Next-generation sequencing carrier screen for Alpha Thalassemia identifies both common and rare variants

Jared R. Maguire, Kevin M. D'Auria, Henry H. Lai, Xin Wang, Clement S. Chu, Imran S. Haque, Eric A. Evans, H. Peter Kang, Dale Muzzey

Counsyl

All Counsyl posters available online at **research.counsyl.com**

All common deletions plus SNPs and indels underlying the most-prevalent disease among Southeast Asians are resolved via high-throughput NGS assay

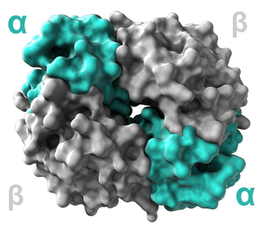
Abstract

Alpha thalassemia (AT)—one of the most common recessive diseases—is characterized by deficient production of alpha globin from the *HBA1* and *HBA2* genes, which have identical tandem coding sequences. Carrier status depends not just on the total number of HBA copies in the genome, but also on copy number per chromosome. Current screening methods for AT are low-throughput and high cost, plus they either cannot reflect the underlying locus architecture or have a low signal-to-noise ratio that makes them error prone. We developed an NGS-based assay that identifies all of the major large deletional variants—including the phase and copy number of *HBA1* and *HBA2* genes—as well as SNPs and short indels, giving it a detection rate near 100%. The assay is entirely automated and could, in principle, process thousands of AT samples on a single Illumina RapidRun flow cell.

What is AT?

Hemoglobin

- Delivers oxygen to cells
- Tetramer of globin proteins — α, β, γ, δ, ε, ζ — each with different oxygen-binding kinetics
- 95% is $\alpha_2\beta_2$



When α chains scarce/mutated

↳ Hemoglobin composition changes (e.g., β_4)
↳ Oxygen binding compromised
↳ α-thalassemia

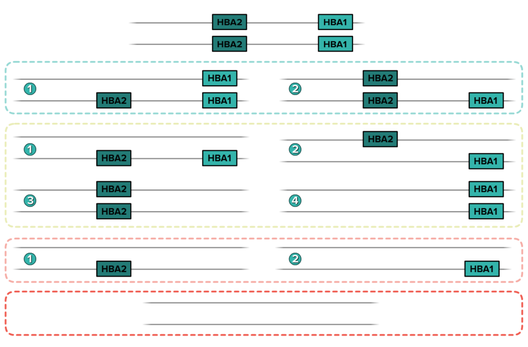
Why screen for AT in DNA?

- AT has carrier rate of up to 1 in 7 among Southeast Asians; 1 in 200 lethally affected pregnancies

Frequency of severe recessive disease in Southeast Asians

- Biochemical assays (e.g., hemoglobin electrophoresis) may not assess heritability of AT, since they do not directly report on the phase of the *HBA1* and *HBA2* genes that encode α globin

HBA copy number drives AT



4 copies
normal

3 copies
silent carrier

2 copies
carrier (cis) or silent carrier (trans)

1 copy
HbH affected

0 copies
HbH Bart's affected

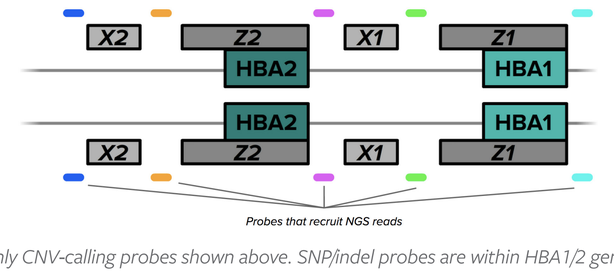
What is standard practice?

- MLPA
- Bead array
- qPCR
- Hemoglobin electrophoresis

Low-throughput, noisy, possibly inaccurate, and/or expensive

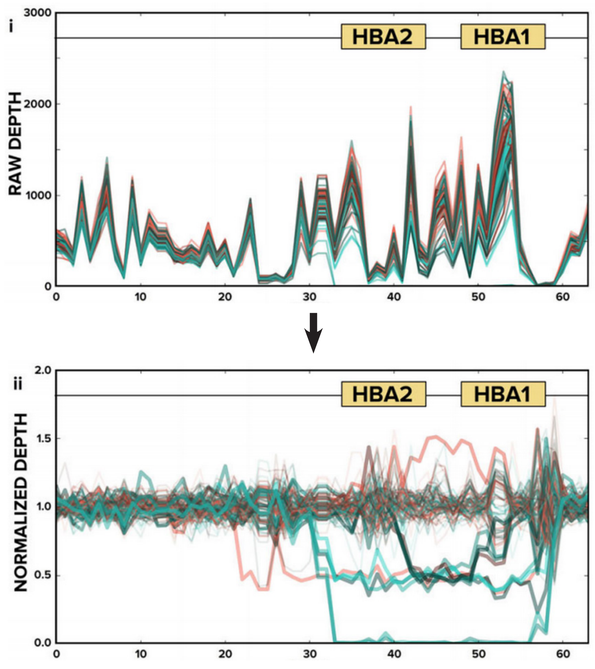
Calling phased CNVs in AT

1. Probe design: Target hybrid-capture probes to areas of uniqueness*, avoiding highly homologous regions X1/2 and Z1/2.



**Only CNV-calling probes shown above. SNP/indel probes are within HBA1/2 genes*

2. Normalization: Scale data across samples and probes to transform depth into copy number.



3. Calling: Compare profile to known common deletions, pick maximum-likelihood match, compute confidence scores, prepare plots for review of calls and no-calls.

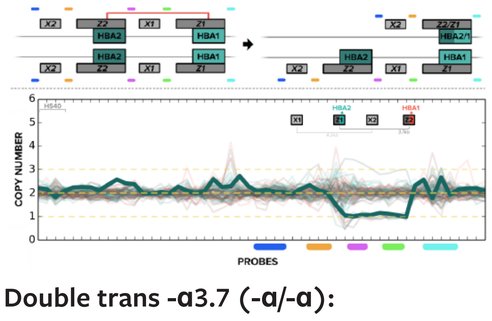
Calling SNPs and Indels in AT

- The majority of AT carriers have CNV mutations, but there are common SNPs and indels, too.
- Disambiguating *HBA1* from *HBA2* is challenging due to their identical coding regions → must use few existing intron and UTR variants.
- Critically, must do CNV analysis first to determine which allele-balance steps to expect (e.g., 50% for CN=2, 33% for CN=3, etc.)
- Recent test of 2500 samples identified 12 samples with deleterious SNPs, ~10% of the number of single - α3.7 deletions.

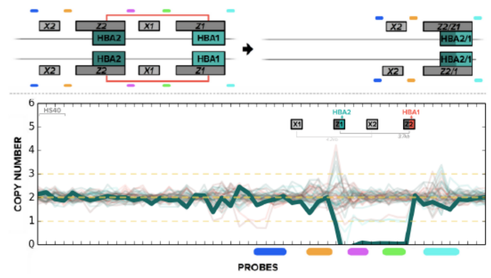
Exon	Gene	Ref	Alt	Variant	Curation
1	*HBA1*	T	C	Start codon	Deleterious
1	*HBA1*	T	A	Start codon	Deleterious
1	*HBA2*	A	G	Hb I	Deleterious
1	*HBA2*	G	A	Hb Caserta	Deleterious
3	*HBA2*	T	C	Hb Quong Sze	Deleterious
3	*HBA2*	G	C	Hb Sun Prairie	Deleterious
3	*HBA2*	T	C	Hb Ethiopia	Deleterious
3	*HBA2*	T	C	Hb CS	Deleterious
3	*HBA2*	T	C	Hb CS	Deleterious
2	*HBA1*	G	A	Hb Riccarton	Deleterious
2	*HBA1*	G	A	Hb Riccarton	Deleterious
3	*HBA1*	C	T	Hb Groene Hart	Deleterious

Sample patient data

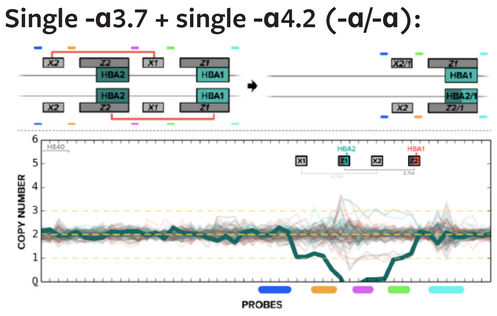
Single -α3.7 (-α/αα):



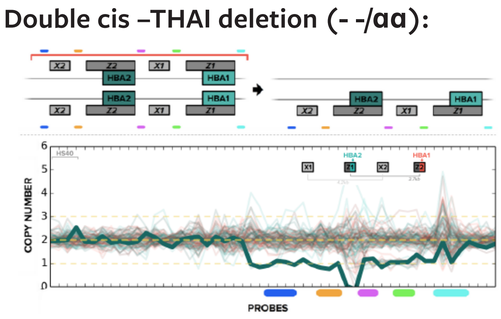
Double trans -α3.7 (-α/-α):



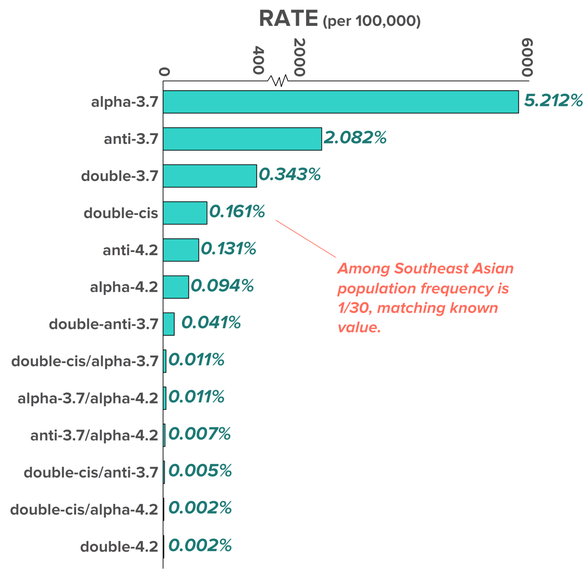
Single -α3.7 + single -α4.2 (-α/-α):



Double cis –THAI deletion (- -/αα):



Bulk frequencies



Conclusion

Carriers of the CNVs, SNPs, and indels underlying alpha thalassemia can be assessed at low cost and high throughput via an NGS-based assay. We have screened more than 100,000 patient samples for CNVs in AT, finding variant frequencies matching those expected in the literature.