Children's Mercy Kansas City
SHARE @ Children's Mercy

Posters

3-2023

Evaluating the Impact of Long Read Genomes in Rare Disease: A systematic analysis of 1000 HiFi Genomes

Emily Farrow
Children's Mercy Hospital

Isabelle Thiffault
Children's Mercy Hospital

Ana S A Cohen
Children's Mercy Hospital

Tricia N. Zion
Children's Mercy Kansas City

Adam Walter
Children's Mercy Hospital

See next page for additional authors

Let us know how access to this publication benefits you

Follow this and additional works at: https://scholarlyexchange.childrensmercy.org/posters

Part of the Medical Genetics Commons

Recommended Citation

Farrow, Emily; Thiffault, Isabelle; Cohen, Ana S A; Zion, Tricia N.; Walter, Adam; Gibson, Margaret; Bi, Chengpeng; Cheung, Warren A.; Johnston, Jeffrey J.; and Pastinen, T, "Evaluating the Impact of Long Read Genomes in Rare Disease: A systematic analysis of 1000 HiFi Genomes" (2023). *Posters*. 312.
https://scholarlyexchange.childrensmercy.org/posters/312

This Poster is brought to you for free and open access by SHARE @ Children's Mercy. It has been accepted for inclusion in Posters by an authorized administrator of SHARE @ Children's Mercy. For more information, please contact hlsteel@cmh.edu.

Authors

Emily Farrow, Isabelle Thiffault, Ana S A Cohen, Tricia N. Zion, Adam Walter, Margaret Gibson, Chengpeng Bi, Warren A. Cheung, Jeffrey J. Johnston, and T Pastinen

This poster is available at SHARE @ Children's Mercy: https://scholarlyexchange.childrensmercy.org/posters/312

Evaluating the Impact of Long Read Genomes in Rare Disease: A Systematic Analysis of 1000 HiFi Genomes

Emily Farrow[1,3], Isabelle Thiffault[2,,3], Ana S.A. Cohen[2,3], Tricia Zion[1,3], Adam Walter[3], Margaret Gibson[3], Chengpeng Bi[1,3], Warren A. Cheung[1,3], Jeff Johnston[1,3], Tomi Pastinen[1,3]

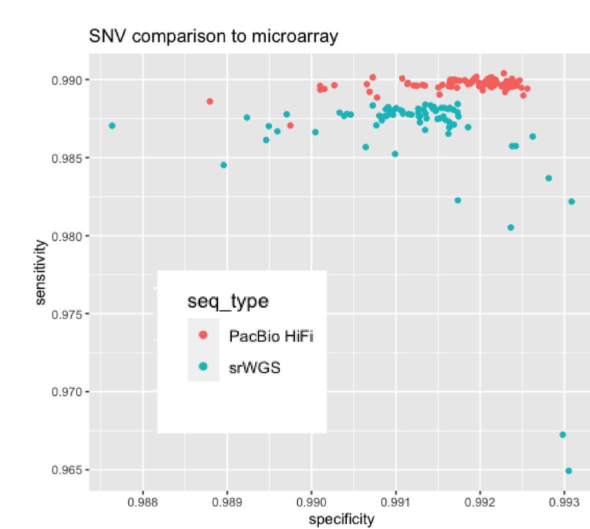
1) Department of Pediatrics, 2) Department of Pathology and Laboratory Medicine, 3) Genomic Medicine Center, Children's Mercy Hospital, Kansas City, Missouri, USA

Introduction and Methods

Genomic technologies continue to advance at a rapid rate, leading to continued novel gene-disease discoveries. However, despite the exponential increase in new gene discoveries, diagnostic rates in rare disease continue to range from 30-50%. To evaluate the impact of long read HiFi genome sequencing (lrGS) in a rare disease cohort, lrGS was implemented systematically in an institution-wide research program, Genomic Answers for Kids (GA4K). Individuals enrolled in GA4K, with a suspected genetic disorder, that remained undiagnosed after exome or genome sequencing, were submitted for HiFi sequencing. Probands were sequenced to a target depth of 30X coverage. Analyses included copy number, structural variation, single nucleotide variation, repeat expansion, and for a subset of genomes 5-methyl C detection. Clinical variants previously reported were used to assess lrGS variant detection algorithms. Additionally, sensitivity and specificity for lrGS were calculated by comparison to an Infinium Global Screening microarray.

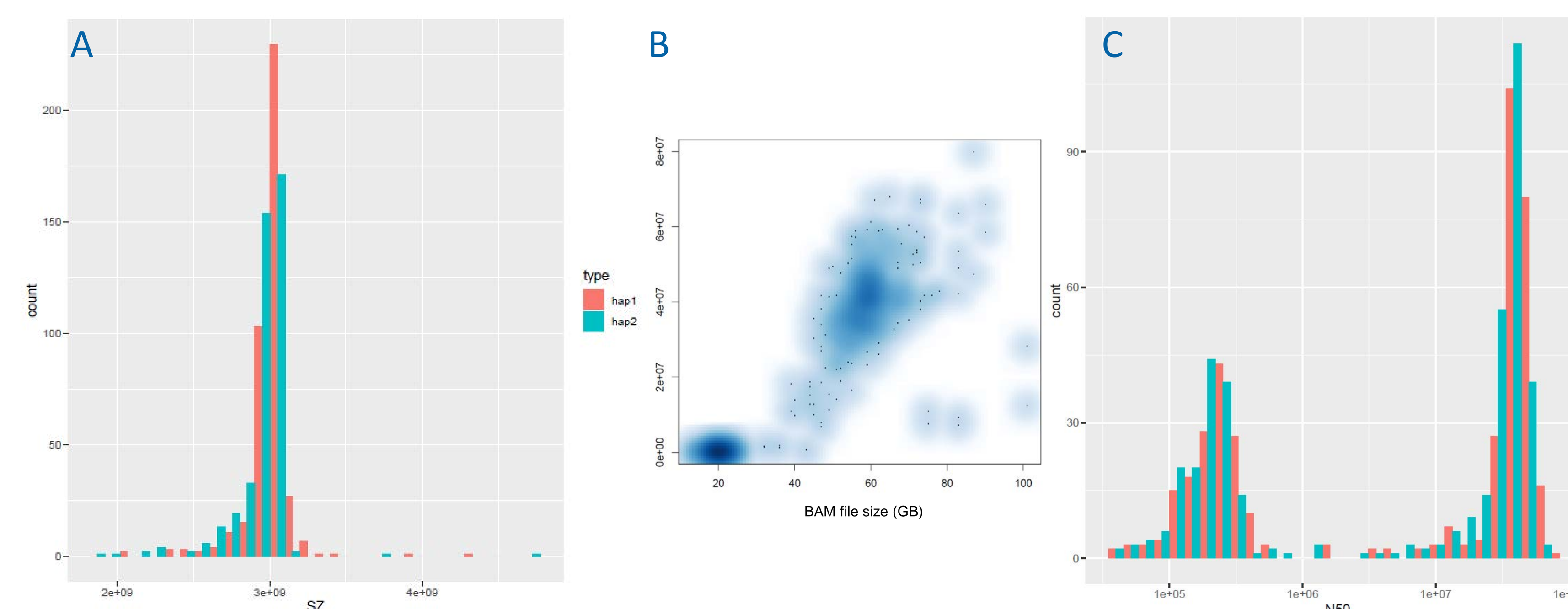
Sequencing Metrics

	TOT	DUP	DEL	INV	BND	INS
srGS	11529	488	4374	N/A	1823	4844
HiFi reads	51020	2789	22626	79	481	25045



HiFi sequencing detected additional variation, both SNV and SV/CNV compared to srGS. When limited by a MAF <0.01%, the average number of variants per sample is 150.

De Novo Assembly Metrics

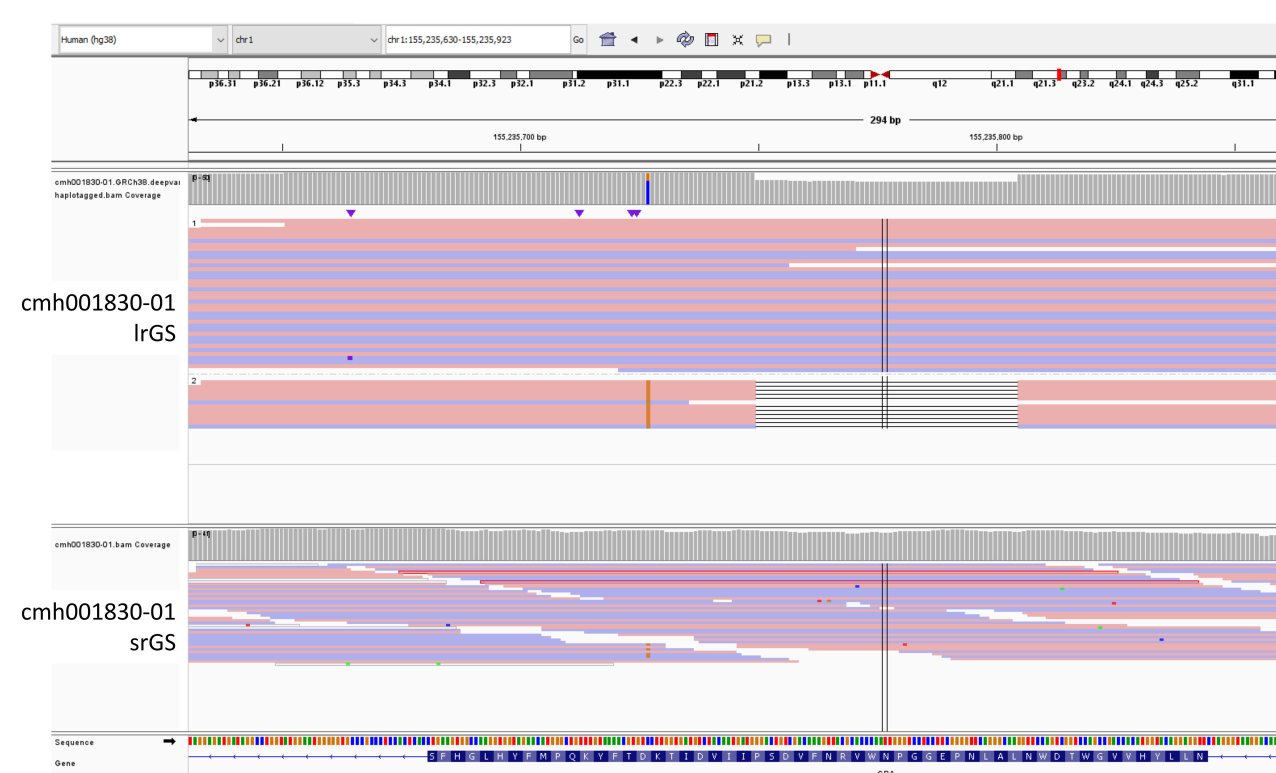


De novo assembly metrics were evaluated for the HiFi cohort, leveraging lrGS to resolve haplotypes. A) The assembled size (SZ) of the HiFi genomes was $3.02x10^9$. B) The N50 was dependent on overall coverage, with increasing coverage resulting in a larger N50. C) Overall, the mean N50 in the HiFi cohort was $2.3x10^7$.

Methylation

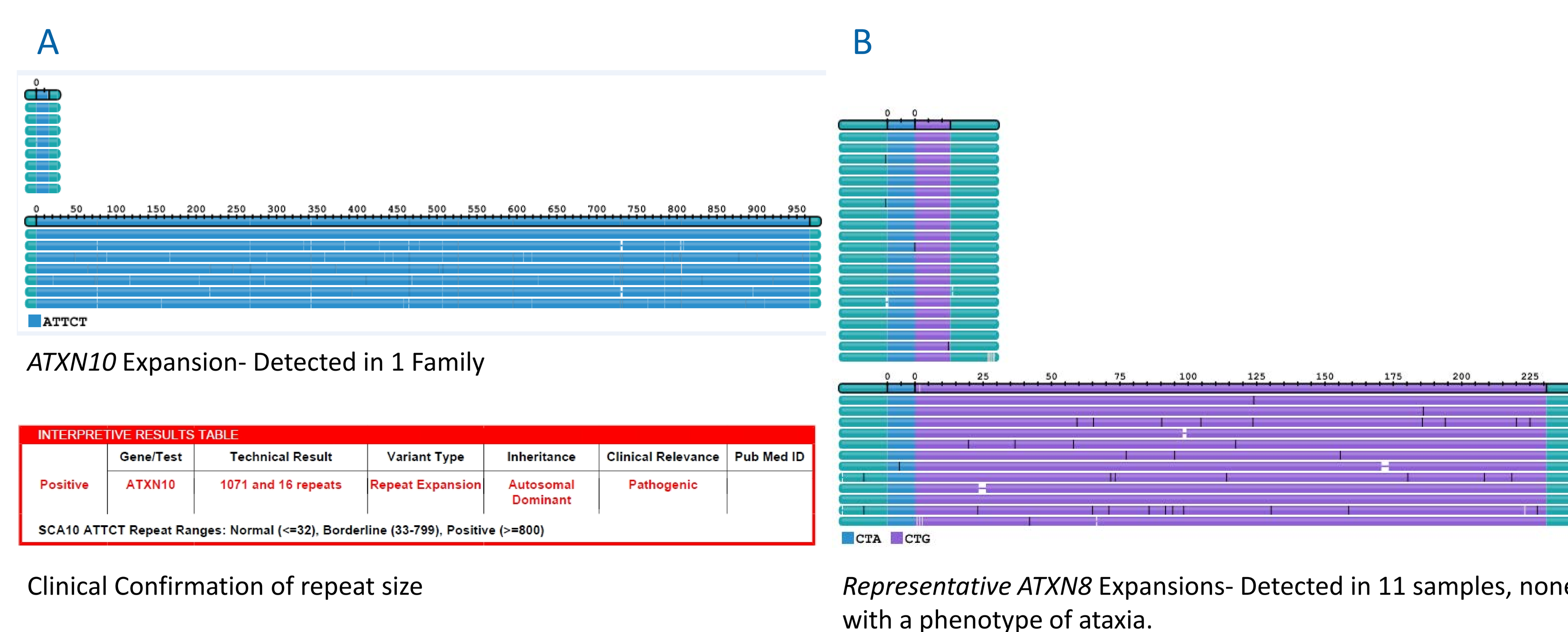
Direct 5-methyl-C detection (5mC-HiFi-GS) results in functional validation of detected expansions, such as in *DMPK* (Hap 1>1600 repeats, Hap 2=5). Additionally, novel hypermethylation events identify candidate disease loci.

Pseudogene Resolution



lrGS is able to resolve known pseudogene regions, such as *GBA*, with high clinical importance. HiFi genome detects a recurrent pathogenic deletion GBA:c.1265_1320del (p.Leu422fs) (top) that was not detected in the srGS (bottom)

Expansion Detection



ATXN10 Expansion- Detected in 1 Family

Clinical Confirmation of repeat size

Representative ATXN8 Expansions- Detected in 11 samples, none with a phenotype of ataxia.

The tandem repeat genotyper (TGRT) was used to interrogate 59 genes associated with repeat expansion disorders. The tool analyzes repeat length and motif. Initially 384 samples were flagged as having an expansion in the pathological range. Upon inspection, 16 were considered pathologic (diagnostic or pre-mutation). A previously undetected full expansion in ATXN10 a proband with congenital ataxia was detected. Clinical TP-PCR confirmed the expansion. B) ATXN8 pathological expansions were identified in 11 patients. However, the patients were either unaffected parents or were not reported to have an ataxia/neurological phenotype, highlighting the importance of larger screens for expansions disorders to better establish population variation, as previous data may be biased due to sample size/selection.

Conclusion

The implementation of HiFi GS in an ES/GS negative cohort resulted in an approximate 10% increase in diagnostic yield. Importantly, previously reported variants were recapitulated, indicating that HiFi GS could be utilized as a first-tier genome test, simplifying genetic testing algorithms and increasing efficiency. Our developing catalog of rare SVs and methylation variants are now giving new handles for unsolved disease in known and novel disease genes. Anticipated improvements in throughput and cost will enable the widespread integration of long read sequencing into clinical care.

Acknowledgements

We would like to thank the family for participating in our study; made possible by the generous gifts to Children's Mercy Research Institute and Genomic Answers for Kids program.

UMKC School of Medicine

KU School of Medicine The University of Kansas

Children's Mercy Kansas City | Research Institute