SCALABLE HUMAN WHOLE GENOME HIFI SEQUENCING FOR RARE AND INHERITED DISEASE RESEARCH
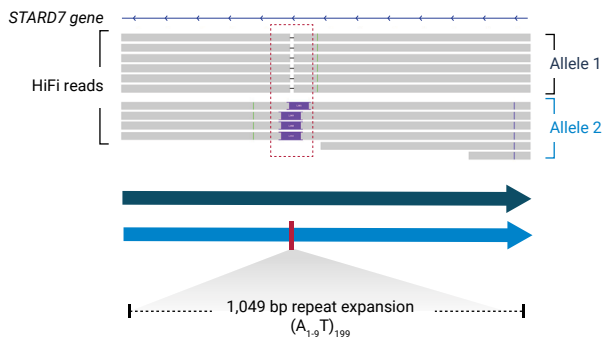
PacBio® highly accurate long reads – HiFi reads – offer a single-platform solution for rare and inherited disease research, elucidating suspected genetic causes of disease in up to ~50% of cases that have not previously been explained using short-read exome or whole genome sequencing.[1,2] Simultaneous 5-methylcytosine (5mC) calling enabled by the platform allows researchers to efficiently evaluate methylation profiles in a single run. PacBio offers an efficient workflow, developed in collaboration with Children's Mercy Kansas City, which provides a scalable solution for sequencing hundreds to thousands of whole human genomes per year on the Sequel® II and Sequel IIe systems.

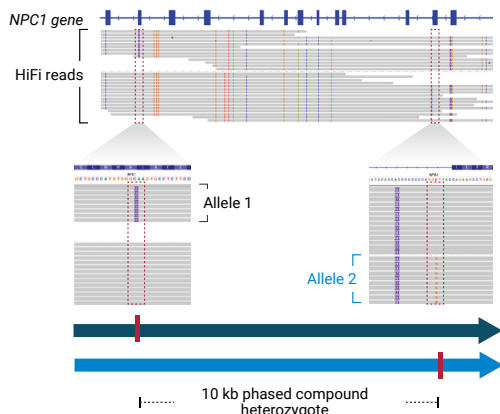## The advantages of HiFi reads for understanding rare diseases

HiFi reads offer per-base accuracy 99.9% and long read lengths, allowing scientists to confidently find causative pathogenic variants and identify novel disease-associated genes in cohorts with:

- Best-in-class variant calling for all variant types including single nucleotide variants (SNVs), indels, and structural variants (SVs)[3,4]

- Complete, accurate, and phased assemblies of the human genome — including regions previously inaccessible to other technologies — for accurate typing of genes including HLA and CYP2D6[5]

HiFi reads had the fewest errors of any sequencing technology in the precisionFDA *Truth Challenge*.[4]
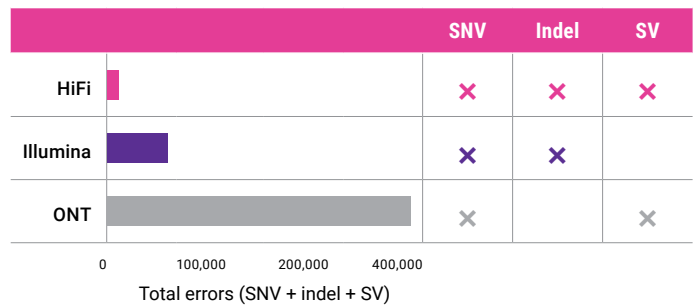
Scientists at Children's Mercy Kansas City use HiFi reads to enable the development of improved rare disease diagnostics

A large repeat expansion identified to segregate with disease in a pedigree with dystonia, seizures, and ataxia in multiple generations.[6,7]
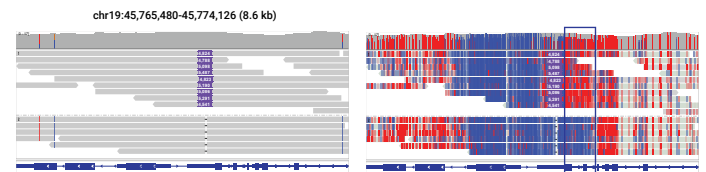

A compound heterozygous variant in the NPC1 gene explains hepatosplenomegaly in a young female.[6,7]

Methylation profiling

In addition to variant detection, HiFi sequencing reads provide 5mC detection, allowing researchers to bypass an additonal whole genome bisulfite sequencing (WGBS) workflow. 5mC detection with HiFi also overcomes GC bias issues and coverage gaps experienced with WGBS while providing phased haplotypes. HiFi powerfully and efficiently enables researchers to access multiomic analysis.

chr19:45,765,480-45,774,126 (8.6 kb)

Example of an individual with myotonic dystrophy due to a pathogenic repeat expansion (CTG triplet) in DMPK.

- Long reads are phased into two alleles

- The top allele has an expansion, which is mosaic (denoted in purple with different repeat lengths)

- The repeat induces hypermethylation of the adjacent region (red)

PacBio

Workflow overview

Sample and library prep

| 3 μg input DNA (blood, saliva, tissue, cell culture) | Shearing | Library prep | Size selection | One technician can produce up to 96 libraries/week |

Starting from a variety of sample types, this efficient, scalable, and automation-friendly workflow enables parallel construction so that one technician can manually produce up to 96 HiFi libraries per week, to facilitate increased throughput in labs with multiple instruments.[8,9,10]
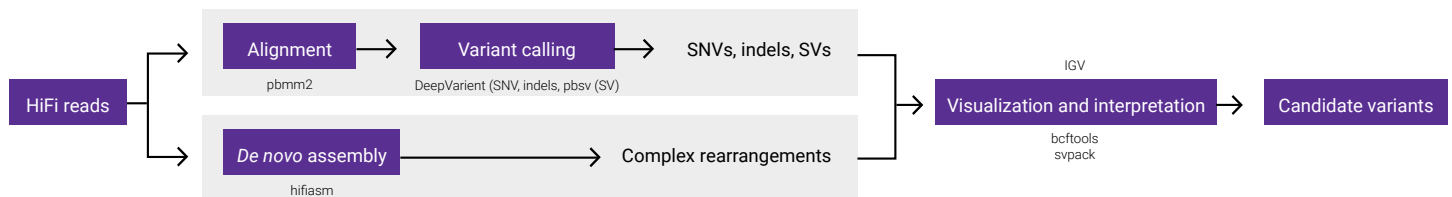
Sequencing

This workflow supports high-volume sequencing of up to 256 samples on one Sequel IIe system or more than 1,500 samples on six Sequel IIe systems per year. For smaller insert libraries, shorter runtimes are available to optimize instrument throughput while maintaining HiFi read accuracy.[9,10]

		HiFi sequencing coverage		
		10-fold	**20-fold**	**30-fold**
Optimal insert size range		15–18 kb		
Number of SMRT® Cells/sample		1	2	3
Variant detection performance (% accuracy, F1)	SNVs	99.0%	99.8%	99.9%
	Indels	92.6%	96.9%	97.9%
	SVs	92.6%	95.7%	95.8%

Data analysis

Using open-source software packages, candidate variants can be efficiently identified from singletons or trios through alignment-based variant calling for SNVs, indels, and SVs, or through assembly-based approaches for complex rearrangements.[11]

HiFi reads → Alignment (pbmm2) → Variant calling (DeepVariant (SNV, indels), pbsv (SV)) → SNVs, indels, SVs

HiFi reads → *De novo* assembly (hifiasm) → Complex rearrangements

→ Visualization and interpretation (IGV, bcftools, svpack) → Candidate variants

KEY REFERENCES

1. Ferreira, C. R. (2019) **The burden of rare diseases.** *American Journal of Medical Genetics.* 179, 885-892.
2. Turro, E. et al (2020) **Whole-genome sequencing of patients with rare diseases in a national health system.** *Nature.* 538, 98-102.
3. Wenger, A. M. et al. (2019) **Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome.** *Nature Biotechnology.* 37, 1155-1162.
4. Olson, N.D. et al. (2021) **precisionFDA Truth Challenge V2: Calling variants from short- and long-reads in difficult-to-map regions.** *bioRxiv* preprint.
5. Nurk, S. et al. (2020) **HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads.** *Genome Research.* 30(9), 1291-1305.
6. Farrow, E. et al. (2021) **Clinical application of long-read sequencing in unsolved rare disease.** ACMG presentation.
7. Miller, N. (2020) **Third generation sequencing for diagnosis of rare disease**. PacBio virtual global summit.
8. **Technical note: Preparing DNA for PacBio HiFi sequencing - Extraction and quality control.** PacBio literature.
9. **Overview – Sequel systems application options and sequencing recommendations.** PacBio documentation.
10. **Procedure & checklist – Preparing HiFi SMRTbell libraries using SMRTbell express template prep kit 2.0.** PacBio documentation.
11. **PB Human WGS workflow.** PacBio GitHub.

Learn about scalable human WGS: **pacb.com/rare-disease**

PacBio