Resolving Complex Genotypes in Residual Clinical Samples with Long Range-PCR and Nanopore Sequencing Assay

Cody Edwards[1], Bryan Killinger[1], Andrew Laurie[2], Lisa Hsu[2], Anne-Sophie Lebre[3], Tony Yammine[4], Chandler Ho[5], Tsoyu Chiang[5] ,Ilona Volkova[5], Monica Roberts[1], Theodore Markulin[1], Mia Mihailovic[1], Eduardo Priego[1], Brian C. Haynes[1], Bradley Hall[1]

[1]Asuragen, Austin, TX, [2]Canterbury Hlth.Lab., Christchurch, New Zealand, [3]CHU Reims et Université Reims Champagne-Ardenne (URCA);Université Paris Cité, Inst. of Psychiatry and NeuroSci. of Paris (IPNP), INSERM, Reims et Paris, France, [4]CHU Reims, Reims, France, [5]Stanford Hlth.Care, Palo Alto,CA

Summary

- Short-read sequencing methods struggle to detect both simple and complex variants in key genes associated with inherited genetic disorders recommended for carrier screening by ACMG due to the presence of highly homologous pseudogenes which are unable to be resolved using NGS.
- We developed a prototype assay based on PCR-enrichment, nanopore sequencing, and machine learning models to enable multiplex detection of diverse variant classes (SNVs, INDELs, SVs, and CNVs) in a single workflow.
- Prototype assay reagents, workflows, and analysis software were tested in four laboratories across the world using previously genotyped residual clinical samples and cell line materials.
- All classes of genetic variation were detected and confirmed by orthogonal methods with >95% agreement with orthogonal results.

Introduction

Traditional 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[1,2]. Described here are results from a single workflow using long-range PCR amplification, long-read nanopore sequencing, and automated analysis software to detect multiple classes of genetic variation within a multigene panel (*F8* intron inversions, *CYP21A2*, *TNXB*,*GBA*). Automated software deconvolves paralogs, classifies and enumerates sequence-based groups, and includes comprehensive variant calling tailored to each target in the assay to detect SNVs, INDELs, copy number variants (CNVs), structural variants (SVs).

Materials and Methods

The assay workflow was utilized by four different laboratories with a total of 260 unique samples including residual clinical samples and cell line material. External laboratories were trained to perform the workflow and review data processed by the automated analysis software. Target regions were multiplexed in a single PCR reaction, barcoded, pooled, and sequenced on MinION™ flow cells (R10.4.1) with a Mk1B (Oxford Nanopore Technologies). Results were compared with orthogonal methods including MLPA, Sanger sequencing, and long-range PCR.

Workflow

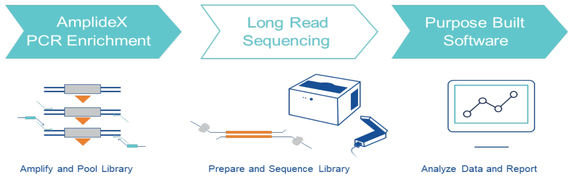
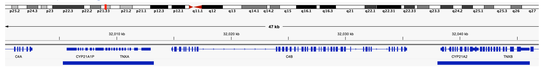


Figure 1. Prototype assay uses a two-step PCR enrichment for target genes and their paralogs and adds barcodes by sample. Samples were pooled and sequenced using a MinION device (Oxford Nanopore Technologies, SQK-LSK114 & R10.4.1). Sequencing data was processed with analysis pipelines that identified and deconvolved paralog allele groups, classified groups based on consensus and determined overall copy number for each group based on sizes of paralog read depth. Each allele group was then processed individually to provide phased variant calling. Prototype software reports line and graphical data.

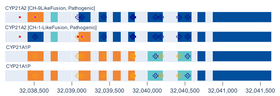
This product is under development. Future availability and performance to be determined.
**Research Use Only. Not for use in diagnostic procedures.*
Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.
Authors affiliated with Asuragen have the financial relationship to disclose: Employment by Asuragen.
Presented at ASHG, Nov 2024.

CYP21A2 Results

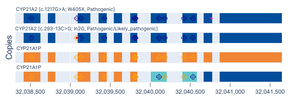
2A) Assay Design



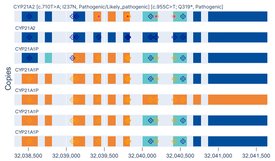
2B) Compound HET CH-1 / CH-10



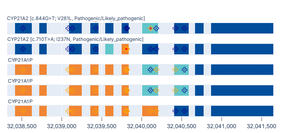
2C) Compound HET In2G / W405X



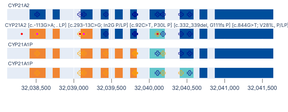
2D) HET, phased SNVs



2E) Compound HET, phased SNVs / V282L



2F) HET large conversion



2G) Normal Control

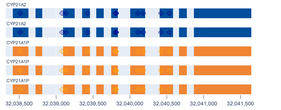
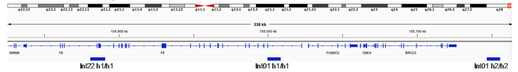


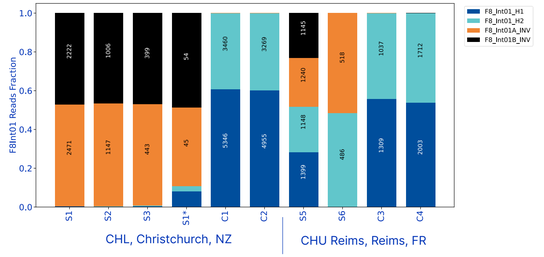
Figure 2. The assay resolves and phases *CYP21A2* and *CYP21A1P* allele groups and reports summary figures in addition to line data. Each horizontal bar describes an allele group (AG). Each row includes boxes associated with exon boundaries that align with *CYP21A2* (Blue) or *CYP21A1P* (Orange). Crosses indicate paralog specific variants (PSVs) associated with *CYP21A2* (Blue) or *CYP21A1P* (Orange). Pathogenic SNVs are indicated with a red dot near the SNV location, in genes and chimeric AGs. Teal boxes represent ambiguous exons or PSVs. Each feature is mapped to Ch 6 on Hg38 shown at the bottom of the image. Of the 260 samples tested 65 were orthogonally characterized for CYP21A2; all 65 were in agreement, with 11 of these requiring manual resolution. An additional 20 unconfirmed variants were called in samples passing QC. **2A)** Assay designs mapped against hg38 **2B)** Orthogonal results indicate two 30-kb dels. One is sub-type CH-10, a newly described attenuated type, the other is CH-1 (w/o P30L). Analysis software reports four allele groups, two AG with the expected atypical large deletion signatures, and two additional chimeric alleles. **2C)** Orthogonal results indicate compound HET c.293-13C>G and c.1217G>A. Analysis software reports four AG identified, two genes each with concordant variants, one pseudogene, and an additional chimeric allele. **2D)** Orthogonal results from Sanger, MLPA, and familial segregation determined that CYP21A2 had two copies of exon 3, and a single copy of exons 4-7, while CYP21A1P had four copies at exon 3 and five or more copies of exons 4-7. The following variants were found to be in phase [1001T>A + 1205G>A + 1382T>A + 1385T>A + 1391T>A + 1764_1765insT + 1996C>T]. Analysis software reports seven AGs, a single functional copy of the gene, two chimeras with one phased SNV concordant with orthogonal data, as well as four copies of the pseudogene. **2E)** Orthogonal results indicate a compound HET c.710T>A; 713T>A; 719T>A and c.844G>T. Analysis software reports five AGs, two copies of the gene with concordant variants, as well as three additional pseudogenes. **2F)** Orthogonal results indicate two copies of both gene and pseudogene, one copy of the gene with a large gene conversion from at least the promotor to exon 7. Analysis software reports four AG, one WT copy of the gene, two of the pseudogene, and the concordant large conversion allele. **2G)** A control sample with two copies of the gene and three copies of pseudogene without any detected variants. Analysis software reports five AGs, two copies of the gene without CYP21A2 variants and two AGs of the pseudogene, with three copies total from the two groups. Occasionally an allele group will not deconvolve, but total CN can be still be determined on the basis of relative read depth across deconvolved groups.

F8 Inversion Results

3A) Assay Design



3B) F8 Intron 01 Inversion Detection



3C) F8 Intron 22 Inversion Detection

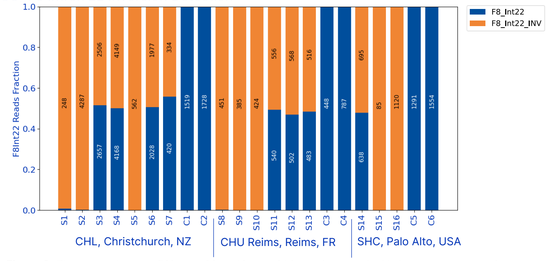
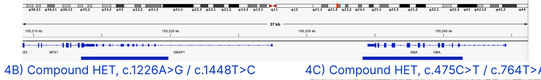


Figure 3. The assay detects *F8* intron 22 and intron 01 inversions on the X chromosome reporting variant zygosity and sample genotype. Read depth of fully spanning reads for each amplicon is provided within each bar and normalized as a fraction of total reads per sample. Blue represents the H1 region of WT intron 22 or intron 01. Teal represents H2 region of WT intron 01. Orange and black represent mutant inversions. No samples with inversion in both introns were observed. A total of 260 unique samples were tested with this module of the assay, 23/25 in agreement with orthogonal methods, 2 failed to amplify, likely due to age and sample quality, three additional intron22 inversions were observed, awaiting confirmation. **3A)** Intron 01 WT H1 and H2 reads were observed in control samples (C1, C2, C3, C4) from two external sites. Six samples with positive detection of *F8* intron 01 inversion are shown, samples S1/S1* (replicates), S2, and S3 (CHL, Christchurch, NZ) were hemizygous male or homozygous female for F8 intron 01 inversion. S1* originally failed QC and was repeated. Heterozygous female (S5) demonstrates reads from all elements indicating both WT and inverted alleles are present. (CHU Reims, Reims, France). S6 represents an uncommon hemizygous *F8* intron 01 inversion, note the pairing of WT H2 reads paired with an H1 inversion. **3B)** Intron 22 WT H1 reads were observed in control samples (C1-C6) from three external sites. Seven samples (S1, S2, S5, S8, S9, S10, S15, and S16) represented hemizygous male or homozygous female Intron 22 inversions. Heterozygous (female) intron 22 inversions were observed in eight samples (S3, S4, S6, S7, S11, S12, S13, S14). Orthogonal methods for inversions were determined using digital droplet PCR or long-range PCR with agarose gel electrophoresis. S1-S13 found concordant with orthogonal results for their inversion status. Orthogonal results are unknown for S14-S16 but represent variants observed in three of seven samples from a cohort of hemophilia positive residual samples which are awaiting orthogonal testing for concordance.

GBA Results

4A) Assay Design

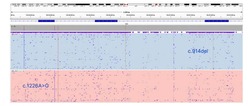


4B) Compound HET, c.1226A>G / c.1448T>C

4C) Compound HET, c.475C>T / c.764T>A

4D) Compound HET, c.259C>T, c.680A>G

4E) Compound HET, c.1226A>G, c.914del



4F) HET, c.1226A>G

4G) Normal Control

Figure 4. The assay resolves and phases *GBA* and *GBAP* allele groups and determines copy number of both gene and pseudogene. Gene-pseudogene chimeras and hybrid alleles could also be resolved, although none were present in the sample panels tested. Shown are IGV[3] views of relevant alleles down sampled to 60 reads per allele. Allele groups of *GBA* were separated and colored by deconvolution tag. Clair3 was used for SNV/indel identification within allele groups. 260 unique samples were tested with this module, with 16/16 in agreement with orthogonal methods, large deletion call required manual resolution, one sample found with unconfirmed pathogenic SNV. Residual samples with two heterozygous pathogenic variants tested at remote site (CHU Reims, Reims France) **4A)** Assay designs mapped against hg38 **4B)** c.1448T>C, c.1226A>G. **4C)** c.764T>A, c.475C>T. **4D)** c.680A>G, c.259C>T. **4E)** c.914del, c.1226A>G. **4F)** Residual gDNA from blood with a single heterozygous pathogenic variant, c.1226A>G, tested at remote site (CHL, Christchurch, NZ). **4G)** Cell line material control gDNA has two copies of *GBA* and no pathogenic variants identified.

Conclusion

- Long-read sequencing coupled with bespoke algorithms enables deconvolution into paralog groups following gene/pseudogene enrichment, sequencing, and alignment. The assay distinguished gene from pseudogene, resolved paralog-related variants, and provided additional insights about the underlying genetics that is difficult with current indirect orthogonal methods.
- Clinical research labs outside the US expressed positive opinions about the ease of familiarization and independent workflow execution.
- Specific samples described in this work contain some of the most challenging and complex variants, traditionally hard-to-decipher with current methods. Overall agreement with orthogonal methods for samples passing QC was >95%, and the single workflow provided additional insights that required multiple orthogonal methods and familial trios to ascertain.

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. James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov: Integrative Genomics Viewer Nature Biotechnology 29, 24–26 (2011)Carrier screening ref general