The Dark Side of Carrier Screening: Illuminating Hard-to-Decipher Genetic Variation with Nanopore Sequencing

Bradley Hall, Bryan Killinger, Christopher J. Fraher, Cody Edwards, Ryan Routsong, Jonathan Turner, Monica Roberts, Pranesh Rao, Bradley Martin, Jon Kemppainen, Adrian Lara, Brennan Greenlee, Jamie Myers, and Gary J. Latham

Asuragen, a Bio-Techne Brand, Austin TX

SUMMARY

- **Cystic Fibrosis (CF), Spinal Muscular Atrophy (SMA), Hemoglobinopathies, and Fragile X Syndrome (FXS) are among the most commonly inherited genetic disorders, each with high carrier rates that often require distinct genotyping methods.**
- **High-prevalence carrier genes, associated with disorders such as Gaucher Disease (GD), congenital adrenal hyperplasia (21-OHD CAH), and Hemophilia A (HA), include complex structural variants and pseudogenes that confound conventional sequencing methods.**
- **We explored whether novel PCR enrichment, nanopore sequencing, and machine learning models could detect multiple classes of variants including SNVs, INDELs, Exon del/dups, SVs, CNVs and STRs in a single workflow.**
- **The assay was optimized with 168 cell-line samples and independently evaluated with 249 whole blood samples across the seven genes to identify potential carriers from presumed normal donors.**

INTRODUCTION

Everyone is a genetic carrier of a disease or condition [1]. Carrier screening (CS) aims to identify couples at risk for having a child with a severe genetic disorder. Although Next-Generation Sequencing (NGS) is a widely used method, 6 of the top 10 CS conditions recommended in professional practice guidelines for CS are challenging for NGS, such as those with tandem repeats, copy number variation, pseudogenes, and structural variation. More broadly, 20.4% of pathogenic/likely pathogenic variants in ClinVar have been reported in “dark” or “camouflaged” regions of the genome that are “technically challenging” to resolve [2]. Many of these genes require specialized techniques and only cover a fraction of carrier risk. Because of technical limitations, some genes with an outsized impact on modeled fetal disease risk, such as *F8*, are missing on many CS panels.

We combined three innovations to address these shortcomings: 1) novel long-range PCR, 2) nanopore sequencing, and 3) customized software and analysis pipelines. Using a single workstream for enrichment and sequencing, we developed a prototype assay panel of 11 genes that are critical for CS, including eight “hard-to-decipher” genes. We estimate that these 11 genes represent ~70% of all pathogenic variants for a severe inherited disorder in at-risk couples compared to gene panels that are at least 16 times larger [3]. Importantly, we designed the assay to be modular to allow the analysis of either the full panel or select genes. Here we describe and evaluation prototype assay and companion bioinformatic software performance across 380 samples, in all 11 genes: *CFTR, SMN1, SMN2, FMR1, HBA1, HBA2, HBB, F8, GBA, CYP21A2,* and *TNXB*.

METHODS

Cell-line (CL) genomic DNA (gDNA) samples (n=235) were obtained from Coriell Cell Repository. Genomic DNA was also isolated from whole blood (WB) donors (n=249). Samples were PCR amplified across four reactions, barcoded per sample, pooled across samples, and prepared using a ligation sequencing kit (LSK110 & LSK114; Oxford Nanopore Technologies, ONT). Sequencing was performed using MinION flow cells (R9.4.1, R10.4.1) on a Mk1B (ONT). Cell-line samples representing all major classes of variants were used to develop custom data analysis pipelines and software. Clair3 was utilized for SNV/INDEL identification [4]. Performance was demonstrated across cell-line and whole blood samples. Orthogonal methods or reporting (e.g. Coriell, 1000 Genomes, melt curve PCR analysis, MLPA, custom PCR/capillary electrophoresis (CE), AmplideX® PCR/CE *CFTR* Kit†, PCR/CE *SMN1/2* Plus Kit†, and PCR/CE *FMR1* Kit†, Xpansion Interpreter®‡ (XI), Sanger sequencing, and qPCR) were utilized to determine comparator results.

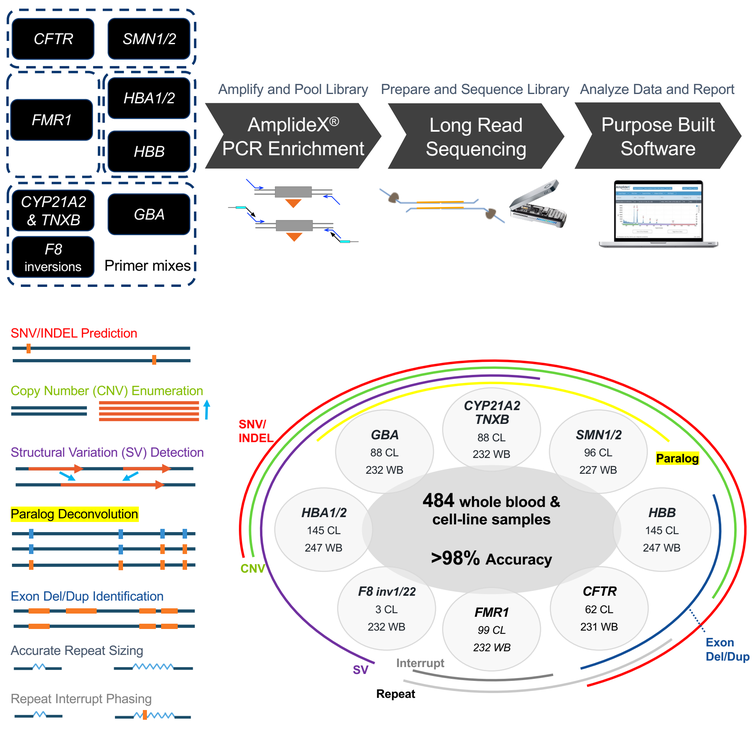

Figure 1. PCR/nanopore Carrier Screening Panel Design and Workflow Identifies Pathogenic Variants for 11 of the Most Common Inherited Genetic Disorders in a Single Workflow. The combination of AmplideX® PCR technology across 4 primer mixes, and nanopore sequencing enables detection of multiple variant classes for each of the 11 genes within the panel. A total of 484 samples were utilized for training and testing. For a subset of samples, only specific gene data was analyzed and compared, especially if a variant was known in the gene of interest. Highlights of the streamlined workflow are shown under the graphic.

This product is under development. Future availability and performance to be determined.
Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.
†For Research Use Only. Not for use in diagnostic procedures. ‡Xpansion Interpreter® is a laboratory-developed test. Analytical and clinical performance have not been reviewed by the FDA. All authors have the financial relationship to disclose: Employment by Asuragen.
Presented at London Calling, 2023

RESULTS

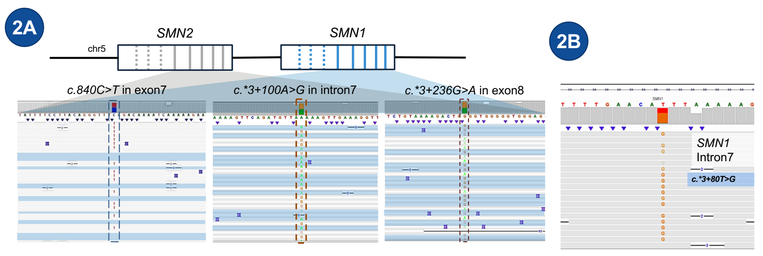

Figure 2. Sequence Data Differentiates *SMN1* and *SMN2* and Informs Silent Carrier Status and Disease Modifiers. A) Differentiation and assignment of reads to *SMN1* and *SMN2* by three different paralog-specific variants align reads to each gene and inform CNV prediction model. **B)** Silent carrier (SC1; *c.*3+80T>G*) variant alignments. SC2 (*c.*211_212del*) and disease modifier (DM; *c.859G>C*) not shown.

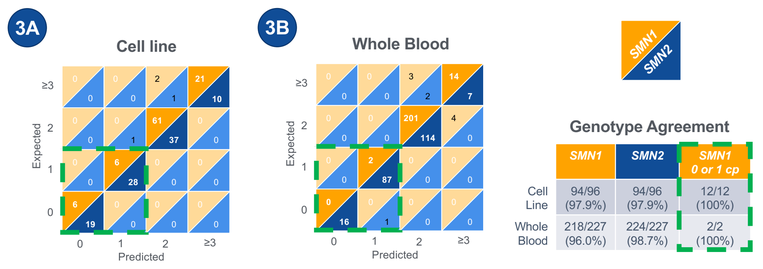

Figure 3. *SMN1/2* PCR/Nanopore Assay Accurately Classifies Carrier Status. Calling accuracy for *SMN1* and *SMN2* copy numbers in **A)** 96 CL and **B)** 227 WB samples using R9.4.1 flow cells. Hyperparameters for the decision tree model were selected using an 80:20 train:test split in a stratified randomly selected five-fold cross validation scheme on an independent set of 349 samples (62 CL and 287 WB). Carriers were identified with 100% accuracy (green dashed outline).

4A

Cell Line	Orthogonal Data			Overall
Sample Level Accuracy	Allele1/Allele2	WT/WT	MUT/WT	MUT/MUT	Genotype Agreement
CFTR nanopore	WT/WT	48	0	0	48/48 (100%)
	MUT/WT	0	27	0	27/27 (100%)
	MUT/MUT	0	0	28¥	28/28 (100%)

4B

Whole Blood	Orthogonal Data			Overall
Sample Level Accuracy	Allele1/Allele2	WT/WT	MUT/WT	MUT/MUT	Genotype Agreement
CFTR nanopore	WT/WT	223	0	0	223/223 (100%)
	MUT/WT	0	9¥	0	9/9 (100%)
	MUT/MUT	0	0	2	2/2 (100%)

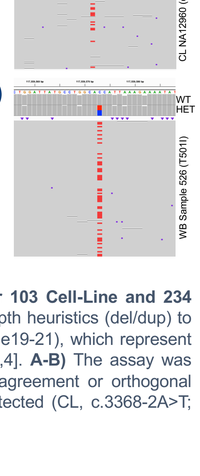

Figure 4. *CFTR* Sample Level Agreement with Orthogonal Data for 103 Cell-Line and 234 Whole Blood Samples. The assay used Clair3 (SNV/indel) and read depth heuristics (del/dup) to detect 57 unique variants, including two del/dup (CFTRdele2,3, CFTRdele19-21), which represent 88.9% prevalence of variants in the ethnically diverse US population [1,4]. **A-B)** The assay was performed with both R10.4.1 (n=290), and R9.4.1, (n=47) with 100% agreement or orthogonal methods. **C-D)** Nanopore read pile-ups of two variants previously undetected (CL, c.3368-2A>T; WB, T501I). ¥Sanger sequencing verification pending.

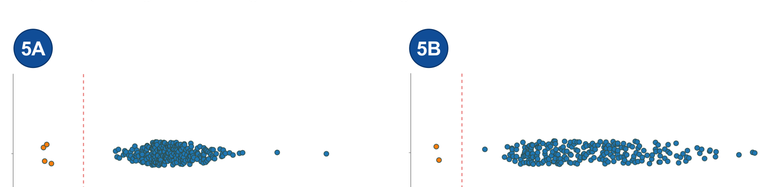

Figure 5. Large Exon Deletions Detected at 100% PPV and Sensitivity. Normalized amplicon coverage distinguishes exon deletion (orange) from wildtype genotypes (blue) for **A)** CFTR2,3dele and **B)** CFTRdele20. Samples with normalized coverage below the threshold (vertical dashed line) were classified as heterozygous for the large exon deletions.

Table 3. *FMR1* Categorical Agreement with Orthogonal Genotypes for 99 Cell-Line and 232 Whole Blood Samples. Using R9.4.1, ACMG categorical genotype boundaries are included for reference. All samples fell within expected categories based AmplideX PCR/CE *FMR1* precision metrics (± 1: 0-70 repeats, ± 3: 71-119). All expanded samples, including full mutations up to 940 CGG repeats, were flagged correctly. Additionally, CGG sizing was accurate within precision for 321/331 (97.0%) samples and 443/454 (97.5%) alleles. In 7/11 samples, the algorithm accurately called one of two alleles when two similar sized alleles (1-3 repeats difference) were present. The algorithm identified a previously unidentified minor mosaic allele in the remaining three samples that did not change the categorical call.

Sample level categorical accuracy	Normal <45 CGG	Intermediate 45-54 CGG	Premutation 55-200 CGG	Full Mutation >200 CGG	Sensitivity	Specificity
Training	61	18	55	19	100%	100%
Cell line	76	5	14	4	100%	100%
Whole Blood	225	7	0	0	100%	100%

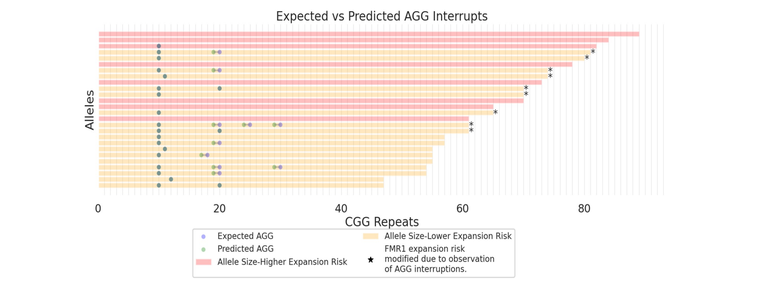

Figure 6. Predicted Risk of *FMR1* Expansion Based on AGG Interruption Status. A cohort of 26 intermediate and premutation alleles were assessed using Asuragen Xpansion Interpreter® (XI) and PCR/nanopore using a custom algorithm. Genotypes were in 100% agreement with XI for the absolute number of AGG interruptions and within ±1 for the absolute position of each AGG interruption within the CGG repeat. Asterisks denote samples where AGG status modified the risk for a full mutation in the next generation compared to CGG repeat information alone.

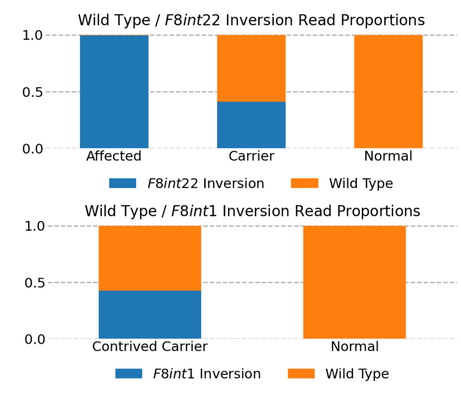

Figure 7. Prototype Assay Amplifies and Identifies >100 kb *F8* Inversions. *F8* inversions were successfully amplified and aligned across 3 cell lines with R10.4.1. These include a homozygous intron 22 inversion “affected”, a heterozygous intron 22 inversion “carrier”, and a contrived heterozygous intron 1 inversion “carrier” created by mixing an inversion g-Block and a wild-type sample. The results show the proportion (y-axis) of *F8* reads that aligned to either the reference genome (Wild Type) or an *F8* inversion. Inversion reads were not identified in the wild type sample. Both intron 1 and intron 22 inversions were verified by orthogonal methods (data not shown).

Gene	Variant	Type	Positive Samples
HBA	SEA	α^0-thal HBA1/2 del	11 HET, 1 HOM, 2 COMP
HBA	FIL	α^0-thal HBA1/2 del	1 HET, 1 COMP
HBA	THAI	α^0-thal HBA1/2 del	1
HBA	3.7del	α^+-thal HBA2 del	14 CL & 4 WB HET, 3 CL HOM
HBA	anti-3.7	HBA2 dup	5 CL & 8 WB HET, 1 CL COMP
HBA	4.2del	HBA2 del	2
HBA2	c.427T>C (Hb Constant Spring)	SNV/indel	1
HBA1	c.207C>G (Hb G-Philadelphia)	SNV/indel	1
HBB	(Hb Lepore-Baltimore)	Exon 1, 2 del	2
HBB	(Sicilian [Δβ 0-Thal)	Exon 1-3 del	1
HBB	c.19G>A (Hb C)	SNV/indel	2
HBB	c.79G>A (Hb E)	SNV/indel	5
HBB	c.20A>T (Hb S)	SNV/indel	3
HBB	c.126_129del	SNV/indel	2
HBB	c.93-21G>A	SNV/indel	2
HBB	c.-137C>G	SNV/indel	1
HBB	c.-138C>T	SNV/indel	1
HBB	c.316-197C>T	SNV/indel	1
HBB	c.82G>T (Hb Knossos)	SNV/indel	1
HBB	c.208G>A	SNV/indel	1

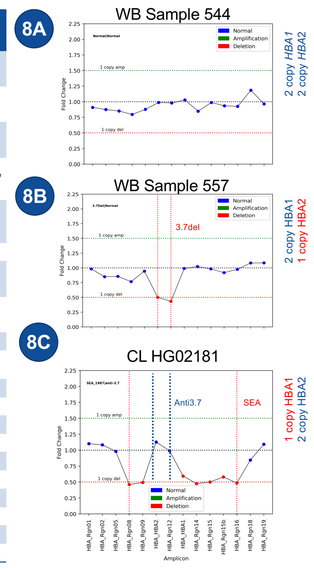

Figure 8. Pathogenic and Likely Pathogenic *HBA1/2* and *HBB* Variants Detected Using the PCR/nanopore Assay. All samples (n=145 CL and 247 WB) tested with the combined *HBA1/2* and *HBB* assay on R10.4.1 called correctly. Of these, 102 CL and 235 WB samples were genotyped as wild-type (aa/aa). Sample variants are heterozygous (HET) unless listed as homozygous (HOM) or compound heterozygous (COMP). Positive samples are all cell line (CL) unless indicated. **A)** Typical WB sample with coverage pattern for two wild-type alleles (aa/aa). *HBA* deletion identification was based on normalized fold change in read depth compared to control sample reference amplicons on the y-axis; assay amplicons are arranged in sequential order on the x-axis. **B** WB silent carrier with one wild-type allele (aa) and one 3.7del allele. **C)** CL sample with a SEA deletion on one allele and an anti-3.7 triplication on the other.

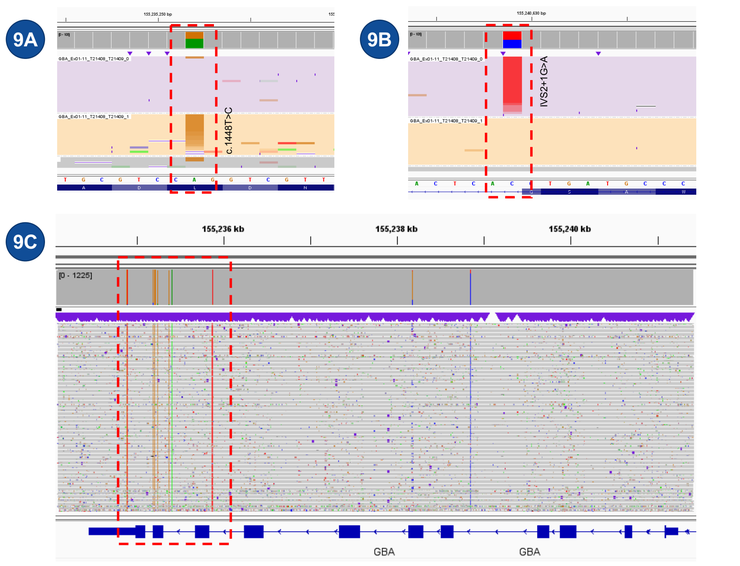

Figure 9. Sequence Data Reveals Diverse Pathogenic Variants Across the *GBA* gene on R10.4.1. Clinically affected cell-line NA20270 is a compound heterozygote and contains two SNVs within the *GBA* gene; **A)** one allele has a T>C transition at nucleotide 1448 in exon 10 (L444P, c.1448T>C), and **B)** the other allele has a splice site mutation in intron 2 (IVS2+1G>A). **C)** A homozygous *GBA-GBAP1* fusion was identified in cell-line NA20273. The variants in the red box are concordant with *GBAP1* paralog-specific variants. All samples were verified with orthogonal methods (data not shown).

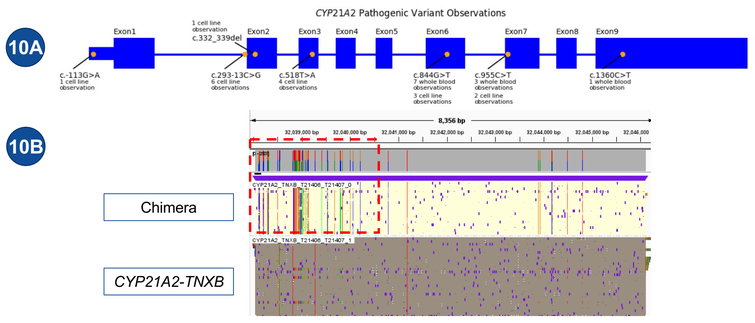

Figure 10. Accurate Resolution of Copy Number and Pseudogene Fusions in the *CYP21A2* Gene Cluster Utilizing Sequencing Deconvolution on R10.4.1. A) 28 pathogenic variants were identified across cell-line and whole blood samples and mapped to their location in *CYP21A2*. **B)** A sample with a *CYP21A2* and *CYP21A1P* fusion, denoted by a 30 kb deletion. The variants in red box align with *CYP21A1P* PSVs.

Gene	# of Carriers	Variants Identified
CFTR	8	5x F508del, wt; R117H, wt; G622D, wt; T501I, wt
SMN1	2	1, 1 and 1, 2 (SMN1, SMN2)
SMN1, SC	3	2, 2 + SC1, SC2†
HBA1/2	4	3.7del/aa
GBA	3	2x N370S (c.1226A>G) R496H (c.1604G>A)
CYP21A2	4	30-kb del; 3x Q318X (c.955C>T);
	8	7x V281L (c.844G>T); P453S (c.1360C>T)
SMN1 & *CYP21A2*	1	1,3 & Q318X (c.955C>T);

Table 4. Detected 34 Carriers (14.7%) in a Presumed Normal Cohort of Whole Blood Samples (n=232) using the Prototype Assay. One donor sample was identified as a carrier for both *SMN1* and *CYP21A2*. *CFTR* T501I is pending confirmation. *FMR1* intermediate expansions were identified in seven samples but not shown. *HBB* carriers (1:43 carrier rate [5]) were not identified. Hemophilia A (*F8 inversion*) carriers were not identified (expected by carrier rate). All *CYP21A2* and *GBA* variants are pathogenic or likely pathogenic, however the lighter grey CYP21A2 variants are non-classical and often go under-diagnosed. All variants were confirmed by orthogonal methods. †SC variants suggest an increased carrier risk.

CONCLUSIONS

- **The prototype PCR/nanopore assay accurately resolves multiple challenging variants across several classes for 11 of the most common gene targets associated with heritable disease.**
- **The assay utilizes a single-platform, streamlined workflow and has potential to greatly reduce carrier screening complexity and turn around times.**
- **Detection of a potential dual carrier (*CYP21A2* and *SMN1*) highlights the importance of a unified carrier screening approach.**
- **In 484 samples tested across the panel, the PCR/nanopore assay agreed with the orthogonal methods for SNVs/INDELs in *SMN1*, *CFTR*, *GBA*, *CYP21A2 HBA1*, *HBA2*, and *HBB* (>99% of samples), *SMN1* CN (96.6%), *SMN2* CN (98.5%), *FMR1* repeat categories (100%), *FMR1* AGG interruptions (100%), and *HBA1/2* SVs (100%).**

REFERENCES

1. Xiao, Q. et al. Npj Genom Medicine 6, 41 (2021).
2. Lincoln SE., et al. Genet Med 23 (2021).
3. Beauchamp, KA., et al. Genet Med 21 (2019).
4. Zheng, Z., et al. Nat Comp Sci 2, 797–803 (2022).
5. Taber, K. J. et al. Genet Med (2021).

Asuragen® a bio-techne® brand