A Comprehensive PCR/Nanopore Based Panel for *CFTR* Population Carrier Screening that Resolves Multiple Classes of Clinically-Relevant Variants

Monica P. Roberts, Ryan Routsong, Pranesh Rao, Christopher J. Fraher, Bradley C. Hall, Gary J. Latham, Brian C. Haynes, and Bryan J. Killinger

Asuragen, a Bio-Techne Brand, Austin, TX

SUMMARY

- ***CFTR*** **gene sequencing can provide a comprehensive analysis of CF-causing variants across ancestries and help address test inequities for carrier screening and molecular diagnostics.**
- **Due to recent improvements to the Oxford Nanopore MinION platform that have resulted in increased sequencing accuracy, we set out to explore the feasibility of the ONT platform to resolve a broad range of *CFTR* pathogenic variants.**
- **We developed a prototype assay comprised of novel PCR enrichment, nanopore sequencing, and bioinformatic variant analysis to simultaneously detect SNVs, indels, large exon deletions, and polyT/TG sizes in the *CFTR* gene.**
- **Assay performance was evaluated with a cohort of 62 cell-line and 231 whole blood samples, resulting in 100% accuracy for SNVs, indels, and large exon deletions and >93% agreement for polyT/TG sizing compared to orthogonal data.**

INTRODUCTION

Next Generation Sequencing (NGS) is routinely used to screen for carriers of inherited diseases. However, the short reads produced by NGS do not allow for the detection and phasing of some complex variants that can be critical in a carrier screening setting. Long-read sequencing (LRS) can produce reads that span much larger regions of the genome, providing the possibility of identifying and phasing variants that are difficult to detect, if not intractable, by NGS. Here we evaluated a versatile combination of PCR enrichment and nanopore LRS to assess both conventional and challenging pathogenic variants in the *CFTR* gene with the potential to readily expand the repertoire of reported variants relevant to Cystic Fibrosis (CF) carrier screening. We applied state-of-the-art bioinformatics methods to identify CF carriers from sequencing data obtained with the latest Oxford Nanopore Technologies (ONT) MinION kits and flow cells to assess performance across common and rare pathogenic *CFTR* variants, including variants adjacent to homopolymers, polyT/TG repeats, and large exon deletions.

METHODS

Cell-line genomic DNA (gDNA) samples (N=62) were obtained from the Coriell Cell Repository that represented 57 unique pathogenic variants, including SNVs, indels, and large exon deletions. Additionally, gDNA was isolated from whole blood of presumed non-carrier donors (N=231). DNA was amplified, barcoded, pooled, prepped by ligation sequencing kit SQK-LSK114 (ONT), and sequenced on R10.4.1 flow cells using ONT's Mk1B platform. Multiple basecalling models available from ONT and variant callers were evaluated for optimal SNV and indel detection. Bespoke algorithms were developed to detect large exon deletions. The AmplideX® PCR/CE *CFTR* Kit†, MLPA, Coriell or 1000 Genomes reporting, and Sanger sequencing were used as comparator methods.

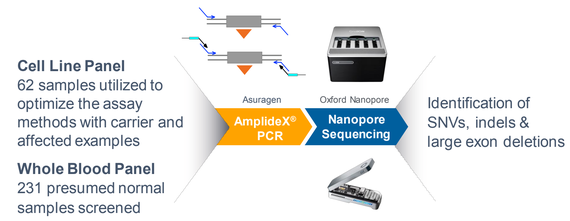


Figure 1. PCR/nanopore *CFTR* carrier screening panel design and workflow identifies multiple classes of clinically-relevant variants in a single workflow. The combination of AmplideX® PCR technology and nanopore sequencing enables detection of pathogenic variants (SNVs, indels, and large exon deletions) across all *CFTR* exons.

RESULTS

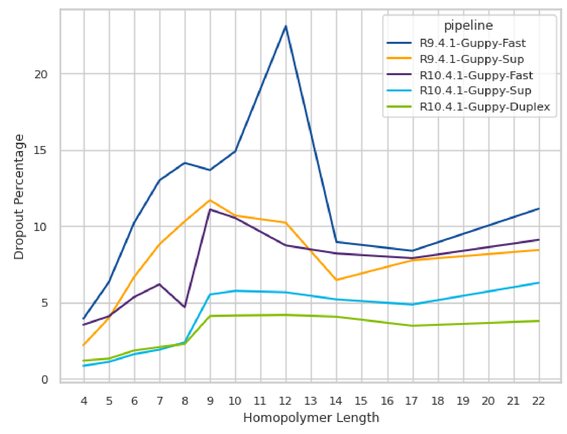


Figure 2. The latest R10.4.1 flow cell with super accurate basecalling can resolve homopolymers with the lowest dropout percentage. Several Guppy basecaller configurations were tested to determine the lowest homopolymer dropout, calculated as the average percent of bases missing of total expected for a given repeat length. Super accurate (Sup) outperformed Fast and Duplex for homopolymers between 4-8bp long. Duplex outperformed Sup and Fast for homopolymers >8bp. Given the majority of homopolymers in *CFTR* are <8bp long, the Sup configuration was used for subsequent analysis. All basecaller configurations showed performance improvement between the R9.4.1 and R10.4.1 flow cells.

Table 1. Deep learning-based LRS variant callers provided 100% sensitivity and PPV for pathogenic SNV and indel identification. Clair3[1] and Pepper-Margin-DeepVariant[2] (PMDV) were utilized to identify known pathogenic variants. Due to higher sensitivity and PPV, Clair3 was chosen for subsequent analysis.

Cell-line Data	Clair3		PMDV	
	Sensitivity	PPV	Sensitivity	PPV
All Variants (N=54)	100%	100%	90.4%	97.4%
Homopolymer-adjacent Variants (N=8)	100%	100%	100%	100%
Non-homopolymer Variants (N=46)	100%	100%	89.3%	97.1%

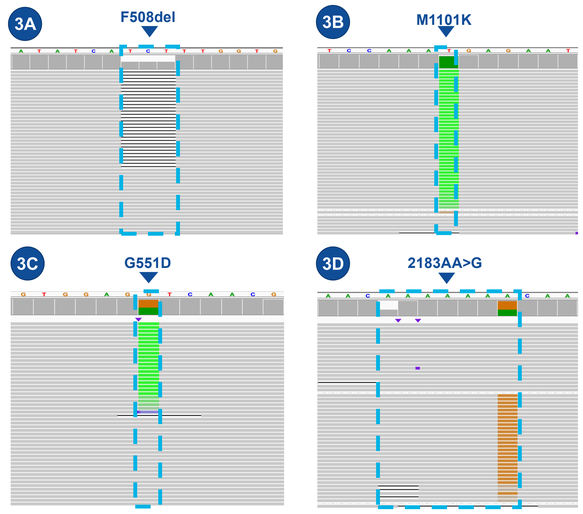


Figure 3. Sequence data reveals diverse pathogenic variants across the *CFTR* gene. A) Heterozygous F508del, a 3-bp deletion resulting in the deletion of phenylalanine at codon 508. **B)** Homozygous missense variant M1101K, a T>A transversion resulting in a protein change from methionine to lysine. **C)** Heterozygous missense variant G551D, a G>A transition resulting in a protein change from glycine to aspartic acid. **D)** Heterozygous 2183AA>G resulting in a frameshift from a two-bp deletion of AA and insertion of G in a homopolymer region of 7 adenosines.

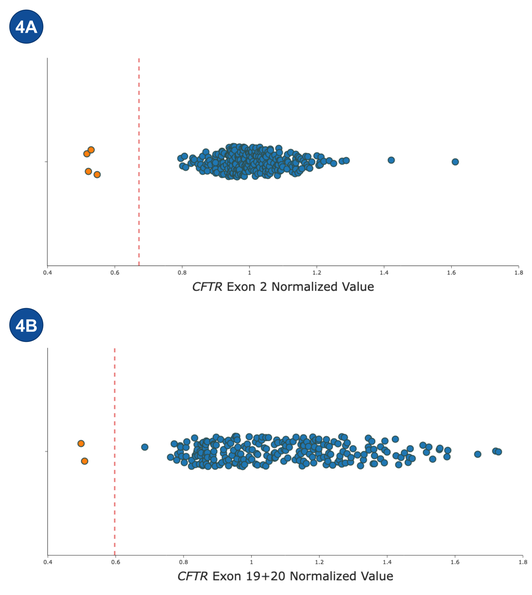


Figure 4. Large exon deletions detected at 100% PPV and sensitivity. Normalized amplicon coverage distinguishes exon deletion (orange) from wildtype genotypes (blue) for **A)** CFTR2,3dele and **B)** CFTRdele20. Samples with normalized coverage below the threshold (vertical dashed line) were classified as heterozygous for the large exon deletions.

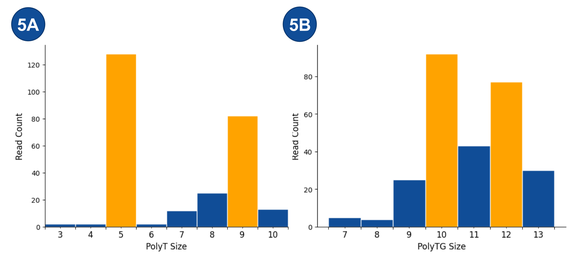


Figure 5. Read coverage identifies allele-specific polyT/TG sizes. A) Elevated read counts corresponding to two distinct polyT alleles (orange) identify the expected 5T/9T genotype. **B)** Read counts corresponding to the polyTG region identify two alleles (orange) correctly matching the expected 10TG/12TG genotype.

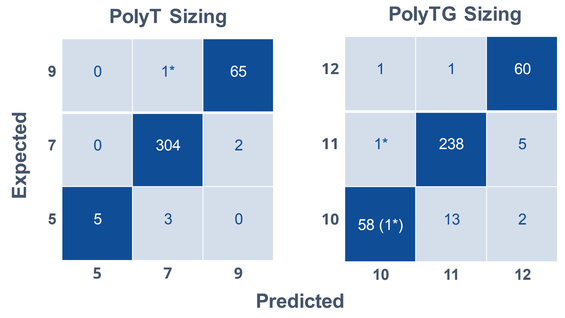
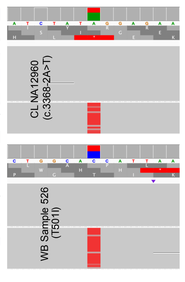


Figure 6. Greater than 93% accuracy in polyT/TG allele sizing was observed compared to reference genotypes. PolyT/TG sizing, upstream of Exon 10, couldn't be performed on 103/293 total samples due to a common SNP in the PCR/ONT primer region resulting in allele-specific loss of reads. The polyT sizing calls ($N_{samples}$=190) agreed with the comparator AmplideX® *CFTR* PCR/CE Kit for 374/380 (98.4%) alleles. The adjacent polyTG region agreed with comparator for 356/380 (93.7%) alleles. The affected primer will be re-designed in future iterations. *Indicates a discordant result with a measured or expected genotype outside of the polyT or polyTG range shown in the corresponding confusion matrices.

Cell-line Samples		Orthogonal Data			Overall
Sample level accuracy	Allele1/ Allele2	WT/WT	MUT/WT	MUT/MUT	Genotype agreement
CFTR PCR/ nanopore	WT/WT	5	0	0	5/5 (100%)
	MUT/WT	0	29	0	29/29 (100%)
	MUT/MUT	0	0	28¥	28/28 (100%)

Whole Blood Samples		Orthogonal Data			Overall
Sample level accuracy	Allele1/ Allele2	WT/WT	MUT/WT	MUT/MUT	Genotype agreement
CFTR PCR/ nanopore	WT/WT	221	0	0	221/221 (100%)
	MUT/WT	0	10¥	0	10/10 (100%)
	MUT/MUT	0	0	0	0/0 (100%)

Table 2. *CFTR* pathogenic variant calls agreed with orthogonal data for 100% of calls. Pathogenic variant calls from PCR/nanopore data of cell-line (top; N=62) and presumed normal whole blood samples (bottom; N=231) using Clair3 for SNV/indels or coverage differences for exon deletion identification. The assay detected 57/57 (100%) unique pathogenic variants. Overall, 293/293 (100%) samples agreed with orthogonal data. ¥Sanger sequencing verification pending for two variants (CL, c.3368-2A>T; WB, T501I), with corresponding nanopore read pile-ups shown on the right.



CONCLUSION

- **The prototype *CFTR* PCR/nanopore assay accurately resolves multiple challenging variants for one of the most common carrier screening genes using a streamlined workflow.**
- **Recently released ONT Kits (LSK-14 & R10.4.1) and software improved sequencing accuracy of challenging regions, including homopolymers.**
- **Both PCR/nanopore and proprietary algorithms provide a comprehensive and scalable carrier screening option, particularly as professional guidelines evolve to recommend increasing variants that address diverse ancestries and populations.**

REFERENCES

1. Zheng, Z., Li, S., Su, J. *et al.* Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci* **2**, 797–803 (2022).
2. Shafin, K., Pesout, T., Chang, PC. *et al.* Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods* **18**, 1322–1332 (2021).

This product is under development. Future availability and performance to be determined.
† For Research Use Only. Not for Use in Diagnostic Procedures. All authors have the financial relationship to disclose: Employment by Asuragen.
Presented at ACMG 2023

Asuragen® a bio-techne® brand