Accessible AmplideX PCR and Nanopore Sequencing For Rare Variant Characterization in 11 Challenging Genes Associated with Hereditary Diseases

Bradley Hall[1], Anne-Sophie Lebre[2], Tony Yammine[2], Cody Edwards[1], Theodore Markulin[1], Bryan Killinger[1], Jacob Ashton[1], Christopher J. Fraher[1], and Brian C. Haynes[1]

[1]Asuragen, a Bio-Techne Brand, Austin, TX, USA; [2]Centre Hospitalier Universitaire De Reims, Reims, France

Summary

- Conventional short-read sequencing methods struggle with detecting complex variants in key genes associated with inherited genetic disorders of high prevalence, necessitating multiple complicated workflows which produce results that often lack comprehensive detection of pathogenic variants.
- We developed a prototype assay based on PCR-enrichment, nanopore sequencing, and machine learning models to enable multiplex detection of diverse variant classes including SNVs, INDELs, Exon del/dups, SVs, gene CNVs and STRs in a single workflow.
- We evaluated performance in collaboration with Dr. Lebre at CHU Reims on a set of 155 unique residual clinical samples collected from centers all over France.
- Sample-level genotype agreement was 100% for all samples that passed QC with variants that were either reported or found to be wild-type with no comparator information.

Introduction

Everyone is a genetic carrier of an inheritable 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 genes with complex pathogenic regions including GC-rich tandem repeats, copy number variation, pseudogenes, and structural variation[2,3]. Therefore, multiple specialized techniques are employed to identify and characterize pathogenic variants of interest.

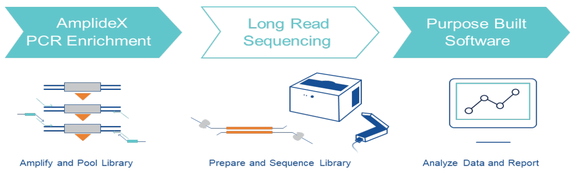
To address this shortcoming for researchers, we combined three innovations: 1) short- & long-range PCR enrichment, 2) nanopore sequencing, and 3) customized software analysis pipelines. Using a single workflow, we developed a modular panel to interrogate 11 genes that represent ~70% of all pathogenic variants associated with inheritable diseases that impact neonates[4]. Here we describe results utilizing this prototype assay to genotype *CFTR*, *SMN1*, *SMN2*, *FMR1*, *HBA1*, *HBA2*, *HBB*, *F8* intron inversions, *GBA*, *CYP21A2*, and *TNXB* from 155 residual clinical gDNA samples collected from centers across France.

Methods

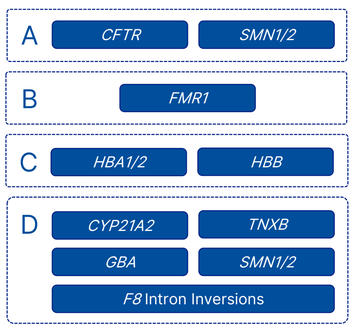
Deidentified and previously genotyped residual clinical genomic DNA samples isolated from whole blood (WB) donors were provided from sites across France (N=155). Target regions were enriched in 1-4 PCR reactions, barcoded, pooled and sequenced on MinION flow cells (R10.4.1) with a Mk1B (ONT) at CHU Reims by Dr. Lebre's group. Bespoke bioinformatics pipelines automated analysis of exon and whole gene gain or loss, gene-pseudogene fusions, and large structural variants. Clair3 and Sniffles2 were utilized for SNV/INDEL and SV identification[5,6]. Performance was demonstrated by comparing automated and manual variant calls to comparator data provided by collaborators.

1A) 155 Residual Clinical Samples Collected from Across France

1B) Single Assay Workstream



1C) 4 Mix Modular Design



1D) Variant Types Analyzed

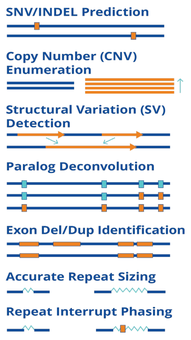
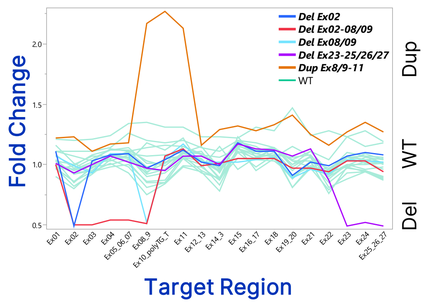


Figure 1. PCR/nanopore Panel Design and Workflow Identifies Pathogenic Variants for 11 genes Associated with Common Inherited Genetic Disorders. **A)** Residual clinical gDNA samples were procured from ten labs across France and tested at CHU Reims by Dr. Lebre's group. **B)** The workstream included PCR enrichment, nanopore sequencing, and automated data analysis for **C)** 11 genes regions in 4 mixes. **D)** Multiple variant classes are analyzed.

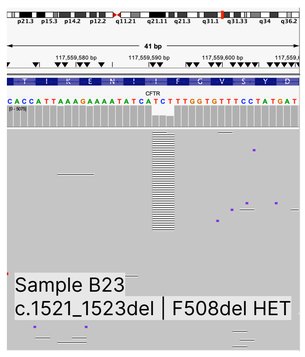
Results

Across *CFTR* (n=24) & *SMN1/2* (n=13) sample-level agreement was 100% for all samples that passed QC (Mix A)

2A) *CFTR* Structural Variant (SV) Accuracy



2B) *CFTR* Indel Example



2C) *SMN1* & *SMN2* Gene-Level Copy Number Agreement

Expected

Measured *SMN1* Cp#	0	1	2	≥3
0	0	0	0	0
1	0	3	0	0
2	0	0	9	0
≥3	0	0	0	1

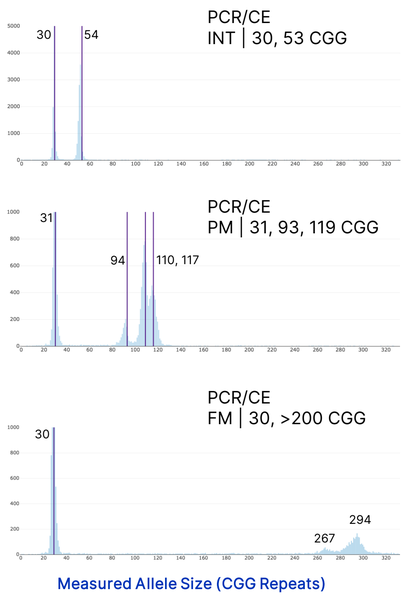
Expected

Measured *SMN2* Cp#	0	1	2	≥3
0	1	0	0	0
1	0	2	0	0
2	0	0	8	0
≥3	0	0	0	2

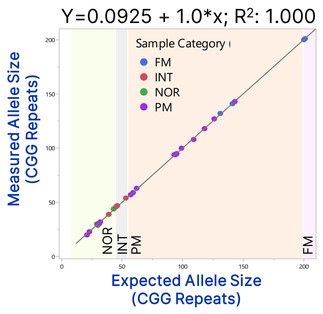
Figure 2. Mix A was 100% Concordant with SNV and Copy Number Across 24 *CFTR* and 13 *SMN1/2* Samples. A) Of 24 samples, 5/5 agreed with exon del/dup comparator data. Two samples did not pass QC for exon del/dup, 17 samples matched wild-type (WT) or had no comparator information. **B)** Example of a heterozygous (HET) F508del variant classified as Pathogenic/Likely Pathogenic (P/LP) by ClinVar. 17/17 samples with HET P/LP variants agreed with comparator data. One additional P/LP variant was identified in a sample without comparator information. Six WT samples matched or were not reported. **C)** 13/13 samples agreed with copy number (Cp#) reported in the comparator data for both *SMN1* and *SMN2*. Though not previously reported, two samples were identified as hybrids with an allele aligning to *SMN1* with yet c.840T from *SMN2*. One sample previously identified with silent carrier SNVs (c.*3+80T>G HET, and c.*211_*212del HET) matched the reported genotype.

FMR1 3' UTR CGG repeat sizing agreement was 100% for alleles previously identified by PCR/CE (23 samples; Mix B)

3A) *FMR1* Repeat Calling Examples



3B) *FMR1* Repeat Sizing Accuracy



3C) *FMR1* Sample-level Genotype Agreement

Expected

Measured Category	NOR	INT	PM	FM
NOR	4	0	0	0
INT	0	3	0	0
PM	0	0	10	0
FM	0	0	0	6

Figure 3. FMR1 was 100% Concordant with Repeat Sizing Reported by PCR/CE (Precision +/- 1 up to 70 CGG, +/- 3 from 71-120 CGG and +/-5% above 120 CGG) Across the Categorical Range (20-200 repeats) for 23 *FMR1* Samples. **A)** Sizing algorithms identify reads corresponding to CGG repeat sizes for allele identification. Visual mapping displays a histogram of read counts for each repeat size. Examples show 3 female samples, each with 1 normal (NOR) allele along with a categorical intermediate (INT), premutation (PM) or a full mutation (FM) allele. **B)** *FMR1* CGG repeat size correlation plot for 44 alleles colored by sample-level category (background). Four mosaic alleles (3 PM, 1 NOR) were not reported in comparator data and did not change genotyping. **C)** *FMR1* sample-level genotype agreement.

***F8* Intron 1 and 22 inversion sample-level agreement was 100% for all samples that passed QC (11/12 samples; Mix D)**

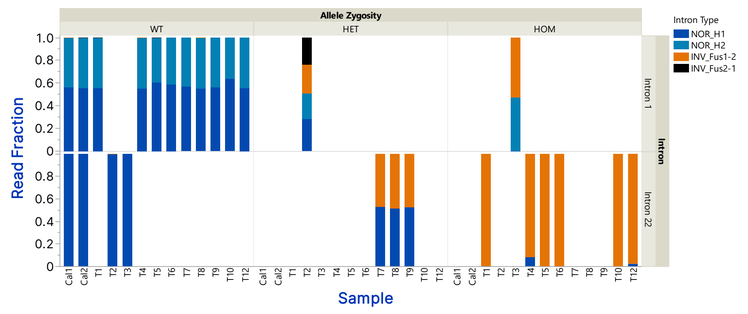
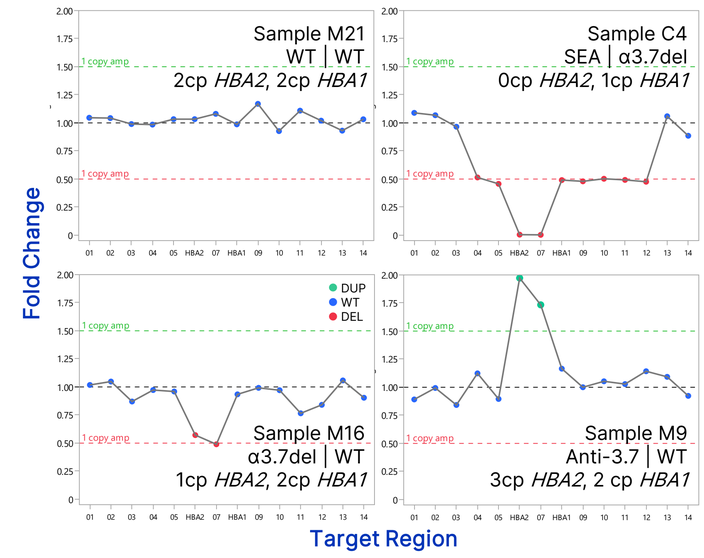


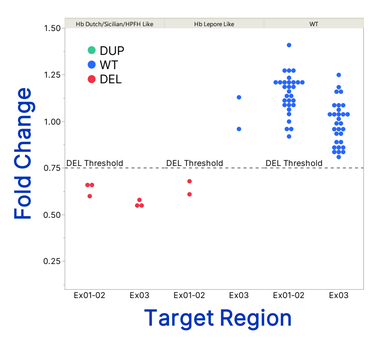
Figure 4. F8 Variant Genotype was 100% Concordant with Comparator Data for Intron Inversion and Zygosity. The algorithm identifies reads associated with the intronic H1 (intron 1 and 22) and H2 (intron 1 only) intergenic regions of the WT sample, or the H1 Fusion to H2 (intron 1) or H2/3 (intron 22) intergenic region indicating an inversion. Of 11 samples that passed QC for read depth at the variant level, 9 were WT for intron 1, 1 each had a heterozygous (HET) or homozygous (HOM) inversion. For intron 22, 2 samples were WT, 3 were HET and 6 were HOM for inversion. At the sample level, an inversion was identified in all *F8* samples. In the dataset, 4 *CYP21A2* and 5 *GBA* (not shown; comparator data unavailable) and 2 Calibrators were WT for both introns.

HBA1/2 & HBB variant level (SV & SNV) agreement was 100% for samples that passed QC (up to 33 samples; Mix C)

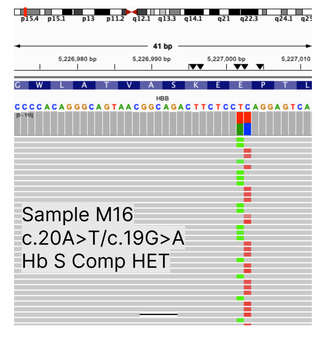
5A) *HBA1/2* SV Examples



5B) *HBB* SV Accuracy



5C) *HBB* SNV Example



5D) *HBA* and *HBB* Sample-Level Functional Copy Number Agreement

Expected

Measured *HBB* Cp#	Measured *HBA* Cp#	0: 2	0: 3	0: ≥4	1: 2	1: 3	1: ≥4	2: 1	2: 2	2: ≥4	Not Rep
0	2	2	0	0	0	0	0	0	0	0	0
	3	0	2	0	0	0	0	0	0	0	0
	≥4	0	0	5	0	0	0	0	0	0	0
1	2	0	0	0	2	0	0	0	0	0	1
	3	0	0	0	0	1	0	0	0	0	0
	≥4	0	0	0	0	0	10	0	0	0	2
2	1	0	0	0	0	0	0	1	0	0	0
	2	0	0	0	0	0	0	0	1	0	0
	≥4	0	0	0	0	0	0	0	0	4	0

Figure 5. Mix C was 100% Concordant with Functional Gene Copies of HBA and HBB Reported in Comparator Data. A) A set of 14 "sentinel" amplicons target regions of the hemoglobin alpha cluster to differentiate known common breakpoints. Amplicon fold change patterns identify WT, compound HET SEA | 3.7del, HET α3.7del | WT, and anti-3.7 | WT. *HBA1/2* SNV and SV variant level classification was 100% for samples that passed QC (n=31/33). **B)** Two amplicons cover the *HBB* gene and differentiate two classes of SVs: Sicilian/HPFH-like (full-gene deletion) from Hb Leopore-like (only Exon 1-2 deletion). **C)** An example of a compound HET *HBB* Hb S "sickle cell" SNV. Sample M16 has 0 functional allele copies of *HBB*, 1 copy of *HBA2* and 2 copies of *HBA1*. **D)** Functional allele copies were assessed for 28/31 samples and 100% agreed with comparator data in both *HBA1/2* and *HBB* or were WT but had no comparator information. We identified two SNVs and an Exon 01-02 del in *HBB* that were not reported in comparator data.

Conclusion

- The prototype PCR/nanopore assay accurately resolves multiple challenging variants across several variant classes for 11 of the most common gene targets associated with heritable disease. Eight genes with comparator data collected are represented.
- The assay utilizes a single-platform, streamlined workflow, and has potential to greatly reduce variant detection complexity, reflex testing, and turn around times compared to current workflows.
- Accompanying software simplifies data navigation, provides QC metrics, provides variant call information, and allows in-depth investigation of sequencing data and analysis results.
- Data for 105 of 155 samples are presented here. Over 95% (100/105) passed QC and 97/100 had comparator data or were WT with no comparator information. 97/97 (100%) agreed with expected sample-level genotypes.

Acknowledgements:
Feyereisen L[1], Yammine T[1], Landais E[1], Moret H[2], Audrezet M-P[3], Bonello N[4], Rouzier C[5], Chaussenot A[5], Pissard S[6], Gueguen P[7], Metay C[8], Froissart R[9], Bekri S[10], Bellanné-Chantelot C[11], Lebre AS[12]

[1]CHU Reims, Laboratoire de génétique, Reims, France
[2]CHU Reims, Plateforme PFBM, Reims, France
[3]CHU Brest, Laboratoire de Génétique Moléculaire, Brest, France
[4]APHM, Hôpital de la Timone, M2GM Labortaoire de génétique moléculaire, Marseille, France
[5]Service de génétique médicale, Centre de référence des maladies mitochondriales, CHU Nice, Université Côte d'Azur, CNRS, INSERM, IRCAN, Nice, France
[6]GHU Henri Mondor, Département de génétique, Créteil, France et Université Paris-Est Créteil (UPEC), Créteil, France
[7]Service de Génétique, CHRU de Tours - Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBraiN, Tours, France
[8]AP-HP Sorbonne Université, Hôpital Pitié-Salpêtrière, Centre de Génétique Moléculaire et Chromosomique, UF de Cardiogénétique et Myogénétique Moléculaire et Cellulaire, Paris, France
[9]CHU de Lyon HCL - GH Est, Service Biochimie et Biologie Moléculaire, Bron, France
[10]CHU Rouen Normandie, Laboratoire de Biochimie métabolique, Rouen, France et Université Rouen Normandie, UNIROUEN, INSERM Rouen, France
[11]AP-HP Sorbonne Université, Hôpital Pitié-Salpêtrière, Département de Génétique Médicale, Paris, France
[12]CHU Reims et Université Reims Champagne-Ardenne (URCA), Reims, France et Université Paris Cité, Institute of Psychiatry and Neuroscience of Paris (IPNP), INSERM, Paris, France.

References:

1. Xiao, Q. et al. Npj Genom Medicine 6, 41 (2021).
2. Woo, EG. et al. Front in Genet 12 (2021).
3. Seaby, EG. et al. Brief in Funct. Genom 19(4), 243–258 (2020).
4. Beauchamp, KA., et al. Genet Med 21 (2019).
5. Zheng, Z., et al. Nat Comp Sci 2, 797–803 (2022).
6. Smolka, M., et al. Nat Biotech, 1-10 (2024).
7. Taber, K. J. et al. Genet Med (2021).

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.

Authors affiliated with Asuragen have the financial relationship to disclose: Employment by Asuragen.

Presented at London Calling, May 2024.