Combining AmplideX® PCR Chemistries, Nanopore Sequencing, and Automated Analysis Software to Genotype Complex, Clinically-relevant Variants in 11 Challenging Targets Associated with Hereditary Diseases

John N. Milligan, Bryan J. Killinger, Cody Edwards, Christopher J. Fraher, Jon Kemppainen, Monica Roberts, Jonathan Turner, Theodore Markulin, Brennan Greenlee, Eduardo Priego, Juliette A. Baker, Mahesh Yarasi, Connor Parker, Mia K. Mihailovic, Ashima Sharma, Brian C. Haynes, and Bradley Hall
Asuragen, a Bio-Techne Brand, Austin, TX

Poster Number: P21.015.B

Summary

- Conventional sequencing methods struggle with detecting complex variants in key genes associated with inherited genetic disorders of high prevalence, necessitating complicated workflows and producing results that often lack comprehensive detection of pathogenic variants.
- We developed a novel PCR enrichment/Nanopore sequencing assay and accompanying analysis software that detects both simple and complex variants across *CFTR*, *SMN1/2*, *FMR1*, *HBA1/2*, *HBB*, *GBA*, *CYP21A2*, *TNXB*, and *F8* intron inversions with a streamlined end-to-end workflow.
- Analysis performance was optimized and assessed using a mixed sample set consisting of 371 cell-line samples and 141 gDNA samples isolated from whole blood, which demonstrated >95% agreement across all variant categories and genes included in the assay.

Introduction

Detection of pathogenic variants associated with severe genetic disorders is critical for diagnostic and screening applications. Although NGS is a widely used method for genetic characterization, it fails to resolve many variants in genes of clinical interest that have complex pathogenic elements like repeats, structural variation, and pseudogenes[1,2]. This traditionally necessitates use of multiple specialized workflows that only cover a fraction of pathogenic alleles[3].

To address this shortcoming for researchers, we combined short- and long-range PCR enrichment with nanopore sequencing in a kit-based modular panel to accurately genotype traditionally NGS addressable genes (*CFTR* and *HBB*) and hard-to-decipher genes (*SMN1/2*, *FMR1*, *HBA1/2*, *GBA*, *CYP21A2*, *TNXB*, and *F8* intron inversions) in a single streamlined workflow, representing ~70% of all pathogenic variants associated with inheritable diseases that impact neonates[3]. We developed bespoke software and algorithms to automate analysis of SNVs, indels, copy number gain or loss, gene-pseudogene fusions, and large structural variants. Here we describe results using this prototype assay to genotype 512 total samples (297 unique) across multiple variant classes. We also present key elements planned or already implemented in the accompanying software interface that allow users to easily navigate sequencing results, including quality control metrics, variant detection, and results visualization that inform the user with context for different classes of variant calls.

Methods

The assay was optimized and evaluated using 512 sample measurements, including cell-line genomic DNA (gDNA) samples (n=371 total/187 unique) and whole blood gDNA from residual clinical samples (n=141 total/110 unique). The selection of cell-line and whole blood samples represent all major variant classes and were used to train and/or evaluate analysis performance. Target regions were enriched in 4 PCR reactions, barcoded, pooled, and sequenced on MinION flow cells (R10.4.1) with a Mk1B or GridION (Oxford Nanopore Technologies, ONT). Software was developed to automate analysis and reporting. Multiple orthogonal methods provided comparator data for determining assay performance.

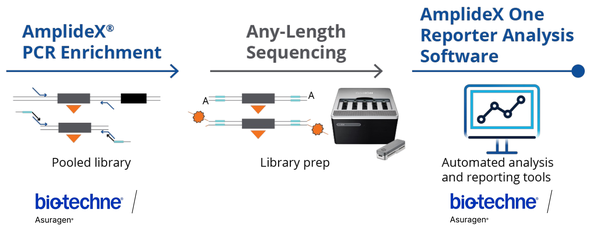


Figure 1. Summary overview of PCR enrichment, sequencing, and automated analysis software. Targeted PCR enriches panel-specific genetic regions. The amplicons (~300bp – 10kb in length) are PCR barcoded by sample, pooled, and sequenced on an ONT (Oxford Nanopore Technologies) device. Raw sequencing data is analyzed with automated software to produce quality control metrics, annotated variant calls, and results visualization.

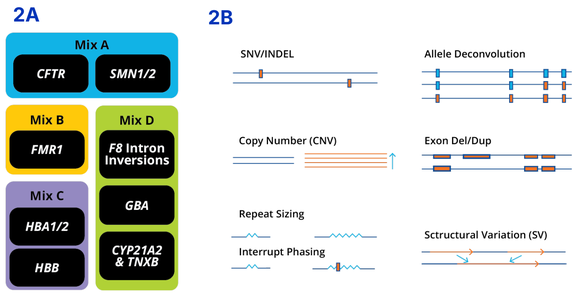


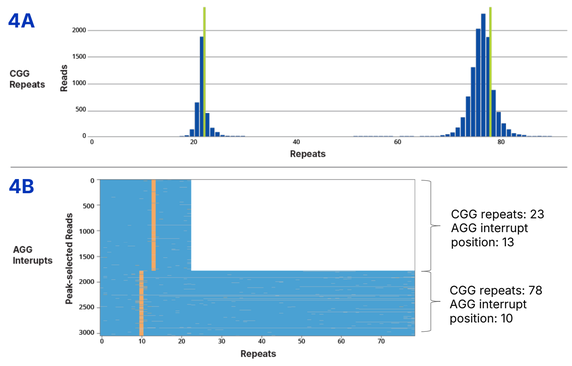
Figure 2. Gene content and variant types automatically analyzed. (A) Four PCR mixes are used to amplify targets across 11 genes. **(B)** Multiple variant classes are identified and reported simultaneously by the analysis software.

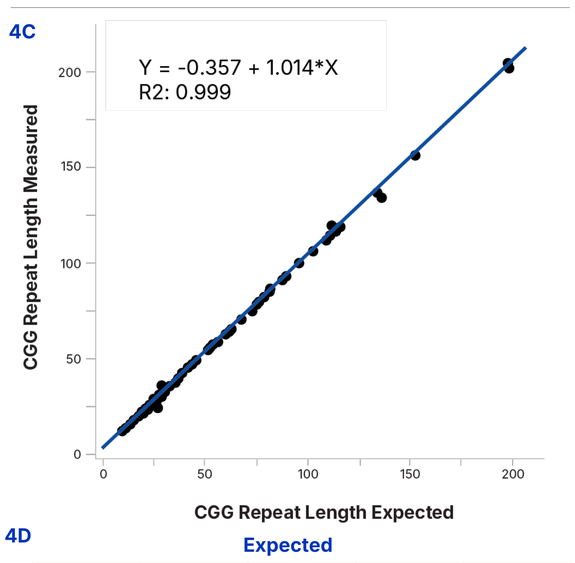
Results

Sample	Barcode	Filtered Variants	Mixes Analyzed	QC
NA06905	BC56	*FMR1*: 78 CGG [10 AGG] (Premutation)	A B C D	All pass
NA18668	BC38	*CYP21A2*: c.923dup (pathogenic) *CYP21A2*: c.955C>T (pathogenic) *CYP21A2*: c.1069C>T (pathogenic) *CFTR*: c.1521_1523del (pathogenic)	A B C D	All pass
HG02353	BC71	*HBA1*: $\alpha^{0(SEA)}$ (pathogenic) *HBA2*: $\alpha^{0(SEA)}$ (pathogenic)	A B C D	All pass
NA23255	BC57	*SMN1*: Copy number loss (pathogenic) (2x)	A B C D	All pass

Figure 3. Customizable high-level analysis results allow users to quickly identify desired variants. A simple interface allows users to filter samples in a variety of ways, including ClinVar pathogenicity annotations. Color-coded PCR mixes and summarized quality control information allows users to quickly identify pertinent issues related to an analysis. The presence of all green PCR "Mixes Analyzed" and "All pass" QC entries indicate all analyses across all mixes were analyzed without QC flags raised for any of these samples.

FMR1 CGG/AGG





4D

Measured \ Expected	*FMR1* Genotype	Normal	Inter-mediate	Pre-mutation	Full Mutation
	Normal	376	0	0	0
	Inter-mediate	6*	14	0	0
	Pre-mutation	0	0	29	0
	Full Mutation	0	0	0	8

Figure 4. Signal processing visuals allow users to thoroughly investigate *FMR1* CGG repeats and AGG interrupts. (A) The "CGG Repeats" panel displays a histogram of read counts for each repeat size, which is used by the algorithm to determine the number of CGG repeats in each *FMR1* allele (green line) and are identified in this sample (NA06905) at 23 and 78 repeats. **(B)** Phased "AGG Interrupts" will be graphically displayed to allow for investigation of interrupt patterns (orange) within each allele. AGG interrupts were detected at the 13th and 10th repeat positions for each respective allele and indicate the risk of *FMR1* repeat expansion. **(C)** *FMR1* CGG repeat size correlation plot for 627 alleles across 451 sample measurements using 251 unique samples. 16 sample measurements were excluded due to lack of orthogonal data. One allele call was excluded due to an extra call in the software (98/136 expected, 98/122/136 called). **(D)** *FMR1* sample-level genotype agreement was 98.6% (427/433) for exact sizing. All six discrepancies (*) occurred in samples near the repeat boundary that differed by 1 CGG repeat, and all were within size tolerance according to EMQN guidelines (+/- 5% of repeat size). Accounting for these tolerances, agreement was 100%. 4 samples were excluded due to no call (QC flag). 14 samples were excluded due to lack of orthogonal data.

SMN1/2 Copy Number

5A

Measured \ Expected	*SMN1* Cp#	0	1	2	≥3
	0	17	0	0	0
	1	0	18	0	0
	2	0	0	320	2
	≥3	0	0	4	101

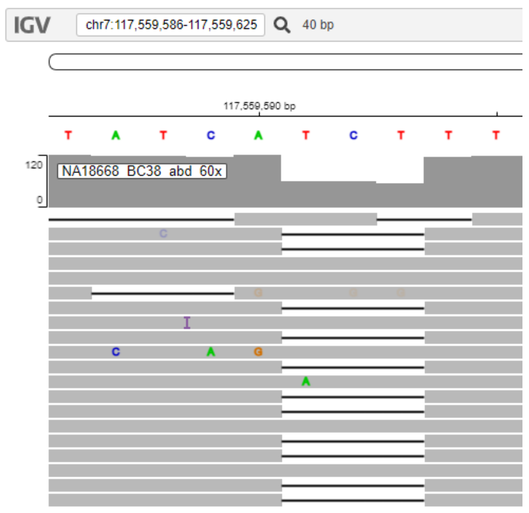
5B

Measured \ Expected	*SMN2* Cp#	0	1	2	≥3
	0	63	0	0	0
	1	0	145	1	0
	2	0	0	205	1
	≥3	0	0	1	43

Figure 5. PCR/Nanopore assay accurately quantifies *SMN1/2* copy number. Calling accuracy for **(A)** *SMN1* and **(B)** *SMN2* copy numbers across 463 measurements from 255 unique samples (178 cell line, 77 residual clinical). Overall percent agreement was 98.7% (456/462) for *SMN1*, and 99.3% (456/459) for *SMN2*. For genotypes associated with SMA (*SMN1* E7 Cp = 0) and SMA carriers (*SMN1* E7 Cp = 1), agreement was 100%. Two samples were excluded for QC failure, 8 samples were excluded due to known or suspected issues with orthogonal reference method data. Hybrid genes were manually interpreted; automated analysis will be implemented later in software design. There were 3 *SMN1* and 6 *SMN2* clinical samples removed from analysis due to lack of orthogonal data. For SNVs and Indels associated with *SMN1* gene duplication and *SMN2* disease modifiers, agreement was 100% (data not shown).

CFTR SNVs, INDELs, and Large Exon Deletions

6A



6B

Measured \ Expected	*#CFTR* Variants	0	1	≥2
	0	302	2	0
	1	0	44	0
	≥2	0	0	59

Figure 6. PCR/Nanopore assay identifies pathogenic SNVs, indels, and exon deletions in *CFTR*. (A) The software's BAM/VCF browser allows users to review the aligned sequencing data at the relevant genomic location. The *CFTR* c.1521_1523del (F508del) variant in sample NA18668 is shown here with a clear signal of a heterozygous 3-base pair deletion in the .bam pileup. **(B)** Calling accuracy for *CFTR* across 407 sample measurements. Overall sample-level agreement was 99.5% (405/407), including for three large exon deletions (*CFTR* dele2,3, *CFTR* dele19-21, *CFTR* dele4-11). Both misses were due to missed detection of variant 3876delA, located in a low complexity homopolymer region. This variant was correctly called in three additional sample measurements.

HBA1/2 Copy Number

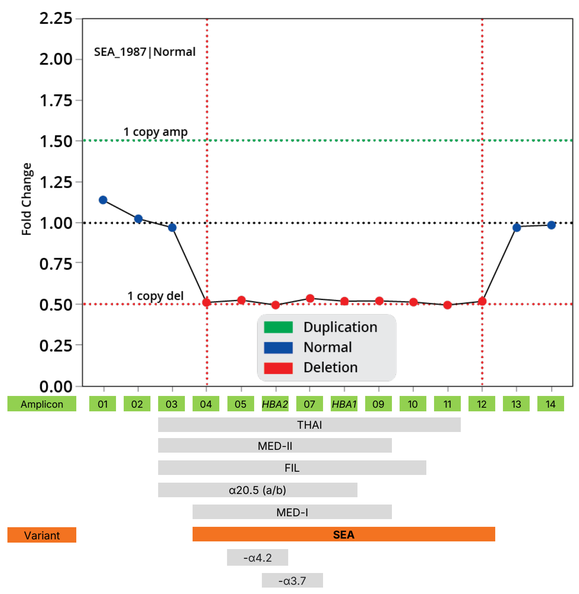


Figure 7. *HBA1/2* amplicon fold-changes provide signal for a diverse set of variant genotypes. A set of 14 "sentinel" amplicon target regions of the hemoglobin alpha cluster differentiate known common breakpoints. Reduced fold-change signal (~0.50) across a set of neighboring amplicons (4-12) indicates a heterozygous SEA deletion in the sample shown here (HG02353). An algorithm classifies the genotype as $\alpha\alpha/\alpha^{0(SEA)}$ with only one copy of *HBA1* and *HBA2,* indicating a α^0 "α-thalassemia-trait" genotype.

Cyp21A2 Structural Variation and Paralog Deconvolution

8A

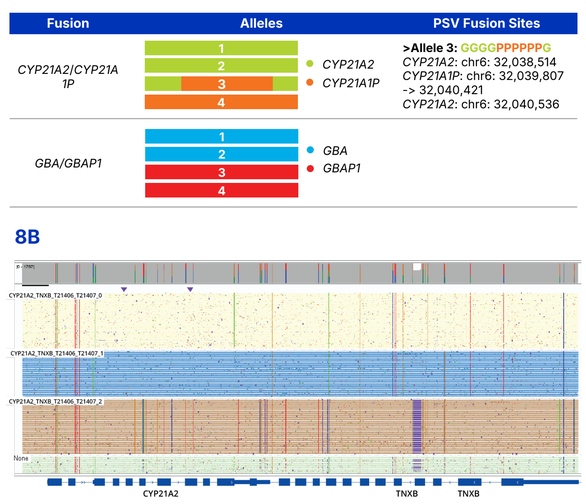


Figure 8. Detection and investigation of *CYP21A2*/*CYP21A1P* gene fusions in sample NA18668 with allele deconvolution and IGV shows clear fusion signal. (A) Fusion sites corresponding to both (G)ene and (P)aralog entries in *CYP21A2*/*CYP21A1P* allele 3 indicate the detection of paralogous sequence variants. In contrast, no fusion is detected in *GBA*/*GBAP1 (data not shown)*. **(B)** Users can thoroughly investigate .bam file alignments for these regions by leveraging the use of "deconvolution group" read tags that identify unique alleles (shown grouped and colored).

Table 1. Performance metrics across variant classes. All variant classes for *CFTR, SMN1/2*, *FMR1*, *HBA1/2*, *HBB*, *CYP21A2/CYP21A1P*, *GBA/GBAP*, and *F8* intro inversions were detected at ≥95% agreement with orthogonal data. Performance is evaluated based on samples and alleles with orthogonal comparator data (N). Overall percent agreement (OPA) refers to the accuracy of the analysis. CNV calls are evaluated with performance from 0 to ≥3 copies. Includes external evaluation testing performed by Canterbury Health Laboratories where indicated (see footer a).

Gene	Variant Class	N	Accuracy/OPA
CFTR	PolyT/TG[a]	866 alleles (433 samples)	PolyT: 99.4% PolyTG: 99.0%
	SNV/Indel[a]	51622 variants (413 samples)	100%
	Large Exon Deletions[b,c]	95 samples	100%
SMN1/2	CNV[a]	468 samples	*SMN1*: 98.7% *SMN2*: 99.4%
	SNV/Indel	1413 variants (471 samples)	100%
FMR1	AGG[d]	58 alleles (24 samples)	100%
	CGG[a]	641 alleles 441 samples	100%
HBA1/2	CNV[a]	536 alleles (266 samples)	96.1%
	SNV/Indel	2 variants (1 sample)	100%
HBB	CNV[a,e]	26 variants (19 samples)	96.2%
	SNV/Indel[a]	94 variants (39 samples)	100%
CYP21A2/CYP21A1P	CNV[d]	109 samples	95%
	Gene/Pseudogene Fusions	59 alleles (27 samples)	100%
	SNV/Indel	620 variants (27 samples)	100%
GBA/GBAP	CNV[d]	7 samples	100%
	Gene/Pseudogene Fusions	4 alleles (4 samples)	100%
	SNV/Indel	135 variants (9 samples)	100%
	Structural Variants	4 variants (4 samples)	100%
F8	Intron 1 & 22 Inversions[a]	10 samples	100%

a. Includes internal data presented above plus 37 additional samples tested at an external site; n additional samples per gene are as follows: *CFTR* (8), *SMN1/2* (6), *FMR1* (8), *HBA* (5), *HBB* (6), *F8* (4). These samples were analyzed in a newer pipeline (v14); QC failures were excluded.
b. Performance evaluated on a recently optimized set of Mix A primers with a different sample set consisting of 5 known deletions and 90 assumed WTs.
c. Performance was evaluated on a subset of data used for testing and not model development. All other analyses show performance for all available data.
d. Sparse orthogonal annotations confound performance metrics as context is needed for each sample (e.g., fusions may be orthogonally annotated as deletions).
e. Performance evaluation does not include cell-line samples used as calibrators.

Conclusions

- Our PCR/Nanopore assay detects SNVs, indels, copy number gain or loss, short tandem repeat sizes, large structural variants, and gene-pseudogene fusions with ≥95% overall percent agreement to orthogonal results across *CFTR*, *SMN1/2*, *FMR1*, *HBA1/2*, *HBB*, *GBA*, *CYP21A2*, *TNXB*, and *F8* intron inversions.
- The assay utilizes a single-platform, streamlined workflow with the potential to greatly reduce assay complexity and turnaround times for characterization of complex genotypes associated with important and prevalent inherited disorders.
- Accompanying software allows users to navigate quality control metrics, view variant calls, and perform in-depth investigation of sequencing data and analysis results with a simple user interface.
- In >500 samples tested, 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 (98.7%), *SMN2* CN (99.4%), *FMR1* repeat categories (100%), *FMR1* AGG interruptions (100%), *HBA1/2* deletions (100%), CYP21A2 fusions (100%), GBA fusions and structural variants (100%), and *F8* inversions (100%).

References:

1. Woo, E. G., Tayebi, N., & Sidransky, E. (2021). Next-Generation Sequencing Analysis of *GBA1*: The Challenge of Detecting Complex Recombinant Alleles. *Frontiers in genetics*, *12*, 684067.
2. Seaby, E. G., & Ennis, S. (2020). Challenges in the diagnosis and discovery of rare genetic disorders using contemporary sequencing technologies. *Briefings in functional genomics*, *19*(4), 243–258.
3. Beauchamp, K.A., Muzzey, D., Wong, K.K., Hogan, G.J., Karimi, K., Candille, S.I., Mehta, N., Mar-Heyming, R., Kaseniit, K.E., Kang, H.P., *et al.* (2018). Systematic Design and Comparison of Expanded Carrier Screening Panels. *Genet. Med. 20*, 55–63.

This product is under development, including the software interface that may differ from the partially mock representation shown here.

Future availability and performance to be determined.

All authors have the financial relationship to disclose: Employment by Asuragen.

Presented at ESHG 2024.

We would like to thank Lisa Hsu and Andrew Laurie from Canterbury Health Laboratories for providing residual clinical samples and orthogonal method results included in Table 1.