Long-read Genome Sequencing Resolves Genetic Cases Missed by Short-read Sequencing

Greta Pitsava[1], Krista Bluske[2], Ivan De Dios[1], Rebekah Barrick[1], Sami Belhadj[2], Erica Smith[2], Miguel Almalvez[1], Arthur Ko[3], Lightning Auriga[1], UCI-GREGoR Consortium, Rachid Karam [2], Changrui Xiao[4], Emmanuèle C. Délot[1], Seth Berger[2], Eric Vilain[1]

UC Irvine

[1] Institute for Clinical and Translational Science, University of California, Irvine, CA, USA
[2] Ambry Genetics, Aliso Viejo, CA, USA
[3] Center for Genetic Medicine Research, Children's National Research Institute, Washington, DC, USA
[4] Department of Neurology, University of California, Irvine, CA, USA

Background

Long-read genome sequencing (LR-GS) has the potential to increase the diagnostic yield of genetic testing by overcoming limitations of short-read genome sequencing (SR-GS) such as the detection of variants in highly homologous regions and the detection of complex structural variants. However, the diagnostic yield of LR-GS is not well-characterized. We aimed to evaluate the diagnostic utility of LR-GS in individuals suspected to have a genetic disorder but with no molecular diagnosis despite extensive prior genetic workup including chromosomal microarray testing, exome sequencing, and SR-GS.

Methods

We performed PacBio HiFi LR-GS in 73 families (193 participants): 55 trios, 10 parent-proband duos and 8 proband-only cases. Proband phenotypes were converted to Human Phenotype Ontology (HPO) terms. Variant interpretation was carried out using the Geneyx platform combined with manual curation and filtering. Variant effects on RNA were confirmed with short-read sequencing of whole blood.

Results

We identified a diagnosis in 2 out of 73 cases, representing an incremental diagnostic yield of 2.7% (95% CI: 0.3%-9.6%) after negative SR-GS. Below we present those 2 solved cases.

Case 1:
15-year-old male with short stature, autoimmune enterocolitis, autoimmune hepatitis (which led to acute liver failure and later liver transplant), splenomegaly, reactive airway disease, developmental delay, recurrent skin abscesses, poor weight gain in the first month of age along with intractable vomiting and diarrhea, eczema, dry skin, and dysmorphic facial features (flat nasal bridge, narrow spaced eyes and asymmetric facies).

Previous genetic testing included primary immunodeficiency panel (429 genes) and exome sequencing that did not reveal any variants. SR-GS that identified a heterozygous maternally inherited stop-gain variant NM_000285.4:c.769G>T, p.Gly257* in *PEPD,* associated with autosomal recessive prolidase deficiency. The p.Gly257* variant is located in exon 11/15 and classified as pathogenic in ClinVar (ID: 2580569).

LR-GS revealed a paternally inherited 13.4kb (Chr19:33471512-33484890) deletion including the in-frame exon 7 in *PEPD*. **Both breakpoints demonstrated sequence identity, explaining why short-read copy number variant callers were unable to confidently identify the deletion.**

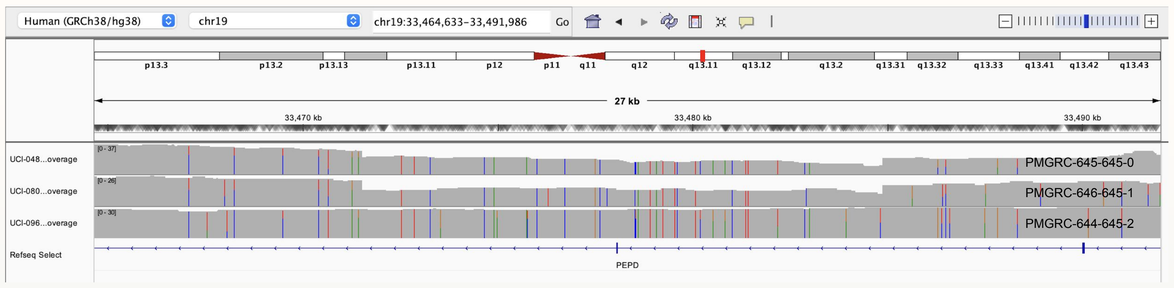

Figure 1. IGV snapshot demonstrating the paternally inherited 13.4kb deletion in the affected proband (PMGRC-645-645-0). PMGRC-646-645-1 represents the father and PMGRC-644-654-2 represents the mother.

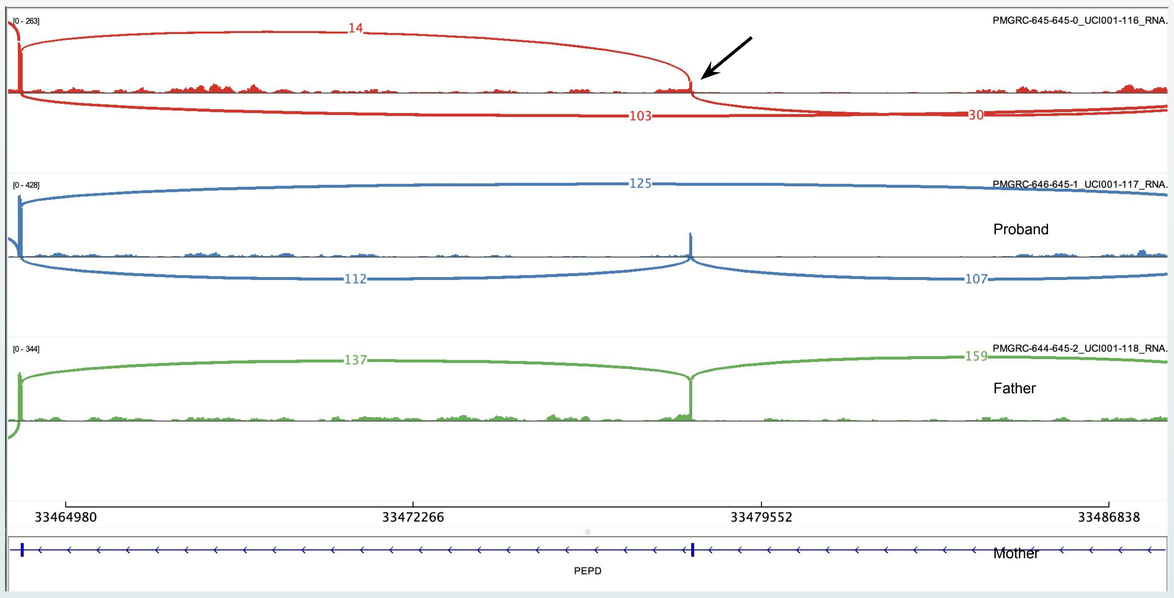

Figure 2. Sashimi plot (from RNA-seq) illustrating the abnormal PEPD expression in the proband (PMGRC-645-645-0).

To confirm the diagnosis a urine amino acid analysis was done that demonstrated imidodipeptiduria.

Proband's sister was also affected with symptoms of autoimmune enterocolitis and eczematous skin. Urine amino acid analysis was performed and confirmed the same diagnosis.

Case 2:
12-year-old female with multiple osteochondritis dissecans lesions, short stature (currently on growth hormone) and ADHD. Family history notable for father with osteochondritis and short stature and paternal grandmother with similar bone breakdown and very early osteochondritis. Previous testing which included chromosomal microarray, exome sequencing and SR-GS was negative.

LR-GS identified a paternally inherited nonsense variant (c.3676G>T / p.Gly1226*) in *ACAN* associated with autosomal dominant short stature and advanced bone age, with or without early-onset osteoarthritis and/or osteochondritis dissecans. **The variant was previously missed because it fell within a 1-kb "dead zone" in exon 12 where SR-GS reads align with zero mapping quality because of identical matches elsewhere in the genome.**

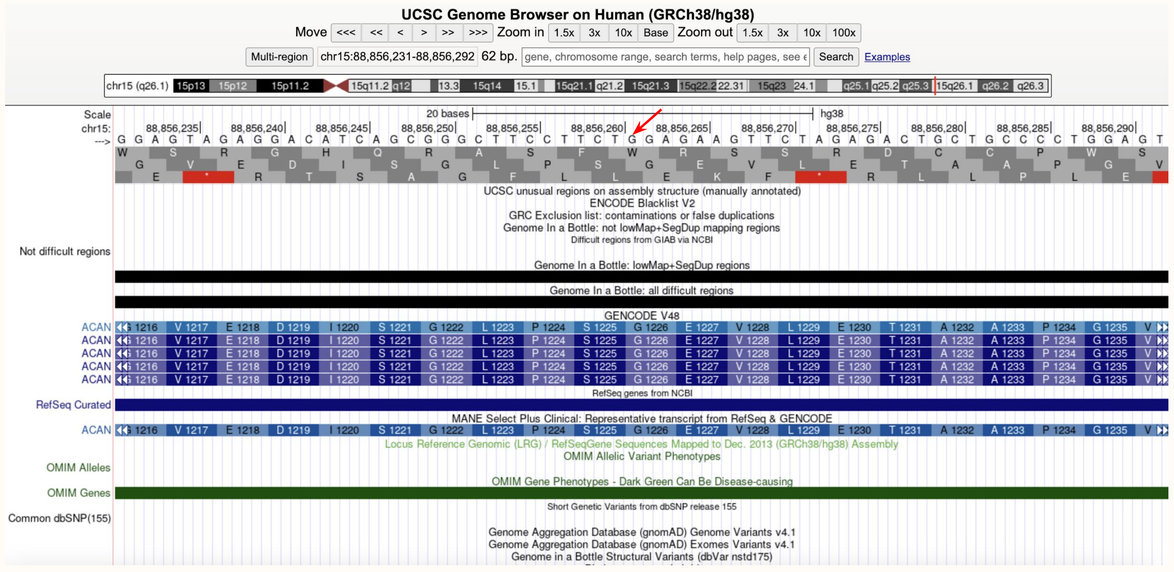

Figure 3. UCSC browser snapshot highlighting that the causative variant occurs in a region problematic for SR-GS.

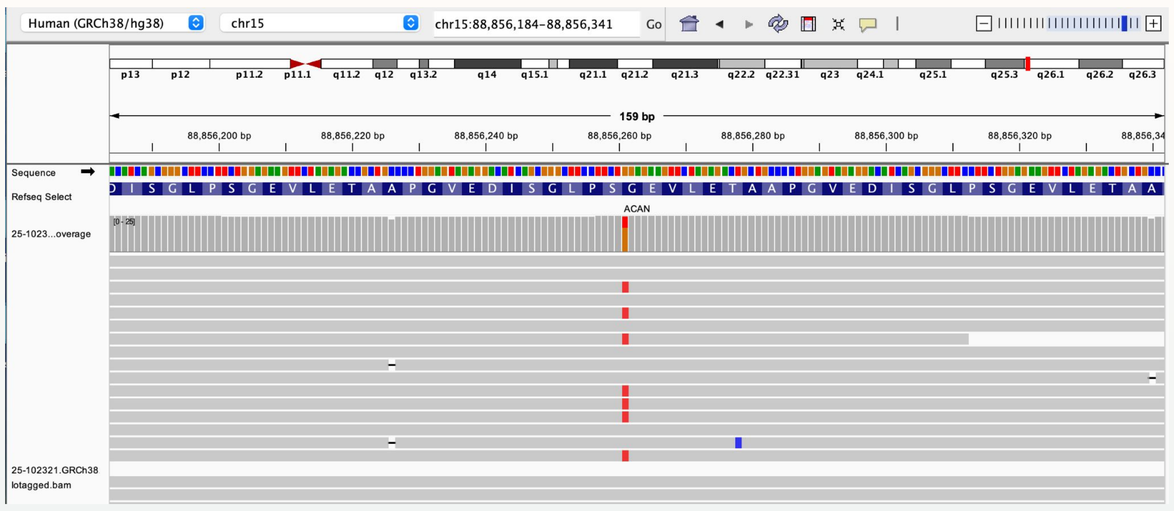

Figure 4. IGV snapshot sequenced via LR-GS showing the region containing the pathogenic variant in our participant.

Figure 5. IGV snapshot of a random sample sequenced via SR-GS showing the low mappability of the region containing the pathogenic variant in our participant.

We also identified a candidate variant in 14 additional cases. Three of these cases (4.1%) were deemed by the study team as "probably solved" based on compelling molecular characteristics, phenotype, and mode of inheritance match, though additional confirmatory studies are needed. All 3 of these could have been detected with a reanalysis of SR-GS. Of the 11 remaining cases, two had variants undetectable with SR-GS. The first was an inversion in *trans* with a previously identified missense variant. The second case was a compound heterozygote with two missense variants, neither of which were called by SR-GS. The remaining candidate variants could all have been detected via reanalysis of SR-GS but were filtered out by the automated pipeline.

Discussion

When focusing on variant detection, LR-GS was able to diagnose clinically important variants not detected in SR-GS through overcoming issues related to homology and detection of structural variation.