A rapid and robust single-tube SLIMamp NGS assay for detection of mutations associated with Thalassemia

Xiaoxi Wu[1], Akuah Kontor[1], Yue Ke[1], Hainan Yang[2], Hao Wang[3], Michael Liu[1], Sean Polvino[1], Erin Petrilli[1], Geoffrey Richman[1] and Zhaohui Wang[1]

[1]Pillar Biosciences, Inc. Natick MA [2] Ethan Medical Ltd, China [3]Zhengu Bio, Shanghai, China

Abstract

Introduction: Next-generation sequencing (NGS) has been widely used in clinical settings for carrier screening tests (CST) for inherited monogenic disorders. However, variant detection, especially structural variant (SV) detection, remains a challenge in highly homologues genes such as HBA1 and HBA2 for α-thalassemia. HBA1 and HBA2 share the same coding sequences and are located on the alpha globin cluster with most of the pathogenic mutations being deletions. On the other hand, β-thalassemia is caused by HBB genetic mutations which are mostly SNV and small indels (sIndel). Most CST still need to perform two different assays for deletion detection and SNV/sIndels separately. To enable the simultaneous detection of SNV, sIndels and SV, we developed a single-tube SLIMamp multiplex PCR (mPCR) assay, the INHERIT/Reveal Thalassemia Panel, to cover the extended regions for HBA1, HBA2 and HBB. Also, we developed a proprietary CNV caller in our NGS analysis tool (PiVAT) to enable the detection of relevant SNV, sIndels and SV. Additionally, some SVs are also detected with gap-PCR for NGS in the same pool that enables the precise identification of the genomic breakpoints.

Methods: The panel contains 131 amplicons with the sizes ranging from 130 to 174 bp. The CNV caller integrated various statistical and bioinformatics approaches with a series normalization steps of the amplicon coverage and the heterozygosity of common SNVs to minimize the coverage variations caused by different reagent lots and different batch of sample-prep. A set of 102 samples with 42 known positive SVs in the HBA region and 60 SV-normal samples were used for CNV caller development. To verify the performance of the caller, a different set of 20 blinded samples with 15 HBA CNVs were then used as the validation set. Out of the total 122 samples, 45 samples had pathogenic SNVs in the three genes. All the mutations were confirmed by Sanger sequencing and conventional gap-PCR-gel assays for SNV-sIndels and large SVs respectively. The sample libraries were prepared with known SV-normal samples as batch process references for accurate CNV calling. To characterize and assess the assay performance, representative samples were tested with different input amounts and run in multiple batches. All libraries were sequenced with the 2x150bp protocol on Illumina MiSeq or NextSeq and the NGS data were analyzed by PiVAT 2020.1.

Results: Overall, the INHERIT/Reveal Thalassemia Panel performed well with a mapping rate of 99.7% ± 0.1%, on-target rate of 99.5% ± 0.1%, and coverage uniformity >0.2x mean coverage of 98.5% ± 2.2%. The CNV caller detected all 42 HBA SVs in the training set and all 15 HBA SVs in the validation set. In addition, all the detected SVs were correctly identified and annotated with the conventional CNV types (e.g. SEA, Thai, etc). All 69 confirmed SNVs and sIndels were detected.

In the reproducibility study from 3 different runs, except for one case with a-3.7dup due to the high noise, all 7 HBA-CNV types were correctly identified. The DNA input study showed that the CNVs were correctly identified over a range of input amounts from as little as 5ng. In addition, the CNV caller was robust with DNA input amounts ranged within 2-3 folds of each other in one batch. The average of 400 coverage depth was sufficient for the variant calls, suggesting high throughput levels can be achieved with possible hundreds of samples in each sequencing run.

Conclusions: The INHERIT/Reveal Thalassemia Panel provides a solution for the robust detection of CNVs, SNVs and small Indels in a rapid high-throughput NGS assay. Furthermore, the Pillar NGS analysis software PiVAT has optimized variant detections for various variant types and incorporated the identification for conventional HBA SV annotation (e.g. a-3.7).

Methods and Assay Design

Panel Summary	
Sample Type	gDNA from blood cells
DNA Input	5-80 ng (20 ng is recommended)
Amplicon Size	130-174 bp (Average size 155 bp)
Primer Pool	Single pool
Amplicon#	131
Variant Type	SNV, Small Indels, CNV
Mapping Rate	99.7% ± 0.1%
On-Target Rate	99.5% ± 0.1%
Coverage Uniformity	%>0.20 x Mean: 98.5% ± 2.2% %>0.15 x Mean: 99.2% ± 1.7% (excluding GAP-PCR amplicons.)
Variant Frequency	> 10%
Multiplex Level	343 samples on a MiSeq V3.

Table 1 – INHERIT/Reveal Thalassemia Panel overview.

Panel Design Highlight	
3 Primary Genes	HBB, HBA1, HBA2
NM Numbers	NM_000518(HBB), NM_000558(HBA1), NM_000517(HBA2)
Total number of amplicons	131 (including 7 Gap-PCR amplicons)
Gap-PCR covered Thalassemia Types	α-THAI, α-QINZHOU, α-SEA, α-HONGKONG, β-SEA-HPFH
Major Thalassemia Types Covered by the Panel	α-3.7, α-4.2, anti-α-3.7, anti-α-4.2, α-THAI, α-QINZHOU, α-SEA, α-HONGKONG, β-SEA-HPFH, α-FIL, α-MED, α-20.5, etc. (In total, 40 types are covered in our panel design.)
Hotspot Mutations	* HBA1 and HBA2 exon 3 differentiable SNVs. * HBB SNVs covers pathogenic mutations in promoter, deep intronic, and 3'- UTR regions.

Table 2 – Highlights of the INHERIT/Reveal Thalassemia panel, specially designed for detecting structure variants.

Results and Conclusions

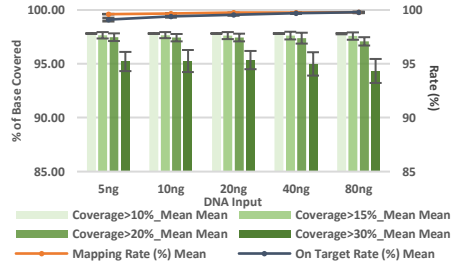

Figure 1- Mapping rates, Effective On Target rates and Coverage Uniformity across range of inputs. All inputs have >99% Effective On-Target and Mapping Rate. >94% of bases are covered at >10%x ~ >30%x mean for all inputs.

Input Range	# of α-3.7 Called/ # of α-3.7 Tested	# of α-4.2 Called/ # of α-4.2 Tested
5ng	2/2	2/2
10ng	2/2	2/2
10ng~40ng	2/2	2/2
20ng~60ng	2/2	2/2
10ng~60ng	2/2	0/2

Table 3 – Mixed input study results for Thalassemia type specific calls. For 5ng and 10 ng inputs, the caller requires the same input for one analysis run. Starting from 20ng, the caller keeps the sensitivity when input varies within +/- one-fold.

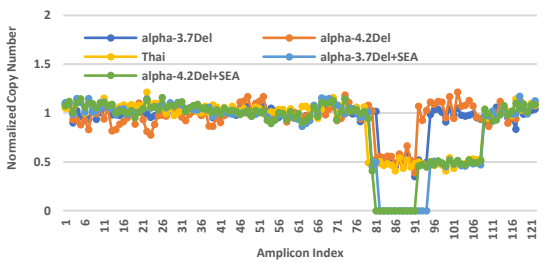

Figure 2 - Plots of normalized copy numbers of 5 Thalassemia Structure Variant positive samples with different breakpoints.

		PiVAT Call +	PiVAT Call -	PiVAT Sensitivity (CI95%)	PiVAT Specificity (CI95%)
Training Dataset	Gap/Digital PCR +	42	0	100% (91.6-100)	91.7% (81.9-96.4)
	Gap/Digital PCR -	5	55		
Validation Dataset	Gap/Digital PCR +	15	0	100% (79.6-100)	60.0% (23.1-88.2)
	Gap/Digital PCR -	2*	3		

Table 4 - Thalassemia type specific calls summary based on 122 samples including 57 samples with confirmed structure variants. (*The false positive calls are false deletion calls on chr11 due to poor sample quality.)

Results and Conclusions

Mutation Type	Gene	Common Name	HGVSC	HGVSP	Zygosity	Unique Sample#	Pillar VAF Mean (std)
SNV	HBA1 | HBA2	αWS	369C>G	His123Gln	HETER	5	50.14(0.61)%
		αQS	377T>C	Leu126Pro	HETER	2	48.98(0.49)%
		αQS	377T>C	Leu126Pro	HOMO	1	99.63(0.04)%
		αCS	427T>C	Ter143GlnextTer31	HETER	6	49.22(1.04)%
		αCS	427T>C	Ter143GlnextTer31	HOMO	3	99.93(0.05)%
	HBB	-28(A>G)	-78A>G	-	HETER	6	50.17(1.01)%
		-28(A>G)	-78A>G	-	HOMO	1	99.92%
		-29(A>G)	-79A>G	-	HETER	1	50.32%
		CD43(GAG>TAG)	130G>T	Glu44Ter	HETER	1	51.69%
		IVS-II-654(C>T)	316-197C>T	-	HETER	6	46.48(1.46)%
		CD17(AAG>TAG)	52A>T	Lys18Ter	HETER	12	50.79(0.68)%
		CD17(AAG>TAG)	52A>T	Lys18Ter	HOMO	1	99.41%
		βE CD26(GAG>AAG)	79G>A	Glu27Lys	HETER	7	49.85(1.04)%
		IVS-I-1(G>T/A)	92+1G>T	-	HETER	2	50.81(0.99)%
Deletion		CD41-42(-CTTT)	126_129del	Phe42LeufsTer19	HETER	10	50.22(0.76)%
		CD41-42(-CTTT)	126_129del	Phe42LeufsTer19	HOMO	1	99.75%
Insertion		CD71-72(+A/+T)	217dup	Ser73LysfsTer2	HETER	4	49.16(0.95)%

Table 5- SNV-sIndel variants in clinical samples called in the INHERIT/Reveal Thalassemia panel. All SNVs, deletions and insertions confirmed in the 45 samples were correctly called.

Thalassemia Type	Unique Sample #	# of Repeats Called/# of Repeats Tested	Sample Input
α-4.2/wt	2	8/8	20ng
α-SEA/wt	4	13/13	20ng
α-3.7/wt	5	18/18	20ng
α-THAI/wt	1	2/2	20ng
α-4.2/α-SEA	2	7/7	20ng
α-3.7/α-SEA	1	3/3	20ng
anti-α-3.7 (dup)/wt	1	2/3	20ng

Table 6 - Reproducibility study based on 7 different Thalassemia HBA Structure Variants from 16 samples in three independent runs. All SV are correctly detected with expected type except one anti-α-3.7 duplication.

Thalassemia Panel Summary:

- The INHERIT/Reveal Thalassemia panel is a robust and sensitive assay for the detection of Thalassemia SNVs, small indels and structure variants.
- The assay can accurately detect various structure variant types with high reproducibility and incorporated the identification for conventional HBA SV annotations.
- The workflow is streamlined, allowing for same-day loading of finished libraries in less than 8 hours if starting from isolated DNA.
- The Thalassemia SV caller accepts DNA input as little as 5ng and starting from 20ng, the input range can vary +/- one-fold for one analysis run. To ensure reliable SV calls, we recommend users to provide 2~3 wild type in-run controls with similar sample condition.
- The PiVAT development for sample QC and false positive filter is in progress. The caller specificity will be better quantified as we collect more validation data.

Disclosure Slide

Financial Disclosure for:

Xiaoxi Wu, Akuah Kontor, Yue Ke, Michael Liu, Sean Polvino, Erin Petrilli, Geoffrey Richman and Zhaohui Wang are all employees of Pillar Biosciences, Inc. Natick MA

Hainan Yang is an employee of Ethan Medical Ltd, China

Hao Wang is an employee of Zhengu Bio, Shanghai, China