Diagnosis of haemoglobinopathies: New scientific advances

Cornelis L. Harteveld

Laboratory for Diagnostic Genome analysis (LDGA), Dept of Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands

Abstract

The molecular defects underlying haemoglobinopathies are both deletions and point mutations in the alpha- or beta-globin genes or gene-clusters. To detect point mutations causing alpha- or beta-thalassaemia, direct sequencing is the method of choice to detect the widest spectrum of molecular defects. The most established approach in DNA diagnostics to screen for the most common deletion defects causing alpha-thalassaemia or beta-thalassaemia is gap-PCR, Multiplex Ligation-dependent Probe Amplification (MLPA) and Sanger Sequencing technology to detect breakpoint sequences of previously uncharacterized deletions/duplications. We demonstrate the recent advances in the determination of duplications and deletions causing alpha- or beta-thalassaemia, using Next Generation Sequencing, array Comparative Genome Hybridization and Target Locus Amplification.

We present three cases in which the use of advanced technologies allow the diagnosis of unexpected disease genotypes.

Introduction

The thalassaemias are a diverse group of disorders of haemoglobin synthesis, all of which result from a reduced output of the alpha- or beta-globin chains of the adult haemoglobin. Despite the great technological advances in mutation detection, the screening of haemoglobinopathies still requires the combined use of haematological and molecular techniques to arrive at an accurate diagnosis. Specialist knowledge is required of genotype/phenotype relationships because of the multitude of complex phenotypes which result from interactions between genotypes and co-inherited globin gene disorders. Recent advances in technology, such as array Comparative Genome Hybridization (aCGH), Target Locus Amplification (TLA) and Next Generation Sequencing (NGS), may help to determine deletion/duplication breakpoints to the sequence level, and provide more insight in rare disease mechanisms.

The NGS approach is now embedded in many genomic laboratories to detect small sequence changes, but is less adapted to deter-

mine structural variants because of the relatively short sequence reads. Target enrichment methods, such as bait capture, are used to obtain more of the breakpoint spanning fragments and use NGS more effectively (Clark *et al.* 2017, 2016). Mechanically sheared genomic DNA is used to create a library of adapter containing fragments (Agilent SureSelect Library preparation), hybridised to biotinylated RNA probes designed to hybridise fragments from the region of interest, selected by streptavidin coated magnetic beads and subsequently subjected to Illumina sequencing (MiSeq).

The arrayCGH cytscan HD array (Affymetrix, Thermo Fisher Scientific, Santa Clara, CA) contains ~2.67 million markers of which ~743000 SNP probes and ~1950000 copy number probes (average spacing 1 probe per 1.1 kb). ArrayCGH technology has been embedded in the cytogenetic lab as a tool to visualize deletions which otherwise are not detectable by microscopic investigation or difficult to detect by FISH analysis. Array CGH has largely replaced these techniques in the genetic lab. By looking at SNP's on the array a distinction can be made in large deletions in chromosomes of paternal or maternal origin.

Targeted sequencing by proximity ligation is used for comprehensive variant detection and local haplotyping. TLA works without prior knowledge of the locus of interest other than the specific primer sequence information necessary to design inverse PCR primer set to amplify and sequence tens to hundreds of kilobases of surrounding DNA. This enables the detection of single nucleotide variants, deletion/duplication/inversion breakpoint sequences and local haplotyping of neighbouring stretches of DNA, which have been cross-linked, digested and ligated to form anchor-containing DNA circles amplified by inverse PCR. The library obtained is subsequently sequenced by NGS (de Vree *et al.* 2014).

Materials and Methods

Patients EDTA blood samples were collected and analysed by standard haematological and biochemical methods. The Hb fractions were separated by HPLC (Trinity Biotech, Menarini) and Capillary Electrophoresis (Capillarys Flexpiercing, Sebia, France). DNA was isolated from leucocytes and samples diluted to a standard concentration of 50 ng/ul.

MLPA kit P140C2 and P102B (MRC-Holland, Amsterdam) were used to detect rearrangements in the alpha- and beta-globin gene clusters resp.

Next Generation sequencing was partly performed in King's College London (MiSeq, Agilent Sure select) and partly in Leiden (HiSeq, GenomeScan, Leiden, The Netherlands). ArrayCGH was performed using the Cytoscan HD array (Affymetrix, Thermo Fisher Scientific provides, Santa Clara, CA, USA); 250ng of DNA was used in the CytoScan™ Automated Target Preparation Solution on NIMBUS™. The array-chip contains ~2.67 million markers of which ~743000 SNP probes and ~1950000 copy number probes (average spacing 1 probe per 1.1 kb).

TLA technology was used from Cergentis BV (Utrecht, NL) and NGS was performed at the Leiden Genome Technology Center (LTGC, Leiden, NL).

Correspondence: Cornelis L. Harteveld, Laboratory for Diagnostic Genome analysis (LDGA), Dept of Clinical Genetics, Leiden University Medical Centre, building 2, S6-P, PO box 9600, 2300 RC Leiden, The Netherlands.
Tel. +31 (0) 71 526 9800; Fax +31 (0) 71 526 8276.
E-mail: c.l.harteveld@lumc.nl ; <http://www.lumc.nl/klingen>

This work is licensed under a Creative Commons Attribution 4.0 License (by-nc 4.0).

©Copyright C.L. Harteveld, 2018
Licensee PAGEPress, Italy
Thalassemia Reports 2018; 8:7473
doi:10.4081/thal.2018.7473

Results

Case #1. This was a collaborative study of Barnaby Clark, Claire Shooter and Swee Lay Thein (King's college, London) in which a family of Syrian ancestry was studied. Mother was a carrier of HBB:c.135delC (cd44(-C)), father showed normal haematology and CE/HPLC results. Two daughters were shown to carriers of beta-thalassemia, but the haematology was more severe than seen in the mother. MLPA analysis revealed a duplication of the alpha-globin gene cluster including the Major Conserved Regions 1 to 4. To determine the orientation of the duplication (head-head, head-tail or tail-tail or translocated) and determine if any intervening sequences between the duplication breakpoints were present, an enrichment step was introduced making a fragment library using bait capture by biotinylated RNA oligo's specific for the alpha-globin gene region, followed by Next Generation Sequencing using the Illumina MiSeq. After breakpoint primers were designed the amplified fragment was sequenced to contain the breakpoint. A head-to-tail arrangement was determined with three ambiguous basepairs between the two duplicated segments of 120,500 bp in length (Clark *et al.* 2016, 2017).

Case #2. A female born as a healthy carrier who inherited just the HBB:c.315+1G>A (IVS2-1g>a) beta-thalassemia mutation from her father presented with a transfusion dependent beta-thalassemia major phenotype later in life. The DNA analysis showed a mosaic for an almost complete hemizygosity for the mutated allele from father. Array-CGH using an array-chip containing ~2.67 million markers of which ~743000 SNP probes and ~1950000 copy number probes (average spacing 1 probe per 1.1 kb), showed the presence of a 5Mb deletion (1,313,791-6,287,277; hg19 / grch37) on the tip of the short arm of chromosome 11, which takes away the maternal allele containing the normal beta-globin gene, alongside with other genes among which the maternally imprinted erythrocyte growth factor (EGF-1) and H19 genes. The mosaic presence suggests a somatic event and the deletion of the maternal allele may explain the growth advantage of cells containing the deletion and the mutated beta-gene in hemizygosity (Traeger-Synodinos *et al.* manuscript in preparation).

Case #3. A Dutch Caucasian family with a novel alpha0-thalassemia deletion was found by haematological and MLPA analysis. The deletion was between 10 and 20 kb based on MLPA results. Targeted Locus Amplification (TLA, Cergentis) was used in combination with capture baits designed to recognize fragments from the alpha- and beta-globin gene clusters to enrich for the breakpoint fragment in isolated DNA of the carrier, followed by NGS. The breakpoint sequence was confirmed by designing gap-PCR primers and subsequent amplification and direct sequencing of the breakpoint fragment. This novel alpha0-thalassemia deletion is called —Jc, it is 16.7 kb in length involving the HBM, HBA1 and HBA2 gene (NG_000006.1:g.151901_168673del16772). (Harteveld *et al.*, manuscript in preparation).

Discussion

The molecular defects underlying these disorders are both

deletions and point mutations in the alpha- or beta-globin genes or gene-clusters. To detect point mutations causing alpha- or beta-thalassaemia, direct sequencing is the method of choice to detect the widest spectrum of molecular defects. The most established approach in DNA diagnostics to screen for the most common deletion defects causing alpha-thalassaemia or beta-thalassaemia is gap-PCR, Multiplex Ligation-dependent Probe Amplification (MLPA), and automated Sanger sequencing. The new technologic advances such as micro-array and Next Generation Sequencing technology make possible new scientific advances in the identification of disease mechanisms involved in haemoglobinopathy. This is demonstrated by three cases presented.

Complex interactions involve duplications of the alpha-globin gene cluster which by overexpression of the alpha-globin genes influences the disease phenotype in beta-thalassemia carriers expressing a beta-thalassaemia intermedia. Recent advances in technology including Next Generation Sequencing (NGS) were used to characterize the duplication breakpoint.

Late onset beta-thalassemia is an extremely rare phenomenon, in which a healthy beta-thalassemia carrier develops a transfusion dependent thalassaemia intermedia during life. DNA analysis reveals a mosaic loss of heterozygosity in the majority of blood cells which is absent in DNA isolated from hair or buccal cells. This is suggestive of a somatic deletion of the wild type maternally inherited beta-gene in a (subset of) hematopoietic stem cell(s) which gradually replaced the heterozygous cells in time. By using array Comparative Genome Hybridization (aCGH) the deletion length, the level of mosaicism and the positioning of the deletion to the maternal chromosomal could be identified.

In the last example, we presented a novel technology used to enrich for fragments containing the deletion breakpoint of a rare novel deletion found in a family of Dutch origin. Target Locus Amplification (TLA) was used to crosslink neighbouring DNA stretches surpassing tens to hundreds of kb, in order to target the deletion breakpoint by inverse PCR and subsequent Next Generation Sequencing.

References

- Clark BE *et al.* Beta thalassaemia intermedia due to co-inheritance of three unique alpha globin cluster duplications characterised by next generation sequencing. Br J of Haematol. 2016, doi:10.1111/bjh.14294
- Clark BE *et al.* Next-generation sequencing as a tool for breakpoint analysis in rearrangements of the globin gene clusters. Int J Lab Hem 2017; 39(suppl.1):111-120.
- Traeger-Synodinos J *et al.* EMQN: Best Practice Guidelines for Hemoglobinopathies carrier detection and prenatal diagnosis. Eur J of Hum Genet 2015.
- De Vree PJP *et al.* Target sequencing by proximity ligation for comprehensive variant detection and local haplotyping. Nature Biotechnology; published on-line 17 August 2014; doi:10.1038/nbt.2959