Clinical Validation of a Haplotyped Whole Genome Sequencing Assay Using PacBio HiFi Sequencing

David Corney, Carrie Ziemniak, Wei Wang, Yu Qiu, Christopher Mozdzierz, Jeffrey Otto, Haythem Latif

GENEWIZ from Azenta Life Sciences, South Plainfield, NJ 07080

Abstract

Next generation sequencing (NGS) has revolutionized genomics and enabled unprecedented insights into the molecular basis of health and disease. Short-read NGS technologies have become integral to the diagnosis of genetic diseases, and in supporting clinical trials. Unfortunately, while current short-read technologies are more cost-effective and scalable than ever before, short read-lengths are limited in resolving complex genomic regions, detecting structural variations, and covering repetitive sequences. Additionally, short-reads may not be able to accurately differentiate between trans and cis compound heterozygosity. This is critical in human germline diploid studies, since whether variants are present on the same or different chromosomes may confer vastly different phenotypes. While such challenges can be mitigated by long-read sequencing approaches, widespread clinical adoption is limited due to challenges in sample handling, scalability, and cost.

To address these limitations, we report the CLIA/CAP validation of a clinical long-read WGS sequencing assay for the comprehensive genome-wide detection and phasing of SNVs, and INDELs. High molecular weight DNA was extracted from frozen EDTA whole blood, followed by DNA fragmentation to ~15-20 kb by automated pipette-mediated mechanical shearing, and SMRTbell® library preparation. After sequencing each sample on a PacBio® Revio™ SMRT® Cell, ~20-30x average genome-wide coverage was obtained. Variant calling accuracy was assessed by comparing variant calls in healthy donors to those obtained by Illumina® PCR-free WGS assay, as well as analysis of Genome In A Bottle (GIAB) well-characterized reference samples. Overall, the variant calling concordance F1 score was 0.99 and 0.98 for SNVs and INDELs, respectively. Phasing accuracy was performed by trio analysis on specific genes of interest.

This validated long-read WGS assay overcomes key limitations of short-read technologies, enabling comprehensive genome-wide detection and phasing of all variant classes. By leveraging automation and high-throughput workflows, this assay provides a powerful tool for precision genomic medicine, with the potential to improve diagnostic accuracy, treatment selection, and patient outcomes.

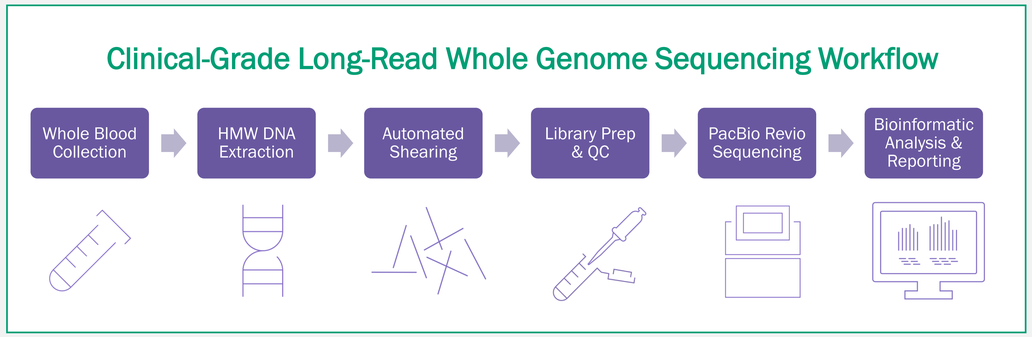


Figure 1. Schematic of clinically validated PacBio WGS workflow. High molecular weight (HMW) DNA is obtained from whole blood followed by automated DNA shearing using a Hamilton® STAR™ to target ~15-20 kb average insert size. Subsequently, libraries are generated and sequenced on a Revio sequencer followed by analysis for variant calling, and phasing of SNPs and INDELs.

A.

SAMPLE ID	YIELD (ug)	A260/280 RATIO	$GQN_{10\ kb}$
Sample 1	6.6	1.85	9.3
Sample 2	7.5	1.87	9.6
Sample 3	5.3	1.87	9.2
Sample 4	8.4	1.84	9.7
Sample 5	9.6	1.81	9.8

B.

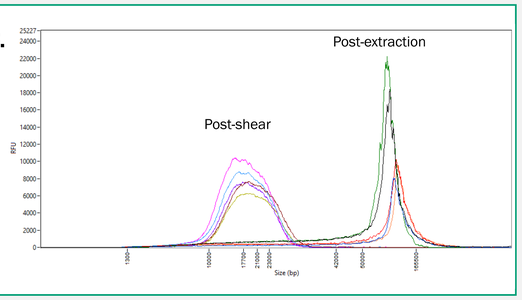
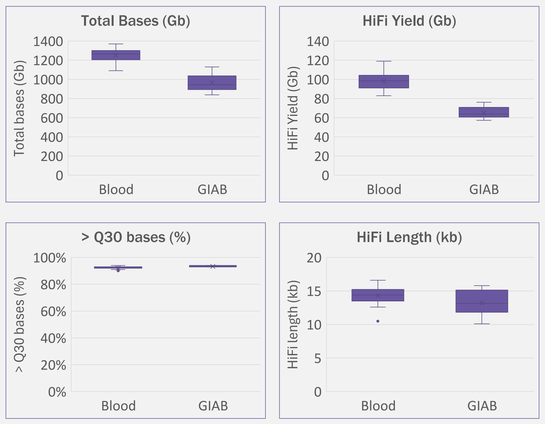


Figure 2. HMW DNA QC and automated pipette shearing. (A) Following extraction, nucleic acids are QC'd by NanoDrop™, Qubit™, and Femto Pulse™. QC results from five representative whole blood samples is shown - showing DNA with high yield, quality, and integrity obtained from 400 µl whole blood. (B) Femto Pulse plot showing HMW DNA with average ~100 kb sizing post-extraction ("Post-extraction"), and uniform ~15-20 kb sized DNA following automated pipette-mediated shearing using Hamilton Microlab STAR ("Post-shear"). A single shearing condition was used for all samples.

A.



B.

SAMPLE ID	ACCURACY		
	SNV	INDEL	SNV/INDEL COMBINED
Sensitivity	99.7%	91.4%	98.5%
Specificity	98.1%	88.2%	96.3%
Accuracy	99.1%	90.0%	97.6%
Positive Predictive Value (PPV)	98.7%	90.4%	97.5%
Negative Predictive Value (NPV)	99.6%	89.4%	97.8%
F1 Score (GIAB)	99.9%	98.0%	N/A

C.

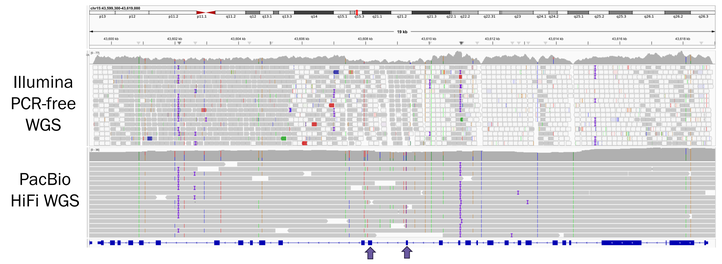
SAMPLE ID	INTRA-BATCH CONCORDANCE	
	SNV	INDEL
NA24385	98.7%	82.5%
NA24149	98.6%	82.5%
Sample 1	98.8%	86.8%

D.

SAMPLE ID	INTER-BATCH CONCORDANCE	
	SNV	INDEL
NA24385	98.7%	83.7
NA24149	98.6%	83.0%
Sample 1	98.8%	87.7%

Figure 3. Summary of analytical validation of HiFi WGS using PacBio Revio sequencer. (A) Raw data output per Revio SMRT Cell. Values are shown for HMW DNA obtained from blood, as well as partially degraded DNA obtained from Coriell Genome In A Bottle (GIAB) samples. (B) Summary of analytic performance characteristics. All metrics were obtained by comparing SNV and INDEL calls by this assay to Azenta's clinical-grade PCR-Free Illumina WGS assay across 20 donor samples, except the F1 score which was calculated compared to public GIAB data. (C, D) Concordance between intra- and inter-batch triplicate reactions.

A.



B.

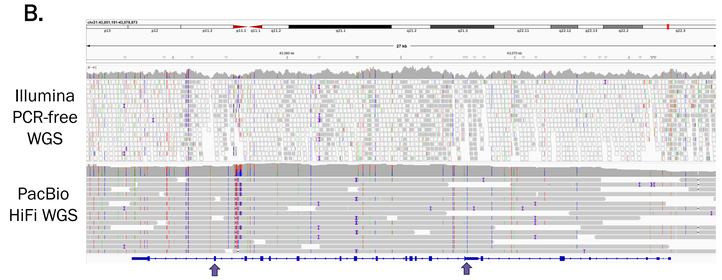


Figure 4. PacBio HiFi data enables detection of variants within challenging regions of the genome compared to short-read WGS. Two examples of medically-relevant genes which have been reported to be challenging to sequence by short-read platforms are shown. Arrows indicate exonic regions with low short-read coverage. (A) STRC, which is involved in deafness. (B) CBS, which has been associated with homocystinuria.

A.

SAMPLE ID	PHASED VARIANTS IN TRUTHSET	CORRECTLY PHASED VARIANTS	PHASING CONCORDANCE
NA12878	190	190	100%
NA24385	138	138	100%
NA24631	72	72	100%
NA19238	229	227	99.1%
NA19239	178	178	100%

B.

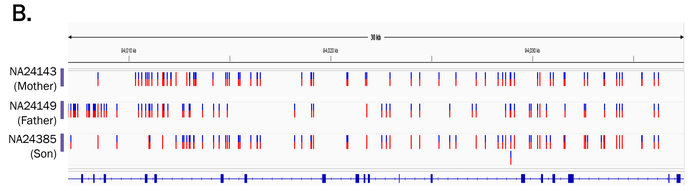


Figure 5. The analysis pipeline accurately phases SNV and INDEL variants. (A) Summary of phasing accuracy in a ~120 kb gene of interest compared to public data. (B) Example of variant phasing across a 30 kb region within a GIAB Trio (NA24143/NA24149/NA24385).

Conclusions

- Long-reads enable detection of challenging variants and allows for haplotyping which may be especially critical in rare disease.
- Validation studies demonstrate up to 99% variant calling accuracy compared to reference datasets, and phasing concordance.
- Regulatory approval has been granted and the test is on Azenta's clinical test menu.

Acknowledgements: We are grateful to Nevil Patel, Ping Xu, Mia Sutton, and Ye Wang (Azenta) for automation and sequencing assistance, and Jay Morris, Gregory Young, and Andrew Mason (PacBio) for technical support.