Democratizing Carrier Screening: A 35-Gene Panel of Routine and Challenging Targets Resolved Using a Single Workflow

James Hernandez, Bradley Hall, Jon Kemppainen, Julie Thibert, Cody Edwards, Liangjing Chen, Ryan Routsong, Chris Fraher, Theodore Markulin, Melissa Church, Jacob Wisotsky, Jonathan Turner, Justin Janovsky, Stela Filipovic-Sadic, Gary J Latham, Brian C Haynes and Sarah Statt

Asuragen, a Bio-Techne Brand, Austin, TX

Poster Number: eP361

Summary

- The 2021 ACMG practice resource on screening for autosomal recessive and X-linked conditions recommends equitable carrier screening using an expanded set of genes through which diverse populations can benefit from new technologies.
- There are significant technical hurdles for laboratories to accommodate the many challenging genes and variants that require non-NGS workflows.
- Here we demonstrate a single accessible workflow using long-read sequencing sequencing that can genotype variants in technically challenging genes such as *FMR1, F8, HBA1/2, SMN1*, and *CYP21A2* along with 360 loci from 29 autosomal recessive, and X-linked genes.

Introduction

The American College of Medical Genetics (ACMG) recently replaced its preconception and prenatal expanded carrier screening position statement[1]. The new practice resource considers the impact of recent sequencing innovations and, critically, emphasizes the need for equity and social justice in screening. To achieve diverse and inclusive clinical coverage, this resource recommends offering "Tier 3" screening for all pregnancies, which includes 113 genes comprising 97 autosomal recessive genes with a frequency of at least 1/200 along with 16 X-linked disease genes. Conventional methods, like short-read sequencing (SRS), fail to adequately cover many of these genes due to reasons including extreme GC content or homology to other genomic regions. In fact, 6 of the top 10 genes contributing to at-risk couples in this setting are difficult, if not intractable, by SRS, and more broadly speaking, 20.4% of pathogenic/likely pathogenic variants in ClinVar have been reported to be "technically challenging"[2]. Consequently, analysis of problematic, high-prevalence carrier screening genes requires multiple bespoke non-SRS methods and highly trained personnel. Access to such methods is further exacerbated by the substantial financial commitment required for many SRS technologies. This leads to inefficient workflows and subpar detection rates and often limits testing to sophisticated high throughput centralized labs in a send-out model.

To help address these challenges, we created a prototype for a distributable assay solution that incorporates PCR enrichment for a 35-gene, 32 condition carrier screening panel, long-read sequencing (LRS) on Oxford Nanopore Technologies' (ONT) Mk1B or Mk1C, and a custom software application to manage and analyze data. This panel includes content aligned with criteria recommended in ACOG committee opinion number 690 as well as many challenging Tier 3 genes recommended in the latest ACMG carrier screening practice resource. The results indicate that this prototype has potential to enable analysis of the full spectrum of Tier 3 variants using a single, accessible workflow. This would allow labs to expand their carrier screening offering in a measured and manageable fashion.

Materials and Methods

More than 200 cell-line, residual whole blood, and contrived samples were evaluated using the 35-gene panel, including loci that are routinely genotyped by SRS and challenging genes such as *CYP21A2, FMR1, F8, GBA, HBA1, HBA2,* and *SMN1* often associated with non-SRS workflows. Samples were split across 4 PCR tubes for target-specific enrichment using novel long-range amplification reagents based on AmplideX® PCR chemistry (Asuragen). Genomic DNA was amplified, barcoded, pooled, prepped by ligation sequencing kit (ONT) and run on R9.4.1 flow cells (ONT) using the Mk1B or Mk1C. Data analysis was managed and initialized using Asuragen's software platform and proprietary bioinformatics pipeline.

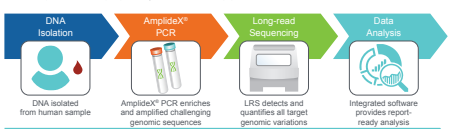


Figure 1. Prototype Workflow using Long-range PCR, LRS and Bioinformatic Analysis. DNA isolates are enriched using optimized PCR panels and input into ONT's Mk1B or Mk1C LRS platform. Upon sequencing completion, data are analyzed using a proprietary bioinformatics pipeline. Variants are identified using off the shelf and bespoke algorithms for repeat analysis, allele deconvolution, sequence mutations, and copy number calculations.

*Proof-of-concept data only. Future availability and performance cannot be ensured.
Presented at ACMG 2022

Results

We were able to detect multiple classes of pathogenic variation with >90% accuracy per class across >200 cell-lines, residual whole blood, and contrived samples. These variations include short tandem repeats (STRs), single nucleotide variants (SNVs), insertions/deletions (INDELs) and copy number variation (CNV) and structural variants (such as inversions). We also achieved target read balance (>97% amplicons within 5x median coverage) using a high-density multiplex PCR amplification across 360 amplicons.

		Predicted			
		Normal	Intermediate	Premutation	Full Mutation
Expected	Full Mutation	0	0	0	12
	Premutation	0	0	48	0
	Intermediate	2	17	0	0
	Normal	36	0	0	0

Table 1. Bioinformatic Algorithms Accurately Assigned FMR1 Genotype Category. *FMR1* categorical labels were assigned with >98% accuracy. Results were generated across two sequencing runs with a total of 120 samples. CGG sizing precision bounds based on ACMG recommendations[3] were considered in categorizing samples.

A

Gene	Expected Pathogenic Variant Count: SNV	INDEL	Total
ASPA	2	0	2
CFTR	11	1	12
GALT	3	0	3
GBA	15	1	16
PAH	2	0	2
SMN1	13	13	26
SMPD1	2	1	3
Total	48	16	64

B

Variants Detected	Variants Missed
58	6

Table 2. Pathogenic SNV and INDEL Variants of Interest Across 7 Genes Were Detected with 91% Sensitivity. A) 64 variants (48 SNVs and 16 INDELs) were evaluated across 7 genes. **B)** Pathogenic SNV and INDEL variants were identified with 91% sensitivity across 30 samples. The 6 missed variants consist of: 3 (all SNVs) *CFTR*, 1 (SNV) *GALT*, and 2 (1 SNV, 1 INDEL) *SMN1*.

A

SMN1 Performance		Predicted Copy: 0	1	2	3	>=4
Expected Copy	>=4	0	0	0	0	9
	3	0	0	0	7	2
	2	0	0	38	1	0
	1	0	7	0	0	0
	0	3	0	0	0	0

B

SMN2 Performance		Predicted Copy: 0	1	2	3	>=4
Expected Copy	>=4	0	0	0	0	2
	3	0	0	0	4	0
	2	0	1	28	0	0
	1	0	18	0	0	0
	0	14	0	0	0	0

Table 3. Prototype PCR/NP Assay and Bioinformatic Algorithms Detected SMN1 and SMN2 Copy Number Alterations. *SMN1* **A)** and *SMN2* **B)** copy number was identified with 97% accuracy across 67 samples. Expected copies and sequence variants were determined using Asuragen's AmplideX PCR/CE *SMN1/2* Plus Kit.

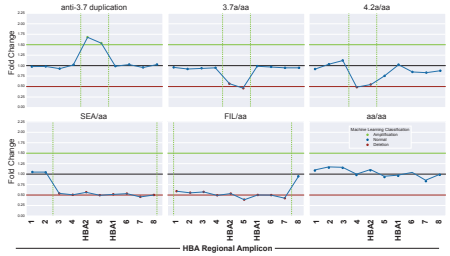
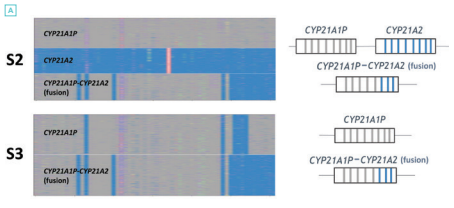


Figure 2. Copy Number Variation Within Alpha Hemoglobin Cluster (*HBA1/2*) Were Detected and Categorized Based on the Copy Number Change Profile. Copy number alterations were accurately identified for a variety of amplicons throughout the alpha hemoglobin cluster region. Based on the regions that have detected copy number changes, a deletion/amplification type can be assigned (e.g. SEA, FIL, anti-3.7) and are depicted between the green vertical lines. Horizontal lines represent expected fold changes for single copy gain (green) or loss (red). Individual regions were determined to contain copy number alteration events using a machine learning classifier to differentiate true copy number changes from noise.



Sample	Copy Number (Expected | Predicted): CYP21A2		CYP21A1P		CYP21A2-CYP21A1P	
S1	1	1	2	2	0	0
S2	1	1	1	1	1	1
S3	0	0	1	1	1	2
S4	3	3	2	2	0	0

Figure 3. Copy Number Variations and Fusions for CYP21A2 and CYP21A1P Were Accurately Calculated and Identified. A) Copy number values were determined using paralog-specific variants (PSVs) for read clustering and proportional read counts. Images show samples S2 and S3 read clustering and annotation using PSVs along with graphical representation of the expected alleles. Blue regions represent *CYP21A2* and grey region represent *CYP21AP* **B)** Using the method described above, four samples with known copy numbers and fusions were assessed using. Fusions were correctly identified for 2/2 samples and copy number values were accurately calculated for 11/12. The one over-estimation of fusion copy number was a borderline 1 or 2 copy call.

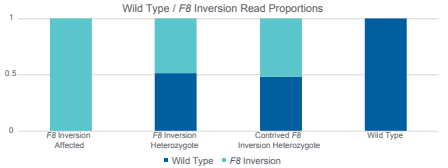


Figure 4. Long-amplicon PCR/NP Prototype Assay Amplified and Identified >100 kb *F8* Inversions. *F8* inversions were successfully amplified and aligned across 3 samples (1 affected intron 22 inversion, 1 intron 22 heterozygote, and 1 contrived intron 1 heterozygote). The results show the proportion of *F8* reads that aligned to either the reference genome (wild type) or *F8* inversion genome. The wild type sample showed no inversion reads.

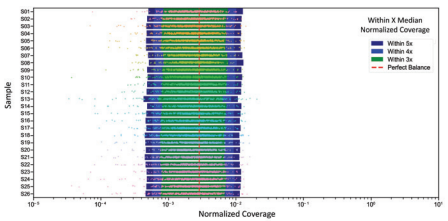


Figure 5. A Single-tube PCR Achieved Targeted Coverage Metrics. 26 samples were assessed using a single-tube PCR targeting 29 genes across 360 amplicons. For all samples, over 97% of those amplicons within 5x of the median normalized coverage. Normalized coverage values represent (amplicon fully spanning reads/total fully spanning read count). Perfect balance line represents where reads are even distributed across all amplicons.

Conclusions

- *FMR1* categorical accuracy was >98% across 120 samples including 60 with premutation or full-mutation repeat expansions.
- Structural alterations and copy number alterations were accurately identified for multiple "hard-to-sequence" genes such as *SMN1/2, HBA1/2, CYP21A2* and *F8*.
- The accuracy in detecting pathogenic SNVs, indels, copy number changes and fusions was 95%; accuracy for calling all pathogenic variants, including *FMR1* CGG repeats, was 96%.
- High-density multiplex PCR provided uniform sequencing coverage for 29 genes and 360 amplicons in a single tube.
- This prototype assay system demonstrates an accessible, single-platform workflow for both challenging and conventional genes with potential to support equitable, decentralized carrier screening of individuals with diverse ancestries.

References

1. Gregg AR, et al. *Genet Med* 2021;23:1793-1806
2. Lincoln SE, et al. *Genet Med* 2021;23:1673-1680
3. Spector E, et al. *Genet Med* . 2021;23(5):799-812

Asuragen
a bio-techne brand