VARIANT DETECTION USING WHOLE GENOME SEQUENCING WITH HIFI READS – BEST PRACTICES

With highly accurate long reads (HiFi reads) from the Sequel® II or IIe systems, you can comprehensively detect variants in hundreds to thousands of genomes in a year. HiFi reads provide high precision and recall for single nucleotide variants (SNVs), indels, structural variants (SVs), and copy number variants (CNVs), including in difficult-to-map repetitive regions.

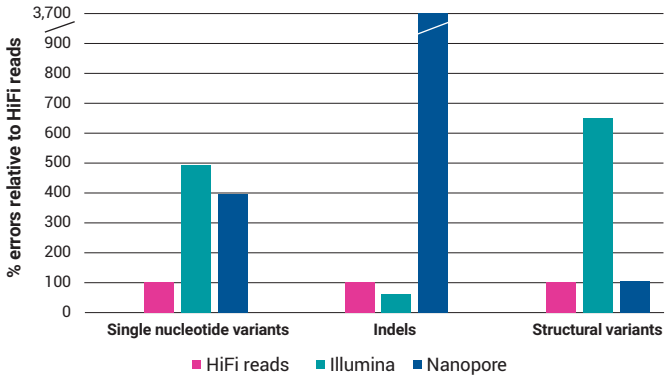
Why choose HiFi reads?

- High accuracy
- Even coverage
- Genome completeness
- Allele resolution, long-range phasing
- Best performance for all variant types

Did you know?

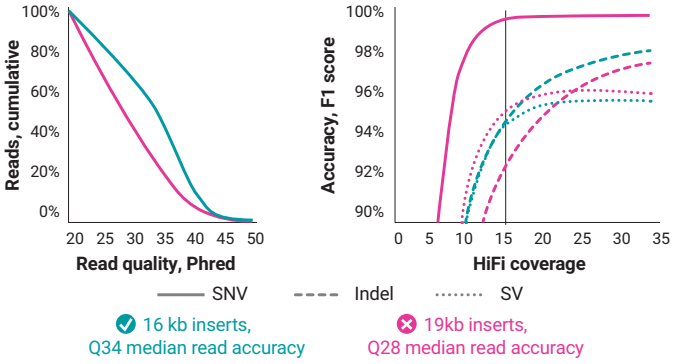
In the *precisionFDA Truth Challenge V2*, PacBio® HiFi reads delivered the highest precision and recall in all categories.[1]

Low error rates for all variant types



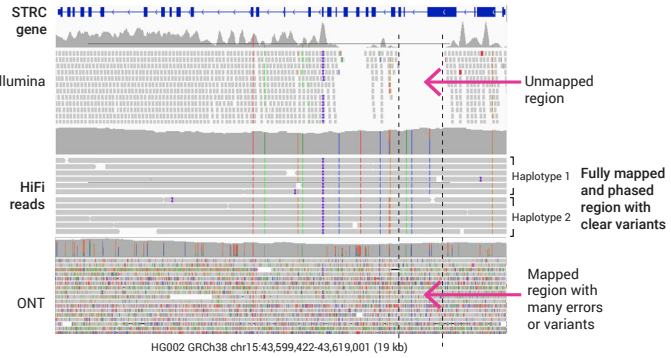
Variant calling performance against *Genome in a Bottle* benchmarks for PacBio HiFi reads (35-fold, Sequel II system, 2.0 chemistry); Illumina (35-fold, NovaSeq); Oxford Nanopore (60-fold, PromethION R9.4.1).

High precision and recall at 15-fold coverage



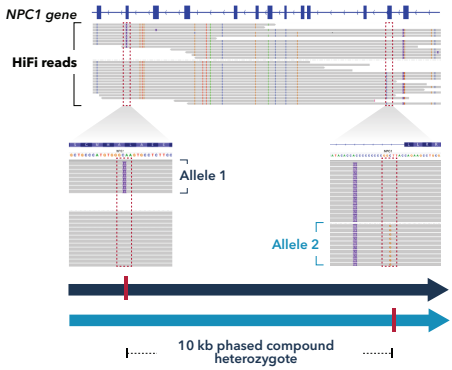
Variant calling performance against *Genome in a Bottle* benchmark v4.2 (Sequel II system, 2.2 chemistry; DeepVariant v1.1). Read lengths 15–18 kb are recommended to achieve the highest read and variant calling accuracy.

Detect more variants in medically relevant genes[2]



STRC gene alignments from *Genome in a Bottle* (GIAB), HG002_NA24385_son.

Phase all variants into haplotypes[3]



NPC1 gene showing a phased compound heterozygote.

PacBio

HIFI SEQUENCING WORKFLOW RECOMMENDATIONS

From DNA to comprehensive variant detection

Library prep

- Prepare SMRTbell® libraries for HiFi sequencing of up to 16 samples with manual prep, or 96 samples using this automation-friendly workflow[4,5]
  - Start with unamplified genomic DNA (>5 µg input) from any sample type (blood, tissue, cell lines)[6]
  - Enrich for 15–18 kb inserts with size selection. Inserts larger than this range may reduce read and variant calling accuracy

SMRT sequencing

- Use the Sequel II or IIe system and SMRT® Cell 8M to sequence to desired coverage depth
  - Recommend 2 SMRT Cells 8M to achieve >15-fold coverage of a human genome for comprehensive variant detection for $2,600*

*Read lengths, reads/data per SMRT Cell 8M and other sequencing performance results vary based on sample quality/type and insert size. Prices, listed in USD, are approximate and may vary by region. Pricing includes library and sequencing reagents run on a Sequel II or IIe system and does not include instrument amortization or other reagents.

Data analysis

- Call structural variants with pbsv[7], available through *structural variant calling analysis* in SMRT® Link[8]
- Call small variants with DeepVariant[9] using the PacBio model
- Use joint calling in pbsv and DeepVariant for multiple samples
- Phase variants with WhatsHap[10]

Learn about variant detection using whole genome sequencing with HiFi reads: **pacb.com/variant**

Connect with PacBio for more info:
North America: **nasales@pacb.com**
South America: **sasales@pacb.com**
EMEA: **emea@pacb.com**
Asia Pacific: **apsales@pacb.com**

This scalable workflow allows for sequencing hundreds to thousands of genomes per year

KEY REFERENCES

1. Olson, N.D. et al. (2021) **precisionFDA Truth Challenge V2: Calling variants from short- and long-reads in difficult-to-map regions.** *bioRxiv preprint.*
2. 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.
3. Farrow, E. et al. (2021) **Applications of Third Generation Sequencing in Unsolved Disease.** ACMG conference presentation.
4. **Procedure + checklist — Preparing HiFi SMRTbell libraries using SMRTbell express template prep kit 2.0.** PacBio documentation.
5. **Overview — Sequel systems application options and sequencing recommendations.** PacBio documentation.
6. **Technical note: Preparing DNA for PacBio HiFi sequencing — Extraction and quality control.** PacBio literature.
7. **pbsv — PacBio structural variant (SV) calling and analysis tools**. PacBio GitHub.
8. **Product brochure: SMRT Link — Explore and analyze your data with confidence.** PacBio literature.
9. Poplin, R. et al. (2018) **A universal SNP and small-indel variant caller using deep neural networks.** *Nature Biotechnology.* 36, 983–987.
10. Martin, M. et al. (2016) **WhatsHap: fast and accurate read-based phasing.** *bioRxiv preprint.*

102-193-604 REV02 19APR2022