

Stanford Team Shows How Long Reads Can Diagnose Disease Caused by Structural Variants

Jan 18, 2017 | [Monica Heger](#)

NEW YORK (GenomeWeb) – Sequencing with long reads can help diagnose rare disease caused by variants that are difficult to identify with shorter reads, according to a recently published study.

Researchers from Stanford University's Clinical Genomics Service reported last month in the pre-print server bioRxiv that they used Pacific Biosciences' Sequel instrument to sequence the whole genome of an individual with an unknown disease and were able to find the causative mutation, which had been missed using short-read sequencing technology.

Euan Ashley, senior author of the study, said in an interview that the price of sequencing on the Sequel is "at the point where within a clinical context it's reasonable to consider" using the technology.

Ashley, who is also an associate professor of medicine and genetics and director of Stanford's Clinical Genomics Services, said that the group is currently in the midst of sequencing a number of other cases on the Sequel system. Currently, he said, PacBio performs the sequencing for the Stanford cases, but said that the clinical genomics lab is considering purchasing its own instrument.

For the first case, Ashley said the group identified an individual who had symptoms that were suspicious of Carney complex, a disorder characterized by an increased risk of benign heart tumors called myxomas, as well as tumors within the endocrine glands. The patient had suffered myxomas throughout his childhood but physicians had been unable to confirm Carney complex with a molecular diagnosis.

Single-gene sequencing of the PRKAR1A gene, the most commonly mutated gene in the disorder, was negative as was whole-genome sequencing on the Illumina HiSeq 2500. Ashley said that the patient was being considered for a heart transplant, but physicians wanted to confirm the diagnosis before performing the surgery. "We don't want the issues to come back after the transplant," he said.

That's when the team decided to use the Sequel to do whole-genome sequencing with longer reads, which would help identify structural variants.

Sequencing was performed to 10x coverage, deep enough to call structural variants, but not so deep to be prohibitively expensive. Since short-read sequencing had already been performed, the researchers focused on calling deletions and insertions, rather than SNVs. An initial variant call identified nearly 7,000 deletions and insertions each. The researchers then applied various filters to narrow down the lists, eliminating variants in a control individual, and then focusing on those that overlapped with those in the Online Mendelian Inheritance in Man database. The filters narrowed the list down to three deletions and three insertions, which were manually reviewed.

One of the six variants was a heterozygous deletion in the PRKAR1A gene, which was confirmed with Sanger sequencing.

Ashley said the case represents the potential of using long-read sequencing to improve the diagnostic rate of current clinical sequencing pipelines, which has hovered between 25 percent and 35 percent for the last several years. Sequencing with long reads could help bump that up since it can better identify structural variants; however, he noted that it is still too expensive to do routinely on every case and to a high depth of sequencing to identify SNVs. Sequencing at 10x coverage on the Sequel cost around \$5,000, he said. That price point and level of coverage would enable structural variant identification for select cases, he said.

"We're not quite ready to make it routine," he said, but "we're interested in introducing long-read sequencing into our Clinical Genomics Services lab for cases [where] we have a strong suspicion that structural variation plays a prominent role in that disease," like neurodevelopmental disorders and cancer, for instance.

"Increasingly, we want to be looking at multiprong strategies to give us the best results," he said. Although sequencing on the Sequel is still more expensive than short-read sequencing, the real question is: What is the cost of a diagnosis? "It doesn't matter how cheap a short-read genome is if it doesn't find the answer," Ashley said.

Whole-genome sequencing on the Sequel also revealed the extent to which short-read sequencing technology misses structural variants. The sheer number of structural variants identified and the number of structural variants that are not annotated in databases reveals that discrepancy, he said. Fortunately, in this case, the filtering strategy was straightforward since there was a variant in a well-known gene, but that is unlikely to happen for every case, Ashley said.

In addition, he said that the researchers have not yet examined closely the five other structural variants that were linked to genes in the OMIM database. As more and more genomes are sequenced using longer reads, more unknown structural variants will be identified. "We'll have to get more population data to give us a sense of the context that we need to view those variants in," he said.