A Streamlined PCR/Nanopore Sequencing Carrier Screening Panel for Cystic Fibrosis, Spinal Muscular Atrophy, and Fragile X Syndrome

Jon Kemppainen, Pranesh Rao, Julie R Thibert, Walairat Laosinchai-Wolf, Ryan Routsong, Ted Markulin, Monica Roberts, Bradley Martin, Bryan Killinger, Chris Fraher, Gary J Latham, and Bradley Hall
Asuragen, a Bio-Techne Brand, Austin, TX

Poster # G057

Summary

- Cystic Fibrosis (CF), Spinal Muscular Atrophy (SMA), and Fragile X Syndrome (FXS) are three of the most common inherited genetic disorders, each with high carrier rates that often require distinct genotyping methods.
- We developed a prototype assay comprised of novel PCR enrichment and Nanopore sequencing to simultaneously detect SNV/indels, copy number variation, and repeat sizing in a unified workflow, without the need for manual analysis.
- The assay utilizes amplicon read depth and machine learning models to automate and streamline identification of key genetic variants specific to each disease.
- Assay performance was evaluated with a cohort of 94 cell-line and 207 whole blood samples for all three genes resulting in 96-100% accuracy for each variant class.

Introduction

Cystic Fibrosis (CF), Spinal Muscular Atrophy (SMA), and Fragile X Syndrome (FXS) are three of the most common inherited genetic disorders, each with high carrier rates (~1/30 for CF, ~1/50 for SMA and ~1/250 for FXS). Historically, screening has required separate assays for each gene, often with inefficient workflows and uneven detection rates across ethnicities. Additionally, genes for SMA and FXS (*SMN1* and *FMR1*, respectively) are technically challenging to characterize due to homologous or GC-rich sequences. Repetitive polymorphisms, exon deletions and other *CFTR* variants can also be problematic to resolve.

Here, we demonstrate a prototype long-read sequencing (LRS) trio assay on Oxford Nanopore Technologies' (ONT) MinION platform. The assay identifies pathogenic variants across all exons of *CFTR*, exons 3-8 in *SMN1*/2, and sizes the CGG repeats in *FMR1* with phased AGG status.

Materials and Methods

Cell-line genomic DNA samples were obtained from Coriell Cell Repository including HapMap controls, 0 to ≥4 copies of *SMN1*/2, 48 unique pathogenic SNVs/indels in *CFTR* representing >87% carrier prevalence[1] in the US population, and all major repeat expansion categories for *FMR1*. Genomic DNA was isolated from whole blood of presumed healthy donors. DNA was amplified in two PCR enrichment reactions, barcoded, pooled, prepped by ligation sequencing kit (ONT), and run on R9.4.1 flow cells using the Mk1B (ONT). Data were analyzed using custom-developed software. The AmplideX® PCR/CE *CFTR* Kit[1], PCR/CE *SMN1/2* Plus Kit[1], PCR/CE *FMR1* Kit[1], Xpansion Interpreter[RUS], Sanger sequencing, and 1000 Genomes were used to provide reference genotypes where appropriate.

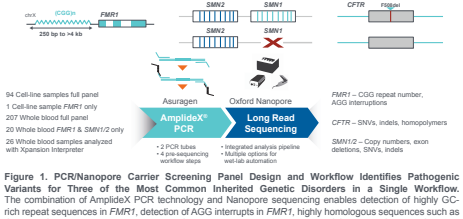

Figure 1. PCR/Nanopore Carrier Screening Panel Design and Workflow Identifies Pathogenic Variants for Three of the Most Common Inherited Genetic Disorders in a Single Workflow. The combination of AmplideX PCR technology and Nanopore sequencing enables detection of highly GC-rich repeat sequences in *FMR1*, detection of AGG interrupts in *FMR1*, highly homologous sequences such as *SMN1* or *SMN2*, and pathogenic variants (SNVs and indels) across all exons of *CFTR* and *SMN1*/2. One cell-line sample was included for *FMR1* only as an edge case for >900 CGGs. Whole blood samples (n=20) were included for *FMR1* and *SMN1*/2 only since *CFTR* orthogonal data was unknown. Highlights of the streamlined workflow are shown under the graphic.

Results

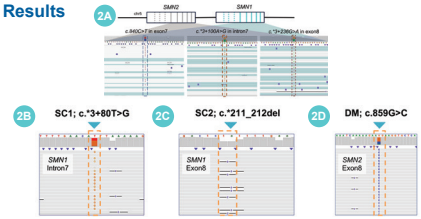

Figure 2. Sequence Data Differentiates *SMN1* and *SMN2* and Informs Silent Carrier Status and Disease Severity. A) Differentiation and assignment of reads to *SMN1* and *SMN2* by three different paralog-specific variants align reads to each gene and inform copy number variation prediction model. **B-D)**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

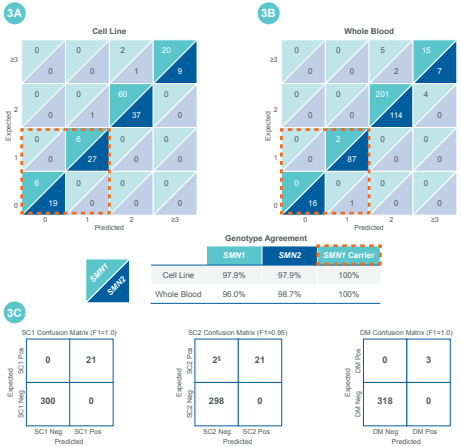

Figure 3. *SMN1/2* PCR/Nanopore Assay Classifies Carrier, Silent Carrier, and Disease Modifier Status with 96-100% Accuracy. Calling accuracy for *SMN1* and *SMN2* copy numbers in **A)** 94 cell-line and **B)** 227 whole blood samples. Hyperparameters for the decision tree model were selected using an 80:20 train:test split in a stratified randomly selected five-fold cross validation scheme on an independent set of 431 samples (79 cell lines and 352 whole blood). All carriers were identified with 100% accuracy (orange dashed outline). **C)** Variant allele frequency at each SNV associated with silent carrier variant SC1 (c.*3+80T>G, left panel), silent carrier variant SC2 (c.*211_212del, middle panel), and disease modifier mutation DM (c.859G>C, right panel) were used to predict status of silent carrier or disease modifier. Grid search using F1 scores determined the optimal allele frequency threshold, with samples above the threshold being classified as "Predicted Positives" and samples below the threshold as "Predicted Negatives". The algorithm accurately identified SC1 (n=21) and DM mutation (n=3). [1]Investigating evidence of two erroneous orthogonal calls suggests 100% agreement across SC1, SC2, and DM. Overall, 310/321 (96.6%) *SMN1* and 316/321 (98.4%) *SMN2* sample copy number agreed with the orthogonal method or copy number (CN) between 0 and >=3 gene copies.

Figure 4. Sequence Data Reveals Diverse Pathogenic Variants Across *CFTR* Exons. A) Heterozygous F508del, a 3-bp deletion at GRCh38 chr7:117,559,591 resulting in the deletion of phenylalanine at codon 508. **B)** Homozygous missense variant M1101K, a T>A transversion at position chr7:117,587,806 resulting in a protein change from methionine to lysine. **C)** Heterozygous missense variant G551D, a G>A transition at position chr7:117,587,806 resulting in a protein change from glycine to aspartic acid. **D)** Heterozygous 2183AA>G in region chr7:117592212- 117592219 resulting in a frameshift from a two-bp deletion of AA and insertion of G.

Table 1. *CFTR* Sample Agreement with Orthogonal Data for 94 Cell-line Samples (top) and 207 Presumed Normal Whole Blood Samples (bottom) Using Clair3[2] or Coverage Differences for Exon Del/Dup Variant Identification. Assay detected 47/48 (97.9%) unique variants including two del/dup (*CFTR*dele2,3, *CFTR*dele19-21). The false negative (miss) was 3876delA in two cell-line samples called wild-type by the algorithm. Whole blood samples were accurately called, except for one potential false positive. ¥Evidence of the *CFTR*dele19-20 call in one sample could not be verified by an orthogonal method. Overall, 298/301 (99.0%) samples agreed with orthogonal data.

Cell Line	AmplideX PCR/CE *CFTR*			Overall
Sample level accuracy	Allele1/ Allele2	WT/WT	MUT/WT	MUT/MUT	Genotype agreement
CFTR PCR/ Nanopore	WT/WT	48	2	0	48/50 (96.0%)
	MUT/WT	0	20	0	20/20 (100%)
	MUT/MUT	0	0	24	24/24 (100%)

Whole Blood	AmplideX PCR/CE *CFTR*			Overall
Sample level accuracy	Allele1/ Allele2	WT/WT	MUT/WT	MUT/MUT	Genotype agreement
	WT/WT	200	0	0	200/200 (100%)
CFTR PCR/ Nanopore	WT/MUT	0¥	7	0	7/7 (100%)
	MUT/MUT	0	0	0	0/0 (100%)

Table 2. *FMR1* Categorical Agreement with Orthogonal Data for 95 Cell-Line and 227 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: 71-120). Expanded samples up to 940 CGG repeats were all flagged correctly. Additionally, CGG sizing was accurate within precision for 312/322 (96.9%) samples and 454/462 (98.3%) alleles. In 7/10 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.

Sample Level Categorical Accuracy	Normal <45 CGG	Intermediate 45-54 CGG	Premutation 55-200 CGG	Mutation >200 CGG	Sensitivity	Specificity
Training	61	18	55	19	100%	100%
Cell line	72	5	14	4	100%	100%
Whole Blood	220	7	0	0	100%	100%

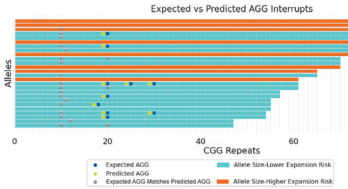

Figure 5. Predicted Risk of *FMR1* Expansion Based on AGG Interruption Status. A cohort of 26 intermediate and premutation alleles were assessed using Asuragen Xpansion Interpreter (XI) and PCR/ Nanopore using a custom algorithm. Genotypes were in 100% agreement with XI for the absolute number of AGG interruptions and within ±1 for the absolute position of each AGG interruption within the CGG repeat.

Table 3. Seventeen Carriers (8.2%), Including Those with Intermediate or Larger *FMR1* Expansions, were Identified in a Presumed Normal Cohort (n=207) Using the PCR/Nanopore Panel. One donor sample (SID530 in bold) was a carrier for both *CFTR* and *SMN1*. All variants were confirmed by orthogonal assay.

Sample ID	*FMR1*	*CFTR* Allele 1	*CFTR* Allele 2	*SMN1*	*SMN2*	DM	SC1	SC2	Identified Carriers
SID537	46	WT	WT	2	0	(-)	(-)	(-)	*FMR1* Intermediate
SID542	47	WT	WT	2	0	(-)	(-)	(-)	*FMR1* Intermediate
SID497	48	WT	WT	2	1	(-)	(-)	(-)	*FMR1* Intermediate
SID428	30,54	WT	WT	2	2	(-)	(-)	(-)	*FMR1* Intermediate
SID535	39,47	WT	WT	2	2	(-)	(-)	(-)	*FMR1* Intermediate
SID611	24,46	WT	WT	3 EXP 2 PRED	2	(-)	(-)	(-)	*FMR1* Intermediate
SID481	31	F508del	WT	2	1	(-)	(-)	(-)	*CFTR*
SID562	30	F508del	WT	2	2	(-)	(-)	(-)	*CFTR*
SID569	30	F508del	WT	2	2	(-)	(-)	(-)	*CFTR*
SID637	30	F508del	WT	2	3	(-)	(-)	(-)	*CFTR*
SID489	30	R117H	WT	2	1	(-)	(-)	(-)	*CFTR*
SID420	31	R74W (VUS)	WT	3	1	(-)	Positive	Positive	*CFTR*
SID530	**30**	**F508del**	**WT**	**2**	**1**	**(-)**	**Positive**	**Positive**	***CFTR, SMN1* SC**
SID461	20,30	WT	WT	1	1	(-)	(-)	(-)	*SMN1*
SID546	29	WT	WT	1	2	(-)	(-)	(-)	*SMN1*
SID578	20,33	WT	WT	2	2	(-)	Positive	Positive	*SMN1* SC
SID589	30	WT	WT	2	2	(-)	Positive	Positive	*SMN1* SC

Conclusions

- The prototype PCR/Nanopore assay accurately resolves multiple challenging variants for three of the most common carrier screening genes using a streamlined workflow.
- At least 96 samples can be combined in a single run using only two PCR enrichment reactions. This has potential to reduce carrier screening turn around times when paired with other similar assays (see Poster # G055).
- *FMR1* CGG repeats and AGG interruptions associated with expansion risk from mother-to-child were detected in the same workflow.
- Detection of a dual carrier (*CFTR* and *SMN1*) highlights the importance of a unified carrier screening approach.
- In over 300 samples tested across each gene, the PCR/Nanopore assay agreed with the orthogonal method for *SMN1* SNVs/indels (>99% of samples), *SMN1* CN (96.6%), *SMN2* CN (98.4%), *CFTR* SNVs/indels/ exon deletions (99.0%), *FMR1* repeat categories (100%) and *FMR1* AGG number (100%).

References

1. Beauchamp, K. A. et al. Sequencing as a first-line methodology for cystic fibrosis carrier screening. Genet Med 37, 773–8 (2019).
2. Zheng, Z. et al. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Biorxiv 2021.12.29.474431 (2021).

Asuragen
a biotechne brand

This product is under development. Future availability and performance to be determined. [1]For Research Use Only. Not for Use in Diagnostic Procedures. [2]Xpansion Interpreter[RUS] 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 AMP 2022