

Deploying long-read sequencing to bridge diagnostic and equity gaps in rare disease

Authors:

Emmanuèle C. Délot (University of California, Irvine), Greta Pitsava (Ambry Genetics), Rebekah Barrick (University of California, Irvine), C Hadley King (University of California, Irvine), Ivan deDios (University of California, Irvine), Krista Bluske (Ambry Genetics), Nida Karra (Ambry Genetics), Arthur Bandeira de Mello Garcia (University of California, Irvine), Arthur Ko (Children's National Hospital), Stuti Joshi (Ambry Genetics), Sami Belhadj (Ambry Genetics), UCI GREGoR Center, Seth Berger (Ambry Genetics), Changrui Xiao (University of California, Irvine), Rachid Karam (Ambry Genetics), Eric Vilain (University of California, Irvine)

Abstract:

Long-read sequencing (LRS) is emerging as a transformative tool for resolving previously intractable rare disease diagnostic challenges. However, translating LRS into routine clinical practice requires understanding the specificities and limitations of each tool deployed in workflows. We launched an LRS-first approach for participants in the PMGRC-UCI genomics research center, a founding site of the GREGoR Consortium. We have previously reported on deploying specialized tools to detect variants in segmental duplications, phasing biallelic variants in proband-only cases, and validating deep intronic variants using syndrome-specific methylation patterns.

HiFi sequencing was performed on a Revio (PacBio) at Ambry Genetics, generating one of the largest single-center HiFi LRS datasets (>1,200 to date). Variant prioritization was performed on the GeneYX platform. Phenotype and family structure were uploaded from the UCI-ICTS Dashboard, a collaborative platform we developed to integrate the project from phenotype collection and biobank specimen tracking to analysis result sharing. Nanopore-based LRS was applied to a smaller cohort.

Across a cohort of 144 families receiving both genome short-read sequencing (SRS) and LRS, LRS yielded an additional 10% diagnostic rate to date, with a third of the findings undetectable with SRS. Those include a pathogenic stop-gain variant located within a variable-number tandem repeat region with zero mapping quality on SRS; a 13.4 kb single-exon deletion, undetected by seven structural variant callers on SRS; a large intronic repeat expansion diagnostic for FGF14-related ataxia.

We were able to deploy LRS both on frozen DNA for new diagnoses on a historical ataxia cohort and on non-invasive oral samples, allowing countrywide recruitment of participants. However, we also identified limitations of current workflows. Using different versions of the tools across trio members introduced artifactual de novo calls, and both aligner tool and human genome reference directly affected variant discoverability.

Emerging tools (such as Paraphase for segmental duplications or duoNovo for de novo calling in incomplete trios) and multi-modal data layers such as DNA methylation or RNA sequencing currently require manual analysis. While LRS offers the promise to close technical and equity gaps in rare disease diagnostics, our study identifies bioinformatic infrastructure improvements needed to translate this potential into routine clinical practice.