Accurate Genotyping of Complex, Clinically-Relevant Variants in 11 Hard-to-Decipher Genes by Combining Novel PCR Chemistries with Any-length Nanopore Sequencing for Carrier Screening

G030

Bradley Hall, Bryan Killinger, Christopher J. Fraher, Cody Edwards, Ryan Routsong, Jonathan Turner, Monica Roberts, Jon Kemppainen, Brennan Greenlee , 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 inherited genetic disorders.
- 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 233 cell-line samples and performance evaluated with 347 whole blood, 1 saliva, and 5 buccal samples across the eleven genes with a subset set of 233 whole blood to identify potential carriers from presumed normal donors.

Introduction

Nearly everyone is a genetic carrier of a disease or condition[1]. Carrier screening (CS) identifies couples at risk for having a child with a severe genetic disorder. Although Next-Generation Sequencing (NGS) is a widely used method, it fails to resolve many problematic genes recommended in professional practice guidelines due to GC-rich 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.

To address these shortcomings, we combined three innovations : 1) short- & long-range PCR, 2) any-length nanopore sequencing, and 3) customized software analysis pipelines. Using a single workstream, we developed a modular panel of 11 genes critical for CS, including nine "hard-to-decipher" genes. We estimate the panel represents ~75% of at-risk couple detection compared to panels that are 15 times larger[3]. Here we describe results utilizing this prototype assay to genotype *CFTR*, *SMN1*, *SMN2*, *FMR1*, *HBA1*, *HBA2*, *HBB*, *F8* intron inversions, *GBA*, *CYP21A2*, and *TNXB* across more than 400 samples.

Methods

Cell-line (CL) genomic DNA (gDNA) samples (n=233) were obtained from Coriell Cell Repository. Genomic DNA was also isolated from presumed normal or clinical donors (n=353). One contrived and two commercial control samples were utilized for *F8* intron inversion. Target regions were enriched in four PCR 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). A cohort of 12 cell lines and 232 presumed normal whole blood samples were analyzed by the entire panel. Remaining samples were tested by a subset of primer mixes or genes analyzed for variants. 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 *SMN*1/2 Plus Kit†, and PCR/CE *FMR*1 Kit†, Xpansion Interpreter*‡† (XI), Sanger sequencing, and qPCR) were utilized to determine comparator results.

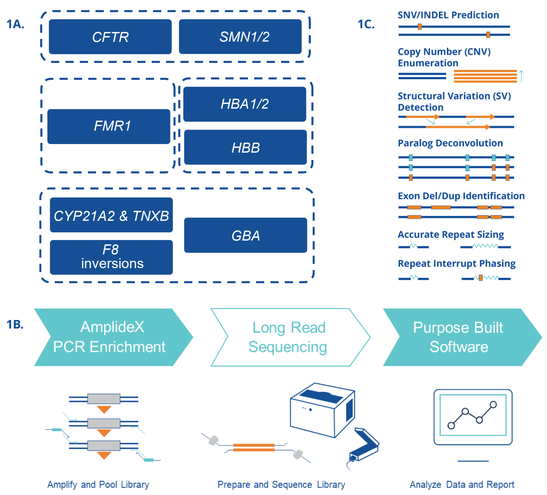


Figure 1. PCR/nanopore Carrier Screening Panel Design and Workflow Identifies Pathogenic Variants for 11 Genes Responsible for Common Inherited Genetic Disorders and/or Challenging to Assay by Convention Methods. The combination of AmplideX® PCR technology across **A)** 4 primer mixes, and **B)** nanopore sequencing enables detection of **C)** multiple variant classes for each of the 11 genes within the panel in a single workflow.

Results

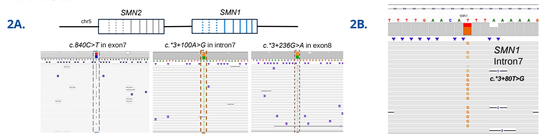
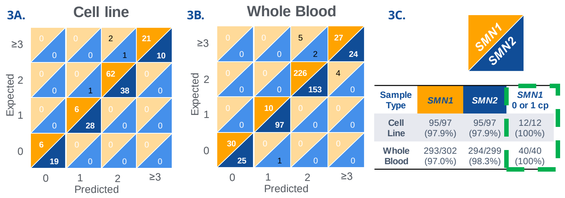


Figure 2. Sequence Data Differentiates *SMN1* and *SMN2* and Informs Silent Carrier Status and Disease Modifiers. A) Differentiation and assignment of reads to *SMN1* and *SMN2* by three different paralog-specific variants align reads to each gene and inform CNV prediction model. **B)** Silent carrier (SC1; *c.*3+807>G*) variant alignments. SC2 (*c.*211_212del*) and disease modifier (DM; *c.859G>C*) not shown.



Sample Type	*SMN1*	*SMN2*	*SMN1* 0 or 1 cp
Cell Line	95/97 (97.9%)	95/97 (97.9%)	12/12 (100%)
Whole Blood	293/302 (97.0%)	294/299 (98.3%)	40/40 (100%)

Figure 3. *SMN1/2* PCR/Nanopore Assay Accurately Classifies Carrier Status. Calling accuracy for *SMN1* and *SMN2* copy numbers in **A)** 97 CL and **B)** 232 presumed normal and 70 clinical 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). **C)** Carriers were identified with 100% accuracy (green dashed outline). There were 3 clinical samples removed from analysis for *SMN2* due to lack of orthogonal data.

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	28	0	28/28 (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%)

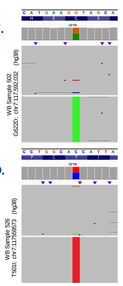


Figure 4. *CFTR* Sample Level Agreement with Orthogonal Data for 104 Cell-Line and 234 Whole Blood Samples. The assay used Clair3 (SNV/indel) and read depth heuristics (del/dup) to detect 65 unique variants, including three del/dup (*CFTR*dele2,3, *CFTR*dele19-21, *CFTR*dele4-11). **A-B)** The assay was performed with both R10.4.1 (n=291), and R9.4.1, (n=47) with 100% agreement or orthogonal methods. **C-D)** Nanopore read pile-ups of two variants previously undetected in whole blood (G622D, T501I; ¥Sanger sequencing verification pending).

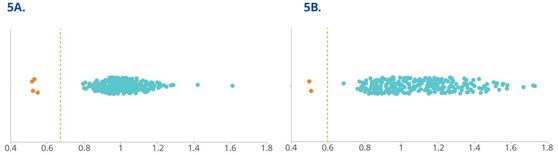


Figure 5. Large Exon Deletions Detected at 100% PPV and Sensitivity. Normalized amplicon coverage distinguishes exon deletion (orange) from wildtype genotypes (teal) for **A)** *CFTR*2,3dele and **B)** *CFTR*dele20. Samples with normalized coverage below the threshold (vertical dashed line) were classified as heterozygous for the large exon deletions.

Table 1. Pathogenic and Likely Pathogenic *HBA1/2* and *HBB* Variants Detected in 147 Cell-Line and 283 Whole Blood or Clinical Samples. The sample set contained 247 presumed normal WB, 30 clinical WB, 5 buccal, 1 saliva tested with the combined *HBA1/2* and *HBB* assay on R10.4.1 flow cells. Of these, 94 CL and 238 WB samples were genotyped as wild-type (aa/aa). All but 3 samples were concordant with orthogonal data. The three discordant samples were miscalled for anti-3.7 (2 FN, 1 FP) and do not affect carrier status.

Gene	Variant Type	Unique Variants Detected	Sample Agreement
HBA	Structural	SEA, FIL, THAI, 3.7del, anti-3.7, and 4.2del	72/75 (96%)
	SNV/Indel	c.427T>C (Hb Constant Spring), c.207C>G (Hb G-Philadelphia), c.237del, c.142G>C, c.95+2_95+6del	8/8 (100%)
HBB	Structural	(Hb Lepore-Baltimore), (Sicilian [Aβ 0-Thal), HPFH-1, HPFH-2	8/8 (100%)
	SNV/Indel	c.19G>A (Hb C), c.79G>A (Hb E), c.20A>T (Hb S), c.126_129del, c.93-21G>A, c.-137C>G, c.-138C>T, c.-138C>T, c.118C>T, c.315+1G>A, c.316-197C>T, c.82G>T (Hb Knossos), c.316-2A>G c.17_18del, c.-79A>G, c.404T>C, c.27dup, c.364G>C, c.92+5G>C, c.92+6T>C, c.92G>C, and c.208G>A	44/44 (100%)

Table 2. *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%

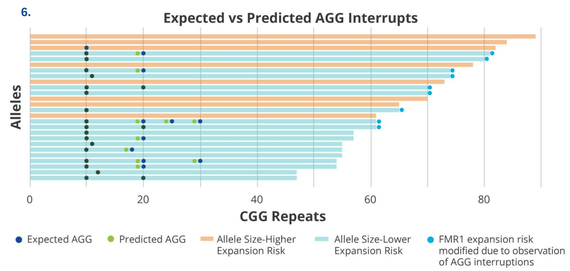


Figure 6. Predicted Risk of *FMR1* Expansion Based on AGG Interruption Status. A cohort of 26 intermediate and premutation alleles were assessed using Asuragen Xpansion Interpreter* (XI) and PCR/nanopore using a custom algorithm. Genotypes were in 100% agreement with XI for the absolute number of AGG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

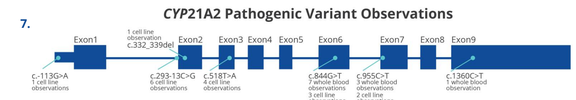


Figure 7. Accurate Resolution of Copy Number and Pseudogene Fusions in the *CYP21A2* Gene Cluster Utilizing Sequencing Deconvolution on R10.4.1. 28 pathogenic variants were identified across cell-line and whole blood samples and mapped to their location in *CYP21A2*. We did not find evidence of *TNXB* in samples tested.

Figure 8. Prototype Assay Amplifies and Identifies >100 kb *F8* Inversions. *F8* inversions were observed in three samples using 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).

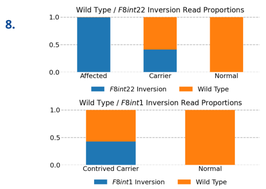
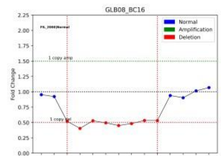


Figure 9. Pathogenic and Likely Pathogenic *HBA1/2* and *HBB* Variants Detected Using the PCR/nanopore Assay. *HBA* deletion identification was based on normalized fold change in read depth compared to control sample reference amplicons on the y-axis; assay amplicons are arranged in sequential order on the x-axis. WB sample with a FIL deletion on one allele .



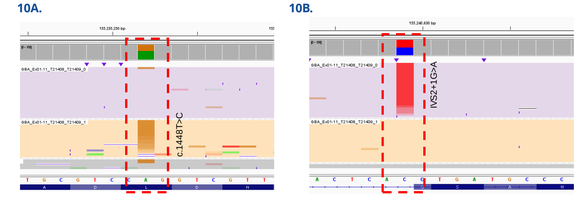


Figure 10. Sequence Data Reveals Diverse Pathogenic Variants Across the GBA gene on R10.4.1. Clinically affected cell-line NA20270 is a compound heterozygote and contains two SNVs within the *GBA* gene; A) one allele has a T>C transition at nucleotide 1448 in exon 10 (L444P, c.1448T>C), and B) the other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

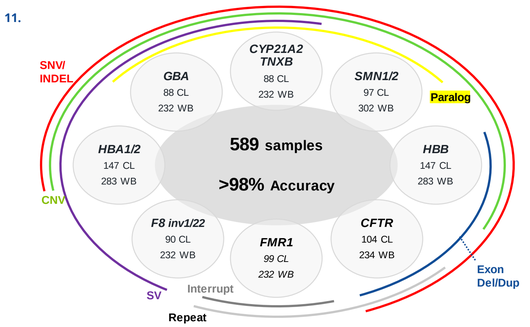


Figure 11. PCR/nanopore Carrier Screening Panel Design and Workflow Identifies Pathogenic Variants for 11 Genes Responsible for Common Inherited Genetic Disorders and/or Challenging to Assay by Convention Methods. The combination of AmplideX® PCR technology across 4 primer mixes, and nanopore sequencing enables detection of multiple variant classes for each of the 11 genes within the panel. A total of 589 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.

Table 3. Detected 37 Carriers (15.9%) 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. Hemophilia A (*F8* intron 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.

Gene	# of Carriers	Variants Identified
CFTR	8	5x F508del, wt; R117H, wt; G622D, wt; T501I, wt
SMN1	2	1, 1 and 1, 2 (*SMN*1, *SMN*2)
SMN1, SC	3	2, 2 + SC1, SC2‡
HBA1/2	4	3.7del/aa
HBB	3	c.404T>C; c.20A>T (Hb S); c.-79A>G
GBA	3	2x N370S (c.1226A>G); R496H (c.1604G>A)
CYP21A2	4 8	30-kb del; 3x Q318X (c.955C>T); 7x V281L (c.844G>T); P453S (c.1360C>T)
SMN1 & CYP21A2	1	1,3 & Q318X (c.955C>T)

Conclusions

- The prototype PCR/nanopore assay accurately resolves genotypes across multiple classes of variation for 11 of the most common and challenging genes 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 dual carrier (*CYP21A2* and *SMN*1) demonstrates the potential of a unified workflow to address targets that are conventionally analyzed through disparate assays.
- In 589 samples tested across the panel, the PCR/nanopore assay agreed with the orthogonal methods for SNVs/INDELs in *SMN*1, *CFTR*, *GBA*, *CYP21A2, HBA1*, *HBA2*, and *HBB* (>99% of samples), *SMN*1 CN (97.2%), *SMN*2 CN (98.2%), *FMR*1 repeat categories (100%), *FMR*1 AGG interruptions (100%), and *HBA1/2* deletions (100%).

We would like to thank Jaime L. Lopes, Marcus Cannon, Charles T. Quinn, and D. Brian Dawson from Cincinnati Children's for providing residual clinical samples and orthogonal method results (*HBA/HBB*). We would like to thank Ahmad Tayoun and Sawsan Yaslam from Al Jalila Children's Specialty Hospital, Dubai, UAE for testing residual clinical samples and orthogonal method results (*SMN1/2*).

†This product is under development. Future availability and performance to be determined.
‡Xpansion Interpreter is a laboratory-developed test.
All authors have the financial relationship to disclose: Employment by Asuragen.

Presented at AMP 2023

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).

bio-techne®
Asuragen®