Collaborative Force: Advancing Rare Disease Diagnosis with Long-Read Sequencing

Introduction

In 2024, Ambry Genetics was selected by the University of California, Irvine (UCI) and the Genomics Research to Elucidate the Genetics of Rare diseases (GREGoR) Consortium to support the Pediatric Mendelian Genomics Research Center (PMGRC) program. This combined initiative aims to broaden our understanding of the biological mechanisms underlying rare diseases.

The GREGoR Consortium is a collaborative effort funded by the National Human Genome Research Institute (NHGRI), working to transform the landscape of Mendelian disease research by identifying the underlying genetic cause of rare disease in patients for whom prior genomic analysis did not yield answers. By combining the expertise, data, and resources of leading institutions, this partnership seeks to accelerate rare disease diagnostics and uncover answers that no single organization could achieve alone. Over a three-year period, this ambitious research collaboration uses long-read sequencing technology to sequence and analyze whole genomes with a focus on developing new insights into the causes of rare disease.

“There remain a multitude of rare diseases that are difficult to diagnose, and for which effective treatments remain elusive,” said Eric Vilain M.D., Ph.D., Associate Vice Chancellor for Scientific Affairs, Health Affairs at UCI. “Our research endeavors aim to shed light on these complexities, revealing insights that legacy technologies struggle to uncover. Collaborating with our partners at Ambry Genetics and PacBio, we are poised to enhance our comprehension of rare diseases and in the future revolutionize diagnostic capabilities. This collaborative effort is designed to offer hope not only to families in our study, but to all families looking to unlock answers for children facing rare diseases.”

This pioneering initiative has united leading genomics researchers working collaboratively to incorporate innovative methods for understanding the biology of rare disease including phenotyping, variant identification, and functional analysis of both coding and non-coding sequence alterations. By using highly accurate 5-base, long-read sequencing technology, the researchers are working to discover new rare variants and to understand the role of epigenomics on disease manifestation. In addition to identifying new Mendelian gene variations, researchers are also categorizing previously-identified variants of unknown significance by building new analysis pipelines for genomic and epigenomic data.

Long-Read Whole Genome Sequencing

As genomic technologies evolve, the field of genetics gains tools that enable researchers to understand the genome with greater depth and precision. Long-read sequencing is a prime example of this, paving the way for enhanced diagnostic strategies in rare disease.

For decades, short-read sequencing techniques (Figure 1) (DNA fragments that are 50 to 300 bases long) have been the predominant technology used to provide genomic answers for patients with rare disease. While short-read technologies are effective at identifying single nucleotide variants (SNVs) and small insertions and deletions, they are limited in their detection of structural variants, copy number variants, and repeat expansions. With short-read technology, analyzing repetitive parts of the genome is like trying to solve a puzzle where all the pieces look the same. To overcome these limitations, Ambry and UCI chose a 5-base long-read whole genome sequencing method to provide the necessary accuracy.

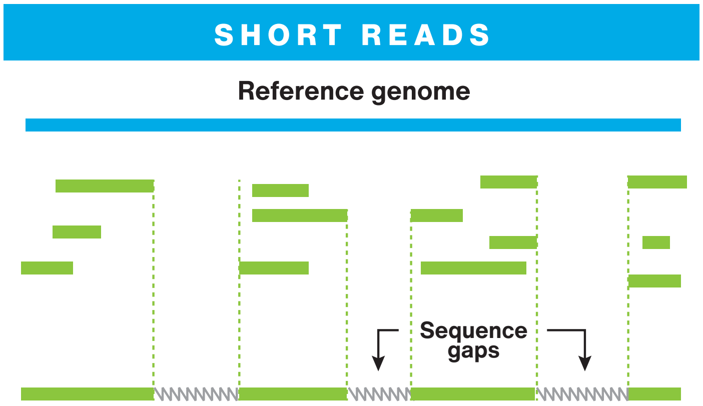


Figure 1. Short reads (50-300 base pairs) result in missing sequence data and limit variant detection.

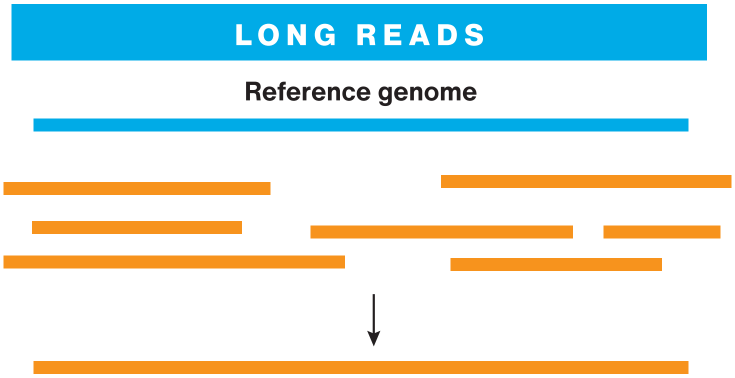


Figure 2. Long reads (1,000-20,000+ base pairs) bridge regions and provide comprehensive variant detection.

Benefits of Long-Read Whole Genome Sequencing

Longer Sequence Reads: Long-read technology (Figure 2) reads much longer stretches of DNA at a time, from 1,000 to 20,000 bases or more. This is particularly useful when analyzing repetitive sections of the genome that short reads often miss. It's like having bigger, more unique puzzle pieces, making it easier to see how they all fit together.

Exceedingly Accurate: Long-read sequencing reads each DNA sequence multiple times then compares those readings to generate a highly accurate "consensus" sequence, ensuring both length and precision in variant detection.

Superior Coverage: By eliminating amplification bias, long-read sequencing provides more complete and uniform coverage across the genome, enabling analysis of regions that were previously very difficult or impossible to analyze. This includes regions with high sequence repetition, or areas rich in AT or GC pairs.

5-Base Sequencing: Traditional sequencing detects the standard four DNA bases: A, C, G, and T. Five-base sequencing also detects 5-methylcysteine (5mC), a key epigenetic marker allowing for the incorporation of epigenetic data (referred to as an episignature), into interpretation of genetic results.

In short, this technology delivers a more complete, accurate, and detailed view of the genome leading to the detection of multiple variant types with a single assay (Figure 3). Using this approach, Ambry Genetics and the UCI GREGoR Consortium are better equipped to uncover the genetic etiologies of rare disease and bring clarity to families searching for answers.

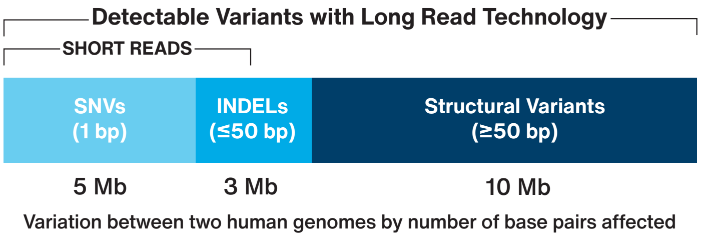


Figure 3. Long reads can excel at detecting multiple variant types such as single nucleotide variants, indels (insertions and deletions) and structural variations including duplications, inversions, translocations and complex rearrangements.

Illustrative Examples

Here, we present a series of patient cases that powerfully illustrate the benefits of long-read whole-genome sequencing in resolving previously intractable clinical challenges and providing a comprehensive view of complex genomic variation.

These cases show how improved variant mapping, better variant resolution and integrating data types supported by long-read sequencing have led to meaningful diagnoses and ended the diagnostic odyssey for patients participating in the collaboration.

CASE 1

A 12-year-old female presented with short stature, low birth weight, and multiple osteochondritis lesions. Family history was notable for osteochondritis in the patient's father. Previous genetic testing included non-diagnostic chromosomal microarray and whole exome sequencing.

Trio long-read genomic analysis identified a pathogenic variant in the ACAN gene, c.3676G>T (p.G1226*), in the proband and her father. This confirmed a diagnosis of short stature and advanced bone age with or without early-onset osteoarthritis and/or osteochondritis dissecans (SSOAOD).

Why Long Read? This variant falls within a region that contains variable number tandem repeats (VNTR), which is unmappable on short read genome. Long-read technology enabled accurate detection and interpretation of this clinically relevant variant.

CASE 2

A 14-year-old male presented with short stature, developmental delay, poor weight gain requiring nasogastric feeding, reactive airway disease, splenomegaly, and recurrent skin abscesses. He also had autoimmune enterocolitis, hepatitis, and liver failure requiring a liver transplant. Family history was significant for a potentially affected younger female sibling.

Extensive prior testing, including STAT3 single gene analysis, a primary immunodeficiency panel, whole exome sequencing, and short-read whole genome sequencing, was non-diagnostic across both clinical and research settings.

Trio long-read genomic analysis identified two pathogenic variants in the PEPD gene: a maternally-inherited c.769G>T (p.Gly257*) variant, and a paternally-inherited 13.4 kb deletion of exon 7. Biochemical testing confirmed a diagnosis of Prolidase deficiency in both the proband and his sibling.

Why Long Read? Although the c.769G>T variant was identified on short-read genome, none of the seven tested structural variant callers identified the 13.4 kb deletion using short-read genome data. Long-read sequencing provided the resolution necessary to detect this large structural variant, resulting in a diagnosis.

CASE 3

A 5-year-old female presented with developmental delays, regression, hypotonia, white matter abnormalities, and cardiac findings including atrial septal defect, aneurysm of the septum primum, and patent ductus arteriosus. Dysmorphic features included thick lips, widely spaced teeth, up-slanting palpebral fissures, low-set posteriorly rotated large ears, and frontal bossing. Family history was unremarkable. Previous testing included non-diagnostic chromosomal microarray and whole exome sequencing.

Trio long-read genomic analysis identified a de novo pathogenic variant in the ARID1B gene, c.3025+700C>G, located deep within an intronic region. Methylation profiling via 5-base sequencing identified an episignature consistent with a diagnosis of Coffin-Siris syndrome.

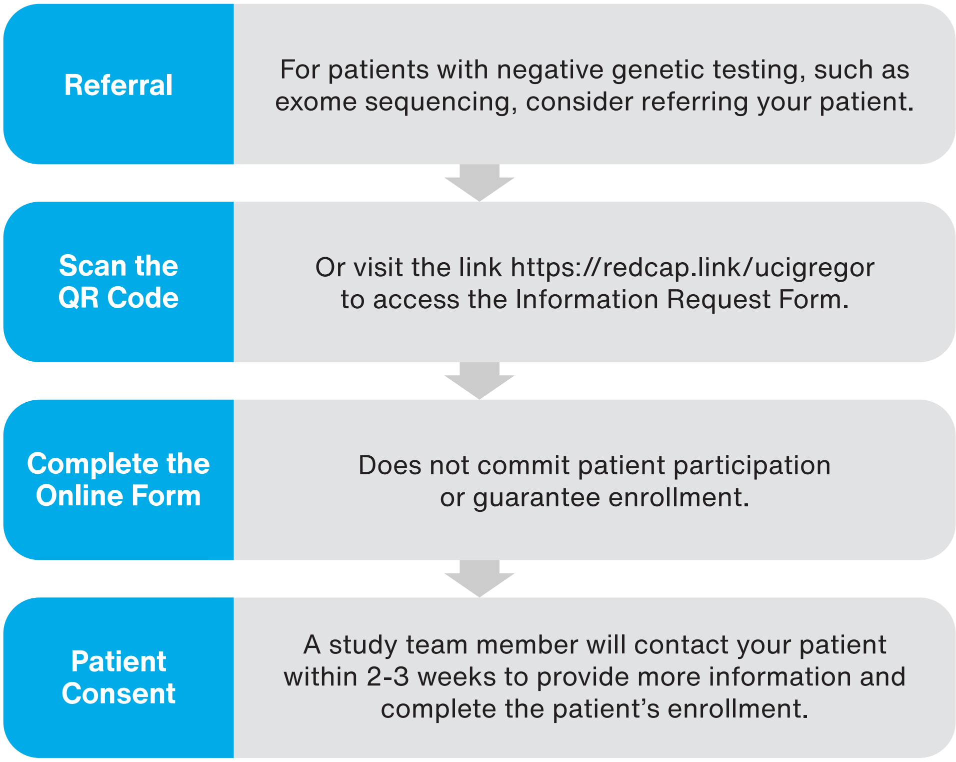
Why Long Read? This deep intronic variant would not be captured by whole exome sequencing, and likely would have been considered a variant of uncertain significance (VUS) using traditional whole genome sequencing. The integration of epigenomic data provided important phenotypic data that supported pathogenicity of this variant and confirmed this patient's diagnosis.

Conclusions

When genomic experts work in partnership, their impact isn't constrained by the boundaries of their expertise. The medical genomic experts at UCI interpret genomic data in the context of disease biology and patient care. The diagnostic experts at Ambry know how to translate those insights into precise assays. And, the long-read technology ensures the sequencing and analysis tools deliver accurate, scalable, and cost-effective results. This powerful collaboration came together so that more patients can have answers based on today's pioneering technology.

Referring Patients to the GREGoR/PMGRC Study

The study is seeking participants with suspected genetic conditions. To qualify, the patient must have received genetic testing, such as exome sequencing, that resulted in a negative or uncertain result. Patients may be enrolled from anywhere in the United States.



Referral — For patients with negative genetic testing, such as exome sequencing, consider referring your patient.

Scan the QR Code — Or visit the link https://redcap.link/ucigregor to access the Information Request Form.

Complete the Online Form — Does not commit patient participation or guarantee enrollment.

Patient Consent — A study team member will contact your patient within 2-3 weeks to provide more information and complete the patient's enrollment.

Study Participant Information

Patients participating in the study will possibly need to provide a blood or saliva sample. The consenting and sample collection process should take no more than 1 hour to complete.

This study is based at UC Irvine School of Medicine, 1003 Health Sciences Road, Suite 308, Irvine CA 92617-3054, however travel is not required and both consent and sample collection can take place locally to participants.

To Learn More

Visit https://gregorconsortium.org to learn more about participating in GREGoR Consortium research.

Ambry Genetics®
A TEMPUS COMPANY

1 Enterprise, Aliso Viejo, CA 92656 USA Toll Free +1.866.262.7943 Fax +1.949.900.5501 **ambrygen.com**