An Amplification-Based, Long-Read Sequencing Assay* Accurately Identifies Clinically-Relevant SNVs, Duplications, and Deletions in Hemoglobinopathy Relevant Genes

Jonathan Turner, Sarah Statt, Ryan Routsong, Chris Fraher, Gary J Latham, and Bradley Hall

Asuragen, a Bio-Techne Brand, Austin, TX

Poster # G055

Summary

- Pathogenic thalassemia variants, particularly in *HBA1/2*, are difficult to detect.
- Here, we explored whether novel PCR enrichment and nanopore sequencing could detect both large structural variants and SNVs/INDELs in *HBA1/2* and *HBB* simultaneously.
- Our assay design uses a coverage-based copy number approach to identify large *HBA1/2* and *HBB* variants while also identifying relevant SNVs/INDELs.
- Large rearrangements and SNVs/INDELs were detected with 100% and >95% sensitivity, respectively, and 100% PPV at the allele level in known pathogenic samples; large *HBA1/2* rearrangements were detected with 100% agreement at the sample level during screening.

Introduction

Hemoglobinopathies, including thalassemias and abnormal structural variants such as HbS, HbE, and HbC, are among the most common inherited genetic diseases; at least 5% of the global population carries a pathogenic variant. Deletions affecting the four hemoglobin α subunit gene copies (*HBA1* and *HBA2*) cause 90% of α-thalassemia, whereas *HBB* mutations account for 99% of β-thalassemia. Homologous gene sequences and complex recombination events cause deletions and duplications that are challenging to resolve and often require distinct workflows. Here, we combine scalable PCR enrichment with long-read nanopore sequencing in a prototype assay that accurately identifies common pathogenic SNVs, deletions, and duplication events across *HBA1/2* and *HBB*.

Materials and Methods

Genomic DNA from whole blood samples and cell lines (Coriell Cell Repository) was PCR amplified in a single tube, barcoded, pooled, and prepared using a ligation sequencing kit (Oxford Nanopore, ONT). Sequencing was performed using R9.4.1 flow cells and a Mk1B or Mk1C (ONT). Custom software was developed to manage projects, start sequencing, review run-time, automate data analysis, and report genotypes. A set of 47 samples was used to train the algorithm to identify deletion and duplication events. Performance was tested using a second set of 58 unique challenge samples bearing additional variants. Two hundred previously untested whole blood samples were then screened using the prototype assay. SNVs/INDELs of interest were identified by filtering those identified by our analysis for the following clinical significance annotations per ClinVar; Pathogenic, Pathogenic/Likely pathogenic, and Conflicting interpretations of pathogenicity (when the majority of the significance detail citations were Pathogenic, Likely pathogenic, and/or Uncertain significance). Comparator methods (melt curve analysis[1], GAP-PCR[2,3], Sanger sequencing) were used to determine concordance for SNVs/INDELs and *HBA1/2* copy number.

Results

Table 1. Pathogenic and Likely Pathogenic *HBA1/2* and *HBB* SNV/INDEL Variants Detected in Challenge Samples Using the PCR/Nanopore Assay. *HBB* SNVs and short deletions were detected in 12 cell line samples and three whole blood samples. One cell line sample was homozygous for Hb E and one was homozygous for -138C>T; all other SNVs were heterozygous. Five cell line samples were heterozygous for two different *HBB* SNVs, and four cell line samples were heterozygous for one *HBB* SNV. In total, 22 out of 23 expected *HBB* variant alleles were detected by the PCR/Nanopore assay (95.7% sensitivity, 100% PPV). One cell line sample was heterozygous for one *HBB* and one *HBA2* SNV; another was heterozygous for one *HBB* and one *HBA1* SNV.

Gene	Variant	Positive Cell Line Samples	Positive Whole Blood Samples
HBB	c.19G>A (Hb C)	3	0
HBB	c.79G>A (Hb E)	3	0
HBB	c.20A>T (Hb S)	3	1
HBB	c.126_129delCTTT	2	0
HBB	c.93-21G>A	2	0
HBB	c.-137C>G	1	0
HBB	c.-138C>T	1	0
HBB	c.316-197C>T	1	0
HBB	c.82G>T (Hb Knossos)	1	0
HBB	c.404T>C (Hb Yaounde)	0	1
HBB	c.-79A>G	0	1
HBA1	c.207C>G (Hb G-Philadelphia)	1	0
HBA2	c.427T>C (Hb Constant Spring)	1	0

**This product is under development. Future availability and performance to be determined.*

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

Presented at AMP 2022

Table 2. Pathogenic and Likely Pathogenic Large *HBA1/2* and *HBB* Variants Detected in Challenge Samples Using the PCR/Nanopore Assay. Large *HBA1/2* variants were detected in 19 cell line samples and one whole blood sample. Four cell line samples had two large *HBA1/2* variants each (two 3.7del/3.7del, one FIL/SEA, one SEA/anti3.7); the rest had one. Large *HBB* variants were detected in three cell line samples. The sample with the Sicilian *HBB* deletion was homozygous for that variant. In total, 24 of 24 expected and four of four expected large *HBA1/2* and *HBB* variant alleles, respectively were detected (100% PPV and sensitivity).

Gene	Variant	Positive Cell Line Samples	Positive Whole Blood Samples
HBA1/2	SEA	7	0
HBA1/2	3.7del	8	0
HBA1/2	anti3.7	3	1
HBA1/2	4.2del	1	0
HBA1/2	FIL	1	0
HBA1/2	THAI	1	0
HBB	NG_000007.3:g.63564_70978del (Hb Lepore-Baltimore)	2	0
HBB	NG_000007.3:g.64336_77738del13403 (Sicilian [deltabeta]0-Thal)	1	0

Figure 1. HBA Deletion Identification In Challenge Samples Based on Amplicon Coverage Using a Machine Learning Algorithm. Normalized fold change in coverage compared to control sample reference amplicons is shown on the y axis; assay amplicons are arranged in sequential order on the x axis. **A)** Cell line sample A with an SEA deletion on one allele and an anti-3.7 triplication on the other. *HBA2* and Rgn12 amplicons receive "wild-type" copy number calls because the combination of variants results in the same coverage levels as in a wild-type sample. **B)** Cell line sample B with two 3.7del alleles. Note the unexpected deletion in Rgn16 does not affect genotype call. **C)** Whole blood sample A with one wild-type allele (aa) and one anti-3.7 triplication allele. **D)** Whole blood sample B showing a typical coverage pattern for two wild-type alleles (aa/aa).

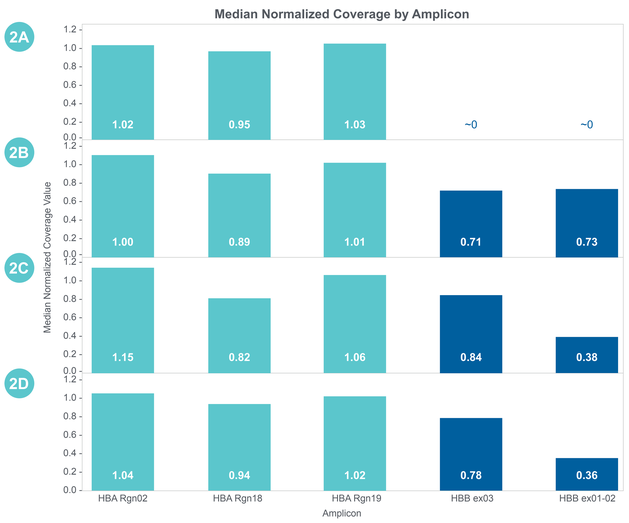

Figure 2. *HBB* Amplicon Coverage in Samples with Large Structural Variants. HBA Rgn02, Rgn18, and Rgn19 coverage, which are not affected by the deletions this assay detects, are used to calculate normalized coverage values for *HBB* amplicons. **A)** Cell line sample which is homozygous for a large deletion including both the *HBB* ex01-02 and ex03 amplicons. **B)** An *HBB* wild-type cell line sample shown for reference. **C)** and **D)** Two cell line samples with the Lepore-Baltimore variant. The breakpoint for that variant is within the *HBB* ex01-02 amplicon; coverage of that amplicon is therefore reduced by approximately half in the two heterozygous samples compared to the wild-type sample. Normalized coverage values are shown at the base of the bars.

Table 3. Machine Learning Algorithm Accurately Identifies *HBA1/2* Variants Based on *HBA1/2* Copy Number in Whole Blood Samples. A) PCR/Nanopore assay call agreed with melt curve copy number results for all 12 3.7del/aa samples and for 186 of 187 wild-type samples. Known limitations of the melt curve assay preclude reliable detection of three *HBA2* copies in samples with two *HBA1* copies (as shown by gray color); the accuracy of anti-3.7/aa ONT calls was therefore evaluated using GAP-PCR. **B)** Anti-3.7 and 3.7del GAP-PCR product detection agreed with ONT genotype for all 205 screening samples. Overall agreement between ONT and comparator method results for *HBA* copy number variants was 204/205 = 99.5%.

A

		HBA1/HBA2 copy number (melt curve)			
		2/3	2/2	2/1	1/1.5
ONT assay genotype call	anti3.7/aa	2	4	0	0
	aa/aa	0	186	0	1
	3.7del/aa	0	0	12	0

B

		GAP-PCR products detected		
		anti3.7	none	3.7del
ONT assay genotype call	anti3.7/aa	6	**0**	0
	aa/aa	0	187	0
	3.7del/aa	0	0	12

Figure 3. Pathogenic and Likely Pathogenic *HBB* SNVs Identified Among Screening Set Whole Blood Samples. Fully-spanning reads are shown; the darker gray bars at the top of each row show a rollup of the sequencing reads at each position in histogram form, while the lighter gray bars below represent a subset of the individual reads (each line is one read). **A)** Whole blood sample 1 is a carrier of c.118C>T. **B)** Whole blood sample 2 is a carrier of c.-79A>G. Whole blood sample 3 is heterozygous for two SNVs, **C)** c.364G>C and **D)** c.92+1G>T.

Conclusions

- The PCR/Nanopore assay can detect large rearrangements and SNVs/INDELs simultaneously in *HBB* and *HBA1/2* with high accuracy, potentially replacing at least three independent assays required to obtain equivalent information.
- Unlike current gold standard methods, this assay enables detection of deletions with non-standard breakpoints that result in loss of *HBA1/2* copies while also identifying common deletions.
- The simple workflow allows multiplexing of at least 96 samples in a single run and could greatly reduce thalassemia carrier screening turnaround times.

References

1. Turner A *et al.* (2015) BMC Med Genet 16, 115.
2. Tan, ASC *et al.* (2001) Blood 98(1):250-251.
3. Oron-Karni, V *et al.* (1998) Am J Hematol 58:306-310.

Asuragen
a biotechne brand