PacBio

HUMAN GENOMICS

Unlock your next great discovery with HiFi sequencing

Get the most from your samples the first time. Choose HiFi sequencing.

PacBio® HiFi sequencing delivers highly accurate, single-molecule long reads for analyzing both DNA and RNA. The unique combination of accurate base calling and long read length of HiFi enable the phasing of haplotypes, coverage of high-GC regions, and bisulfite-free detection of 5-methylcytosine.

Together, these features allow you to explore the full extent of human genetic variation with a more comprehensive view of the genome, transcriptome, and epigenome than ever before. The benefits of HiFi sequencing, and its numerous applications in human genomics, are ready to enable your next great discovery.

PacBio

Whole genome sequencing

Use the length, accuracy, and unbiased coverage of HiFi sequencing to generate more complete, phased human genome assemblies.

HiFi sequencing delivers whole genome sequence information with stunning accuracy and detail. Based on PacBio Single Molecule, Real-Time (SMRT®) technology, HiFi reads are typically around 15 kb in length with an accuracy of 99.9% and can sequence through GC-rich regions and even repeats in segmental duplications and centromeres — all while delivering simultaneous 5-methylcytosine (5mC) calls for a commanding view of the genome. These state-of-the-art advantages led the *Telomere-to-Telomere Consortium* and the *Human Pangenome Reference Consortium* to select HiFi sequencing as the sequencing technology of choice for the creation of the first gapless human genome[1] and diploid genome.[2]

HiFi sequencing enables some of the highest precision and recall for calling variants of all types: single-nucleotide variants (SNVs), small insertions and deletions (indels), structural variants (SVs), tandem repeats, and expansions. Whole genome sequencing with HiFi can be performed at different depths, depending on research goals.

	10-fold coverage	20-fold coverage	30-fold coverage
Structural variants F1	✓ 93.3%	✓ 94.9%	✓ 95.0%
SNV F1	✓ 98.2%	✓ 99.9%	✓ 99.9%
Indel F1	92.1%	✓ 98.2%	✓ 99.3%
Phasing Phase block N50	✓ 164 kb	✓ 247 kb	✓ 252 kb
Methylation Correlation to WGBS	✓ 0.89	✓ 0.92	✓ 0.93

Table 1. Typical performance metrics from HiFi whole genome sequencing of HG002 at various read depths.

10-fold depth

8- to 10-fold depth HiFi genomes offer a cost-effective method to meaningfully augment and complement pre-existing short-read data sets. Structural variants, a phased 5mC profile, and repeat expansions can be reliably detected at this depth.

20- to 30-fold depth

HiFi WGS at this depth offers an extraordinary ability to interrogate all variant types and uncover genomic regions missed by short-read sequencing. In applications such as rare disease research, the ability to look beyond small variants has been shown to boost discovery and explanation rates.[3]

WGS variant pipeline

A single workflow of consolidated state-of-the-art tools allows you to achieve the most benefits from HiFi data.

The PacBio WGS variant pipeline for human genome consolidates 11 robust secondary analysis tools into a best practice workflow for alignment, variant calling, joint calling, and optional annotation.

Tandem repeats

Tandem repeats are one of the most abundant types of variation in the human genome, and due to their repetitive nature, they have the highest mutation rate in the genome. This genomic instability is a major driver of disease.

There are more than 50 disorders caused by expansions of short tandem repeats (STRs), and several variable number tandem repeats (VNTRs) have been shown to be involved in common complex disease like Alzheimer's, autism, epilepsy, ALS, and others. As part of the recent gapless human genome sequencing efforts, it has been estimated that 50–80% of SVs are actually tandem repeats.[3]

HiFi sequencing enables a more complete and accurate calling of repeat expansions and identification of medically relevant interruption sequences along with methylation profiles.

Genotyping and visualizing

Accurate analysis of tandem repeats — including repeat expansion detection — poses unique requirements from a complete, end-to-end solution. With HiFi sequencing, you can sequence the entire expanded repeat in one read, enabling identification of the repeat size, sequence interruptions, and methylation. Combining HiFi with the *Tandem repeat genotyping tool* (TRGT), *Tandem repeat visualizer* (TRVZ), and other components of our tandem repeat analysis software suite makes it possible to perform a genome-wide analysis of genetic and epigenetic changes in tandem repeat regions, including discovery of novel repeat alleles and methylation patterns.

Segmental duplications

This tool enables high-throughput and comprehensive profiling of genes affected by segmental duplications, including identifying full-length haplotypes, determining gene copy numbers, and calling phased variants using long-read HiFi data. Typical target genes are *SMN1/2*, *PMS2*, *STRC*, *NEB*, and *F8*.

Variant detection

SNVs are the most numerous variant type (~4–5 M/person) followed by indels. The importance of SVs and tandem repeats may seem insignificant because there are far fewer. However, because of their size (>50 bp), they account for more genomic variation between individuals than SNVs and indels combined.

HiFi sequencing offers accurate, small variant calls while also offering significantly more SV detection.[4] Fuller resolution of these variant types enables researchers to better study human genetic diversity and disease association.

Known pathogenicity of variant types is important to consider in disease association. Analysis of *ClinVar* shows the percentage of pathogenic variants in SVs, indels, and tandem repeats is much higher than SNVs. HiFi sequencing enables researchers to more comprehensively study all variation types, better elucidate function, and potentially improve the relatively static solve rates of short-read WGS and whole exome sequencing (WES).

Epigenetics

See a new dimension to human genomes with the extraordinarily accurate sequencing reads and DNA methylation in a single experiment.

HiFi sequencing provides accurate DNA base calls and simultaneous 5mC detection in CpG context without any additional library preparation. This feature enables the resolution of methylation profiles with phased haplotyping. Human genome researchers may also use this capability to interrogate imprinting disorders and methylation abnormalities associated with tandem repeats.

In the example shown below, a repeat expansion on one allele can be seen to impact adjacent methylation status.

PacBio HiFi sequencing simultaneously calls the four DNA bases and 5mC from untreated genomic DNA. Achieve genome-wide detection and phasing of genetic and epigenetic variants from a single, standard HiFi library prep with long and accurate reads.

Figure 1. HiFi sequencing phases and identifies hypermethylation of the region adjacent to a mosaic 5 kb *DMPK* expansion in a sample with myotonic dystrophy (Children's Mercy Kansas City). Magenta indicates methylation, while blue indicates unmethylated bases.

Targeted sequencing

Easily and cost-effectively sequence genomic regions of interest at scale — even those that are challenging to characterize.

HiFi targeted sequencing advantages

Targeted sequencing with HiFi reads allows you to sequence only the genomic regions you care about — easily and cost-effectively at scale. Hybrid capture and amplicon workflows are available to suit your needs, giving you all the benefits you expect from HiFi reads including accurate haplotype resolution and more comprehensive detection of all variant types. The length and accuracy of HiFi reads allow you to access genomic dark regions such as areas of high homology, GC-rich and repetitive regions, and to resolve complex gene families like HLA and pharmacogenes like *CYP2D6*.

Custom hybrid capture and amplicon workflows

- Robust coverage of difficult-to-sequence or difficult–to-map regions
- Unambiguous haplotype resolution through direct phasing
- Consolidate legacy molecular workflows into one assay
- Cost-effective and flexible scaling

HiFi target enrichment

Target large or small gene panels or individual loci with a flexible and scalable workflow using custom DNA probes from Twist Bioscience. HiFi target enrichment can span regions of the genome inaccessible to other technologies and comprehensively call both small variants and structural variants. Tiled HiFi reads enable long-range phasing and accurate haplotype resolution.

HiFi amplicon

Target medically relevant, individual genes or small gene panels using a PCR-based approach. If you can amplify it, HiFi sequencing can span your amplicon in a single read, giving unambiguous haplotype resolution through direct phasing. HiFi amplicon sequencing allows you to replace multiple assays with one fast, easy, scalable, and economical workflow.

RNA sequencing

Easily and affordably sequence more complete transcript isoforms in genes of interest or across the entire transcriptome.

From RNA to full-length isoforms

Changes in transcript splicing and abundances are important clues in deciphering how genomic variants drive the phenotypic differences between health and disease. Short-read sequencing methods often struggle to capture or assemble full-length transcripts, especially longer transcripts or those with complex alternative splicing.

The PacBio Iso-Seq® method allows for capture of full-length transcripts without assembly or complicated algorithms, providing an unambiguous view of the transcriptome. This enables a higher-resolution isoform-level information, which is crucial for understanding functional cellular diversity and dynamic expression in human biology and disease.

Use the Iso-Seq method to:

- Profile whole transcriptomes exhaustively at the isoform level for both bulk tissue and single-cell RNA samples
- Characterize alternative start and polyadenylation sites as well as exon-skipping events
- Identify allele-specific splicing events
- Direct open reading frame (ORF) prediction to infer expressed proteins due to splicing changes
- Target genes of interest within a sample

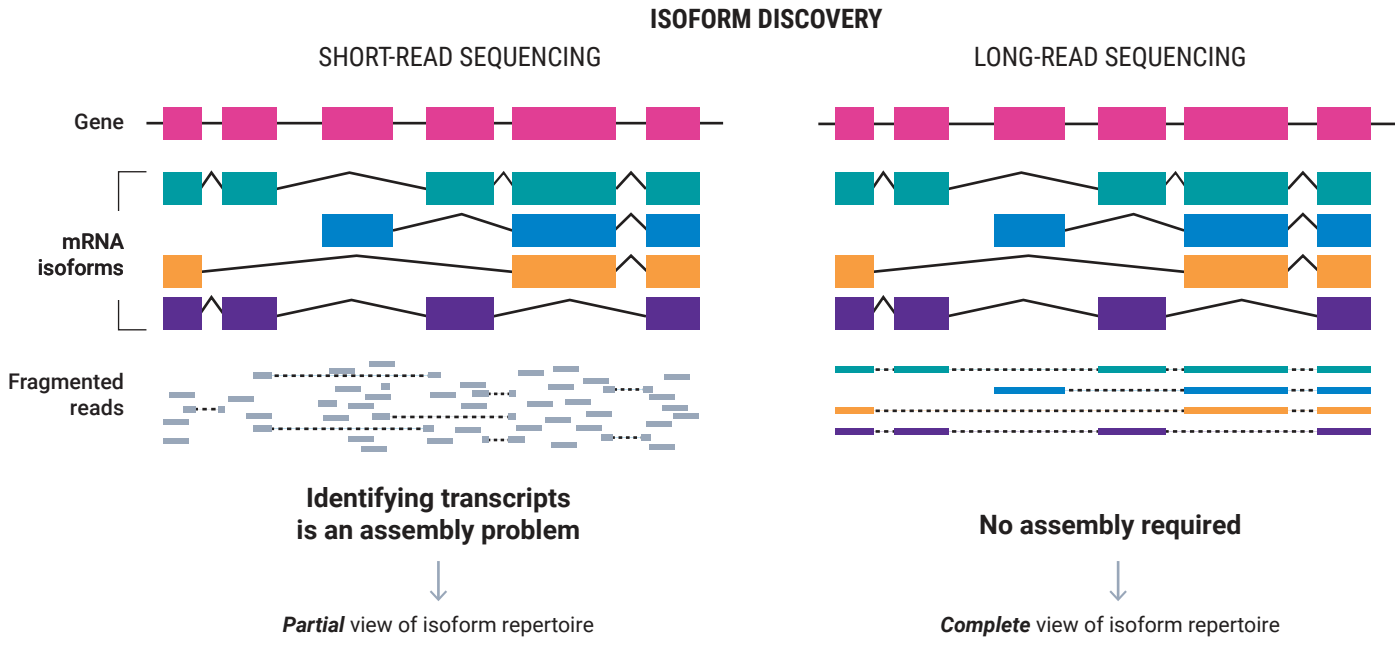

Figure 2. The HiFi figure on the right demonstrates the correct sequencing of the full length isoforms. In contrast, the short-read sequencing figure shows multiple exons sequenced. These exons must be assembled via software to recreate and approximate the original isoforms.

Kinnex™ full-length RNA kit

Reveal hidden isoform diversity with full-length isoform information for single-cell RNA sequencing resolution.

Get more with Kinnex single-cell RNA kit

With the introduction of the Kinnex method[5], PacBio Iso-Seq workflows are now higher throughput, for even greater transcriptome coverage. The Kinnex method is a concatenation technique that joins together cDNA molecules into longer concatenated fragments, allowing for multiple transcripts to be sequenced on a single HiFi read.

The result is higher throughput and reduced sequencing needs for cost-effective bulk and single-cell isoform sequencing. PacBio Iso-Seq workflow analysis is streamlined and fully supported in SMRT® Link, with outputs (gene- and isoform-level count matrices) that are compatible with tertiary analysis software such as Seurat,[6] Scanpy,[7] or Kana.[8]

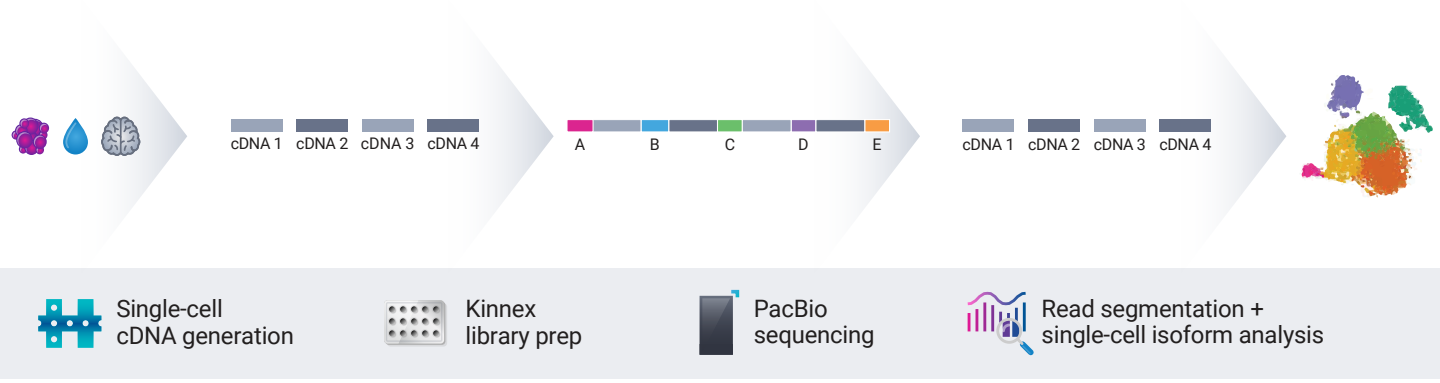

Figure 3. Kinnex for single-cell isoform sequencing. Single-cell cDNA molecules are concatenated into a larger insert library and sequenced. PacBio HiFi reads cover the entire insert, allowing for post-sequencing bioinformatics segmentation into reads representing the original cDNA molecule. Single-cell Iso-Seq workflow produces a gene- and isoform-level count matrix that can be used with tertiary analysis software for cell type analysis.

PacBio HUMAN GENOMICS

Headquarters
1305 O'Brien Drive
Menlo Park, CA 94025
United States
Phone: 1.650.521.8000

Customer service
1.877.920.PACB (7222), option 1
Fax: 1.650.618.2699
orders@pacb.com

Technical support
1.877.920.7222, option 2
support@pacb.com

We have offices in countries around the world. Visit **pacb.com/contact** for contact info.

READY TO GET STARTED WITH HIFI SEQUENCING?

Learn more about human genomics:
pacb.com/human-genomics

Connect with us for more info:

North America:
nasales@pacb.com

South America:
sasales@pacb.com

EMEA:
emea@pacb.com

Asia Pacific:
apsales@pacb.com

Connect with a PacBio scientist to get started:
pacb.com/scientist

KEY REFERENCES

1. Wang, T. et al. (2022) The Human Pangenome Project: A global resource to map genomic diversity.
2. Cohen, Ana S. A., et al. 2019. Genomic answers for children: dynamic analyses of >1000 pediatric rare disease genomes. *Genetics in Medicine* 1336–48.
3. Estimated by Adam English (Baylor College of Medicine) with Truvari software; SV from 36 haplotype-resolved assemblies spanning at least 50 bp; TR definitions from the UCSC *SimpleRepeats* track
4. Ebert, P. et al. (2021) Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science*.
5. Procedure & checklist – Preparing Kinnex libraries using Kinnex for 10x *Single Cell 3'* kit
6. Seurat: https://satijalab.org/seurat/
7. Scanpy: https://scanpy.readthedocs.io/
8. Kana: https://www.jkanche.com/kana/

102-326-536 REV04 18OCT2023

PacBio