



SCALE HUMAN DISEASE RESEARCH WITH HIFI SEQUENCING

Reveal the complexity of the human genome to power human disease research

The human genome holds a wealth of knowledge that reveals the underpinnings of human health. Despite accomplishing a sequence map of much of the genome in the early 2000s, scientists could not declare the human genome as “complete” until recent years. With the initial draft primarily restricted to coding-heavy regions, this reference lacked coverage of genes and regulatory elements that can play significant roles in health and disease. Only now, with the advancement of highly accurate long-read sequencing, could these missing pieces be resolved and fully complete the genome.

PacBio® HiFi sequencing gave rise to this achievement and has dramatically improved understanding of the human genome by:

- Resolving gaps that are unattainable by other sequencing methods
- Accurately mapping complex regions
- Enabling more contiguous genome assemblies

Achieving a complete map of the human genome is crucial for accurate identification of disease-causing genetic variants, understanding evolutionary relationships, and developing targeted therapies. By bridging these gaps, scientists can improve the understanding of genetic drivers of disease and disease risk, ultimately leading to improved human health.

2000

The first human genome is sequenced by the Human Genome Project

2022

This initial reference genome is made complete by the Telomere-to-Telomere Consortium with gapless chromosome assemblies

2023

The sequence of the remaining Y chromosome is completed

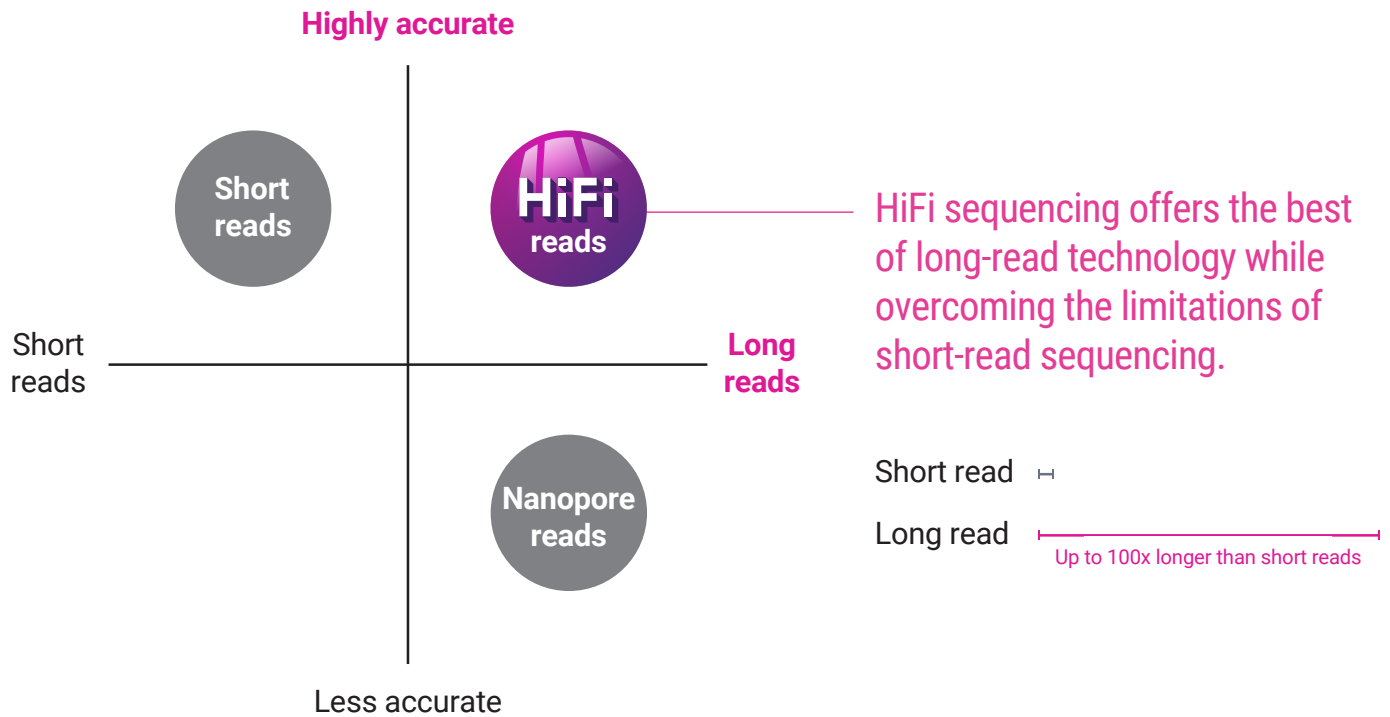
2023

Long-read sequencing, the critical technology for these milestones, is named *Nature Methods* “Method of the Year”



What is HiFi sequencing?

PacBio HiFi sequencing unites long reads and accuracy, giving you the highest quality genomic data for human health research. When it comes to understanding the genetic drivers of disease, why compromise with incomplete genomes that provide limited information?



The benefits of HiFi reads



Long read lengths
up to 20 kb



Easy library
preparation



Low coverage
requirements



Small file sizes to
minimize compute time



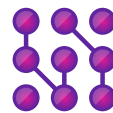
Supported analysis tools
consolidated into a
single pipeline



High read
accuracy



Compatible ecosystem
partners to enable an
end-to-end workflow

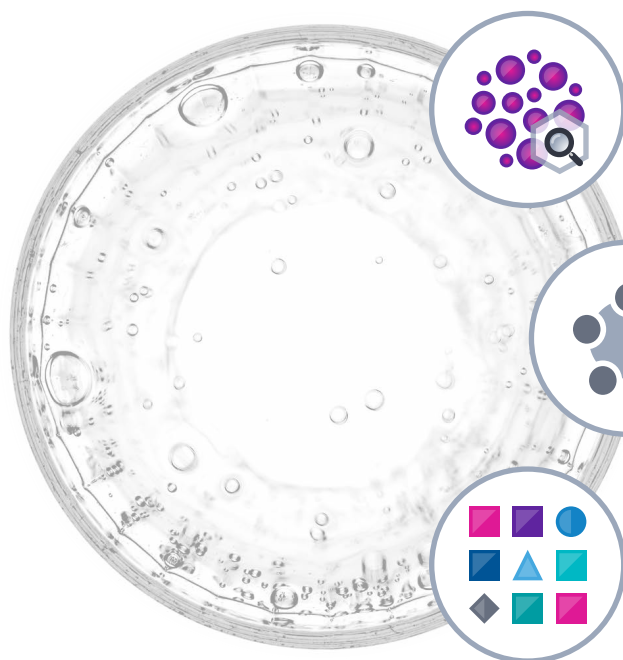


A single
technology solution



HiFi data in action

PacBio HiFi sequencing is uniquely suited to help solve problems in fields of human disease research:



Human biology and disease etiology



Clinical and translational
research of inherited disorders



Population health and precision
medicine research

Applications to power translational and human disease research



Whole genome sequencing

Generate highly accurate whole genomes to identify potential causative variants of disease



RNA sequencing

Capture the full transcriptome with isoform-level resolution using single-cell and full-length Kinnex™ kits



Structural variant calling

Characterize large structural variants that cause rare or inherited diseases



Repeat expansion detection

Comprehensively genotype tandem repeats with the PureTarget™ repeat expansion panel and TRGT computational tools



Epigenetics

Simultaneously detect atypical methylation patterns that contribute to rare diseases with on-board 5mC calling, no additional workflow required



Targeted sequencing

Help detect medically relevant gene variants missed by traditional short-read sequencing with the PureTarget repeat expansion panel and Twist Alliance *Dark Genes* panel



Research in human biology and disease etiology

The HiFi advantage

- **Enable** cutting-edge discovery with the extraordinary accuracy and genome quality of HiFi sequencing
- **Characterize** complex causative variants of disease with accurate long read lengths
- **Accelerate** research with tools at the forefront of genomics

Reveal the complexity of the human genome to better understand human biology and the development of disease

The power of a complete reference genome

Since its original release in 2000, the human reference genome has enabled important insights into human molecular biology but has remained incomplete by only covering coding-heavy regions of the chromosomes (International Human Genome Sequencing Consortium, 2001). Now, with the advent of highly accurate long-read sequencing, the Telomere-to-telomere (T2T) consortium has finally achieved a complete reference genome using HiFi sequencing to generate gapless assemblies for the full length of all chromosomes (Nurk et al., 2022), including the Y chromosome (Rhie et al., 2023). This historic milestone marks a turning point for research into human health, as all progress in identifying genetic drivers of disease and potential therapy targets rests on the strength of the reference genome. With the extraordinary accuracy and read lengths afforded by HiFi sequencing, the quality of this newly complete reference genome assures the excellence of these discoveries.

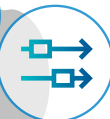
Whole genome:

Reveal the variants associated with Alzheimer's disease using highly accurate WGS



SV detection:

Identify novel mutations associated with retinal disease with detection of structural variants



RNA sequencing:

Create a cell-specific isoform atlas for colorectal cancer with single-cell RNA sequencing

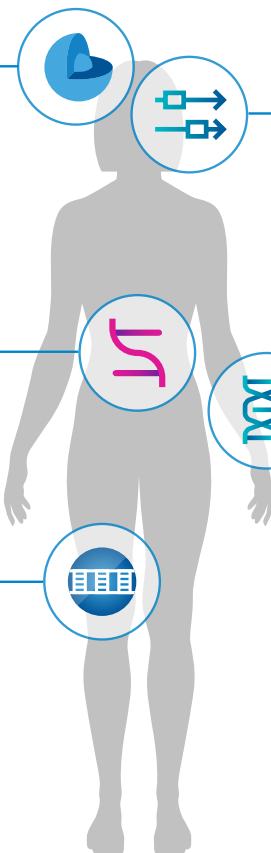


Methylation:

Detect 5mC methylation in myotonic dystrophy gene *DMPK*

Repeat expansions:

Characterize repeat expansions in ALS gene *C9orf72*



The impact of highly accurate WGS and detection of complex variants

Now with the completion of the human reference genome, applications like whole genome sequencing (WGS) and the detection of complex variants can be employed to their full potential to reveal important insights into human disease. Where traditional technologies were limited to the use of whole exome sequencing for human disease research, HiFi WGS is now leading the way for a more complete picture of causative variants for various diseases including Alzheimer's disease (Tesi, Salazar, et al., 2024), rare pediatric disease (Cohen et al., 2022), and structural variants in Mendelian disease (AlAbdi et al., 2023; Merker et al., 2018).

Not limited to only the genome, HiFi RNA sequencing technology is also making significant strides in its contribution to cell atlases, which are important for understanding developmental biology (Joglekar et al., 2024), as well as linking molecular mechanisms of disease to possible therapies (Rood et al., 2022). Using the PacBio Iso-Seq® method, full-length single-cell isoform sequencing has contributed towards a resolved atlas of cell-type isoforms in the human cortex and for colorectal cancer, the latter of which is being used for research in identifying recurring neopeptides as potential cancer vaccine candidates.

Advances in genomics tools

These advances are only accelerated by the continued development of genomic tools that help detect complex causative variants. This includes a recent groundbreaking multiomics study that combines the genome, CpG methylome, chromatin epigenome, and transcriptome on a single Revio® sequencing run (Vollger et al., 2023). Additionally, the development of advanced variant calling tools allows for the identification of complex disease-causing variants, like tandem repeats, using the Tandem Repeat Genotyping Tool (TRGT, Dolzhenko et al., 2024). These methods are complemented by PacBio computational tools to extract the most value out of long reads, including the WGS Variant Pipeline that consolidates these analysis tools into a single workflow for alignment, comprehensive variant calling, cohort analysis, and annotation.

“The feat [T2T] probably wouldn't have been possible without new sequencing technology from PacBio.”

Karen Miga

University of California, Santa Cruz, T2T member

REFERENCES

1. AlAbdi, L., et al. (2023). Beyond the exome: utility of long-read whole genome sequencing in exome-negative autosomal recessive diseases. *Genome Medicine*, 15(1), 114.
2. Cohen, A. S., et al. (2022). Genomic answers for children: Dynamic analyses of > 1000 pediatric rare disease genomes. *Genetics in Medicine*, 24(6), 1336-1348.
3. Dolzhenko, E., et al. (2024). Characterization and visualization of tandem repeats at genome scale. *Nature Biotechnology*, 1-9.
4. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).
5. Joglekar, A., et al. (2024). Single-cell long-read sequencing-based mapping reveals specialized splicing patterns in developing and adult mouse and human brain. *Nature Neuroscience*, 1-13.
6. Merker, J. D et al. (2018). Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genetics in Medicine*, 20(1), 159-163.
7. Nurk, S., et al. (2022). The complete sequence of a human genome. *Science*, 376(6588), 44-53.
8. Rhie, A., et al. (2023). The complete sequence of a human Y chromosome. *Nature*, 621(7978), 344-354.
9. Rood, J. E., et al. (2022). Impact of the Human Cell Atlas on medicine. *Nature medicine*, 28(12), 2486-2496.
10. Tesi, N., Salazar, A., et al. (2024). Characterising tandem repeat complexities across long-read sequencing platforms with TREAT. *bioRxiv*, 2024-03.
11. Vollger, M. R., et al. (2023). Synchronized long-read genome, methylome, epigenome, and transcriptome for resolving a Mendelian condition. *bioRxiv*.



Clinical and translational research to combat inherited disorders

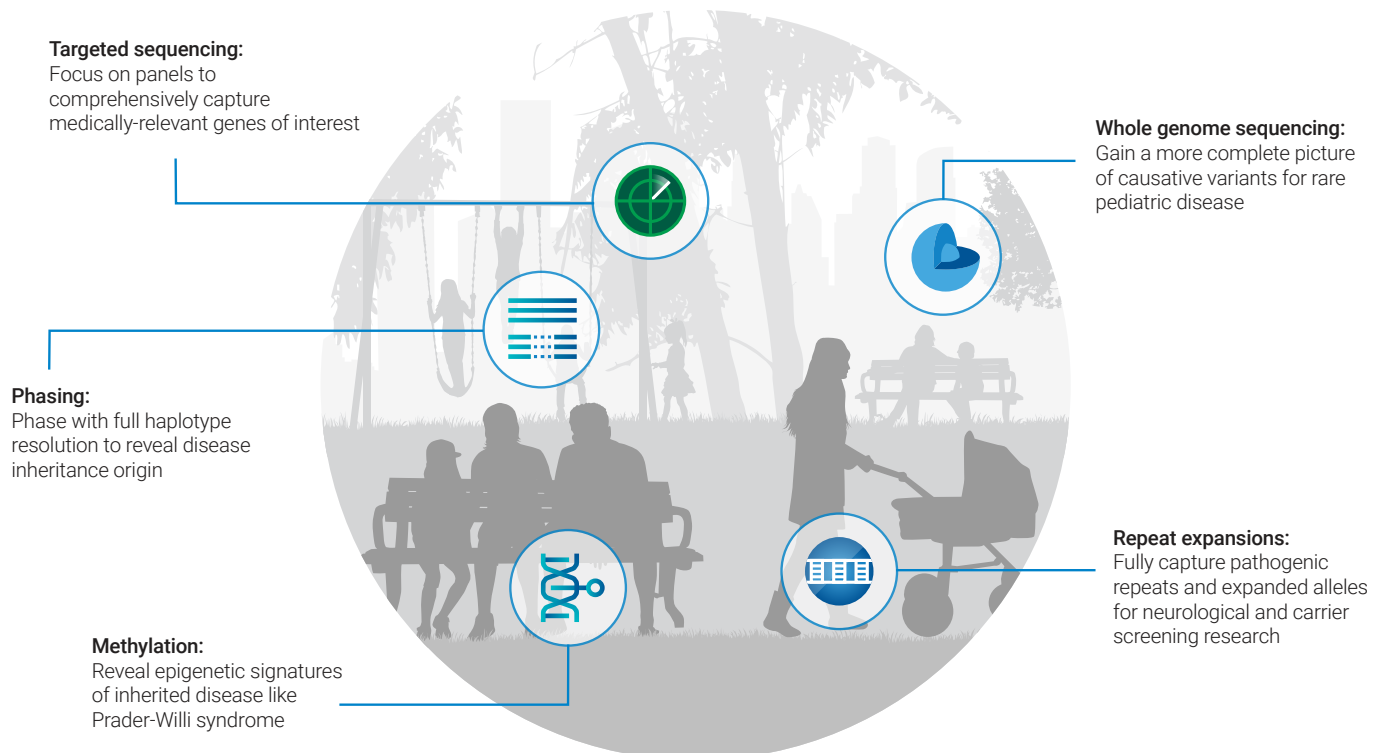
The HiFi advantage

- **Generate** highly accurate, ancestry-agnostic genomes to target genes that may cause rare and inherited disease
- **Resolve** more variants in complex medically relevant genes relative to short-read sequencing
- **Harness** computational tools and analysis partners to better solve rare disease cases
- **Consolidate** multiple assays on one platform, with end-to-end workflows to drive operational efficiency

Drive research on clinically-relevant genes with the most complete human genome

Shining a light on dark genes

Innovations in genome sequencing have led to the flourishing of sequencing for medically relevant genes, but many of these genes still prove to be difficult to characterize in clinical research. The Genome in a Bottle Consortium provides benchmark sets for medically relevant genes, but data for nearly 400 genes was excluded from this original release due to the repetitive and polymorphic complexity of these genes. Recent work by Wagner et al., (2024) however, rectifies this missing data with the characterization of an additional 273 challenging autosomal genes using a HiFi haplotype-resolved whole-genome assembly. This work suggests that assemblies at this resolution may serve as the blueprint for future benchmarks that span the whole genome.



A recent study confirms that not all sequencing methods are created equally when it comes to assessing and understanding clinically relevant genes. Research from the All of Us initiative (Mahmoud et al., 2024) conducts a head-to-head comparison between long- and short-read technologies for sequencing complex medically relevant genes and determines that HiFi sequencing captures the most accurate picture of these genes. This study indicates that HiFi sequencing allows for the identification of more variants at lower depth than legacy methods, as well as the discovery of previously hidden variations – something that has not been achievable with current short-read sequencing. This conclusion signals that HiFi sequencing is surpassing traditional short-read technology in the ability to characterize and understand a wide array of causal variants for rare or inherited disease. This superiority has been borne out in recent research across a broad spectrum of rare diseases (AlAbdi et al., 2023; Cohen et al., 2022; Steyaert, Sagath et al., 2024) and neurodevelopmental disorders (Chen et al., 2023; Hiatt et al., 2024).

“HiFi reads [enable] accurate...detection of all variant classes, and provide a rich genetic context... [advancing] human genetics and... expanding utility of SMRT sequencing.”

Fritz Sedlazeck, PhD

Associate Professor, Human Genome Sequencing Center,
Baylor College of Medicine

Innovation for rare and inherited disorders

For many rare disorders with a lack of functional evidence or where DNA sequencing is uninformative, a growing body of research suggests that RNA sequencing may have diagnostic utility. With the Iso-Seq method, researchers can discover aberrant expression or splicing, prioritize candidate variants, and identify allele-specific expression patterns. The addition of Kinnex RNA sequencing to the sequencing toolbox furthers the understanding of the role of isoforms in disease pathogenesis.

To enable clinical researchers to overcome the challenges posed by traditional sequencing technology, targeted sequencing panels like the Twist Alliance Dark Genes and Twist Alliance Pharmacogenomics panels were designed by collaborators to phase and more comprehensively detect variants in medically relevant genes. Targeted sequencing offers cost-effective scalability on the Revio system, and allows clinical researchers to focus on genes of interest, while still benefiting from all advantages of HiFi sequencing. These panels are now joined by the PacBio PureTarget repeat expansion panel, which captures some of the most challenging repetitive regions of the genome while preserving methylation status. These technologies speak to a growing appreciation of the role of HiFi long-read sequencing in its potential as a single platform for clinical research applications like rare disease, pharmacogenomics, neurology and cardiology research, and carrier screening.

REFERENCES

1. AlAbdi, L., et al. (2023). Beyond the exome: utility of long-read whole genome sequencing in exome-negative autosomal recessive diseases. *Genome Medicine*, 15(1), 114.
2. Chen, X., et al. (2023). Comprehensive SMN1 and SMN2 profiling for spinal muscular atrophy analysis using long-read PacBio HiFi sequencing. *The American Journal of Human Genetics*, 110(2), 240-250.
3. Cohen, A. S., et al. (2022). Genomic answers for children: Dynamic analyses of > 1000 pediatric rare disease genomes. *Genetics in Medicine*, 24(6), 1336-1348.
4. Hiatt, S. M., et al. (2024). Long-read genome sequencing and variant reanalysis increase diagnostic yield in neurodevelopmental disorders. *medRxiv*, 2024-03.
5. Mahmoud, M., et al. (2024). Utility of long-read sequencing for All of Us. *Nature communications*, 15(1), 837.
6. Steyaert, W., Sagath, L., et al. (2024). Unravelling undiagnosed rare disease cases by HiFi long-read genome sequencing. *medRxiv*, 2024-05.
7. Wagner, J., et al. (2022). Curated variation benchmarks for challenging medically relevant autosomal genes. *Nature biotechnology*, 40(5), 672-680.



Population health and precision medicine research

The HiFi advantage

- **Access** blood and saliva samples with lower DNA input requirements using Nanobind kits
- **Sequence** high-volume whole genomes cost effectively at extraordinary accuracy
- **Assess** population-specific variants at a large scale with 15× higher throughput than previous technology
- **Characterize** complex genes with scalable workflows

Capture the full range of genetic diversity to drive individual- and population-level initiatives

Pangenomes to establish population-specific health outcomes

Advances in genomic technology have proven especially important when building a pangenome, which is meant to represent the entire set of possible genes in a genome rather than select genes derived from a reference genome from a few individuals. Human pangenomes are crucial for capturing variation that may be specific to a certain population and can have significant impacts to human health (Ji et al., 2022) like the gene that causes Thalassemia, an inherited blood disorder disproportionately concentrated in the Eastern Hemisphere (Luo et al., 2021). With accurate detection of large structural variants, HiFi sequencing has been used to construct a pangenome reference of Chinese (Gao et al., 2023) and Arab (Uddin et al., 2023) populations, yielding millions of novel variants unique

to these populations and not previously included in the existing HPRC draft human pangenome reference (Liao et al., 2022). These groundbreaking new references now enable genomic analyses that better reflect the full extent of genetic diversity, including populations that have been underrepresented in existing genomes, and allow for actionable population health initiatives. This advance has been made possible by the comprehensive data and remarkable scalability afforded by the high-throughput long-read sequencing technology available on the Revio system.

Whole genome:

Capture challenging variant classes for population-specific reference sequences

Pangenome:

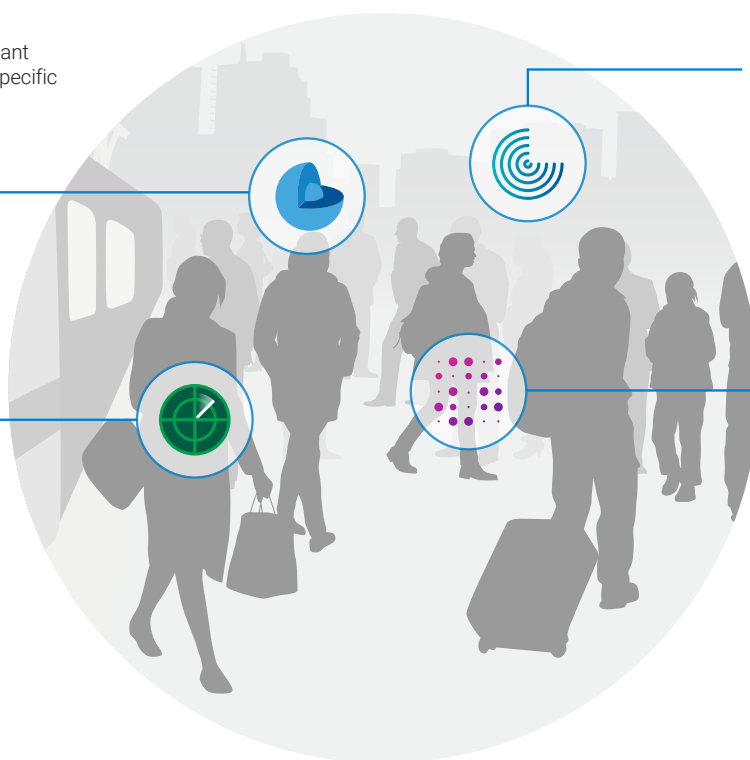
Generate a more comprehensive pangenome to capture the full range of genetic diversity in underrepresented populations

Targeted sequencing:

Enable carrier screening of neurodevelopmental disorders like Fragile X with targeted repeat expansion detection

Ancestry-agnostic:

Understand drug response within a population by capturing genes like *CYP2D6* without bias



Population health initiatives for precision medicine and population screening

A complementary advantage in expanding population representation in genome reference sets is the potential to implement population health programs at scale. These population health efforts, like precision medicine, and carrier and risk screening, aim to understand the impact of genetic variation, environment, and lifestyle on the health of individuals, as well as the role of genetics in the development of complex traits, such as susceptibility to certain diseases. This impact is best understood through analyses of the *CYP2D6* gene (Qiao, Yang, et al., 2015; Buermans et al., 2017), which is responsible for the metabolism of a large portion of pharmaceutical drugs. HiFi sequencing has enabled the characterization of several functional rare and novel

CYP2D6 alleles that affect metabolic traits that are unique to the previously unrepresented Solomon Island population (Charnaud, Munro, et al., 2022).

PacBio helps support this type of pharmacogenomic research through a collaboration with Twist Bioscience. The Twist Alliance Long-Read PGx Panel enables the characterization of haplotypes and star allele architecture of up to 50 clinically relevant PGx genes, with high accuracy and high coverage (van der Lee et al., 2022). Similarly, the PureTarget panel, which supports research in screening of ALS, fragile X syndrome, and similar ataxias, can also be used for large scale screening of populations. Altogether, the ability of HiFi sequencing to generate extraordinarily accurate genomes at scale holds the promise of revolutionizing the future of population health.

“A single test to get all types of variation from the very small to the really large ones”

Professor Lars Feuk
Uppsala University

REFERENCES

1. Buermans, H. P., et al. (2017). Flexible and scalable full-length *CYP2D6* long amplicon PacBio sequencing. *Human mutation*, 38(3), 310-316.
2. Charnaud, S., et al. (2022). PacBio long-read amplicon sequencing enables scalable high-resolution population allele typing of the complex *CYP2D6* locus. *Communications Biology*, 5(1), 168.
3. Gao, Y., et al. (2023). A pangenome reference of 36 Chinese populations. *Nature*, 619(7968), 112-121.
4. Ji, Y., et al. (2022). Characterizing the genetic polymorphisms in 370 challenging medically relevant genes using long-read sequencing data from 41 human individuals among 19 global populations. *bioRxiv*, 2022-08.
5. Liao, W. W., et al. (2023). A draft human pangenome reference. *Nature*, 617(7960), 312-324.
6. Luo, S., et al. (2022). The value of single-molecule real-time technology in the diagnosis of rare thalassemia variants and analysis of phenotype–genotype correlation. *Journal of human genetics*, 67(4), 183-195.
7. Qiao, W., et al. (2016). Long-read single molecule real-time full gene sequencing of cytochrome P450-2D6. *Human mutation*, 37(3), 315-323.
8. Uddin, M., et al. (2023). A draft Arab pangenome reference. *Research Square*. <https://doi.org/10.21203/rs.3.rs-3490341/v1>
9. van der Lee, M., et al. (2022). Design and performance of a long-read sequencing panel for pharmacogenomics. *bioRxiv*, 2022-10.



Sequence on your terms

Whether you need the power and scalability of the Revio system or the simplicity and accessibility of the Vega™ benchtop system, PacBio has you covered with the same HiFi long-read sequencing at exceptional 99.9% accuracy.

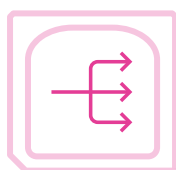
Multimomics in every run

SMRT Cell 1



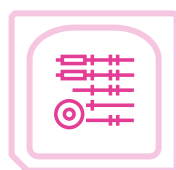
2 whole human genomes (20× coverage) with 5mC and 6mA methylation detection

SMRT Cell 2



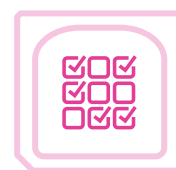
48 samples multiplexed for carrier screening of *FMR1* with PureTarget

SMRT Cell 3



1 sample for Kinnex single-cell isoform sequencing for a human cell atlas project

SMRT Cell 4



96 samples multiplexed Twist Alliance *Long-Read Pharmacogenomic panel*



Revio system

The Revio system with SPRQ™ chemistry delivers an impressive 120 Gb per SMRT® Cell, making it the clear choice for broad-scale population genomics or precision health programs.

With independent sequencing stages, you can run up to 4 SMRT Cells simultaneously, enabling multiple assays on a single platform.

Only 500 ng of sample input allows access to samples and applications previously not possible.

The Revio system is ideal for human genomes, single cell sequencing, full-length isoform sequencing, and human methylation profiles.



Vega benchtop system

The Vega system delivers 60 Gb of premium HiFi data per SMRT Cell — accessibly priced to enable your application of choice.

A single SMRT Cell configuration puts you in full control for streamlined, immediate sequencing.

Innovative Revio technology packed to fit on a benchtop — equivalent HiFi data in a reduced instrument footprint.

The Vega system is ideal for full-length isoform sequencing and targeted panels like the PureTarget™ repeat expansion kit.

US list price is \$995 for sequencing reagents for one Revio SMRT Cell, which has an expected yield of 120 Gb, equivalent to two 20× human genome. US list price is \$1,100 for sequencing reagents for one Vega SMRT Cell, which has an expected yield of 60 Gb, equivalent to one 20× human genome. Your local sales representative can provide detailed pricing in your currency.

More samples, more discovery

Samples per SMRT Cell for human genomics applications:



Vega™ system
HiFi within reach

1 SMRT® Cell



Revio® system with
SPRQ™ chemistry
HiFi at scale

1 SMRT Cell

4 SMRT Cells

Application	Samples per run		
Whole genome sequencing			
Human genome (20×)*	1	2	8
Human methylation profiling (5×)	4	8	32
RNA sequencing			
Kinnex single-cell RNA	1 (3,000–6,000 cells)	1 (6,000–10,000 cells)	4 (6,000–10,000 cells)
Kinnex full-length RNA			
5M reads	6	12	48
10M reads	3	6	24
Targeted sequencing			
Amplicon sequencing	≥1,000	>1,000	>4,000
Target enrichment			
20 Mb panel	12	16	64
2 Mb panel	72	96	384
100 kb panel	288	384	1,536
PureTarget repeat expansion panel	48	48	192

* 40× human genome can be achieved through 1 SMRT Cell on the Revio system with SPRQ chemistry. 2 SMRT Cells are recommended to achieve at least 30× on the Vega system.



Ready to get started with HiFi sequencing?



Learn more about HiFi sequencing:
pacb.com/hifi



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



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

Research use only. Not for use in diagnostic procedures. © 2024 Pacific Biosciences of California, Inc. ("PacBio"). All rights reserved. Information in this document is subject to change without notice. PacBio assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of PacBio products and/or third-party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at pacb.com/license. Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, SBB, Revio, Onso, Apton, Kinnex, PureTarget, SPRQ, and Vega are trademarks of PacBio.

