# A PCR/Nanopore Assay for Accurate Variant Detection in Hard-to-Decipher Carrier Screening Genes with Automated Analysis Software

P620

Bryan J. Killinger, Monica P. Roberts, Theodore J. Markulin, Mia K. Mihailovic, Eduardo Priego, Connor A. Parker, Jonathan Turner, Mahesh Yarasi, Cody W. Edwards, Christopher J. Fraher, Jon N. Kemppainen, Brennan T. Greenlee, Juliette A. Baker, Jacob D. Ashton, Brian C. Haynes, and Bradley Hall

Asuragen, a Bio-Techne Brand, Austin, TX

## Summary

- Traditional carrier screening methods struggle with detecting complex variants in key carrier screening genes, necessitating complicated workflows and producing results that often lack comprehensive detection of pathogenic variants.
- We developed a PCR/Nanopore sequencing assay and accompanying analysis software that detects both simple and complex variants across *CFTR*, *SMN1/2*, *FMR1*, *HBA1/2*, *HBB*, *GBA*, *CYP21A2*, *TNXB*, and *F8* intron inversions with a streamlined end-to-end workflow.
- Analysis performance was optimized and validated on a mixed sample set consisting of 371 cell-line samples and 104 gDNA samples isolated from whole blood of presumed normal donors.

## Introduction

Carrier screening identifies couples at risk for having a child with a severe genetic disorder. Although NGS is a widely used method for carrier screening, it fails to resolve many genes recommended in professional practice guidelines that have repeats, structural variation, and pseudogenes[1,2]. This traditionally necessitates the use of multiple specialized workflows that only cover a fraction of carrier genotypes[3].

To address this shortcoming for researchers, we combined short- and long-range PCR enrichment with nanopore sequencing in a kit-based modular carrier screening panel to accurately genotype traditional NGS addressable genes (*CFTR* and *HBB*) and hard-to-decipher genes (*SMN1/2*, *FMR1*, *HBA1/2*, *GBA*, *CYP21A2*, *TNXB*, and *F8* intron inversions) in a single streamlined workflow. We developed bespoke software and algorithms to automate analysis of SNVs, indels, copy number gain or loss, gene-pseudogene fusions, and large structural variants. Here we describe results using this prototype assay to genotype 475 total samples (260 unique) across multiple variant classes. We also present the accompanying software interface that allows users to easily navigate sequencing results, including quality control metrics, variant detection, and signal processing visuals that inform the user with context for different classes of variant calls.

## Materials and Methods

The assay was optimized and evaluated using cell-line genomic DNA (gDNA) samples (n=371 total/187 unique) and gDNA isolated from presumed normal blood and residual clinical samples (n=104 total/73 unique). The selection of cell-line and whole blood samples represent all major variant classes and were used to train and/or evaluate analysis performance. Target regions were enriched and barcoded in four PCR reactions, pooled across samples, and sequenced on MinION flow cells (R10.4.1) with a Mk1B or GridION (Oxford Nanopore Technologies). Software was developed to automate analysis and reporting. Multiple orthogonal methods provided comparator data for determining assay performance.

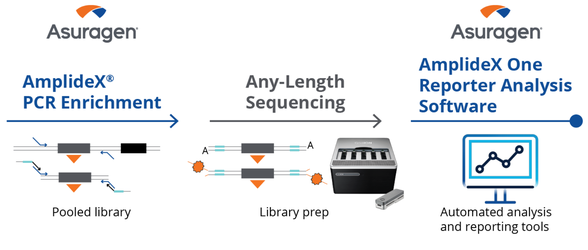


**Figure 1. Summary overview of PCR enrichment, sequencing, and automated analysis software.** Targeted PCR enriches panel-specific genetic regions. The amplicons (~300bp – 10kb in length) are PCR barcoded by sample, pooled, and sequenced on an ONT device. Raw sequencing data is analyzed with automated software to produce quality control metrics, annotated variant calls, and signal processing visuals.

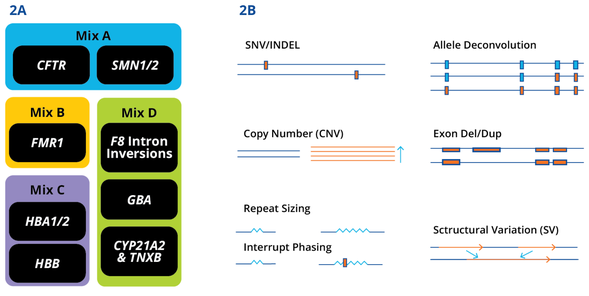


**Figure 2. Gene content and variant types automatically analyzed. (A)** Four PCR mixes are used to amplify targets across 11 genes. **(B)** Multiple variant classes are identified and reported simultaneously by the analysis software.

## Results

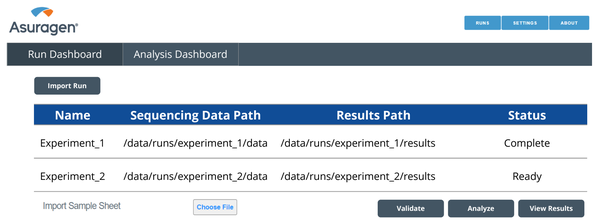


**Figure 3. A simple software interface allows for analysis and review of sequencing data.** After sequencing data is basecalled and demultiplexed, users can initiate software analysis by importing their experiment and uploading a sample sheet. Results can be viewed by selecting the analysis when the status is complete.

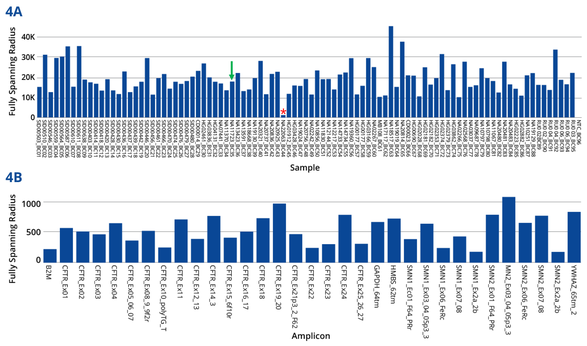


**Figure 4. Coverage reports provide users with quality control information.** Read depth information is used by the software to automatically identify quality control issues and ensure sufficient read depth to make variant calls per variant type. **(A)** A bar chart shows the total number of fully spanning reads across each Mix A sample to easily identify samples with QC issues. In this experiment, NA22063_BC44 (marked with *) may require a re-run due to low coverage. **(B)** A sample-specific amplicon coverage bar chart for Mix A provides users the ability to further investigate sequencing results (sample NA11723 shown and marked with ↓ in panel A).

| Sample | Barcode | Filtered Variants | Mixes Analyzed | QC |
|---|---|---|---|---|
| NA11723 | BC35 | *CFTR*: c.3846G>A (pathogenic) | A B C D | All pass |
| NA18668 | BC38 | *CYP21A2*: c.923dup (pathogenic)<br>*CYP21A2*: c.955C>T (pathogenic)<br>*CYP21A2*: c.1069C>T (pathogenic)<br>*CFTR*: c.1521_1523del (pathogenic) | A B C D | All pass |
| HG03428 | BC46 | *HBB*: c.20A>T (pathogenic)<br>*HBB*: c.19G>A (pathogenic) | A B C D | All pass |
| NA23255 | BC57 | *SMN1*: Copy number loss (pathogenic) (2x) | A B C D | All pass |

**Figure 5. Customizable high-level analysis results allow users to quickly identify desired variants.** A simple interface allows users to filter samples in a variety of ways, including ClinVar pathogenicity annotations. Color-coded PCR mixes and summarized quality control information allows users to quickly identify pertinent issues related to an analysis. The presence of all green PCR "Mixes Analyzed" and "All pass" QC entries indicate all analyses across all mixes were analyzed without QC flags raised for any of these samples.

| Sample | Barcode | Gene | Variant | Annotation | ClinVar ID | Transcript(s) | Consequence |
|---|---|---|---|---|---|---|---|
| NA18668 | BC38 | *CYP21A2* | c.923dup | Pathogenic | 65611 | NM_000500.9 | FRAMESHIFT |
| NA18668 | BC38 | *CYP21A2* | c.955C>T | Pathogenic | 12169 | NM_000500.9 | NONSENSE |
| NA18668 | BC38 | *CYP21A2* | c.1069C>T | Pathogenic | 12152 | NM_000500.9 | MISSENSE |
| NA18668 | BC38 | *CFTR* | c.1521_1523del | Pathogenic | 7105 | NM_000492.4 | INFRAME DELETION |

**Figure 6. Detected variants are reported with relevant annotations.** Variants can be viewed across the cohort in detail with ClinVar annotations, ClinVar IDs, transcripts, and predicted variant effects. Sample NA18668 is shown here to have multiple pathogenic variants that can be further investigated within the software.

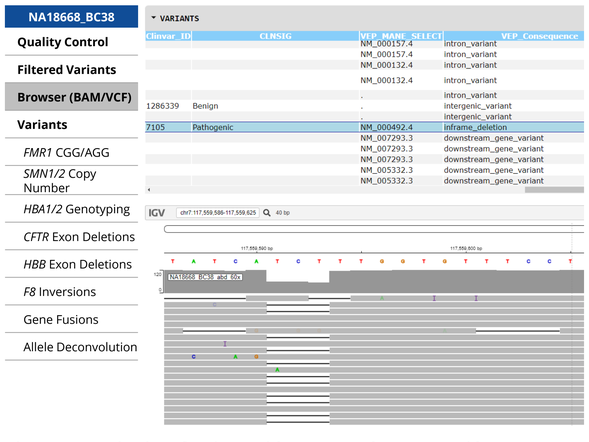


**Figure 7. Investigation of variants with a BAM/VCF browser provides sequencing context for variant calls.** Users can click a variant and review the aligned sequencing data at the relevant genomic location within a BAM/VCF browser. The *CFTR* c.1521_1523del variant in sample NA18668 is shown here with a clear signal of a heterozygous 3-base pair deletion in the .bam pileup. Users can further investigate other analysis results from the menu on the left side of the screen.

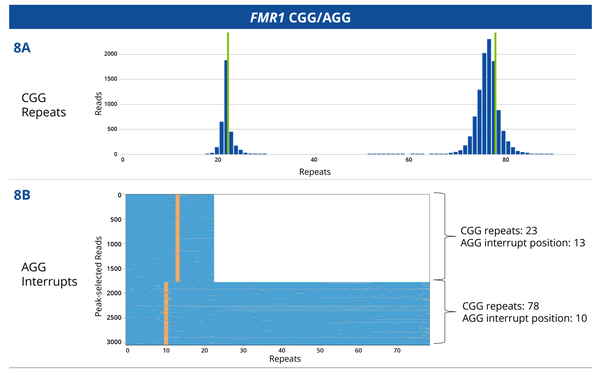


**Figure 8. Signal processing visuals allow users to thoroughly investigate *FMR1* CGG repeats and AGG interrupts. (A)** The "CGG Repeats" panel displays a histogram of read counts for each repeat size, which is used by the algorithm to determine the number of CGG repeats in each *FMR1* allele (green line) and are identified in this sample (NA06905) at 23 and 78 repeats. **(B)** Phased "AGG Interrupts" are graphically displayed to allow for investigation of interrupt patterns (orange) within each allele. AGG interrupts were detected at the 13th and 10th repeat positions for each respective allele and indicate the risk of *FMR1* repeat expansion.

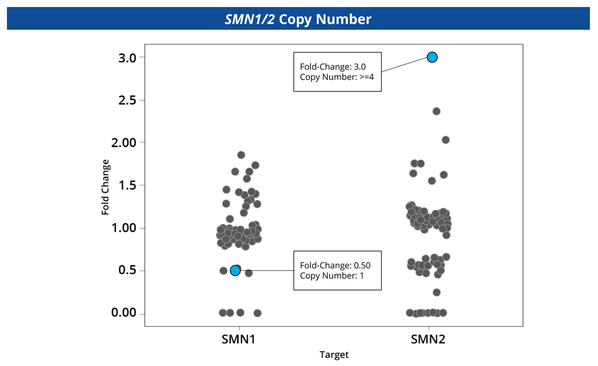


**Figure 9. *SMN1/2* fold-change results indicate copy number variants.** Users can investigate the fold-change of *SMN1* and *SMN2* copies relative to two-copy "wildtype" calibrators for a particular sample (blue) among other experimental samples in the run (gray). Shown here, sample NA03814 is an SMA carrier with a single *SMN1* copy (fold-change 0.50) and >=4 *SMN2* copies (fold-change 3.0).

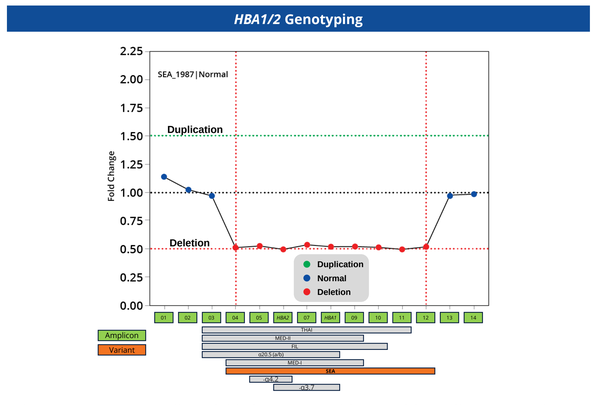


**Figure 10. *HBA1/2* amplicon fold-changes provide signal for a diverse set of variant genotypes.** A set of 14 "sentinel" amplicons target regions of the hemoglobin alpha cluster to differentiate known common breakpoints. Reduced fold-change signal (~0.50) across a set of neighboring amplicons (4-12) indicates a SEA deletion in the sample shown here (HG02353). An algorithm classifies the genotype as αα/$\alpha^{0(SEA)}$ with only one copy of *HBA1* and *HBA2*, indicating a $\alpha^0$ "α-thalassemia-trait" genotype.

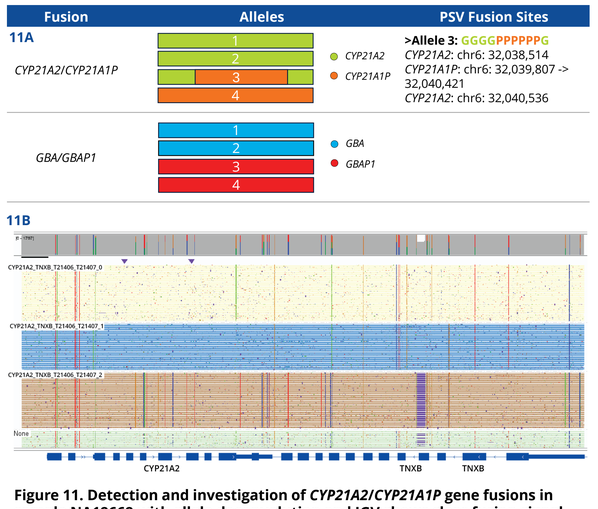


**Figure 11. Detection and investigation of *CYP21A2*/*CYP21A1P* gene fusions in sample NA18668 with allele deconvolution and IGV shows clear fusion signal. (A)** Fusion sites corresponding to both (G)ene and (P)aralog entries in *CYP21A2*/*CYP21A1P* allele 3 indicate the detection of paralogous sequence variants. In contrast, no fusion is detected in *GBA*/*GBAP1*. **(B)** Users can thoroughly investigate .bam file alignments for these regions by leveraging the use of read tags that identify unique alleles (shown grouped and colored).

**Table 1. Performance metrics across variant classes.** All variant classes for *CFTR*, *SMN1/2*, *FMR1*, *HBA1/2*, *HBB*, *CYP21A2/CYP21A1P*, *GBA/GBAP*, and *F8* were detected at >95% agreement with orthogonal data. Performance is evaluated based on samples and alleles with orthogonal comparator data (N). Overall percent agreement (OPA) refers to the accuracy of the analysis. $F_1$-score provides a single metric of predictive performance and is defined as the harmonic mean of precision and recall. CNV calls are evaluated with performance from 0 to >=3 copies.

| Gene | Variant Class | N | $F_1$-score | Accuracy/OPA |
|---|---|---|---|---|
| *CFTR* | PolyT/TG | 860 alleles (430 samples) | 1.0 | PolyT: >99%<br>PolyTG: >99% |
| | SNV/Indel | 51606 variants (405 samples) | >0.99 | >99% |
| | Large Exon Deletions[a,b] | 95 samples | 1.0 | 100% |
| *SMN1/2* | CNV[b,d] | 46 samples | *SMN1*: 1.0<br>*SMN2*: 0.96 | *SMN1*: 100%<br>*SMN2*: 96% |
| | SNV/Indel | 1413 variants (471 samples) | 1.0 | 100% |
| *FMR1* | AGG[c] | 58 alleles (24 samples) | 1.0 | 100% |
| | CGG[b] | 80 alleles (50 samples) | 0.98 | 96% |
| *HBA1/2* | CNV | 522 alleles (261 samples) | 0.91 | 96% |
| | SNV/Indel | 2 variants (1 sample) | 1.0 | 100% |
| *HBB* | CNV[d] | 22 variants (16 samples) | 0.97 | 95% |
| | SNV/Indel | 88 variants (36 samples) | 1.0 | 100% |
| *CYP21A2/ CYP21A1P* | CNV[c] | 109 samples | 0.97 | 95% |
| | Gene/Pseudogene Fusions | 59 alleles (27 samples) | 1.0 | 100% |
| | SNV/Indel | 620 variants (27 samples) | 1.0 | 100% |
| *GBA/ GBAP* | CNV[c] | 7 samples | 1.0 | 100% |
| | Gene/Pseudogene Fusions | 4 alleles (4 samples) | 1.0 | 100% |
| | SNV/Indel | 135 variants (9 samples) | 1.0 | 100% |
| | Structural Variants | 4 variants (4 samples) | 1.0 | 100% |
| *F8* | Intron 1 & 22 Inversions | 6 samples | 1.0 | 100% |

a. Performance evaluated on a recently optimized set of mix A primers with a different sample set consisting of 5 known deletions and 90 assumed WTs.
b. Performance was evaluated on a subset of data used for testing and not model development. All other analyses show performance for all available data.
c. Sparse orthogonal annotations confound performance metrics as context is needed for each sample (e.g., fusions may be orthogonally annotated as deletions).
d. Performance evaluation does not include cell-line samples used as calibrators.

## Conclusions

- Our PCR/Nanopore assay detects SNVs, indels, copy number gain or loss, short tandem repeat sizes, large structural variants, and gene-pseudogene fusions with >95% overall percent agreement to orthogonal results across *CFTR*, *SMN1/2*, *FMR1*, *HBA1/2*, *HBB*, *GBA*, *CYP21A2*, *TNXB*, and *F8* intron inversions.
- Accompanying software allows users to navigate quality control metrics, view variant calls, and perform in-depth investigation of sequencing data and analysis results with a simple user interface.
- Users can tailor reports to their specific needs by filtering variants based on their ClinVar classification and rating or including reports of novel variants with *in silico* variant effect predictions.
- The streamlined end-to-end workflow is designed to scale with ACMG carrier screening guidelines across diverse gene targets while helping labs shift to more efficient and comprehensive variant detection.

### REFERENCES

1. Woo, E. G., Tayebi, N., & Sidransky, E. (2021). Next-Generation Sequencing Analysis of *GBA1*: The Challenge of Detecting Complex Recombinant Alleles. *Frontiers in genetics*, *12*, 684067.
2. Seaby, E. G., & Ennis, S. (2020). Challenges in the diagnosis and discovery of rare genetic disorders using contemporary sequencing technologies. *Briefings in functional genomics*, *19*(4), 243–258.
3. Nan, M. N., Roig, R., Martínez, S., Rives, J., Urgell, E., Espinós, J. J., Tirado, M., Carreras, G., Aulinas, A., Webb, S. M., Corcoy, R., Blanco-Vaca, F., & Tondo, M. (2021). Comprehensive Genetic Testing of *CYP21A2*: A Retrospective Analysis in Patients with Suspected Congenital Adrenal Hyperplasia. *Journal of clinical medicine*, *10*(6), 1183.

*This product is under development, including the software interface that may differ from the partially mock representation shown here.*

*Future availability and performance to be determined.*

*All authors have the financial relationship to disclose: Employment by Asuragen.*

*Presented at ACMG 2024.*

bio-techne®
Asuragen®