



## Scientists Solve One of Genomics' Biggest Challenges by Using HiFi Sequencing to Distinguish Highly Similar Paralogous Genes

March 17, 2025 at 9:05 AM EDT

MENLO PARK, Calif., March 17, 2025 (GLOBE NEWSWIRE) -- PacBio (NASDAQ: PACB), a leading provider of high-quality, highly accurate sequencing platforms, today announced a newly published study in [Nature Communications](#) unveiling a powerful new method for analyzing some of the most complex regions of the human genome. Led by researchers from PacBio, GeneDx, and a global consortium of genomics experts, the study utilizes [Paraphase](#), an informatics tool that, when paired with HiFi long-read sequencing, allows for high-precision variant detection and copy number analysis in 316 previously inaccessible segmental duplication regions, including 9 challenging medically-relevant genes.

Segmental duplications (SDs) are highly similar, duplicated regions of the genome that have posed persistent challenges for genetic analysis. These regions contain hundreds of genes critical to human health—including those implicated in spinal muscular atrophy (SMN1/SMN2), congenital adrenal hyperplasia (CYP21A2), and red-green color blindness (OPN1LW/OPN1MW)—but their high sequence similarity makes accurate mapping and variant detection nearly impossible with short-read sequencing. Paraphase, combined with HiFi sequencing, overcomes these challenges by phasing haplotypes across paralogous gene families, providing a more complete and accurate view of genetic variation. This is enabled by the length and accuracy of reads from HiFi sequencing.

### Study Reveals Previously Inaccessible Regions of the Genome

By applying Paraphase to 160 long (>10 kb) segmental duplication regions spanning 316 genes, the researchers revealed new insights into genetic variation across five ancestral populations.

Among the key findings:

- **Newly Identified De Novo Variants in SDs in Parent-Offspring Trios:** Analysis of 36 trios uncovered 7 previously undetected de novo single nucleotide variants (SNVs) and 4 de novo gene conversion events, two of which were non-allelic—a level of detail not possible with traditional sequencing approaches.
- **Copy Number Variability Across Populations:** The study profiled the copy number distributions of paralog groups across populations, showing high copy number variability in many gene families in SDs. It also provided a new approach for identifying false duplications in the reference genome.
- **Gene Conversion Drives Sequence Similarity between Genes and Paralogs:** The team identified 23 paralog groups with strikingly low genetic diversity between genes and paralogs, indicating that frequent gene conversion and/or unequal crossing-over may have played a role in preserving highly similar gene copies over time.

“For decades, sequencing technologies have struggled to provide reliable data on paralogous genes—some of the most medically relevant but hardest to analyze regions of the genome,” said Dr. Michael A. Eberle, Vice President of Bioinformatics at PacBio and senior author of the study. “With Paraphase and HiFi sequencing, we now have a scalable way to accurately genotype SD-encoded genes across diverse populations, filling in long-standing gaps in genomic research and improving our ability to identify disease-linked variants.”

The study also highlights how Paraphase can disentangle medically important gene families that have long required specialized, multi-step assays like MLPA and Sanger sequencing. For example, in the CYP21A2/CYP21A1P region—where mutations cause congenital adrenal hyperplasia—the researchers characterized a previously overlooked duplication allele carrying both a functional CYP21A2 copy and a CYP21A2(Q319X) pseudogene copy, which could have led to misclassification in standard tests.

“This study demonstrates that when we use HiFi sequencing we see a much richer and more complex picture of genetic variation,” said Dr. Xiao Chen, lead author of the study and principal scientist at PacBio. “Paraphase enables the precise resolution of genetic regions that have been largely inaccessible until now, providing new opportunities for disease research, population genetics, and potentially even clinical testing.”

“Long-read genome sequencing offers the ability to detect variants that are difficult to identify using other testing methods, particularly in regions with highly similar sequence,” said Dr. Paul Kruszka, MD, FACMG, Chief Medical Officer at GeneDx. “This work may enhance variant detection, resolve complex genomic regions, and provide more answers for patients and families, so we are encouraged by the prospect of the data.”

The full study, “[Genome-wide profiling of highly similar paralogous genes using HiFi sequencing](#),” is now available in *Nature Communications*.

### About PacBio

PacBio (NASDAQ: PACB) is a premier life science technology company that is designing, developing and manufacturing advanced sequencing solutions to help scientists and clinical researchers resolve genetically complex problems. Our products and technologies stem from two highly differentiated core technologies focused on accuracy, quality and completeness which include our HiFi long-read sequencing and our SBB® short-read sequencing technologies. Our products address solutions across a broad set of research applications including human germline sequencing, plant and animal sciences, infectious disease and microbiology, and oncology. For more information, please visit [www.pacb.com](http://www.pacb.com) and follow @PacBio.

*PacBio products are provided for Research Use Only. Not for use in diagnostic procedures.*

### Forward Looking Statements

This press release may contain “forward-looking statements” within the meaning of Section 21E of the Securities Exchange Act of 1934, as amended, and the U.S. Private Securities Litigation Reform Act of 1995. All statements other than statements of historical fact are forward-looking statements, including statements relating to the uses, coverage, advantages, and benefits or expected benefits of using, PacBio products or technologies, including in connection with providing a scalable way to accurately genotype SD-encoded genes across diverse populations, fill in long-standing gaps in genomic research, and improve the ability to identify disease-linked variants; enabling precise resolution of genetic regions that were previously largely inaccessible; providing new opportunities for disease research, population genetics, and potential clinical testing; potentially detecting or enhancing the detection of variants difficult to identify using other methods, resolving complex genomic regions, and providing more answers for patients and families; and other future events. You should not place undue reliance on forward-looking statements because they are subject to assumptions, risks, and uncertainties and could cause actual outcomes and results to differ materially from currently anticipated results, including, the difficulty of generating discoveries in complicated areas of biology; potential performance, quality and regulatory issues; and third-party claims alleging infringement of patents and proprietary rights or seeking to invalidate PacBio’s patents or proprietary rights. Additional factors that could materially affect actual results can be found in PacBio’s most recent filings with the Securities and Exchange Commission, including PacBio’s most recent reports on Forms 8-K, 10-K, and 10-Q, and include those listed under the caption “Risk Factors.” These forward-looking statements are based on current expectations and speak only as of the date hereof; except as required by law, PacBio disclaims any obligation to revise or update these forward-looking statements to reflect events or circumstances in the future, even if new information becomes available.

## **Contacts**

Investors and Media:

Todd Friedman

[ir@pacificbiosciences.com](mailto:ir@pacificbiosciences.com)

Media:

[ir@pacificbiosciences.com](mailto:ir@pacificbiosciences.com)