complete diagnosis. Of those ten individuals, all with suspected Mendelian conditions, the team was able to identify pathogenic or likely pathogenic variants in six and variants of uncertain significance in two others.

A scientist uses a pipette to place a sample on a GridION in an Oxford Nanopore lab.

The most exciting thing their data show, according to Miller, is that the long-read sequencing had 100% sensitivity. Although it's a small sample size, he noted, "it was important that we saw everything that we should have seen." Moreover, they were able to end the diagnostic odyssey for some families, where they would have typically run out of clinical testing options.

The team chose ONT's technology, Miller said, largely because of the simplicity of the library prep and cost—it's cheaper on a per sample basis. Although ONT's platform is reported to have higher error rates than other technologies, with sufficient coverage, Miller said the sensitivity may be good enough for clinical testing.

In the long run, Miller predicted that long reads will be a single source of data for genetic testing. In order to make this a reality, the sequencing technology needs to be made more available. But that's not enough. "We also need to understand more information about relevant variants," Miller said. Bainbridge agreed, adding that as databases get further populated by long-read sequencing variant calls, "we'll have a better understanding of what the unsequenceable-by-short-read genome looks like."

Even with the sequence in hand, there may not be enough technical knowledge to solve the case. Miller's long-term goal is to take those cases into the lab and open up new areas of biology.

Length and accuracy without the cost

The technology being developed at the company Quantapore is a mash-up, of sorts, of multiple available sequencing technologies. As such, it is also a mash-up of their qualities.

Martin Huber, Ph.D., formally of Ion Torrent, co-founded Quantapore and invented their nanopore-based sequencing technology. Nanopore technologies pull a single strand (ss) of DNA through a nanopore and infer the underlying sequence by measuring the electrical current in real time as the bases pass through the pore. This method has an error rate that is "still pretty high," suggested Dickinson, a member of the Board of Directors at Quantapore.
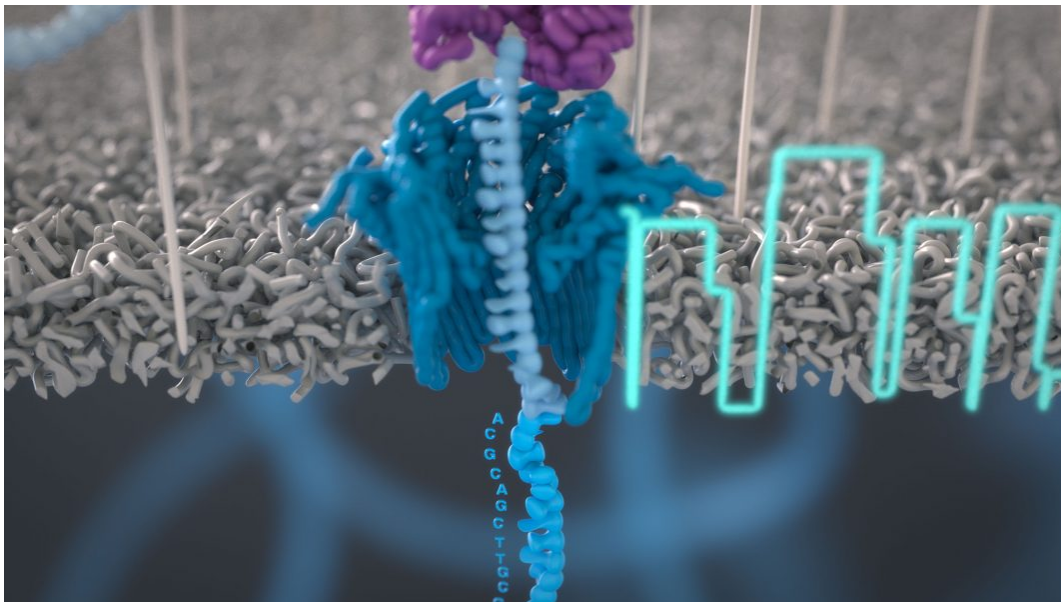
Illustration of a strand being fed through a nanopore with the resulting signal. [Source: Oxford Nanopore Technologies]

By contrast, Quantapore's technology attaches a different colored fluorophore to each base, similar to many short-read sequencing methods. As the ssDNA is pulled through the pore, the fluorophore is sliced off and read by an optical detection system rather than an electrical system. The technology combines what Dickinson says is the best of the nanopore designs while eliminating the weaknesses associated with electrical detection.

Quantapore is not a new company. Founded in 2009, it has been working quietly to develop the platform. The company "just worked out how to make the whole process work," notes Dickinson.

The next phase is to build it into a platform. The small company (under 20 people) told Clinical OMICs that they achieved their proof of concept earlier this year and are targeting 2023 for their launch.

Quantapore's fit in the already-crowded arena of genomic sequencing may be somewhere in between the longest reads and the cheapest platform. Their reads will be 2-5 kb fragments, not the 15-25 kb stretches produced by the long-read technologies. The optical readout, noted Huber, makes the system much more scalable than nanopores using an electrical readout because you need to separate nanopores for an electrical readout—which requires real estate. The technology can also read long stretches of homopolymers in the DNA. And its consumables, the company notes, will cost less than their competitors.

"In theory, long reads are a great idea for many reasons," said Dickinson. But as long as long reads are substantially more expensive than short reads, short reads will dominate. SBS, he adds, is "good enough and always cheapest."

According to Dickinson, PacBio sequencing is about three times more expensive than Illumina sequencing. "You have to have a lot of extra value in those long reads to justify something that is three times more expensive," he said. But the value of diagnosing children—including newborns—with rare genetic diseases has no price tag. Given the momentum this field has, the hope is that long-read sequencing will be embedded into clinical care in the near future, and Kingsmore's cases will no longer make headlines.