Disclosure Slide

Financial Disclosure for:

A. Walker, M. Robinson, J. Smith, Z. Zhang, J. Kearney, R. Heim,
N. Leach, H. Zhu, P. Okamoto

Laboratory Corporation of America® Holdings,
Westborough, MA and Research Triangle Park, NC
All authors are employees of Laboratory Corporation of America Holdings ("LabCorp") and may own LabCorp stock.

An automated α-globin region genotyping tool for the detection of copy number variants by MLPA in a clinical setting

A. Walker, M. Robinson, J. Smith, Z. Zhang, J. Kearney, R. Heim, N. Leach, H. Zhu, P. Okamoto

Laboratory Corporation of America® Holdings, Westborough, MA and Research Triangle Park, NC

Introduction

Alpha-thalassemia is one of the most widespread autosomal recessive hemoglobin disorders in the world.[1] It is caused by one or more deficient α-globin subunits primarily as a result of copy number variations (CNVs) in the *HBA1* and *HBA2* α-globin genes.[1] Multiplex ligation-dependent probe amplification (MLPA; MRC Holland) is an effective and widely used technology for identifying CNVs in the human α-globin.[2] However, analysis of MLPA data is complicated and time-consuming due to both the complexity of the α-globin gene cluster region and the use of lookup tables to manually determine the boundaries and nomenclature of CNVs when they are detected. An automated genotyping algorithm for the HBA MLPA assay was developed and validated, and then applied it to ~45,000 consecutive samples in order to determine the carrier frequency for our testing population as well as carrier and specific allele frequencies for a subset of samples with known ethnic origins.

Methods

In our clinical laboratory, MLPA is used to test for eight common deletion variants (-$\alpha^{3.7}$, -$\alpha^{4.2}$, --$\alpha^{20.5}$, --SEA, –FIL, --THAI, --MED, HS-40del) and one point mutation (HbCS, Constant Spring), as well as unique rare copy number variants. The MLPA process was performed as per the manufacturer's instructions, with small modifications to accommodate high throughput testing. Positive results were confirmed by multiplex gap-PCR, where appropriate, and Sanger sequencing for HbCS. Samples were accrued from throughout the U.S.

The α-globin genotyping algorithm was developed using data from 678 clinical samples, as well as 105 *in-silico* simulations of possible genotype combinations for the variants tested. Additionally, simulation was modified at each MLPA probe to test error tolerance of the algorithm (>3700 additional test cases). After development, the algorithm was tested on 343 clinical specimens.

Overall allele frequency was calculated as the fraction of samples with a heterozygous mutation among all samples analyzed. Allele frequency by ethnicity was based on ~9.2% of the ~45,000 samples for which self-reported ethnic data were available.

Results

The α-globin region is highly complex due to the presence of several pseudogenes and regions of homologous sequence, and thus susceptible to copy number variation (**Figure 1**). In our validation study, the genotyping algorithm was 99% sensitive (86/87 positive genotypes called) and 100% specific (362/362 negative genotypes called). One positive sample was identified for manual review by the algorithm (**Table 1**). Our decision-tree based genotyping algorithm (**Figure 2**) was able to make definitive genotyping calls in >98% of samples tested in both validation and in the >45,000 samples tested for α-globin CNVs. Overall, 9.3% of samples were positive for at least one pathogenic α-globin CNV or point mutation (**Table 2**). In the subsets of samples with ethnic identity information, the positivity rates ranged from 2.0% in the white/ Caucasian cohort to 34.8% for the black/African American cohort. Calculated frequencies for the top five pathogenic α-globin alleles showed that the –$\alpha^{3.7}$ *HBA1* deletion was most common among all ethnic groups followed by the –$\alpha^{4.2}$ *HBA2* deletion (**Table 3**).

Table 1: Contingency table comparing genotype calls made on MLPA data using manual vs. automated methods

Automated Genotype Calling	Manual Genotype Calling					
	Normal/ non-reportable	1 HBA gene del	2 HBA gene del (cis)	Compound del	Manual Review	Total
Normal/ non-reportable	362	0	0	0	N/A	**362**
1 HBA gene del	0	52	0	0	N/A	**52**
2 HBA gene del (cis)	0	0	13	0	N/A	**13**
Compound del	0	0	0	21	N/A	**21**
Manual Review	0	1	0	0	N/A	**1**
Total	**362**	**53**	**13**	**21**	**N/A**	**449**

Table 2: Cumulative carrier frequencies for all deleterious α-globin CNVs and point mutations detected by MLPA

Ethnicity	N	Carrier Frequency
Ashkenazi Jewish	251	0.080
Asian	677	0.066
Black/African American	500	0.348
Hispanic	746	0.082
White/Caucasian	1524	0.020
Other*	557	0.142
All samples (including no ethnic identity)	**46314**	**0.093**

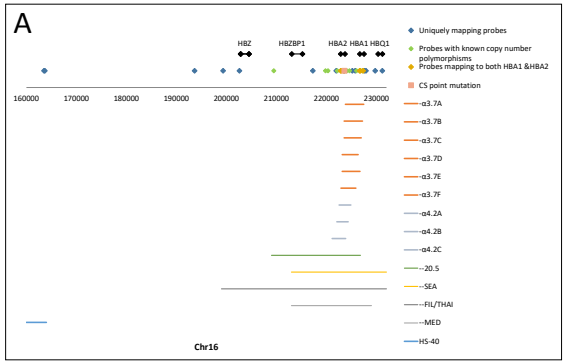
*ethnicity provided but not in one of the above categories

Table 3: Allele frequencies by ethnicity in our cohort for the most commonly detected pathogenic deletions and for HbCS point mutation.

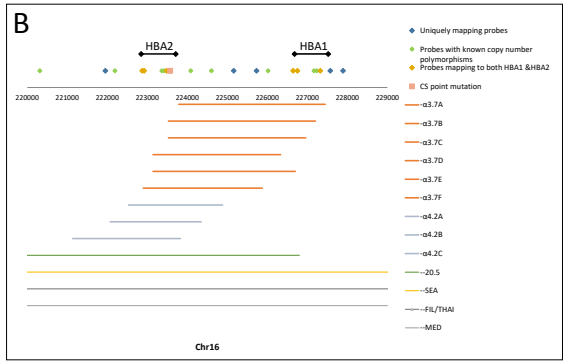
Ethnicity	-α 3.7 deletion	-α 4.2 deletion	--SEA deletion	Hb CS point mutation	HS-40 deletion
Ashkenazi Jewish	**0.0339** (0.0212-0.0536)	**0** (0-0.0076)	**0** (0-0.0076)	**0.0060** (0.0008-0.0127)	**0.0020** (0.0004-0.0112)
Asian	**0.0199** (0.0137-0.0289)	**0.0059** (0.0030-0.0116)	**0.0066** (0.0035-0.0126)	**0.0007** (0.0001-0.0042)	**0** (0-0.0028)
Black/ African American	**0.1790** (0.1565-0.2040)	**0.0020** (0.0005-0.0073)	**0** (0-0.0038)	**0.0010** (0.0002-0.0056)	**0.0010** (0.0002-0.0056)
Hispanic	**0.0416** (0.0326-0.0529)	**0.0013** (0.0004-0.0049)	**0** (0-0.0026)	**0** (0-0.0026)	**0** (0-0.0026)
White/ Caucasian	**0.0095** (0.0066-0.0136)	**0.0007** (0.0002-0.0024)	**0** (0-0.0013)	**0** (0-0.0013)	**0** (0-0.0013)
Other*	**0.0673** (0.0540-0.0836)	**0.0063** (0.0030-0.0129)	**0.0009** (0.0002-0.0051)	**0** (0-0.0034)	**0** (0-0.0034)

*ethnicity reported but not one of the categories listed; numbers shown are **frequencies** (95% CI)

Figure 1: Chr16 α-globin Gene Region, MLPA Probes, and Common Deletions



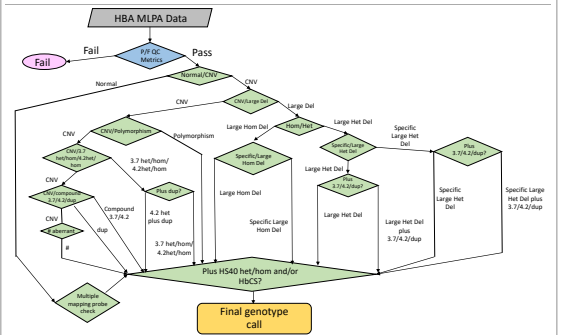
A. Schematic representation of the genes, MLPA probes, and common pathogenic deletions in the α-globin region on chr16:216000-232000 (GRCh37/HG19). The –FIL and –THAI deletions cannot be differentiated by MLPA because they share the same probes and thus, multiplex gap-PCR is used to call these CNVs.



B. Zoomed-in to chr16:222000-229000 to show detail around HBA1 and HBA2 genes.

Figure 2: HBA Genotyping Algorithm Flow Chart

Diagram of the decision tree for the α-globin genotyping algorithm.



Conclusions

- We developed and validated a highly sensitive and specific α-globin genotyping algorithm which was able to rapidly genotype >98% of samples tested by MLPA.
- Of >45,000 samples, 9.3% were positive for at least one pathogenic CNV or point mutation in the α-globin region.
- Based on our cohort, the frequency of pathogenic mutations varied widely by ethnicity. Although our ethnic data were limited to information that was voluntarily provided, the overall allele frequency per ethnic group at 2-34% revealed that variants tested were relatively common in the U.S. population.
- The high rate of positivity supports α-thalassemia testing as part of a universal carrier screening platform.

References:

1. Weatherall DJ, Clegg JB, ed. 2001. *The thalassaemia syndromes, 4th ed*. Blackwell Science, Malden, MA.
2. Product Description SALSA® MLPA® probemix P140-C1 HBA, MRC Holland, April 2017. https://www.mrcholland.com/products/26259/Product%20description%20P140-C1%20HBA-v04.pdf. Accessed April 20, 2018.