



High-throughput HMW genomic DNA extraction from saliva for large-scale PacBio HiFi sequencing projects using WGS or PureTarget workflows

Deborah Moine, Sarah Kingan, Shreyasee Chakraborty, Jeff Zhou, William J. Rowell, Christina Dillane*, Mike Tayeb*, Heather Ferrao
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, *DNA Genotek Inc. Ottawa, ON, Canada

4042W

HMW DNA extraction from saliva samples collected in Oragene™ device



It would be great if I could use saliva samples for HiFi projects because collecting and storing are much easier than blood!

Is it possible to extract High Molecular Weight DNA from saliva samples?



Yes! With saliva collected in Oragene™ devices and DNA extracted with Nanobind kits, we obtain HMW DNA optimal for HiFi sequencing!



In this poster, we demonstrate that high quality PacBio HiFi sequencing results can be reliably generated from saliva samples collected in DNA Genotek Oragene™ devices and extracted using the Nanobind kits. We performed whole genome sequencing (WGS) and PureTraget enrichment sequencing from saliva samples extracted manually or high throughput (HT).

High DNA quality from saliva

- 80 saliva samples were extracted manually or HT from 500 µL of saliva using Nanobind kits
- HT extractions were performed on ThermoFisher KingFisher Apex system (75 samples in 2.5 hours)
- DNA yield = 1 to 28 µg of HMW DNA (modal size > 100 kb).

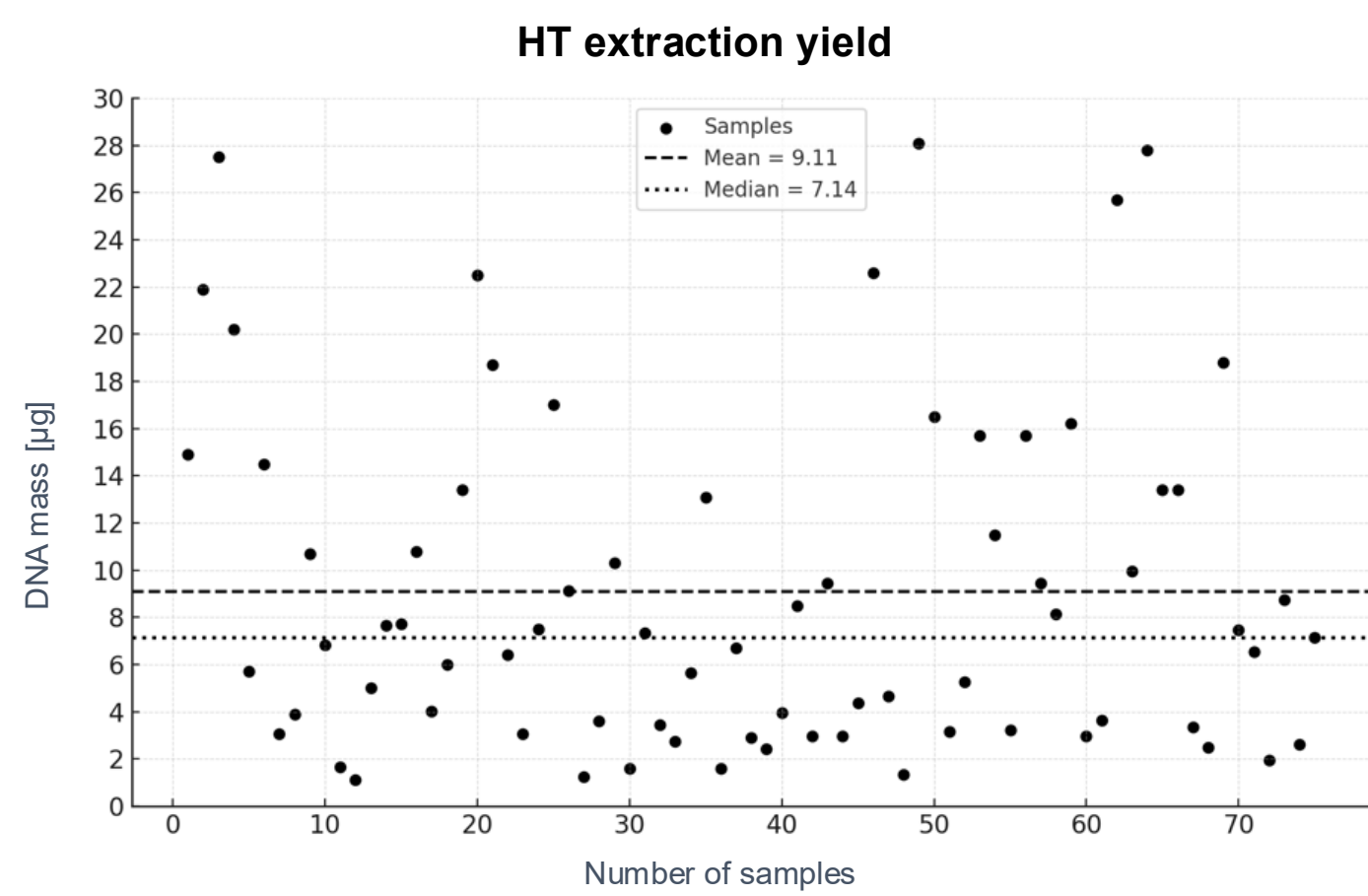


Figure 1. DNA yield from HT extraction (1 to 28 µg)

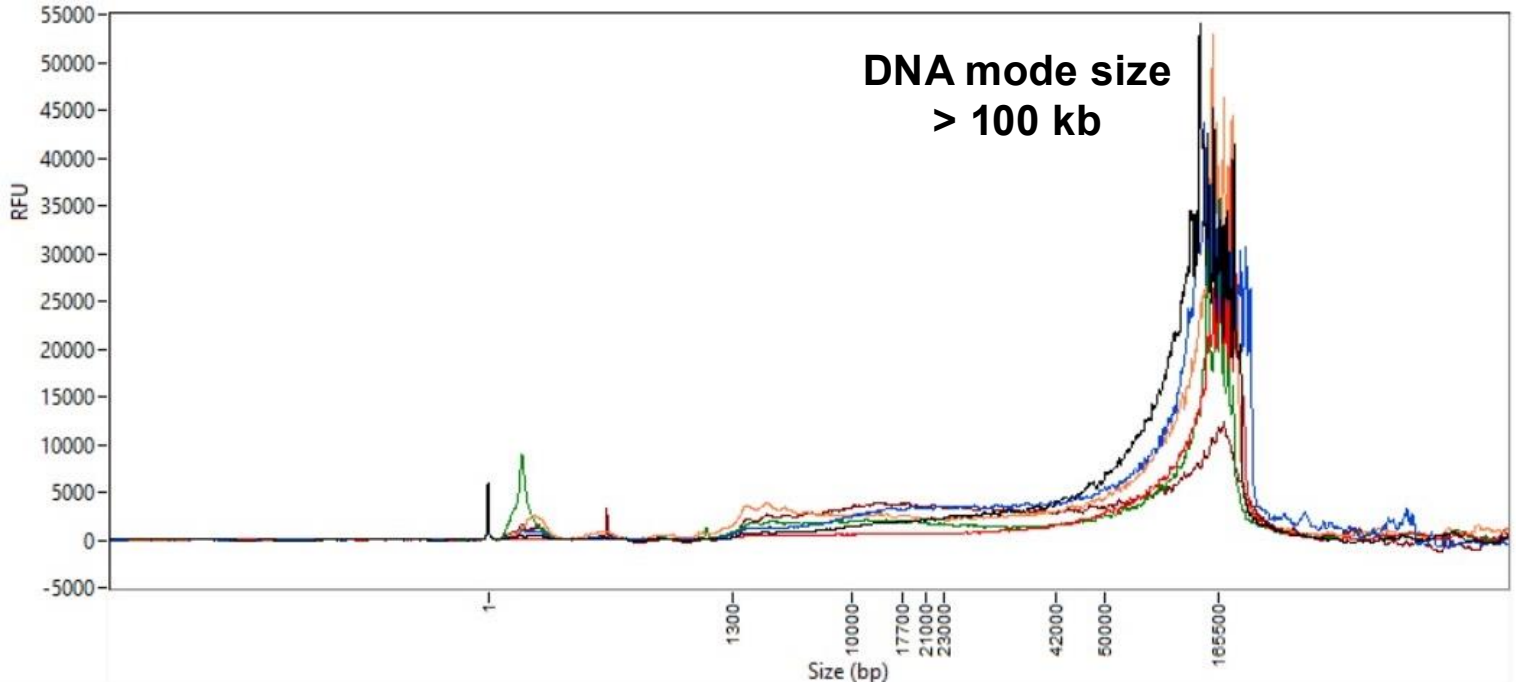


Figure 2. DNA size distribution from 6 samples HT extracted (Femto Pulse system)

Oragene™ is a trademarks of DNA Genotek Inc., registered in various jurisdictions.

Comparison of Whole Genome Sequencing and PureTarget applications (Revio sequencing)

	Whole Genome Sequencing	PureTarget Sequencing
Workflow	HiFi library preparation, native DNA sequencing of whole genome	Library preparation using the CRISPR-Cas9 system to target genes on native DNA
Input material	500 ng of HMW DNA (GQN 10kb > 7)	1 µg of HMW DNA (GQN 30kb > 5)
Result	Comprehensive variant detection and methylation for whole genome 20X coverage	Comprehensive variant detection and methylation for genes targeted by the panel ≥100X coverage on genes of interest
Throughput	1 to 2 samples per SMRT Cell	Up to 96 samples per SMRT Cell
Cost - TAT	~\$500	~\$100
TAT	2 days for 4 to 8 samples	3 days for 96 samples

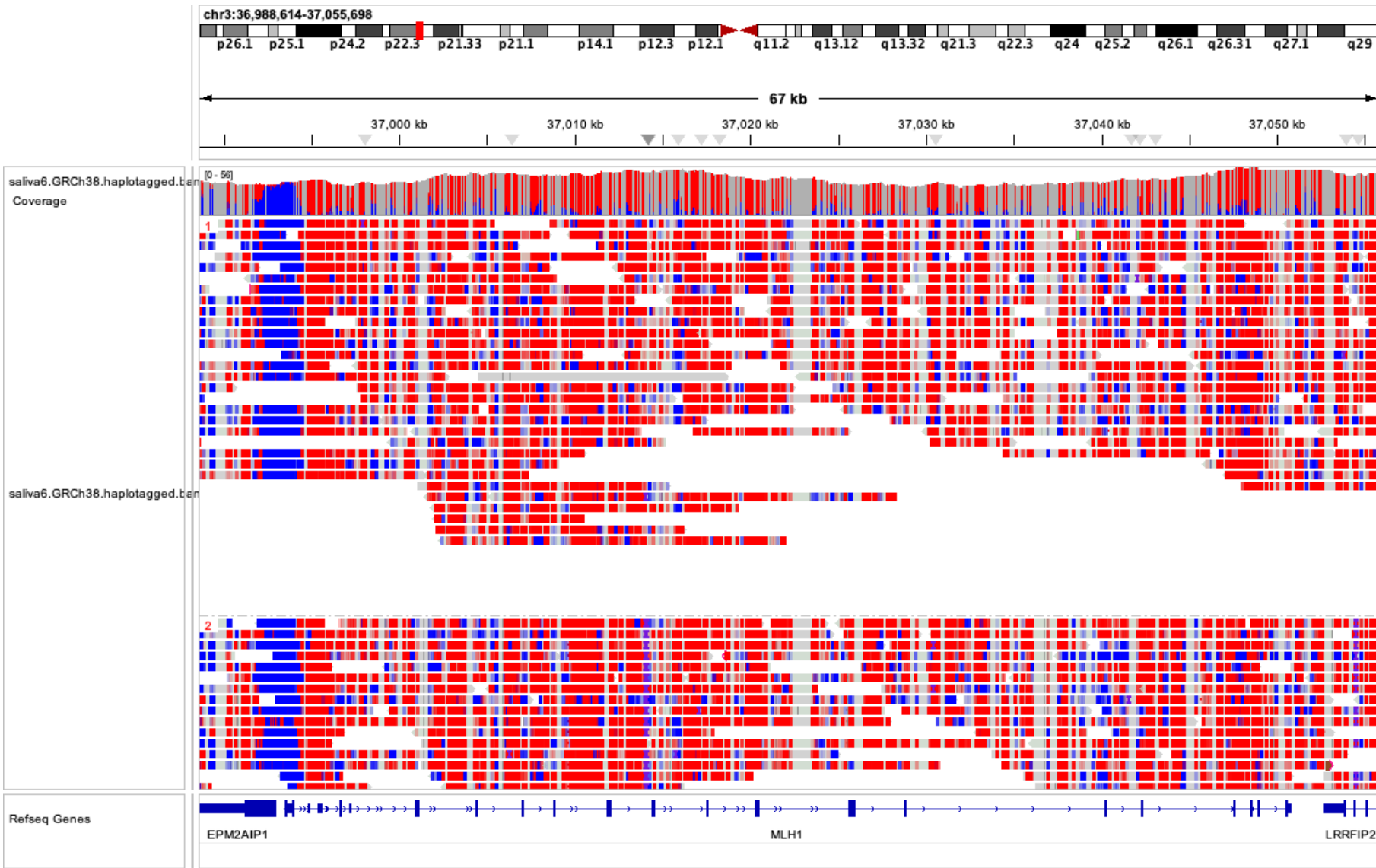
Saliva samples - Whole Genome Sequencing Workflow (WGS)

- HiFi WGS libraries were prepared following standard PacBio procedure and sequenced using SPRQ chemistry on the Revio system (1 sample per SMRT cell).
- HiFi reads were mapped against human genome reference (CRCh38). Variant calling and methylation analysis were performed using SMRT Link analysis pipelines.
- Saliva DNA is typically ~75-95% human, unmapped reads were used to classify the salivary microbiome using sourmash tool (<https://github.com/PacificBiosciences/pb-metagenomics-tools/tree/master/Taxonomic-Profiling-Sourmash>).

A. HiFi summary metrics and mapping for 6 saliva samples.

Sample	HiFi Yield	HiFi Mean Length	Mean QV	Human Mean Coverage	Human mapped read
S-1	135 Gb	15.5 kb	Q35	37x	87%
S-2	125 Gb	13.1 kb	Q36	32x	83%
S-3	119 Gb	12.1 kb	Q35	27x	75%
S-4	134 Gb	14.4 kb	Q34	40x	95%
S-5	135 Gb	13.9 kb	Q35	40x	95%
S-6	125 Gb	14.8 kb	Q35	39x	98%

B. MLH1 region IGV plot colored by methylation status



C. Taxonomic profiling of unmapped reads

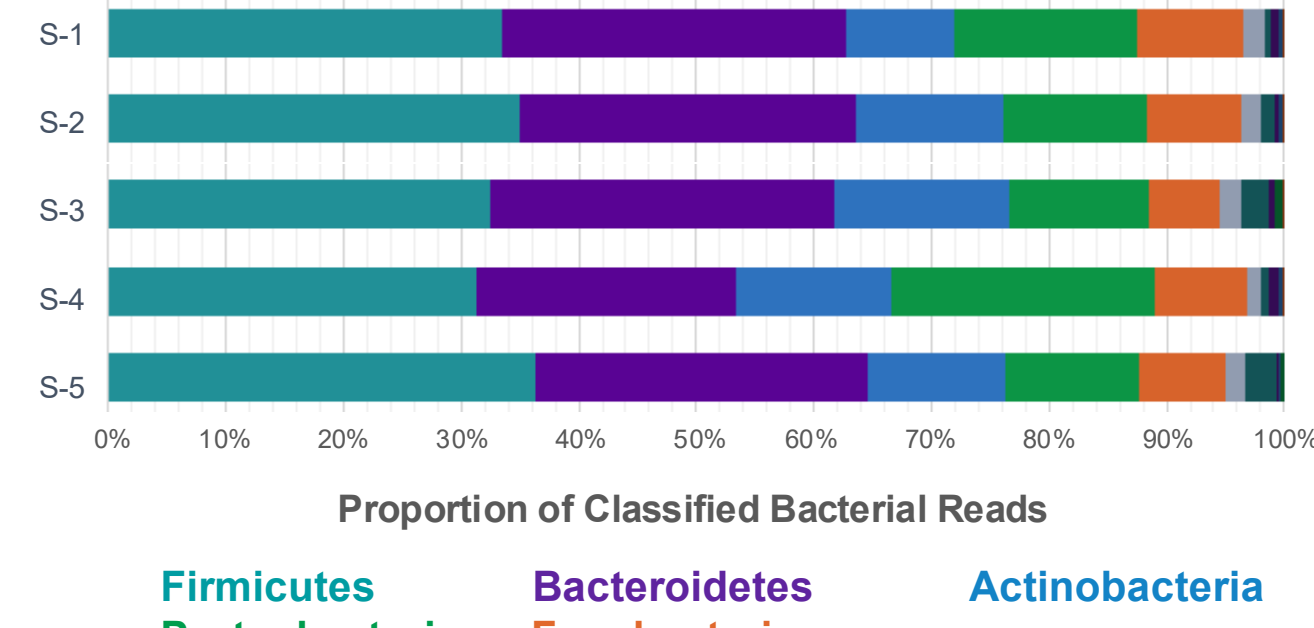


Figure 3.

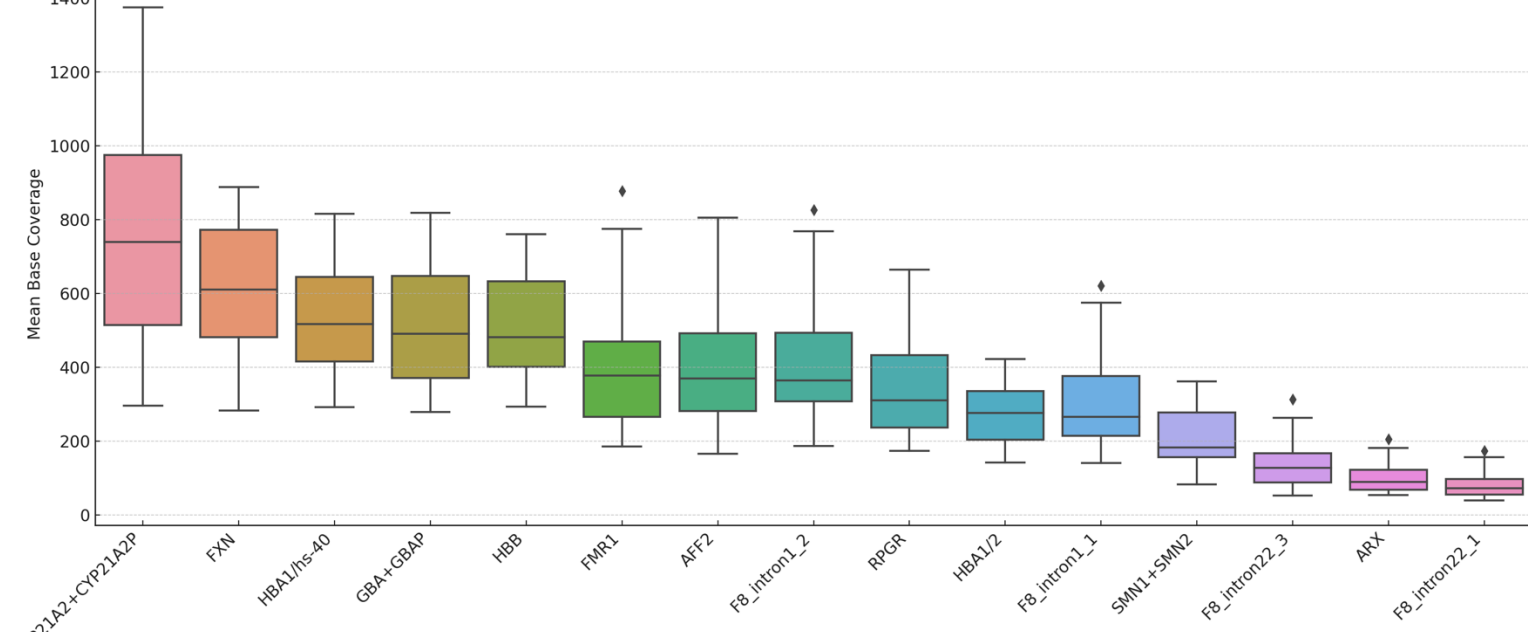
- A)** 6 saliva samples were sequenced on Revio. Samples showing high HiFi yield and RL resulting in 27x to 40x coverage with 75% to 98% of human reads
- B)** Methylation result in MLH1 region reads phased by haplotypes and bases colored by 5mC status. Red indicates high, while blue indicates a low probability of methylation
- C)** Taxonomic analysis of 5 saliva samples show that more than 90% of classified reads are from the five bacterial phyla.

Saliva samples - PureTarget workflow

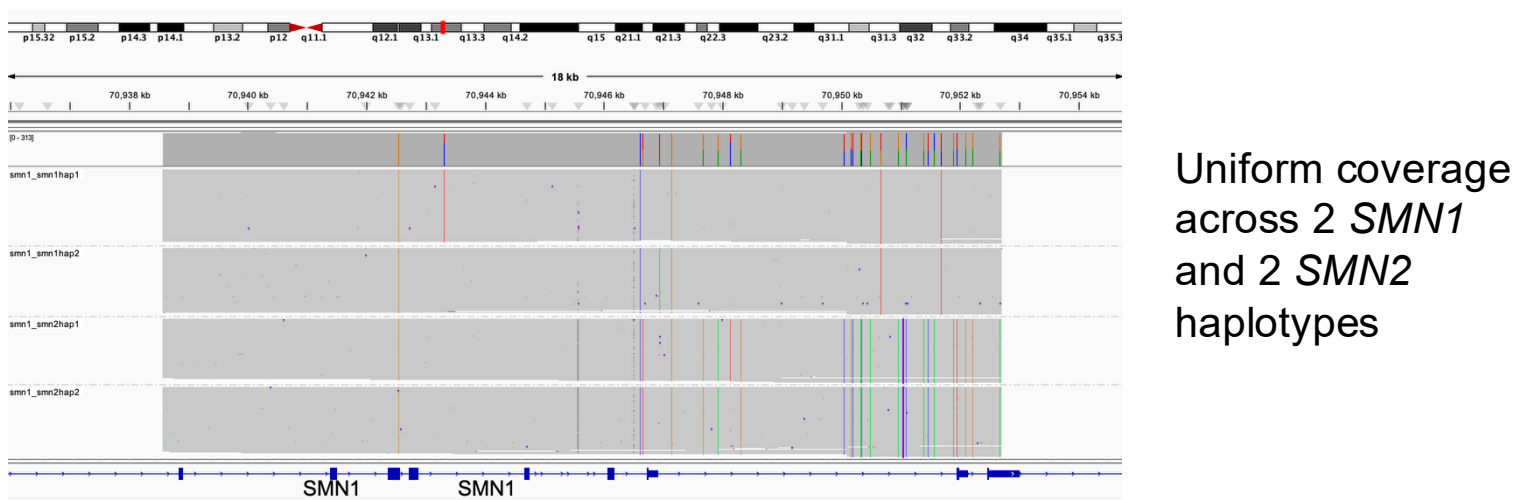
To explore the use of salivary DNA for lower cost targeted approaches, samples were analyzed using the PureTarget carrier panels on the Revio system with SPRQ chemistry. This library prep uses the CRISPR-Cas9 system to generate targeted native DNA libraries. This amplification-free approach retained epigenetic signals and has no PCR artifacts. We multiplexed 24 samples on one Revio (SPRQ chemistry) for a 12-gene repeat carrier panel.

Disease	Gene
Fragile X syndrome, FRAXE type	<i>AFF2</i>
Early-infantile epileptic encephalopathy (EIEE1) and Partington syndrome (PRTS)	<i>ARX</i>
Congenital adrenal hyperplasia (CAH)	<i>CYP21A2/CYP21A2P</i>
Hemophilia A	<i>F8</i>
Fragile-X disease (FXS)	<i>FMR1</i>
Friedreich ataxia (FRDA)	<i>FXN</i>
Gaucher disease	<i>GBA/GBAP1</i>
Alpha thalassemia (AT)	<i>HBA1/2</i>
Sickle Cell Anemia and Beta thalassemia	<i>HBB</i>
X-linked retinitis pigmentosa	<i>RPGR</i>
Spinal muscular atrophy	<i>SMN1/2</i>
Classical-like Ehlers-Danlos syndrome	<i>TNXB</i>

A. Coverage in PureTarget carrier panel across 24 saliva samples (Revio)



B. Paraphase alignment of *SMN1/2* in random donor



C. TRGT waterfall plot of *FMR1* for random male donor



D. Paraphase alignment of *GBA/GBAP1* in random donor

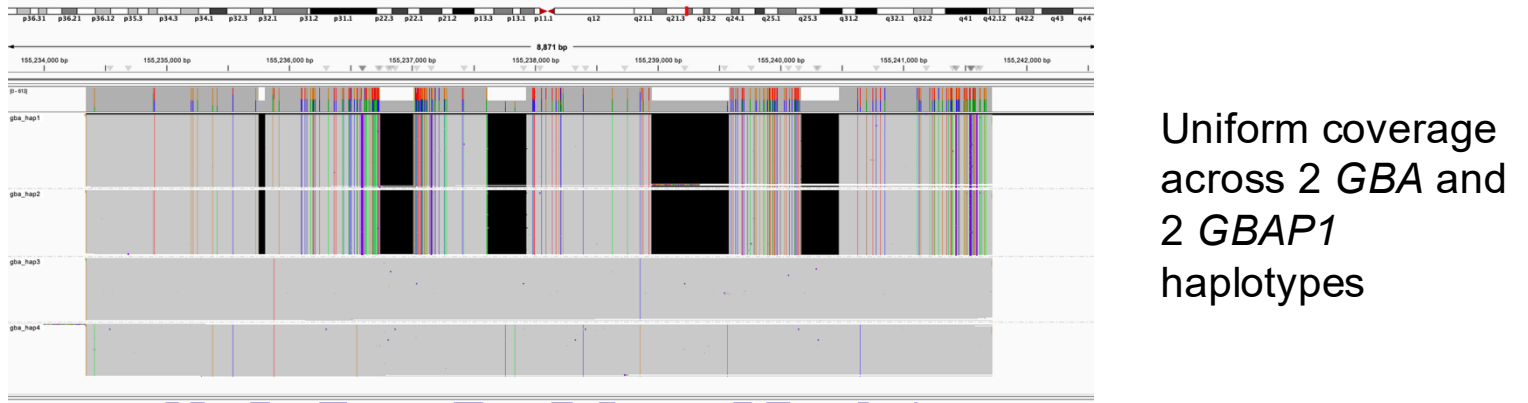


Figure 4. 24 saliva samples were prepared following the manual PureTarget protocol and the carrier panel. 1.5 µg of DNA was used as input per sample and AMPure PB bead cleanup was performed prior to PureTarget library prep (see protocol appendix). HiFi data were analyzed with PureTarget carrier pipeline (PTCP).

Conclusion

- Saliva stabilized in Oragene™ devices and extracted with Nanobind kits** is a convenient sample type for HiFi long-read sequencing and broadly applicable to large-scale genomic studies.
- WGS from saliva samples show optimal HiFi performance** with 20-fold to 40-fold coverage sufficient for comprehensive variant detection and methylation analysis.
- PureTarget from saliva samples enables characterization of complex genes for carrier screening** using an accessible sample type.

Acknowledgements

The authors would like to thank everyone who donated saliva sample to generate data for the poster and Dan Portik for salivary microbiome analysis..