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 CYP2D65

Scientists at Children’s Mercy Kansas City use HiFi reads to enable the development of improved rare disease diagnostics

HiFi reads had the fewest errors of any sequencing technology in the precisionFDA Truth Challenge.4

Methylation profiling

In addition to variant detection, HiFi sequencing reads provide 5mC detection, allowing researchers to bypass an additional 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.

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

- 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)
Workflow overview

Sample and library prep

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 Ile system or more than 1,500 samples on six Sequel Ile systems per year. For smaller insert libraries, shorter runtimes are available to optimize instrument throughput while maintaining HiFi read accuracy.9,10

<table>
<thead>
<tr>
<th>Hifi sequencing coverage</th>
<th>10-fold</th>
<th>20-fold</th>
<th>30-fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal insert size range</td>
<td>15–18 kb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of SMRT® Cells/sample</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Variant detection performance (% accuracy, F1)</td>
<td>SNVs</td>
<td>99.0%</td>
<td>99.8%</td>
</tr>
<tr>
<td></td>
<td>Indels</td>
<td>92.6%</td>
<td>96.9%</td>
</tr>
<tr>
<td></td>
<td>SVs</td>
<td>92.6%</td>
<td>95.7%</td>
</tr>
</tbody>
</table>

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

Key references


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