Blood Cells, Molecules, and Diseases 49 (2012) 121–127

Contents lists available at SciVerse ScienceDirect

Blood Cells, Molecules, and Diseases

journal homepage: www.elsevier.com/locate/bcmd

A novel 506 kb deletion causing εγδβ thalassemia

Helen Rooks [a], Barnaby Clark [b], Steve Best [a], Peter Rushton [b], Matthew Oakley [b], Onn Shaun Thein [b,1], Ann C. Cuthbert [c], Alison Britland [c], Afruj Ruf [c], Swee Lay Thein [a,b,*]

[a] King's College London, Molecular Haematology, James Black Centre, London SE5 9NU, UK
[b] King's College Hospital NHS Foundation Trust, Department of Haematological Medicine, London SE5 9RS, UK
[c] Airedale General Hospital, Keighley, West Yorkshire BD20 6TD, UK

ARTICLE INFO

Article history:
Submitted 30 March 2012
Available online 5 June 2012

(Communicated by Sir D. Weatherall, F.R.S., 4 May 2012)

Keywords:
εγδβ thalassemia
Palindrome
Array CGH
MLPA (Multiplex Ligation-dependent Probe Amplification)

ABSTRACT

We describe a novel deletion causing εγδβ thalassemia in a Pakistani family. The Pakistani deletion is 506 kb in length, and the second largest εγδβ thalassemia deletion reported to date. It removes the entire β globin gene (*HBB*) cluster, extending from 431 kb upstream to 75 kb downstream of the ε globin gene (*HBE*). The breakpoint junction occurred within a 160 bp palindrome embedded in LINE/LTR repeats, and contained a short (9 bp) region of direct homology which may have contributed to the recombination event. Characterization of the deletion breakpoints has been particularly challenging due to the complexity of DNA deletion, insertion and inversion, involving a multitude of methodologies, mirroring the changing DNA analysis technologies.

Introduction

εγδβ thalassemias are rare and found only in the heterozygous form [1]. They are caused by large deletions in the β globin gene (*HBB*) cluster and are classified molecularly into two groups. Group I deletes all or most of the β globin cluster, including the *HBB* gene, while group II deletions remove the upstream β locus control region (β LCR) but leaves the *HBB* gene itself intact. Adult heterozygotes have a hematological phenotype similar to that of β thalassemia trait, but HbA_2 and HbF levels are normal and the red blood cells tend to be relatively more hypochromic and microcytic [2]. Newborns have anemia and hemolysis; in some, the anemia is life-threatening requiring intra-uterine and peri-natal blood transfusions to tie them over the neonatal period.

Here we describe the characterization of a novel εγδβ thalassemia deletion in a Pakistani family, named Pakistani I. The Pakistani I deletion which is 506 kb in size is one of only four deletions in total that are over 400 kb, the other three being Jpn I (1.4 Mb), Eng IV (439 kb), and Dutch II (>400 kb) [2–4]. To date, 29 εγδβ thalassemia deletions (including the present case) have been reported, but many of the early deletions have not been completely characterized because of incomplete sequence data and technical difficulties due to the nature of the deletion breakpoints. Characterization of the Pakistan I deletion first started two decades ago and has encompassed a multitude of methodologies – Southern blot hybridization, quantitative PCR [5], multiplex ligation-dependent probe amplification (MLPA) [3], high resolution array comparative genome hybridization (aCGH) [6], gap PCR, and DNA sequence analysis, which have combined to give a thorough and complete picture of the deletion rearrangement. This case illustrates the technical challenges of defining the exact breakpoints of deletions which occur within repetitive regions and highlights the need for a single comprehensive methodology that can fully characterize large deletions and rearrangements in a single process.

Materials and methods

Blood samples and hematological studies

Whole blood samples were collected in EDTA. Peripheral blood counts and hematological parameters were determined using an automated cell counter. Hemoglobin analyses were performed on HPLC Variant II (Bio-Rad Laboratories, Hercules, CA, USA) using the β Thalassemia Short Program.

Southern blot hybridization and pulsed field gel electrophoresis

DNA was extracted from peripheral blood leucocytes or lymphocyte EBV transformed cell lines using phenol chloroform. Restriction enzyme digestion was performed on the DNA and analyzed by conventional agarose or pulsed field gel electrophoresis, and Southern blot hybridization using probes designed from the chromosome 11p15 sequence made available through the Human Genome Project (http://www.ncbi.nlm.

* Corresponding author at: Molecular Haematology, King's College London, James Black Centre, 125 Coldharbour Lane, London SE5 9NU, UK. Fax: +44 20 7848 5444.
E-mail address: sl.thein@kcl.ac.uk (S.L. Thein).
[1] Present address: University of Oxford, Christchurch College, UK.

1079-9796/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.bcmd.2012.05.010

nih.gov, http://www.ensemble.org). PCR primers were designed using Oligo 6 program and selected regions of the sequence specifically amplified as hybridization probes. The probe sequences were specific and free of repeats. Pulsed field gel electrophoresis (PFGE) followed a previously described methodology [7,8], using high molecular weight DNA isolated in 1% agarose blocks digested with *Sfi*I or *Sal*I.

Quantitative PCR

Using sequence taken from UCSC genome browser (2004, hg17), primers were designed by Oligo 6.1 program to produce 80 bp amplicons positioned at strategic points along the chromosome 11p in the region of *HBB* cluster. The primers were blasted against the human genome sequence to check for specificity and absence of repeats. A 25 μl PCR reaction volume (SybrGreen 12.5 μl, 5 pmol each of the forward and reverse primers, 50 ng DNA, distilled water) was run on the ABI 7900 under the following thermal cycling conditions: 2 min at 50 °C, 10 min at 95 °C, 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Reactions were performed in triplicate and the average C_t value calculated. Data was normalized against a control probe and a control sample, and was converted to a fold change value by the comparative C_t method using the formula $2-\mathrm{ddCt}$ [9] where control sample has a value of 1.0 (represented by the presence of two alleles), a value range of 0.9–1.2 was acceptable as equal to the presence of two alleles. A value range of 0.2–0.6 was accepted as the presence of one allele. Values 0.7 and 0.8 were considered uninformative.

Multiplex ligation-dependent probe amplification (MLPA)

MLPA reactions were performed according to the manufacturers' protocols using MRC Holland kits HP-140 and HP123 for analysis of the alpha and beta globin loci, respectively [3]. Products were separated by capillary electrophoresis on the ABI 3130 (Applied BioSystems) and data analyzed using GeneMarker (SoftGenetics, USA). Threshold ratios for deletion and duplication were set at <0.75 and >1.3, respectively.

Array comparative genome hybridization (aCGH)

DNA extracted using the phenol chloroform method was further cleaned using the Qiagen kit with an RNase step to remove RNA. All DNA was quantified on a nanodrop 1000 prior to analysis.

Array design

The custom oligonucleotide microarray was designed using Agilent e-array software (Agilent Technologies, Santa Clara, CA), by creating 60 mer probes. An 80 kb region containing the β globin gene cluster including the β LCR and the five beta-like globin genes on chromosome 11p was tiled out with 1231 probes. This region was then further extended using validated probe sets, available from Agilent, to cover a 2 Mb region with the 80 kb *HBB* core region at the centre. Repeat regions were avoided by the software in order to keep the data as reproducible as possible. For the alpha globin locus on chromosome 16, a similar approach was taken except the 2 Mb region extended from the tip of chromosome 16p. The design file was then uploaded to the University of California Santa Cruz genome database [Human Genome Browser, March 2006 assembly: hg18, NCBI build 36.1] and the designed probes checked for coverage. Within the alpha globin gene region there were four gaps of more than 5 kb due to the presence of repeat elements. The software was then used to fill each of these gaps with oligonucleotides, resulting in no gaps greater than 2 kb between probes within the globin gene regions. The density of probes within the regions flanking the globin gene was variable but on average there was a probe every 1,000 bp.

Microarray hybridization and analysis

All steps were performed as outlined in the Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis protocol (V6.2). DNA double digested with *Alu*1 and *Rsa*1 was labeled by random priming using the Agilent Genomic DNA Enzymatic Labeling Kit (Agilent Technologies) for 2 h at 37 °C. Patient samples were labeled with Cy3-dUTP and control samples with Cy5-dUTP, and were purified using Micron YM-30 filters (Millipore, Billerica, MA). Patient and control labeled DNA was then mixed and hybridized with human Cot I DNA (Invitrogen) at 37 °C for 30 min prior to hybridizing to the array. Array slides were then loaded with patient/control mixed samples and placed in a hybridization chamber which was placed in a rotary oven at 65 °C for 16 h, turning at 15 rotations per minute. Arrays were washed according to wash procedure B and scanned using an Agilent 0.2 μ scanner. Data analysis was performed using feature extraction software (Agilent v9.1.3.1), and Genomic Workbench software v 3.5 (Agilent Technologies). Copy number aberrations were detected with the built in Aberration Detection Method 2 (ADM-2) algorithm using a threshold value of 6.0. As patient samples were labeled with Cy3 and controls with Cy5 the color coding was reversed by the software so that deletions appeared green and duplications red.

Characterization of the deletion breakpoints by gap PCR and DNA sequencing analysis

Genomic DNA encompassing the deletion breakpoints was amplified by PCR using specific primers in a 20 μl reaction volume containing 300 ng DNA, 10 pmol each of the forward and reverse primer, 10 μl Bioline Bio-X-Act Long mix and 2.0 mmol/L $MgCl_2$. The cycling parameters were 95 °C for 3 min followed by 35 cycles of 95 °C for 45 s, 60 °C for 1 min, 72 °C for 4 min with a final extension at 72 °C for 5 min. The PCR products were resolved by electrophoresis in a 1% agarose gel and the specific fragments purified, cloned, then sequenced using ABI Big Dye Sequencing Chemistry (Terminator version 3.1). The sequence data was analyzed on sequencer version 4.1. All UCSC co-ordinates were taken from 2004 assembly.

Results

Identification of εγδβ thalassemia in a Pakistani family

The proband was a normal full-term male infant of Pakistani descent. Three months after birth, anemia was detected with Hb 8.4 g/dL and normal serum ferritin level of 231 μg/mL. He was referred for further investigation of his anemia at 2 years of age when full blood counts showed Hb 8.3 g/dL, RBC 4.41×10^{12}/L, with MCV 54.4 fL, MCH 18.8 pg, and HbA_2 2.8%. His mother had a similar hematological phenotype with Hb 10.0 g/dL, RBC 5.10×10^{12}/L, with MCV 59.6 fL, MCH 19.6 pg, and HbA_2 3.2%, while the father had completely normal hematological profile with Hb 16.5 g/dL, RBC 5.52×10^{12}/L, with MCV 85.3 fL, MCH 29.9 pg, and HbA_2 2.3%. Globin chain synthesis ratios were imbalanced in proband (non-α/α 0.23) and mother (non-α/α 0.43) but balanced in father (non-α/α 0.95). Thus hematological parameters of the proband and mother resembled the β thalassemia trait except that HbA_2 and HbF levels were within normal limits suggesting heterozygosity for εγδβ thalassemia.

Analysis of restriction fragment length polymorphisms (RFLPs) in the *HBB* complex showed that the proband was homozygous for the presence of *Hind* III-Gγ site (i.e. +/+) having inherited one site from his father (*Hind* III-Gγ +/−) but neither of the maternal *Hind* III-Gγ (−/−) RFLPs. This can be explained by the proband and the mother being hemizygotes for *Hind* III-Gγ site, i.e. proband is (+/) and mother is (−/), in support of heterozygosity for εγδβ thalassemia. Deletion of one copy of the *HBB* complex was confirmed by fluorescence in situ hybridization (FISH) analysis of the mother's cells using *HBB* cosmid probe which showed that hybridization signal was present in only one copy of chromosome 11p (Fig. 1).