bio-techne diagnostics™

Resolution of rare and complex thalassemia genotypes in a single workflow using long range PCR and long read sequencing

P21.022.D

Cody Edwards, Juliette A. Baker, Monica Roberts, Brian C. Haynes, Bradley Hall

Bio-Techne Diagnostics, Austin, TX, USA

Summary

- Variant detection within the alpha and beta globin gene clusters associated with hemoglobinopathies currently requires a multi-platform testing cascade.
- Long-range PCR combined with long-read sequencing offers a unified workflow capable of replacing the cascade testing while delivering improved performance comparable to gold-standard methods.
- The targeted enrichment panel included key globin genes, deletion sentinels, regulatory and modifier regions. The prototype correctly identified and phased multiple classes of pathogenic variants across 51 clinical residual samples extracted from whole blood with 100% orthogonal concordance.

Introduction

Molecular diagnosis of hemoglobinopathy often requires cascaded testing across multiple platforms to detect different classes of variants, which increases turnaround time and laboratory burden. We describe a prototype that detects >95% of all variants associated with thalassemia, including 24 named structural variants (SVs).

Single-tube, multiplex PCR enrichment utilizes primer pairs designed for robust amplification of long and challenging genomic regions. Enrichment targets include single-amplicon coverage of functional globin genes, key regulatory elements, phenotype modifiers, and deletion sentinels across the alpha and beta hemoglobin clusters.

Long amplicon designs coupled with long-read sequencing readouts allow for powerful discrimination between paralogs, hybrids, and allele groups to generate comprehensive genotype results. Automated analysis tools and manual data review were used to comprehensively classify thalassemia-associated variations in a single workflow.

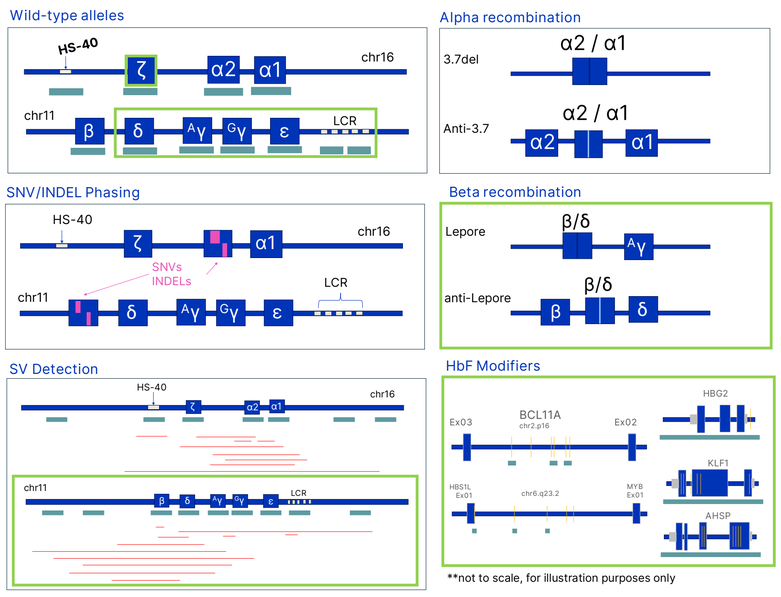

Figure 1. Prototype Assay Design Targets. Coverage of key regulatory regions including HS-40 and LCR-HS1 through LCR-HS5 allow for resolution of rare non-deletion thalassemia types. Long amplicon designs provide paralog discrimination and gene hybrid detection in gene-fusion-like recombination events. Sentinel amplicons targeting strategic locations differentiate common SV signatures in the alpha and beta hemoglobin clusters. Additional coverage of potential phenotype modifying variants found in *KLF1, AHSP*, and various SNVs found in *BCL11A* intron 2 and the intergenic region between *HBS1L* and *MYB* provide additional insight into phenotype severity. Content boxed in green highlights improvements to content and capabilities compared to AmplideX Nanopore Carrier Plus Kit C.

Methods

Genomic DNA from 51 residual clinical samples were amplified using AmplideX® PCR chemistry, followed by barcoding, pooling, and library preparation with the ONT Ligation Sequencing Kit (SQK-LSK114). Sequencing was performed on R10.4.1 MinION flow cells using a GridION® platform. Custom-built software was developed to automate sequencing setup/execution, monitor performance metrics, perform secondary analysis, and generate genotype reports.

Regions-of-interest (ROI) were enriched in a single-tube multiplex. Sequenced reads within each ROI were grouped based on sequence identity, variants are identified and phased, and group depth was used to increase copy number signal confidence. MLPA, Sanger or targeted sequencing, and GAP-PCR were used as comparator methods.

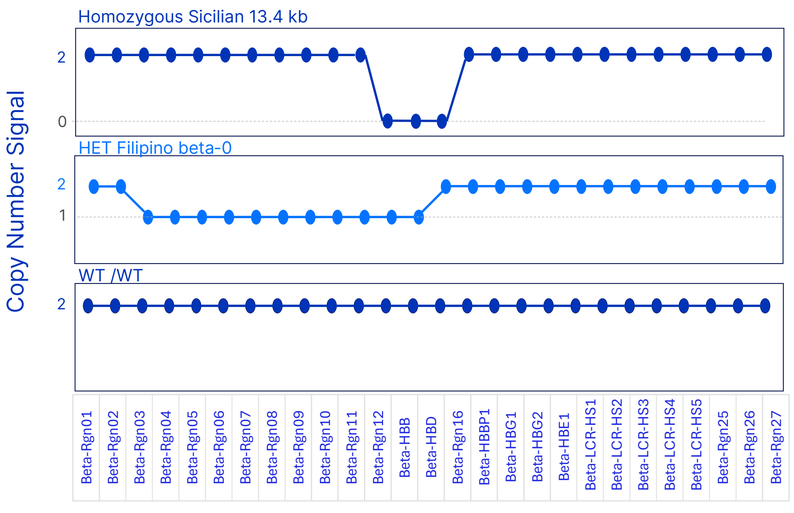

Figure 2. Idealized Copy Number Signal Output. Copy number signal is generated across an array of 27 sentinel amplicons in the beta cluster to determine the extent and zygosity in a region spanning >300 kb proximal to the cluster. The deletion or duplication of sentinel amplicons is characteristic for specific deletion types. The figure shows the signal for various genotypes: top row - HOM Sicilian 13.4 kb deletion; middle row - HET beta-0 Filipino; bottom row - WT sample without copy number variant (CNV) changes. Select SVs in the design generate GAP-PCR like amplicons, allowing for breakpoints to be observed directly. A similar array of 18 sentinel amplicons is utilized for the alpha cluster.

Results

In the 51 samples cohort two samples were confirmed wild type. Six samples in the cohort contained orthogonally confirmed *HBA1/2* pathogenic SNVs, and 26 samples had known SVs in the alpha cluster. 25 samples were found with confirmed SNVs in *HBB*, and an additional 11 samples were confirmed with SVs in the beta hemoglobin cluster. All results were concordant with various orthogonal methods.

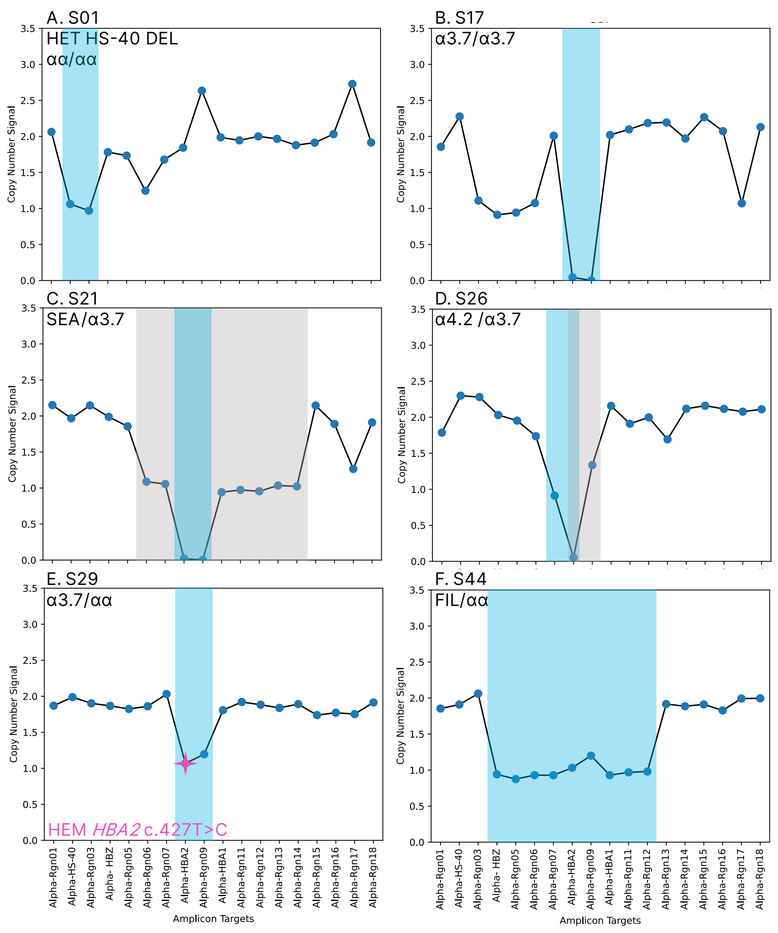

Figure 3. Copy Number Signal in the Alpha Hemoglobin Cluster from 6 Representative Samples in the Cohort. Each blue dot represents a copy number signal determined per target, confirmed CNV signals are highlighted in teal, and compound SVs are shown in teal and grey to indicate overlap. Pathogenic small variants are noted with a pink star in the amplicon where detected. A. HET HS-40 regulatory deletion (aa/aa); B. HOM 3.7del, SV involving *HBZ* to be confirmed; C. SEA/3.7del; D. 4.2del/3.7del; E. HET aa/3.7del, HEM c.427T>C; F. HET FIL/aa.

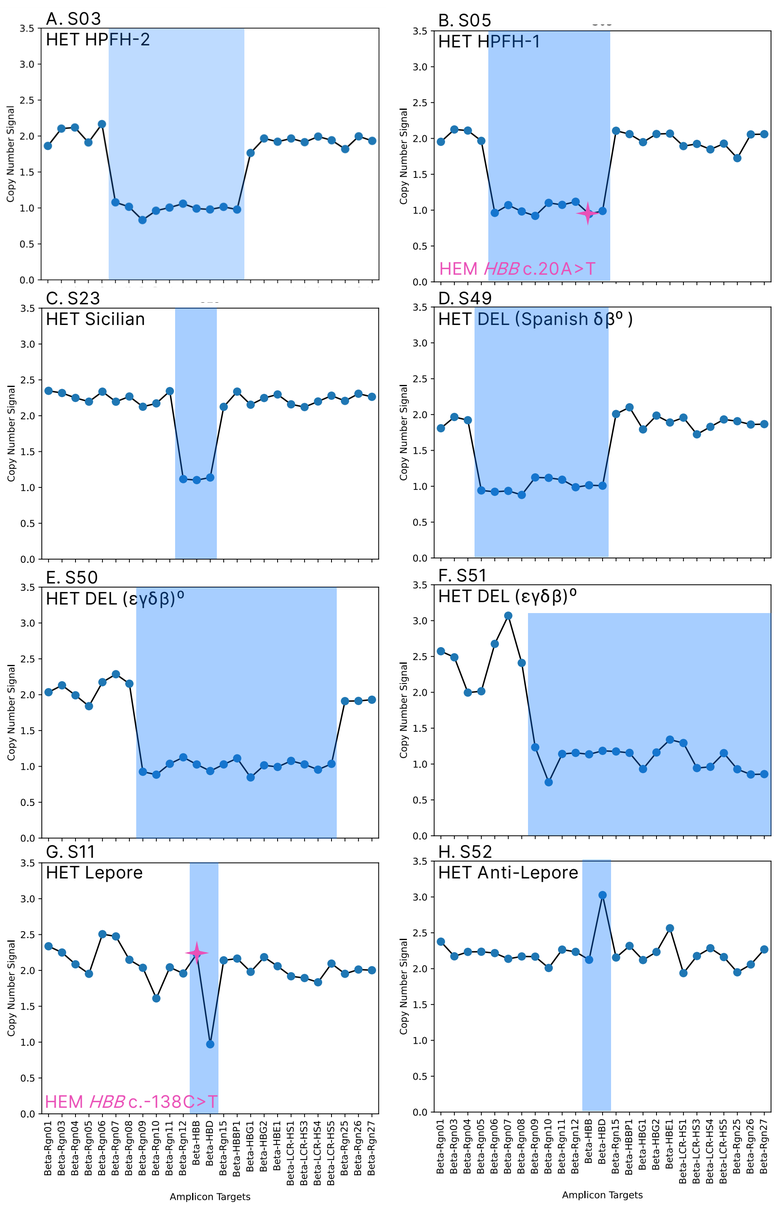

Figure 4. Copy Number Signal in the Beta Hemoglobin Cluster from 8 Representative Samples in the Cohort. Each blue dot represents a copy number signal determined per target. CNV signal changes consistent with a specific SV variant are highlighted in blue. Pathogenic small variants are noted with a pink star in the amplicon where detected. Lepore SV breakpoints are estimated using paralogous sequence variants (PSVs). A. HET HPFH-2; B. HET HPFH-1, HEM c.20A>T; C. HET Sicilian (13.4 kb); D. HET DEL (Spanish δβ⁰); E. HET DEL (εγδβ)⁰; F. HET DEL (εγδβ)⁰; G. HET Lepore, HEM c.-138C>T; H. HET Anti-Lepore duplication. Lepore *HBB*|*HBD* chimeric alleles are enriched; plotted as a CNV loss at *HBD* and contributes to normal CNV at HBB, with duplications being the inverse.

Table 1. Summary of Samples with variants only in the alpha (N=16) or beta (N=18) cluster. Six unique alpha cluster SVs were correctly identified in the sample cohort. Three HET 3.7del SVs were hemizygous or heterozygous with a compound SNV/INDEL. Nine unique beta cluster SVs were correctly identified. Five were differentiated by extent name. Other deletions were detected and zygosity determined, but extent was approximate. Three samples contained a compound SV & hemizygous pathogenic SNV/INDEL variant in *HBB*. Samples and SVs in bold are featured in Figure 3 (Alpha) or Figure 4 (Beta).

	Alpha SV	*HBA* SNV INDEL
S01	αα/αα (HET HS-40 DEL)	
S04		HET *HBA1* c.237del
S08	FIL/αα	
S10	α3.7/αα	HET *HBA1* c.207C>G
S13	α4.2/αα	
S15	SEA/αα	
S16	SEA/αα	
S18		HET *HBA2* c.142G>C
S27		HET *HBA2* c.427T>C
S29	α3.7/αα	HEM *HBA2* c.427T>C
S41	FIL/αα	
S42	FIL/αα	
S43	FIL/αα	
S44	FIL/αα	
S45	FIL/αα	
S55	α4.2 /α3.7	

	Beta SV	*HBB* SNV INDEL
S05	HET HPFH-1	HEM c.20A>T
S09		HOMc.20A>T
S11	HET Lepore	HEM c.-138C>T
S12		HOM c.27dup
S19		HOMc.20A>T
S20	HET HPFH-2	HEM c.-138C>T
S25	HET Sicilian	
S28		HET c.17_18del
S30		HOM c.79G>A
S32		HET c.-79A>G
S34		HET c.92+6T>C
S36		HETc.20A>T, HET c.19G>A
S38		HET c.118C>T
S39		HET c.118C>T
S48	HET DEL	HEM c.92+1G>T
S49	HET DEL	
S50	HET DEL	
S51	HET DEL	
S52	HET Anti-Lepore	

Table 2. Summary of Samples with Complex Variants across Alpha and Beta clusters (N=14). Samples harboring variants in both hemoglobin clusters demonstrate the utility of the streamlined single-tube assay summary data outputs. Samples and SVs in bold are featured as plots in Figure 3 (Alpha) or Figure 4 (Beta).

	Alpha SV	HBA SNV INDEL	Beta SV	HBB SNV INDEL
S02	α3.7/αα			HET c.-138C>T
S03	α3.7/αα		HET HPFH-2	
S07	α3.7/αα			HOMc.20A>T
S14	ααα anti-3.7/αα			HET c.316-2A>G,
S17	α3.7/α3.7			HETc.20A>T, HET c.19G>A
S21	SEA/α3.7			HET c.79G>A
S23	α3.7/αα		HET Sicilian	
S24	ααα anti-3.7/αα			HET c.19G>A, HET c.-138C>T
S26	α4.2/α3.7			HET c.*233G>C
S33	α3.7/αα			HET c.315+1G>A, HET c.92+5G>C
S37	α3.7/αα	HEM *HBA2* c.95+2_95+6del		HET c.364G>C
S40	α3.7/α3.7			HOM c.19G>A
S54	α4.2/αα			HETc.20A>T, HET c.19G>A

Workflow

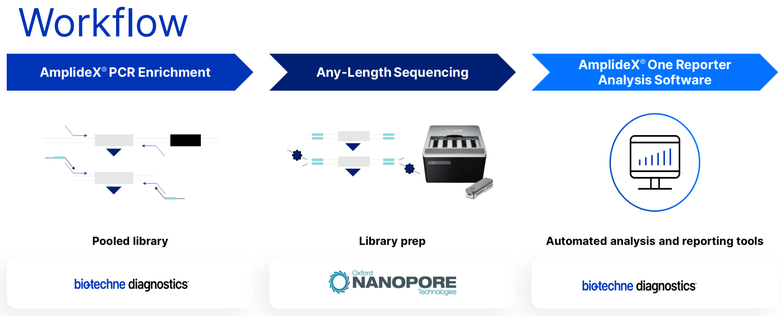

Figure 5. AmplideX Nanopore Carrier Plus Kit Panel Design and Workflow. Carrier Plus kits utilize a harmonized workflow and can be run individually or pooled and sequenced together to cover additional content in a single run.

Conclusions

- The prototype assay accurately genotyped a range of orthogonally confirmed variants of multiple classes found in the α and β-globin clusters using residual genomic DNA from clinical samples.
- 42 SV alleles in 36 samples were successfully detected and phased. Specific deletion SVs were attributed in 37/42 variants, whereas only approximate extant patterns in the five remaining SVs were described.
- 36 pathogenic SNV/INDELs were resolved in 30 samples, agreement 100% including *HBA1/2* (6) and *HBB* (25) both (1)
- These findings indicate that a single-tube long-read sequencing workflow for hemoglobinopathies can detect pathogenic variation across hemoglobin genes, key regulatory elements, phenotype modifiers, and reduce the need for complicated multi-platform testing cascades.

1. Cao, A., & Galanello, R. (2010). Beta-thalassemia. Genetics in Medicine, 12(2), 61–76. https://doi.org/10.1097/gim.0b013e3181cd68ed
2. Harteveld CL, Achour A, Arkesteijn SJG, et al. The hemoglobinopathies, molecular disease mechanisms and diagnostics. Int J Lab Hematol. 2022; 44(S1):28-36. doi:10.1111/ijlh.13885
3. Cato, LD; Li, R; Lu, HY; Yu, F; Wissman, M; Mkumbe, BS; Ekwattanakit, S; Deelen, P; Mwita, L; Sangeda, R; Suksangpleng, T; Riolueang, S. medRxiv March 28, 2023 http://dx.doi.org/10.1101/2023.03.24.23287659

This product is under development; performance characteristics and final product features to be determined.

The prototype assay is not authorized as an IVD medical device in any region

All authors have the financial relationship to disclose: Employment by Bio-Techne Diagnostics