A Single-Assay Solution for Expanded Carrier Screening Relieves Existing Workflow Constraints and Provides More Comprehensive Analysis

Sarah Statt, Jon Kemppainen, Melissa Church, Cody Edwards, Jonathan Turner, Justin Janovsky, Stela Filipovic-Sadic, Julie R Thibert, Bradley Hall, Liangjing Chen, Theodore Markulin, James Hernandez, Jacob Wisotsky, Brian C Haynes, and Gary J Latham
Asuragen, a Bio-Techne brand, Austin, TX

Poster Number: 2021-A-2261-ASHG

Summary

- High-prevalence carrier genes, associated with disorders such as Fragile X Syndrome and Hemophilia A, include GC-rich repeats, complex structural variants, and/or pseudogenes that derail conventional sequencing methods.
- We established a prototype streamlined single platform, multi-gene, sample-multiplexed AmplideX PCR/NP (Nanopore) sequencing assay* that presents a solution for carrier screening, with the potential to replace current fragmented workflows and platforms.
- We leveraged real-time data streaming and improvements in sequencing base quality on the Oxford Nanopore platform to accurately detect complex pathogenic variants, including large structural variants, Short Tandem Repeats (STRs), Insertions/Deletions (INDELs) and Copy Number Variants (CNVs). We demonstrate the utility of this approach across 24 cell line and whole blood samples.

Introduction

More than a million individuals have been tested using Expanded Carrier Screening (ECS) since the first commercial assay was launched in 2009. Conventional methods like short-read sequencing either miss or poorly cover many genes or gene regions; a recent study found that 20.4% of pathogenic/likely pathogenic variants were "technically challenging" by NGS[1]. Furthermore, the analysis of these high-prevalence carrier screening genes requires specialized methods (often ≥ 5 different, disjointed techniques) that burden laboratories with inefficient workflows while often producing subpar detection rates, limiting access to only high throughput sophisticated labs.

Materials and Methods

We sought to create a mid-sized panel that combines routine NGS targets with non-NGS targets in a single workflow that can enable broad access to ECS assays. To do this, we paired novel long-range PCR with long-read nanopore sequencing and algorithms that offer more comprehensive yet streamlined ECS analysis. Samples were initially evaluated using the most challenging genes from the mid-sized panel as a proof-of-concept, namely *FMR1*, *F8*, *CYP21A2*, *HBA1/2*, *GBA*, and *SMN1*. Genomic DNA was amplified, barcoded, pooled, prepped by ligation sequencing kit (ONT) and ran on R9.4.1 flow cells (ONT) using either the MinION or Mk1C. Data was basecalled and demultiplexed using Guppy v4.5.4. *FMR1* associated reads were processed through an in-house pipeline to resolve CGG genotypes and AGG status. *HBA1/2* final copy number status was assigned through a machine learning algorithm to distinguish noise from true copy number changes. Other gene targets underwent a bespoke sequence deconvolution method to resolve paralogous sequences, where applicable, and further analyzed for pathogenic variants.

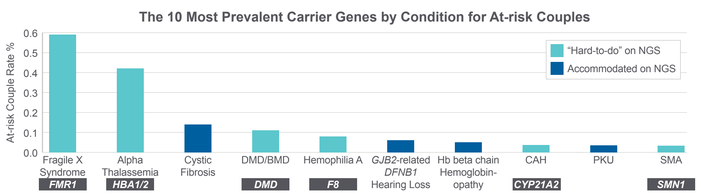

Figure 1. The 10 Most Prevalent Carrier Genes by Condition for At-risk Couples. Dark and camouflaged genes are especially problematic in Expanded Carrier Screening (ECS), where six (shown in teal) of the top ten carrier screening disorders by reproductive risk to couples[2-4] are difficult or intractable using conventional short-read sequencing methods.

**This product is under development. Future availability and performance to be determined.*
†Research use only. Not for use in diagnostic procedures
‡Xpansion Interpreter® is a CLIA Lab Developed Test. Analytical and clinical performance have not been reviewed by the FDA
Presented at ASHG 2021, October 18-22, virtual

Results

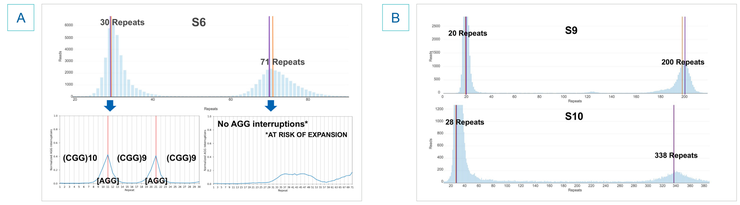

Figure 2. AmplideX PCR/NP (Nanopore) Prototype Assay Resolves *FMR1* CGG Repeats & AGG Interruptions That Inform the Risk of Expansion for Fragile X and Can Identify Full-Mutation Females. The number of AGG interruptions modify expansion risk for premutation (55-200 CGG) samples. As shown in the above **A)** plots, S6 had no AGG interruptions in the 71-CGG allele (high risk for further expansion). Additionally, our automated caller identified two full-mutation female samples, as seen in **B)**. Orange line = reference repeat number using the AmplideX® PCR/CE *FMR1* kit†; Purple line = Automated caller

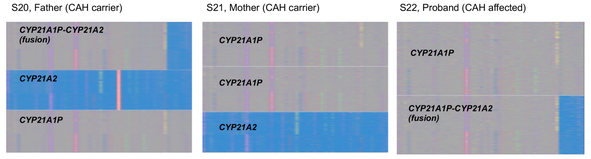

Figure 3. Accurate Resolution of Copy Number and Pseudogene Fusions in *CYP21A2* Gene Cluster. Two Congenital Adrenal Hyperplasia (CAH) carriers are shown above: a mother with a whole gene deletion and a father with a 30kb deletion that results in a fusion of pseudogene (*CYP21A1P*) and gene (*CYP21A2*). Analysis of the proband correctly identifies both pathogenic alleles from the parents. Pink line in the father represents a non-pathogenic SNP.

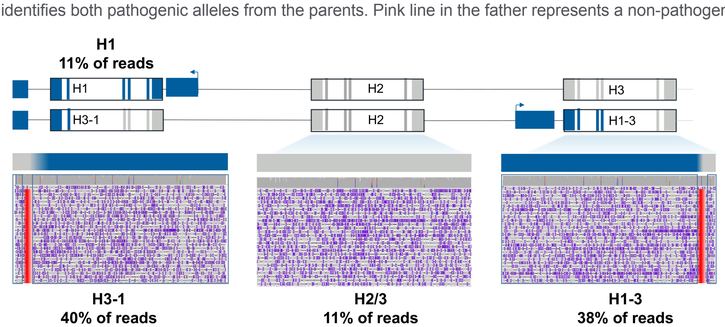

Figure 4. AmplideX PCR/NP Prototype Assay Detects Complex Inversions in a Hemophilia A Carrier. Utilizing sequence deconvolution by Paralog Specific Variants (PSVs), we identified a 0.6-Mb inversion in *F8* resulting in a fusion of Intron 22 H1 and H3. Red line identifies breakpoint between H1 and H3. Blue lines represent H1 specific PSVs while grey lines represent H2/H3 specific PSVs.

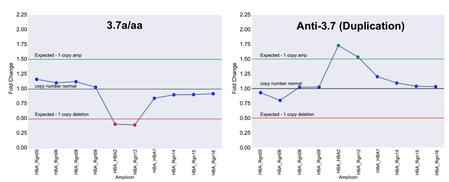

Figure 5. AmplideX PCR/NP Prototype Assay Detects Duplication and Silent Carrier Deletion Within the Alpha Hemoglobin Cluster (*HBA1/2*). Utilizing machine learning classifiers, copy number alternations were identified in regions within the alpha cluster, indicative of a deletion and duplication respectively. Red dots = loss of 1 copy; green dots = gain of 1 copy; blue dots = copy number neutral

Table 1. AmplideX PCR/NP Prototype Assay Results Compared with Reference Genotypes. As seen in **A)**, *FMR1* CGG repeats across all disease categories (intermediate/INT, premutation/PM and full-mutation/FM) were called within previously established AmplideX PCR/CE *FMR1* kit† precision, and AGG interruptions were identified in all samples in agreement with the reference Xpansion Interpreter® assay‡, noting § indicates a sample outside of the PCR/CE assay resolution. As seen in **A)** and **B)**, INDEL and/or SNV carriers and/or affected individuals of *GBA* and *CYP21A2* variants were detected consistent with reference calls. Complex *CYP21A2* gene conversions, pseudogene fusions, and *CYP21A2/CYP21A1P* gene copy number were called using as few as 100 reads per allele. *SMN1* analysis revealed copy numbers ranging from 0 to ≥4 along with gene conversions, and silent carrier-associated variants that were concordant with orthogonal AmplideX PCR/CE *SMN1/2* Plus Kit†. The method identified 0.6-Mb inversions in *F8* despite the challenges in resolving the highly homologous and GC-rich structural perturbation. Lastly, *HBA1/2* copy number was able to be resolved, identifying both a silent carrier and a duplication.

A

Sample #	Sample ID	Sex	Allele 1	Allele 2	Category	Match Reference?
S1	WB#1	F	*FMR1* CGG: 20 | *FMR1* AGG: 10+9	*FMR1* CGG: 46 | *FMR1* AGG: 9+9+26	Fragile X Carrier, INT	YES
S2	WB#2	F	*FMR1* CGG: 38 | *FMR1* AGG: 11+26	*FMR1* CGG: 52 | *FMR1* AGG: 9+9+32	Fragile X Carrier, INT	YES
S3	RU0011	F	*FMR1* CGG: 30 | *FMR1* AGG: 10+9+9	*FMR1* CGG: 56 | *FMR1* AGG: 9+46	Fragile X Carrier, INT/PM border	YES
S4	WB#3	F	*FMR1* CGG: 29 | *FMR1* AGG: 9+9+9	*FMR1* CGG: 68 | *FMR1* AGG: 9+58	Fragile X Carrier, PM	YES
S5	WB#4	F	*FMR1* CGG: 38 | *FMR1* AGG: 9+28	*FMR1* CGG: 70 | *FMR1* AGG: NO	Fragile X Carrier, PM High Risk	YES
S6	WB#5	F	*FMR1* CGG: 30 | *FMR1* AGG: 10+9+9	*FMR1* CGG: 71 | *FMR1* AGG: NO	Fragile X Carrier, PM High Risk	YES
S7	WB#6	F	*FMR1* CGG: 20 | *FMR1* AGG: 10+9	*FMR1* CGG: 80 | *FMR1* AGG: 9+70	Fragile X Carrier, PM	YES
S8	RU004	M	*FMR1* CGG: 84 | *FMR1* AGG: NO	N/A	Fragile X Carrier, PM	YES
S9	NA20239	F	*FMR1* CGG: 20 | *FMR1* AGG: 10+9	*FMR1* CGG: 198 | *FMR1* AGG: 9+9+178	Fragile X, PM/FM border	YES
S10	NA07537	F	*FMR1* CGG: 29 | *FMR1* AGG: 9+9+9	*FMR1* CGG: 321 | *FMR1* AGG: 10+310	Fragile X, FM	YES §
S11	NIBSC 06/204	F	*F8*: Normal	*F8*: Intron22 inversion	Hemophilia A Carrier	YES
S12	NIBSC 07/116	M	*F8*: Intron22 inversion	N/A	Hemophilia A Affected	YES
S13	ND14143	F	*GBA*: Normal	*GBA*: c.1226A>G	Gaucher disease Carrier	YES
S14	NA20270	F	*GBA*: c.115+G>A	*GBA*: c.1448T>C	Gaucher disease Affected	YES

B

Sample #	Sample ID	Sex	Observed SNV(s)/INDEL(s)	Copy Number	Category	Match Reference?
S15	NA03815	M	No Pathogenic Variants	*SMN1*: 1 copy | *SMN2*: 2 copy	Spinal Muscular Atrophy Carrier	YES
S16	NA11067	M	No Pathogenic Variants	*SMN1*: 1 copy | *SMN2*: 2 copy	Spinal Muscular Atrophy Carrier	YES
S17	NA19360	M	c.*3+80T>G, c.*211_*212del, *SMN2* gene conversion	*SMN1*: ≥ 4 copies | *SMN2*: 0 copies	Spinal Muscular Atrophy Carrier	YES
S18	NA20530	F	*CYP21A2* gene conversion (*CYP21A2* + *TNX-A*)	*CYP21A2*: 3 copies *CYP21A1P*: 2 copies	Normal	YES
S19	NA12217	M	*CYP21A2* c.515T>A (p.Ile172Asn)	*CYP21A2*: 1 copy *CYP21A1P*: 2 copies *CYP21A2*:*CYP21A1P*: 1 copy	Congenital Adrenal Hyperplasia Affected	YES
S20	NA14733	M	No Pathogenic Variants	*CYP21A2*: 1 copy *CYP21A1P*: 1 copy *CYP21A2*:*CYP21A1P*: 1 copy	Congenital Adrenal Hyperplasia Carrier	YES
S21	NA14732	F	No Pathogenic Variants	*CYP21A2*: 1 copy *CYP21A1P*: 2 copies	Congenital Adrenal Hyperplasia Carrier	YES
S22	NA14734	M	No Pathogenic Variants	*CYP21A2*: 0 copies *CYP21A1P*: 1 copy *CYP21A2*-*CYP21A1P*: 1 copy	Congenital Adrenal Hyperplasia Affected	YES
S23	WB#7	F	No Pathogenic Variants	*HBA2*: 1 copy | *HBA1*: 2 copies	Alpha Thalassemia Silent Carrier	N/A
S24	WB#8	F	No Pathogenic Variants	*HBA2*: 3 copies | *HBA1*: 2 copies	Duplication	N/A

Conclusions

- The data demonstrate feasibility for a single-platform, multiplexed panel workflow that can accurately resolve different classes of challenging variants, and scale to include more conventional carrier genes.
- Detected pathogenic variants included >300 CGG repeats and AGG interruptions in *FMR1*, 0.6Mb intron 22 inversions in *F8*, large deletions and chimeric fusions in *CYP21A2*, and copy number changes in *SMN1*, *HBA1/2*, and *CYP21A2*.
- By combining nanopore sequencing, tailored PCR reagents, and automated software, we demonstrate proof-of-concept for a carrier screening methodology that may be used in diverse laboratory settings.

References

1. Lincoln, S., et al, One in seven pathogenic variants can be challenging to detect by NGS: an analysis of 450,000 patients with implications for clinical sensitivity and genetic test implementation. Genet Med. 2021 May 18.
2. Balzotti, M., et al., Clinical validity of expanded carrier screening: Evaluating the gene-disease relationship in more than 200 conditions. Hum Mutat, 2020. 41(8): p. 1365-1371.
3. Ben-Shachar, R., et al., A data-driven evaluation of the size and content of expanded carrier screening panels. Genet Med, 2019. 21(9): p. 1931-1939.
4. Capalbo, A., et al., Optimizing clinical exome design and parallel gene-testing for recessive genetic conditions in preconception carrier screening: Translational research genomic data from 14,125 exomes. PLoS Genet, 2019. 15(10): p. e1008409.

Asuragen
a bio-techne brand