An Amplification-based, Nanopore Carrier Screening Panel Resolves Clinically-Relevant Variants in *CFTR*, *SMN1/2*, *HBA1/2*, *HBB*, and *FMR1* in a Unified Workflow

Bradley Hall, Bryan Killinger, Christopher J. Fraher, Bradley Martin, Monica Roberts, Jonathan Turner, Pranesh Rao, Ryan Routsong, 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.**
- **We explored whether novel PCR enrichment and nanopore sequencing could detect multiple classes of variants across these genes in a single workflow.**
- **The assay utilizes PCR enrichment, coverage-based copy number, and machine learning models to automate and streamline identification of SNVs, INDELs, Exon del/dups, SVs, CNVs and STRs specific to each disease.**
- **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

Cystic Fibrosis (CF), Spinal Muscular Atrophy (SMA), Fragile X Syndrome (FXS), and Hemoglobinopathies, including alpha- and beta-thalassemias, are each within the 10 most prevalent inherited diseases by carrier rate. Pathogenic variants in the associated genes can be difficult or intractable to identify using short-read sequencing alone. As a result, genotyping often relies on multiple non-NGS methods, highly trained personnel, and sophisticated or manual data analysis to resolve. Even still, uneven detection rates across ancestries persist.

Genes for alpha-thalassemia, SMA, and FXS (*HBA1/2*, *SMN1*, and *FMR1*, respectively) are technically challenging due to homologous or GC-rich sequences. Repetitive polymorphisms, exon del/dups and other CF variants within *CFTR* can also be problematic to resolve. Here, we combine scalable PCR enrichment with nanopore sequencing in a prototype assay that accurately identifies multiple classes of pathogenic mutations for each of the seven carrier screening genes. The panel interrogates single nucleotide variants (SNVs), insertion/deletions (INDELs), exon del/dups, structural variants (SVs) and enumerates copy number variation (CNVs) and short tandem repeats (STRs) in a single workflow.

METHODS

Cell-line (CL) genomic DNA (gDNA) samples (n=168) were obtained from Coriell Cell Repository. Genomic DNA was also isolated from whole blood (WB) donors (n=249). Samples were PCR amplified across three 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 analyses were performed using custom software. Performance was demonstrated across cell-line and whole blood samples. Orthogonal methods or reporting (e.g. Coriell, 1000 Genomes, melt curve PCR analysis, 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 performance.

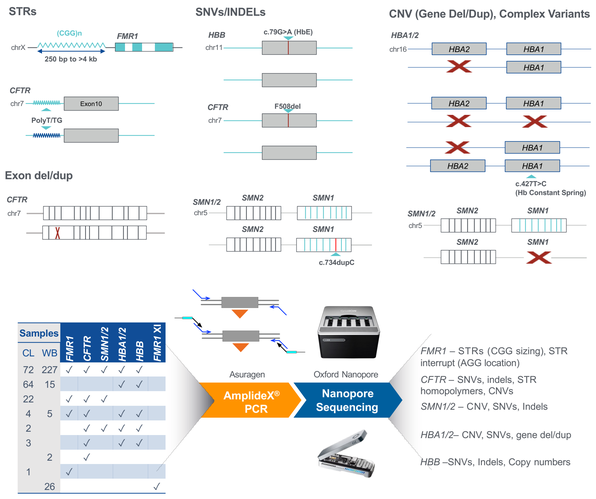

Figure 1. PCR/nanopore Carrier Screening Panel Design and Workflow Identifies Pathogenic Variants for 5 of the Most Common Inherited Genetic Disorders in a Single Workflow. The combination of AmplideX® PCR technology and Nanopore sequencing enables detection of short tandem repeats (STRs), including the highly GC-rich in *FMR1*, PolyT/TG region of *CFTR*, and detection of AGG interruptions in *FMR1;* differentiates and quantitates highly homologous sequences such as *SMN1/2* or *HBA1/2;* and identifies pathogenic or informative variants (SNVs and indels) across *CFTR*, *HBA1, HBA2, HBB,* and *SMN1/2*. The majority of samples were tested across the full panel, especially for whole blood. 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.

RESULTS

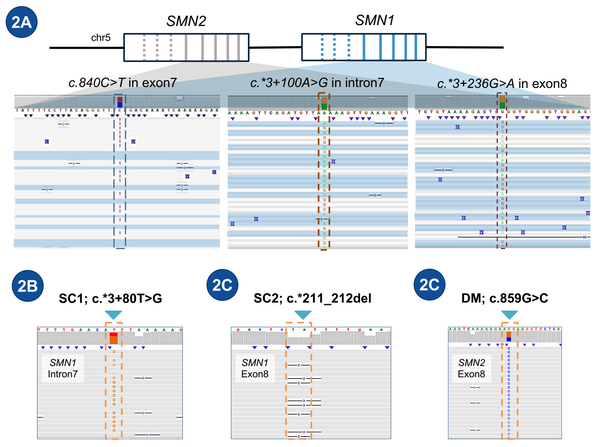

Figure 2. Sequence Data Differentiates *SMN1* and *SMN2* and Informs Silent Carrier Status and Disease Severity. A) Differentiation and assignment of reads to *SMN1* and *SMN2* by three different paralog-specific variants align reads to each gene and inform CNV prediction model. **B-D)** Silent carrier (SC1 or SC2) or disease modifier (DM) variant alignments.

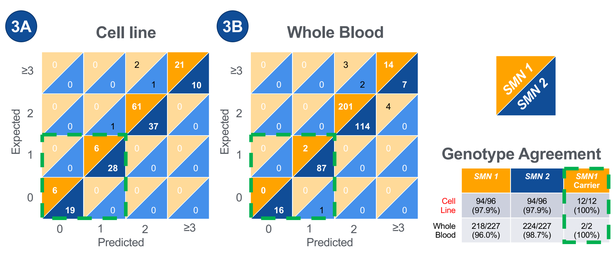

Figure 3. *SMN1/2* PCR/Nanopore Assay Classifies Carrier with 100% accuracy. Calling accuracy for *SMN1* and *SMN2* copy numbers in **A)** 96 CL and **B)** 227 WB samples. Hyperparameters for the decision tree model were selected using an 80:20 train:test split in a stratified randomly selected five-fold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

Table 1. Pathogenic and likely pathogenic *HBA1/2* and *HBB* variants detected using the PCR/nanopore assay. 145 CL and 247 WB samples were tested with the combined *HBA1/2* and *HBB* assay. 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). COMP included SEA/anti3.7 (HG01281), FIL/SEA (NA10797). Positive samples are all cell line unless indicated.

Gene	Variant	Type	Positive Samples	Called Correctly
HBA	SEA	α^0-thal HBA1/2 del	11 HET 1 HOM 2 COMP	100%
HBA	FIL	α^0-thal HBA1/2 del	1 HET 1 COMP	100%
HBA	THAI	α^0-thal HBA1/2 del	1	100%
HBA	3.7del	α^+-thal HBA2 del	14 HET CL, 4 HET WB 3 HOM CL	100%
HBA	anti-3.7	HBA2 dup	5 HET CL, 8 HET WB 1 COMP CL	100%
HBA	4.2del	HBA2 del	2	100%
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	100%
HBB	(Sicilian [deltabeta]0-Thal)	Exon 1-3 del	1	100%
HBB	c.19G>A (Hb C)	SNV/indel	2	100%
HBB	c.79G>A (Hb E)	SNV/indel	5	100%
HBB	c.20A>T (Hb S)	SNV/indel	3	100%
HBB	c.126_129del	SNV/indel	2	100%
HBB	c.93-21G>A	SNV/indel	2	100%
HBB	c.-137C>G	SNV/indel	1	100%
HBB	c.-138C>T	SNV/indel	1	100%
HBB	c.316-197C>T	SNV/indel	1	100%
HBB	c.82G>T (Hb Knossos)	SNV/indel	1	100%
HBB	c.208G>A	SNV/indel	1	100%

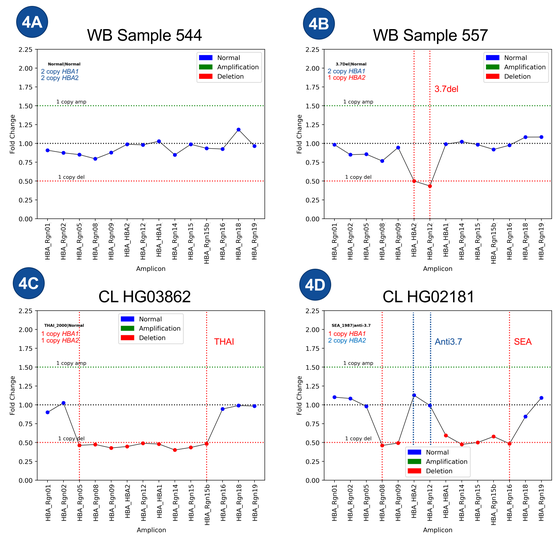

Figure 4. *HBA* Deletion Identification In Samples Based on Amplicon Coverage Using a Machine Learning Algorithm. Normalized fold change in coverage compared to control sample reference amplicons is shown on the y-axis; assay amplicons are arranged in sequential order on the x-axis. **A)** Typical WB sample with coverage pattern for two wild-type alleles (aa/aa). **B)** Typical WB silent carrier with one wild-type allele (aa) and one 3.7del allele leading to an α^+-thal silent carrier where one copy of *HBA2* is effectively deleted (hybrid *HBA1/2* gene). **C)** CL sample with only one copy of *HBA1* and *HBA2* leading to a α^0-thal trait. **D)** CL sample with an SEA deletion on one allele and an anti-3.7 triplication on the other. *HBA2* and Rgn12 amplicons receive "wild-type" copy number calls because the combination of variants results in the same coverage levels as a wild-type sample (1 copy of *HBA1,* 2 copies of *HBA2.*

Table 2. *CFTR* Sample Level Agreement with Orthogonal Data for 103 Cell-Line and 234 Whole Blood Samples. The assay used Clair3[2] (SNV/indel) and read depth heuristics (del/dup) to detect 57 unique variants among 97 total variants, including two del/dup (CFTRdele2,3, CFTRdele19-21), which represent 88.9% prevalence of variants in the ethnically diverse US population.[1] The assay was performed with both R10.4.1 (n=290), and R9.4.1, (n=47). ¥Sanger sequencing verification pending for two variants (CL, c.3368-2A>T; WB, T501I), with corresponding nanopore read pile-ups shown on the right.

Cell Line		Orthogonal Data			Overall
Sample Level Accuracy	Allele1/Allele2	WT/WT	MUT/WT	MUT/MUT	Genotype Agreement
CFTR PCR/nanopore	WT/WT	48	0	0	50/50 (100%)
	MUT/WT	0	27	0	27/27 (100%)
	MUT/MUT	0	0	28¥	28/28 (100%)

Whole Blood		Orthogonal Data			Overall
Sample Level Accuracy	Allele1/Allele2	WT/WT	MUT/WT	MUT/MUT	Genotype Agreement
CFTR PCR/nanopore	WT/WT	233	0	0	232/232 (100%)
	MUT/WT	0	8¥	0	9/9 (100%)
	MUT/MUT	0	0	2	2/2 (100%)

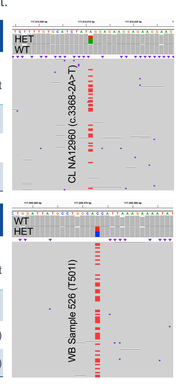

Table 3. *FMR1* Categorical Agreement with Orthogonal Genotypes for 99 Cell-Line and 232 Whole Blood Samples. ACMG categorical genotype boundaries are included for reference. All samples fell within expected categories based AmplideX PCR/CE *FMR1* precision metrics (± 1: 0-70 repeats, ± 3: 71-119). All 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%

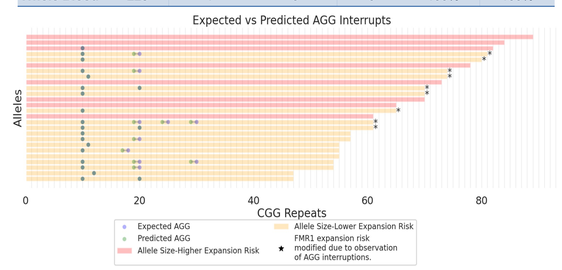

Figure 5. Predicted Risk of *FMR1* Expansion Based on AGG Interruption Status. A cohort of 26 intermediate and premutation alleles were assessed using Asuragen Xpansion Interpreter® (XI) and PCR/nanopore using a custom algorithm. Genotypes were in 100% agreement with XI for the absolute number of AGG 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.

Table 4. Fourteen Carriers (5.7%) were Identified in a Presumed Normal Cohort (n=247) Using the PCR/nanopore Panel. *CFTR* T501I is pending confirmation. *FMR1* intermediate expansions were identified in seven samples. No *HBB* carriers were identified.

Sample ID	Carrier	*FMR1*	*CFTR*	*SMN1, SMN2*	*HBA1/2*
SID489	*CFTR*	30	R117H, WT	2,1	aa/aa
SID502	*CFTR*	23,30	G622D, WT	2,1	aa/aa
SID526	*CFTR*	20,29	T501I, WT	3,2	aa/aa
SID481	*CFTR*	31	F508del, WT	2,1	aa/aa
SID530	*CFTR*	30	F508del, WT	2,1	aa/aa
SID562	*CFTR*	30	F508del, WT	2,2	aa/aa
SID569	*CFTR*	30	F508del, WT	2,2	aa/aa
SID637	*CFTR*	30	F508del, WT	2,3	aa/aa
SID461	*SMN1*	20,30	WT, WT	1,1	aa/aa
SID546	*SMN1*	29	WT, WT	1,2	aa/aa
SID403	*HBA2*	30	WT, WT	2,1	3.7del/aa
SID415	*HBA2*	20,29	WT, WT	2,1	3.7del/aa
SID531	*HBA2*	29	WT, WT	2,2	3.7del/aa
SID557	*HBA2*	30	WT, WT	2,2	3.7del/aa

CONCLUSIONS

- **The prototype PCR/nanopore assay accurately resolves multiple challenging variants across several classes for seven of the most common carrier screening genes.**
- **At least 96 barcodes can be combined in a single run using three PCR enrichment reactions and a unified data processing pipeline.**
- **The assay utilizes a single-platform, streamlined workflow and has potential to greatly reduce carrier screening complexity and turn around times, especially when paired with other similar assays (see Poster #P455, P546).**
- **In over 400 samples tested across each gene, the PCR/Nanopore assay agreed with the orthogonal method for SNVs/INDELs in *SMN1*, *CFTR*, *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. Beauchamp, K. A. *et al.* (2019) Genet Med 37, 773–8.
2. Zheng, Z. *et al.* (2022). Nat Comp Sci 2, 797–803.

This product is under development. Future availability and performance to be determined.
†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 ACMG 2023

Asuragen® a bio-techne® brand