Application brief

COMPREHENSIVE HUMAN GENOMIC VARIANT DETECTION WITH HIFI LONG-READ SEQUENCING

PacBio® HiFi sequencing enables greater variant detection in more regions of the genome than alternative technologies, providing exceptional insight at any depth. Choosing the right level of sequencing is a critical first step for project planning, and HiFi sequencing delivers:

- Industry-leading small and structural variant calling performance
- Excellent small variant (SNV and indel) calling at 20x HiFi coverage
- Highly accurate structural variant detection with as little as 10× coverage

The benefits of HiFi sequencing

Mappability
Access the entire genome

Phasing
Haplotype resolution

Exceptional accuracy
Call more variants with fewer errors

Methylation detection
Single-molecule 5mC insights

The 20× HiFi human genome is best-in-class

Titrating a HG002 dataset sequenced on the Revio® system demonstrates industry-leading variant calling performance of 20× HiFi sequencing compared to 30× Illumina and 60× Oxford Nanopore Technologies (ONT) call sets. The power of a 20× HiFi genome is further supported by a recent independent study showing that 20x HiFi coverage is sufficient to call 96.2% of the clinically relevant germline variants identified at 30×.[1]

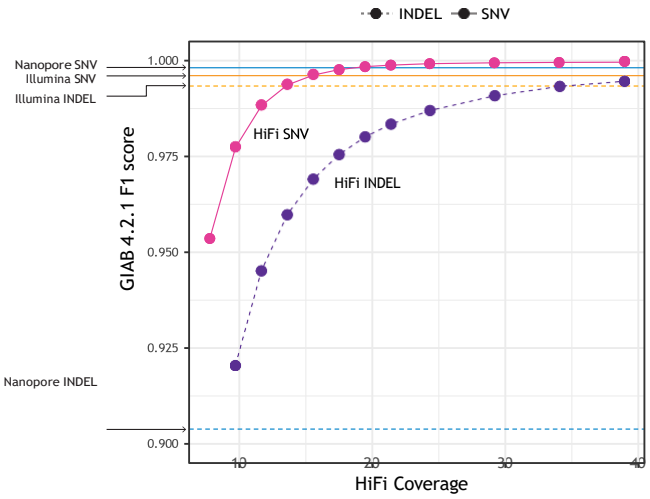

Figure 1. Variant calling performance for single nucleotide variants (SNVs) and insertions and deletions (INDELs, < 50 bp) against the GIAB v4.2.1 benchmark[2]. HG002 PacBio HiFi coverage titration shown in magenta and purple (SNV-solid, INDEL-dashed) with each down-sampled depth marked as a point. Performance for the 30× Illumina DRAGEN 3.7.5[3] and 60× ONT R10.4.1[4] call sets is shown in orange and blue, respectively.

20× HiFi exceeds

- SNV accuracy of 99.8%
- Indel accuracy of 60× ONT

PacBio

Total benchmark errors by dataset

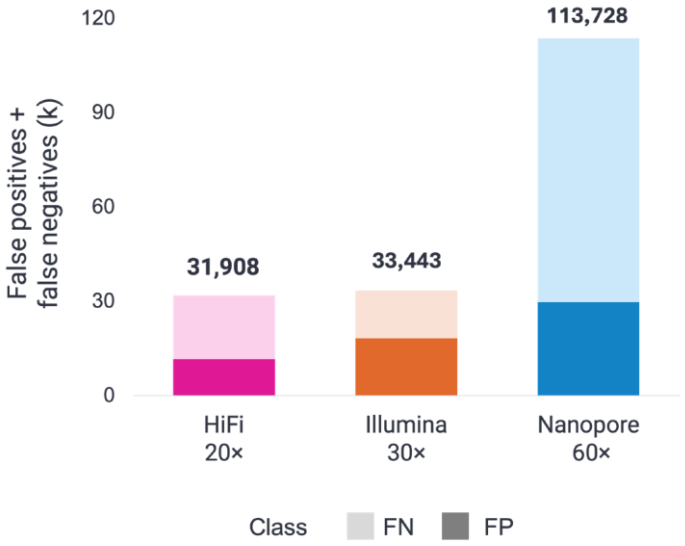

Figure 2. Total small variant (SNV and INDEL) calling errors for each technology. HiFi errors at 20× coverage are shown in magenta, with 30× Illumina DRAGEN 3.7.5 in orange, and 60× ONT R10.4.1 in blue. False negatives are represented as a lighter shade, while false positives are shown in the darker shade.

For structural variants, as little as 10× HiFi sequencing outperforms Illumina and ONT

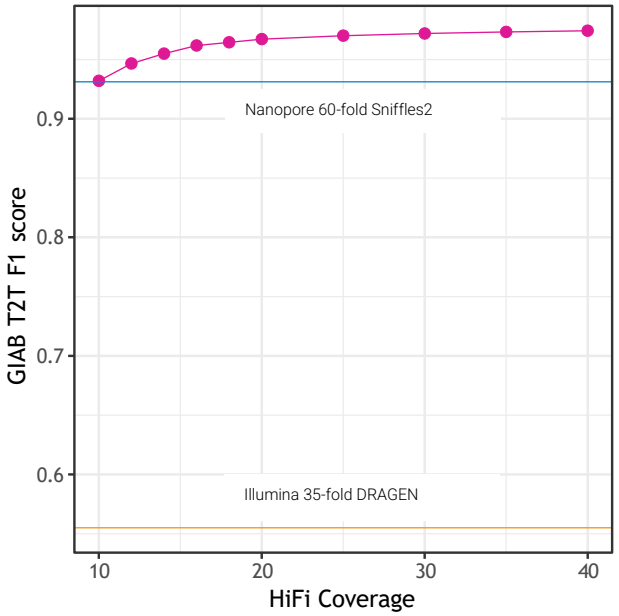

Figure 3. Structural variant (SV) calling performance against the GIAB HG002 telomere-to-telomere (T2T) genome benchmark[5]. Magenta line represents HG002 HiFi performance (F1-score) with each down-sampled data point marked at 10× (93.2%), 12×, 14×, 16×, 18×, 20× (96.7%), 25×, 30× (97.2%), 35×, and 40× (97.4%). Performance for 60× ONT R10.4.1 data is shown in blue (F1- score, 93.1%) and 35× Illumina DRAGEN v4.2.4 data is shown in orange (F1-score, 55.5%).

The PacBio HG002 titrated HiFi dataset was generated with a library prepared using the automated HiFi prep kit 96 workflow[6] and sequenced on a single Revio SMRT® Cell using SPRQ™ chemistry. Sequencing generated 146 Gb of mapped HiFi data and was down sampled to coverage levels ranging from 8× to 40×. HiFi data was processed using the PacBio WGS Variant Pipeline to generate small variant calls (SNVs and INDELs)[7]. Additional call sets, including 30× Illumina DRAGEN v3.7.5 [3] and 60× ONT (Oxford Nanopore Technology) R10.4.1 SUP data[4], were retrieved from publicly available sources. Small variant calls were benchmarked against the GIAB v4.2.1 reference set[2] using the hap.py framework[8]. Structural variant calls were generated from HiFi (Sawfish v0.12.4)[9], 35x Illumina (DRAGEN v4.2.4)[10], and 60x Nanopore R10.4.1 (Sniffles v2.0.7) data, and subsequently benchmarked with Truvari[11] against the T2T Q100 HG002 truth set[12]

Details on the analysis including the Revio SPRQ chemistry aligned bam files can be found here: https://github.com/PacificBiosciences/pb-benchmarks/tree/main/SPRQ-NOV-2024

Visit pacb.com/revio and pacb.com/wgs for more information on workflow and automation options, or connect with a PacBio scientist today.

KEY REFERENCES

1. Höps W., et al. (2024). HiFi long-read genomes for difficult-to-detect clinically relevant variants. medRxiv, 2024-09.
2. Wagner, J., et al. (2022). Benchmarking challenging small variants with linked and long reads. *Cell Genomics*, 2(5).
3. Rossi, M., et al. (2023). Unlocking the full potential of Illumina genomes: The journey to enhanced variant calling quality with DRAGEN informatics and high-quality sequencing. Illumina Genomics Research Hub, published June 29, 2023. (https://human-pangenomics.s3.amazonaws.com/publications/PANGENOME_2022/DeepVariant/samples/HG002/HG002.30x_novaseq_pcrfree.dragen.vcf.gz)
4. Oxford Nanopore Open Data. https://labs.epi2me.io/giab-2023.05/ Accessed September 2024.
5. Telomere-to-telomere consortium HG002 Q100 project: https://github.com/marbl/HG002
6. PacBio Guide & overview – Automated HiFi prep 96 and HiFi ABC for the Hamilton NGS STAR MOA system.
7. PacBio HiFi Human WGS WDL pipeline: https://github.com/PacificBiosciences/HiFi-human-WGS-WDL
8. Illumina Hap.py framework: https://github.com/Illumina/hap.py
9. PacBio Sawfish: https://github.com/PacificBiosciences/sawfish
10. Behera, S., et al. (2024). Comprehensive and accurate genome analysis at scale using DRAGEN accelerated algorithms. *bioRxiv*.
11. English, A.C., et al. (2022). Truvari: refined structural variant comparison preserves allelic diversity. *Genome Biology*, 23(1), 271.
12. HG002 T2T draft genome benchmark: https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/analysis/NIST_HG002_DraftBenchmark_ defrabbV0.018-20240716/

102-326-626 REV02 2024