ORIGINAL ARTICLE

International Journal of Laboratory Hematology   WILEY

Breakpoint characterization of a rare alpha^0-thalassemia deletion using targeted locus amplification on genomic DNA

Quint P. Hottentot[1] 🄳 | Emile de Meijer[2] | Henk P. J. Buermans[2] | Stefan J. White[2] | Cornelis L. Harteveld[1]

[1]Department of Clinical Genetics, Leiden University Medical Centre, Leiden, the Netherlands

[2]Leiden Genome Technology Centre, Leiden University Medical Centre, Leiden, the Netherlands

Correspondence
Cornelis L. Harteveld, Department of Clinical Genetics, Leiden University Medical Centre, Leiden, the Netherlands.
Email: c.l.harteveld@lumc.nl

Abstract

Introduction: The high-sequence homology of the α-globin-gene cluster is responsible for microhomology-mediated recombination events during meiosis, resulting in a high density of deletion breakpoints within a 10 kb region. Commonly used deletion detection methods, such as multiplex ligation-dependent probe amplification (MLPA) and Southern blot, cannot exactly define the breakpoints. This typically requires long-range PCR, which is not always successful. Targeted locus amplification (TLA) is a targeted enrichment method that can be used to sequence up to 70 kb of neighboring DNA sequences without prior knowledge about the target site.

Methods: Genomic DNA (gDNA) TLA is a technique that folds isolated DNA, ensuring that adjacent loci are in a close spatial proximity. Subsequent digestion and religation form DNA circles that are amplified using fragment-specific inverse primers, creating a library that is suitable for Illumina sequencing.

Results: Here, we describe the characterization of a rare 16 771 bp deletion, utilizing gDNA TLA with a single inverse PCR primer set on one end of the breakpoint. Primers for breakpoint PCR were designed to confirm the deletion breakpoints and were consequently used to characterize the same deletion in 10 additional carriers sharing comparable hematologic data and similar MLPA results.

Conclusions: The gDNA TLA technology was successfully used to identify deletion breakpoints within the alpha-globin cluster. The deletion was described only once in an earlier study as the --^{gb}, but as it was not registered correctly in the available databases, it was not initially recognized as such.

KEYWORDS
clinical genetics, hemoglobinopathy, molecular diagnostics, targeted sequencing, thalassemia, TLA

1 | INTRODUCTION

Hemoglobinopathies (HbPs) are the most common monogenic disorders in the world, with α-thalassemia frequencies above 1% in all studied tropical and subtropical populations.[1] α-Thalassemias are characterized by the reduced synthesis of the α-globin chain of hemoglobin A (HbA; α2β2). Carriers of α^+ thalassemia (-α/αα) have mild microcytic hypochromic anemia, while homozygosity for α^+ thalassemia (-α/-α) and heterozygosity for α^0-thalassemia (--/αα) show moderate microcytic hypochromic anemia. The presence of only a single α-globin (HBA) gene (--/-α) will express HbH disease, which is variable in clinical severity, ranging from mild to transfusion-dependent hemolytic anemia. Absence of functional α genes, α^0-thalassemia (--/--), will lead to Hb Bart's Hydrops fetalis syndrome, which is the most severe form of α-thalassemia and is not compatible with life. The high global frequencies, severity of the disease and the serious risk for the pregnant due to in utero death of the hydropic fetus makes early detection and prevention of α-thalassemias of great clinical importance. Furthermore, despite the high prevalence, physicians of nonendemic countries rarely consider HbPs, which is unjustified due to the demographic changes during the last few decades.[2,3]

The most common molecular causes of α-thalassemia are deletions involving one or both of the duplicated α-globin genes,[4] of which the seven most prevalent (-$\alpha^{3.7}$, -$\alpha^{4.2}$, -(α)$^{20.5}$, --^{MedI}, --^{SEA}, --^{FIL}, and --^{THAI}) are screened for by means of gap-PCR during routine molecular diagnostics.[5] Less common large deletions of unknown length were detected with Southern blot or Fluorescent In Situ Hybridization (FISH) analysis in the past, but these methods were largely replaced by Multiplex ligation-dependent probe amplification (MLPA) more than a decade ago.[6] None of MLPA, Southern blot, or FISH gives information about the exact breakpoints of structural deletions. This requires gap-PCR with primers on either side of the deletion to generate a breakpoint fragment for sequencing. Random design of gap-PCR primers is hampered by the distance between the first and last MLPA probe involved in the deletion and the one still present, a distance which can be as large as 10-20 kb depending on the density of the MLPA probes in the region of interest. This approach is inefficient as the result depends on many factors such as the distance between primers, PCR conditions, and nature of the breakpoint region, such as presence of repetitive elements, GC content, and T_m.

During this study, we have used a relatively new technique named targeted locus amplification (TLA) to pinpoint deletion breakpoints in the α-globin-gene cluster, allowing accurate primer design for gap-PCR which can be used as a low cost method for breakpoint determination. TLA results in a highly enriched sequence around the designed inverse primers, allowing detection of all structural variants including deletion breakpoints within 70 kb around the target (gDNA TLA application note: https://cergentis.docsend.com/view/qq2bcwb).

Here, we describe the breakpoint characterization of an α^0-thalassemia deletion found in one patient using gDNA-based TLA and in several patients using gap-PCR.

2 | MATERIALS AND METHODS

DNA was isolated from peripheral blood using a standard salting out procedure.[7] Hematologic data were obtained using standard methodology.[8]

2.1 | Targeted locus amplification

TLA cross-links and fixes loci within close spatial proximity. Due to stochastic variation, subsequent digestion and religation lead to DNA circles with different DNA fragments from around the breakpoint locus. The DNA circles are amplified by means of inverse primers, which are located in the anchor fragment, and the DNA library is prepared using the Nextera XT-sample preparation kit for MiSeq sequencing.[9]

TLA was originally developed for cultured cells, but since no cells were available and fresh material could not be obtained, we have used a proprietary assembly mix from Cergentis to fold the fragmented, isolated DNA prior to the fixation. This means that the obtained range of coverage depends on the fragment length of the degraded DNA, which had been stored for 6 years at 4°C in this particular case. During TLA we wished to capture and analyze DNA segments that surround a selected site (anchor fragment) situated as close to the breakpoint as possible. Custom inverse primers (forward 5′-TGGTGGTACAGCCCTTATCTG-3′ and reverse 5′-CTCAGCACCCATCCTGTCTAC-3′) were carefully designed at the borders of two NlaIII sites using Primer3plus,[10] forming the anchor fragment. The anchor fragment is located at the border of the last MLPA probe at the 5′ breakpoint side still present (chr16:159,469-159,541), to ensure incorporation of the breakpoint-containing fragment in the DNA circles. No inverse primers were designed within the region of the 3′ breakpoint side, as unique primer design was hampered by high sequence homology. Library preparation was conducted using the Illumina Nextera XT library-preparation kit followed by sequencing on the Illumina MiSeq platform. The UCSC Genome Browser[11] and Integrative Genomics Viewer[12] on Human Hg38 assembly were used for localization, visualization of read alignment, and deletion detection.

2.2 | Primer design and breakpoint PCR

Based on the TLA results, the 5′ breakpoint gap-PCR primer was designed upstream of position 161,200, while the 3′ breakpoint gap-PCR primer was designed downstream of position 178,674 (Ju-del forward 5′-AACCACGAGCCACCATGT-3′ and Ju-del reverse: 5′-CCACCACATTTTGTTTACCC-3′, respectively). Amplification was performed with a T-Professional thermal cycler (Biometra) in a final volume of 15 μL using 50 ng DNA template (measured with NanoDrop), 5 pmol of each primer, 3 mM dNTP's (Thermo Fisher Scientific), and 0.6 U DNA polymerase per reaction. A hot start PCR was used running the following program: Hot start: 95°C, 1 cycle

initial denaturation: 10 minutes 95°C, 30 cycles of 45-second denaturation at 94°C, 45-second annealing at 60°C, and 30-second elongation at 72°C, followed by 10 minutes of elongation at 72°C. PCR-products were analyzed using the LabChip GX (PerkinElmer) and sequenced by Sanger sequencing on the ABI 3730XL.

2.3 | Patients

Patients were referred to the Hemoglobinopathy Reference Laboratory of the Department of Clinical Genetics (LUMC) as suspected of carrying thalassemia. After giving informed consent, routine diagnosis for hemoglobinopathies was done using standard hematology, Hb-separation, and molecular analysis as described previously and according to the European Molecular Genetics Quality Network (EMQN) recommendations.[13,14] DNA material was collected from the index patient and the 10 selected patients carrying a similar deletion based on previous hematologic and MLPA analysis.

3 | RESULTS

3.1 | Targeted locus amplification

The index patient was preselected for the α^0-thalassemia phenotype based on a microcytic (MCV 61 fL) hypochromic (MCH 18.7 pg) anemia (Hb 8.9 g%) in the presence of Inclusion Bodies. As all common variants and deletions were excluded, subsequent MLPA analysis was performed using the MRC-Holland P140B kit. This revealed a 5′ breakpoint between chr16:159,541 and chr16:162,720 (3179 bp region) and a 3′ breakpoint between chr16:177,962 and chr16:179,774 (1812 bp region), removing the *HBZP*, *HBM*, *HBA1P*, *HBA2*, and *HBA1* genes while leaving *HBZ* and *HBQ* intact, as depicted in Figure 1A.

An inverse TLA primer pair was designed in the region between the last MLPA probe present and the first absent (chr16:159,469-159,541). The anchor fragment (chr16:159,665-160,999) was as close as possible to the undefined 5′ breakpoint deleted region. Several attempts were made to design an additional primer set at the 3′ end, but due to dependency on *NlaIII* sites for the anchor fragment and the highly repetitive nature of the 3′ breakpoint target sequence, no eligible sites were located.

MiSeq sequence data obtained after TLA showed a normal distribution of reads adjacent to the inverse primers, caused by the religation of fragments within close proximity to the anchor (Figure 2A). While only mapping 8 kb surrounding the anchor fragment, an abrupt drop of reads is visible at 162 250, which align with a *NlaIII* site. This is caused by the lower ligation efficiency of more distal fragments, which are therefore, less likely to be included within the DNA circle.

We were able to identify a reappearing peak at 178 673 kb (Figure 2A). The reappearing peak has a read depth of 100+ reads, suggesting that this fragment is near the anchor during TLA, while the linear genomic distance in the absence of a deletion would have

been approximately 17 kb. This pattern is indicative of the presence of a deletion, of which the reappearance of reads at 178 673 shows the exact 3′ breakpoint.

3.2 | Gap-PCR

Gap-PCR followed by Sanger sequencing was performed to confirm the exact breakpoint sequence. The 5′ breakpoint is located between 161,901 and 161,910, while the 3′ breakpoint is located between 178,672 and 178,681 (±16 771 bp del), with GGCCGGGC as overlapping homologous sequence present at both sides, as shown in Figure 2B. Initial TLA, MiSeq sequencing, gap-PCR, and Sanger sequencing were performed on one patient with a thalassemia phenotype. Ten nonrelated cases with an equivalent α^0-thalassemia trait phenotype and similar MLPA deletions were screened with the same set of gap-PCR primers. Electrophoresis of the PCR fragments (Figure 3) and confirmation by Sanger sequencing demonstrated that all of the 10 unrelated individuals had the exact identical breakpoint sequence, suggesting a common founder for this α^0-thalassemia deletion in these individuals.

3.3 | Dot-plot analysis

Dot-plot analysis of the 5′ breakpoint region vs the 3′ breakpoint region (Figure 1B) revealed an extremely high homology in the α-cluster due to the high density of Short Interspersed Elements (SINEs) (Alu repeats) and Long Interspersed Elements (LINEs). Additionally, nearly all hemoglobin genes in this region are paralogous, contributing to homology between coding and noncoding regions. Interestingly, both breakpoints are located next to an Alu repeat.

4 | DISCUSSION

4.1 | Targeted locus amplification

During this study, we have used the gDNA-based TLA approach on native (isolated) DNA instead of cells to identify the exact breakpoints of a deletion in the α-globin cluster in an index α-thalassemia carrier. Gap-PCR primers were designed to verify the deletion and to confirm the presence of an identical deletion in 10 nonrelated individuals which showed similar MLPA results.

The gDNA-based TLA is an interesting alternative to WGS and long-range PCR, where success can be limited by highly repetitive and GC-rich loci such as the alpha-globin-gene cluster. While WGS generates much more sequence information, overall coverage usually is low, making it especially challenging to detect structural variants in difficult DNA templates. In contrast to the use of RNA fishing baits, TLA requires no prior information on potential sequence variants in the region. The gDNA-based TLA protocol was updated and
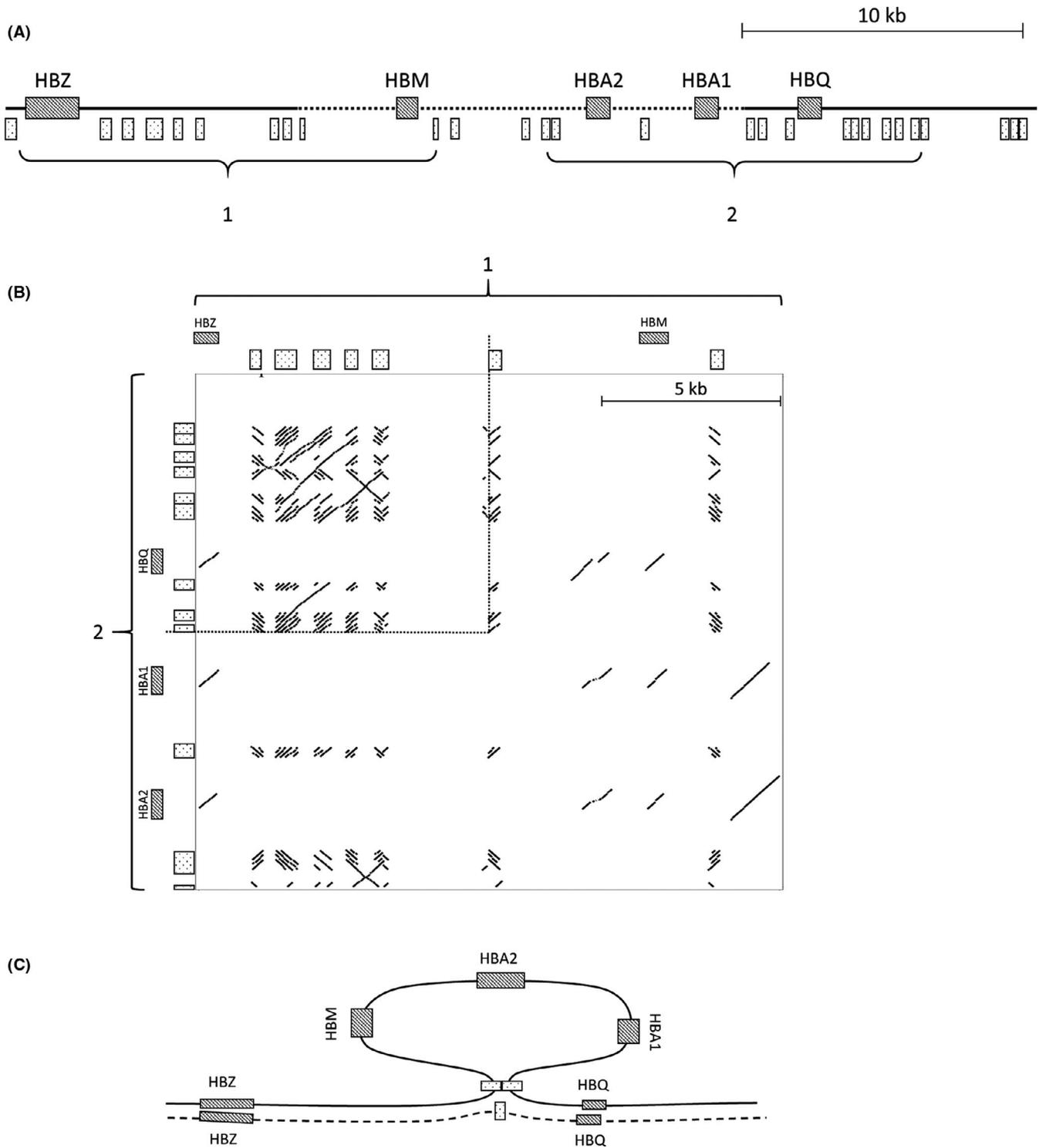
FIGURE 1 A, Schematic representation of the α-globin-gene cluster. B, Dot plot analysis of both the breakpoint sequences. C, A schematic representation of the loop formation during meiosis

fine-tuned during the project, meaning that the current sequence mapping distances for the gDNA protocol could be increased.

The cell-based TLA is preferred when cells are available since the range of coverage can be orders of magnitude greater.[15] DNA samples are typically more readily available, however, especially for historical cases.

Theoretically, TLA should be able to pinpoint the exact breakpoint locations. An additional breakpoint PCR was needed in this particular case, most likely due to the highly repetitive nature of the target sequence that made correct alignment of MiSeq reads challenging. Furthermore, the range of coverage, mapping only 8 kb flanking the anchor, was lower than expected since gDNA TLA
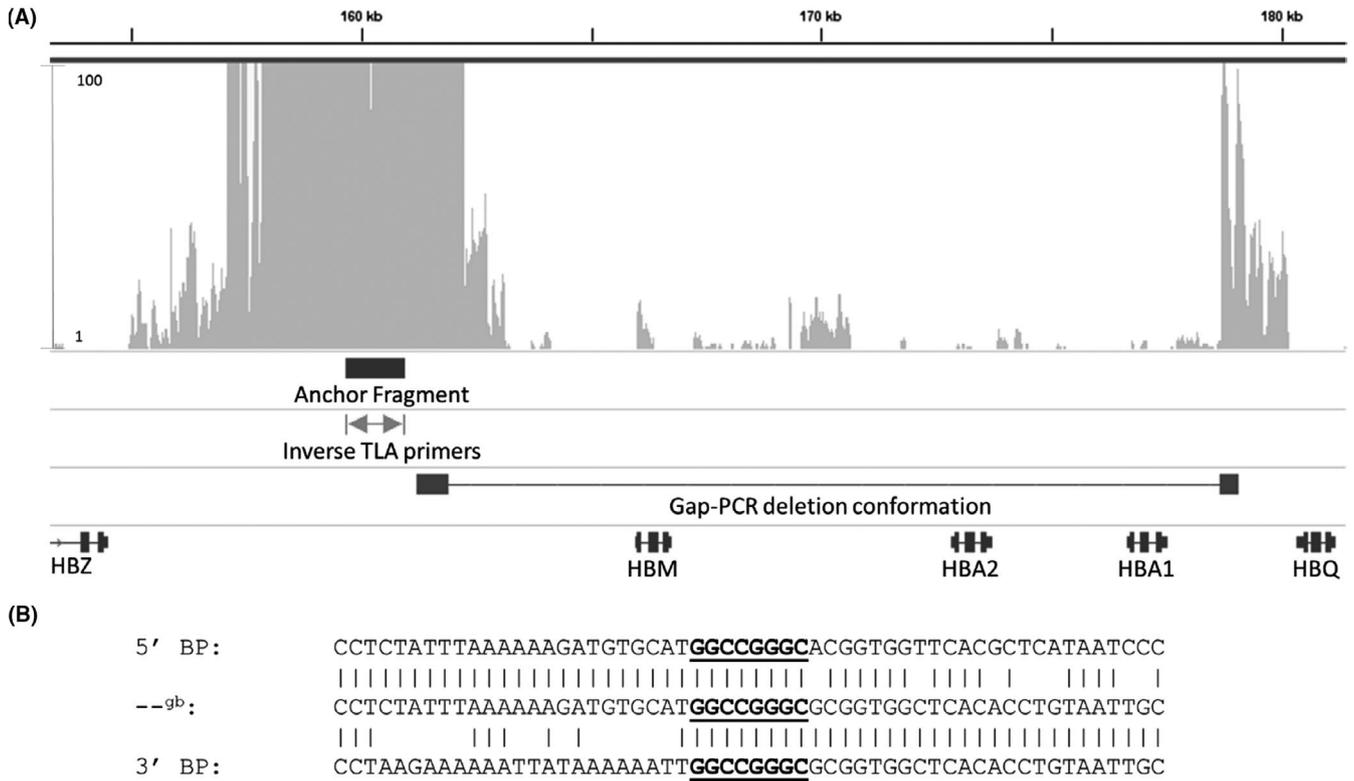
(A)

(B)

```
5' BP:      CCTCTATTTAAAAAAGATGTGCATGGCCGGGCACGGTGGTTCACGCTCATAATCCC
            ||||||||||||||||||||||||||||||||||||| ||||||| ||||| |    |||| |
--gb:       CCTCTATTTAAAAAAGATGTGCATGGCCGGGCGCGGTGGCTCACACCTGTAATTGC
            |||     ||| | |    ||||||||||||||||||||||||||||||||||||||
3' BP:      CCTAAGAAAAAATTATAAAAAATTGGCCGGGCGCGGTGGCTCACACCTGTAATTGC
```

FIGURE 2 A, An IGV view showing MiSeq sequence coverage from TLA. B, The Sanger sequence data is compared to the 5' and 3' breakpoint, showing the breakpoint region. IGV, integrative genomics viewer; TLA, targeted locus amplification
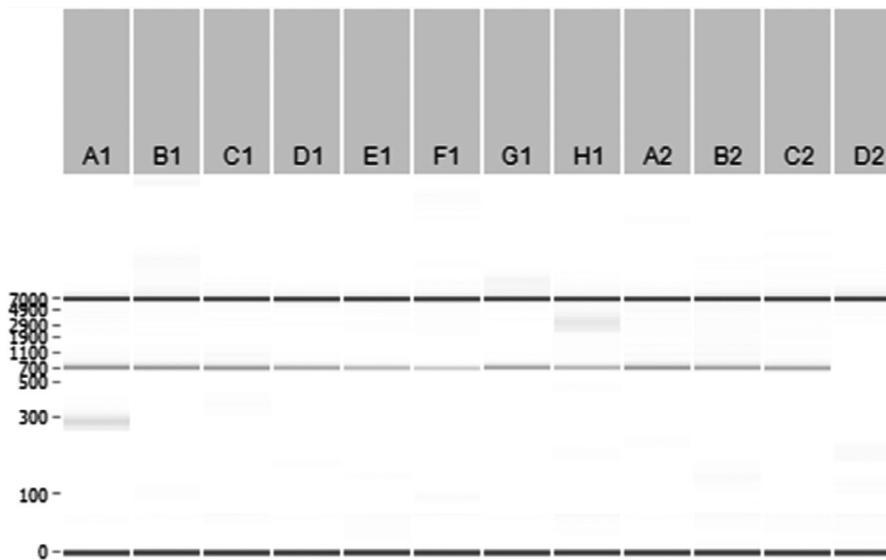

FIGURE 3 LabChip results of amplicons obtained with gap-PCR

should be able to cover 20-50 kb on either side of the primer, as described in the gDNA TLA application note. This might be lower in isolated DNA and is expected to be even lower by the poor quality of the DNA sample, which was stored at 4°C for 6 years and may be partially degraded. This, in combination with the α-globin cluster isochore region impeding correct alignment of reads, may explain the low coverage.

Despite all this, we were able to detect the breakpoints using the gDNA TLA, which is a great advantage over cell-based TLA as for many unresolved genomic deletions only gDNA is stored.

4.2 | Deletion annotation

The deletion initially appeared to be novel, but after the determination of the breakpoint sequence, we have found that this deletion is previously described in earlier articles, where the breakpoints were characterized and the deletion named --^{gb}. While Phylipsen et al,[16] had characterized the --^{gb} deletion to be 16 771 bp in three Dutch individuals, the corresponding Ithanet page: IthaID: 3292, describes a 15 kb deletion. Additionally, Mota et al,[17] found this deletion in a Brazilian patient solely by MLPA and registered this deletion as the 15.2 kb --^{gb} on Ithanet (IthaID: 3296) without determining the breakpoint. Consequently, it is not certain that the deletion found in the Brazilian patient is identical to the --^{gb} deletion. This illustrates the importance of correct registration of novel variants to the appropriate database, as discussed in Giardine et al.[18] This also demonstrates the necessity of a protocol regarding submission of deletions where the breakpoints are undetermined, to prevent duplicated registrations concerning the same mutation.

Since this deletion was not registered in HbVar and incorrectly in the Ithanet database, we were led to believe that we were dealing with a novel deletion, reinforcing the need for timely and accurate submission to the variant databases in general. We have now submitted the breakpoint locations of the --^{gb} deletion to both databases, according to reference sequence GRCh38/Hg38.

4.3 | Deletion mechanism

Alu-sequences belong to the retrotransposon family of SINE, which are noncoding sequences that uses LINEs to replicate and move around the genome by means of an RNA intermediate. Because SINEs have the tendency to cluster in specific regions, nonallelic homologous recombination (NAHR) events are more likely to occur here, resulting in duplications, deletions, inversions, or translocations.[19] This is also the case within the α-cluster, where approximately one-third of all known deletion breakpoints are located between position 177,501-186,001 (Hg38). Of these, atleast 12 break between the α_1- and θ-globin gene, including the --^{gb} deletion. Multiple deletions within the α-globin-gene cluster have their 3' breakpoint in an Alu repeat, such as –^{MedII},[20] --^{Dutch1},[21] --^{CAL},[22] --^{SA},[23] and --^{JB}[24] supporting the hypothesis of NAHR during genesis of the --^{gb} deletion.[25] We suggest that the --^{gb} deletion arose during meiosis where NAHR occurred, leading to recombination of micro-homologous sequences which are far apart on the linear DNA but could be in close proximity within the nucleus. This leads to loop formation of the chromatin that is excised, resulting in loss of genetic material, as is first described by Vanin et al.[26]

Previously the --^{gb} deletion was described as a Dutch deletion.[16] However, since the Netherlands is not endemic to *Plasmodium falciparum*, there is no evolutionary advantage of having this deletion, which makes its selection in the Netherlands unlikely. Furthermore, the cases in this study cohort all had either typical Dutch or

Indonesian family names, suggesting that the --^{gb} deletion may have arisen in Indonesia and, because of the Dutch colonial past, was transferred to the Netherlands during the last 400 years of intermixture and migration of populations.

In conclusion, this study has shown the successful characterization of the 16 771 bp --^{gb} deletion in the α-globin-gene cluster, using gDNA-based TLA on DNA that has been stored for 6 years at 4°C. Despite mapping only 8 kb, which may be caused by the isochoric nature of the α-globin cluster in combination with the quality of the DNA sample, the anchor fragment was within detection range of the deletion breakpoint. Ideally, a cell-based TLA is preferred since the coverage is significantly higher, making it a more reliable method for breakpoint detection and characterization. Nonetheless, the gDNA-based TLA should be able to map up to 70 kb with a single primer pair, which makes it an interesting alternative for long-range gap-PCR when sequence knowledge is necessary for both breakpoint ends. TLA has been shown to be valuable in deletion breakpoint characterization when no cell material could be obtained. The TLA results allowed for effortless gap-PCR primer design, which were used to confirm the --^{gb} deletion in 10 additional patients with similar MLPA results.

CONFLICT OF INTEREST

The authors have no competing interests.

DATA AVAILABILITY STATEMENT

Data are available on request from the authors.

ORCID

Quint P. Hottentot https://orcid.org/0000-0003-2811-4715

REFERENCES

1. Harteveld CL, Higgs DR. α-thalassaemia. *Orphanet J Rare Dis.* 2010;5(13):1-21. https://doi.org/10.1186/1750-1172-5-13
2. Petrou M, Brugiatelli M, Old J, et al. Alpha thalassaemia hydrops fetalis in the UK: the importance of screening pregnant women of Chinese, other South East Asian and Mediterranean extraction for alpha thalassaemia trait. *Br J Obstet Gynaecol.* 1992;99(12):985-989. https://doi.org/10.1111/j.1471-0528.1992.tb13703.x
3. Steinberg MH, Forget BG, Higgs DR, Weatherall DJ. *Disorders of Hemoglobin.* 2nd ed. Cambridge University Press; 2009.
4. Weatherall DJ, Clegg JB. *The Thalassaemia Syndromes.* Blackwell; 1965. https://doi.org/10.1002/9780470696705
5. Liu YT, Old JM, Miles K, Fisher CA, Weatherall DJ, Clegg JB. Rapid detection of α-thalassaemia deletions and α-globin gene triplication by multiplex polymerase chain reactions. *Br J Haematol.* 2000;108(2):295-299. https://doi.org/10.1046/j.1365-2141.2000.01870.x

6. Harteveld CL, Voskamp A, Phylipsen M, et al. Nine un-known rearrangements in 16p13.3 and 11p15.4 causing al-pha- and beta-thalassaemia characterised by high resolution multiplex ligation-dependent probe amplification. *J Med Genet*. 2005;42(12):922-931. https://doi.org/10.1136/jmg.2005.033597

7. Miller S, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1987;16(3):1215. https://doi.org/10.1016/B978-0-7506-7152-1.50011-4

8. Dacie JV, Lewis S. *Practical Haematology*. 7th ed. Churchil Livingstone; 1991.

9. Hottentot QP, Van MM, Splinter E, White SJ. Genotyping. In: White SJ, Cantsilieris S, eds. *Genotyping*. Vol 1492. Humana Press; 2017:185-196. https://doi.org/10.1007/978-1-4939-6442-0

10. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3-new capa-bilities and interfaces. *Nucleic Acids Res*. 2012;40(15):1-12. https://doi.org/10.1093/nar/gks596

11. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;19(12):996-1006. https://doi.org/10.1101/gr.229102

12. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26. https://doi.org/10.1038/nbt0111-24

13. Harteveld CL. State of the art and new developments in molecular diagnostics for hemoglobinopathies in multiethnic societies. *Int J Lab Hematol*. 2014;36(1):1-12. https://doi.org/10.1111/ijlh.12108

14. Traeger-Synodinos J, Harteveld CL, Old JM, et al. POLICY EMQN Best Practice Guidelines for molecular and haematology methods for carrier identification and prenatal diagnosis of the haemoglo-binopathies. *Eur J Hum Genet*. 2014;23(4):426-437. https://doi.org/10.1038/ejhg.2014.131

15. de Vree PJP, de Wit E, Yilmaz M, et al. Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat Biotechnol*. 2014;32(10):1019-1025. https://doi.org/10.1038/nbt.2959

16. Phylipsen M, Prior JF, Lim E, et al. Blood cells, molecules, and dis-eases thalassemia in Western Australia: 11 novel deletions char-acterized by multiplex ligation-dependent probe amplification. *Blood Cells Mol Dis*. 2010;44(3):146-151. https://doi.org/10.1016/j.bcmd.2009.12.011

17. Mota NO, Kimura EM, Ferreira RD, et al. Rare α0-thalassemia deletions detected by MLPA in five unrelated Brazilian patients. *Genet Mol Biol*. 2017;40(4):768-773. https://doi.org/10.1590/1678-4685-gmb-2016-0330

18. Giardine B, Borg J, Higgs DR, et al. Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach. *Nat Genet*. 2011;43(4):295-302. https://doi.org/10.1038/ng.785

19. Harteveld KL, Losekoot M, Fodde R, Giordano PC, Bernini LF. The involvement of Alu repeats in recombination events at the α-globin gene cluster: characterization of two α(o)-thalassaemia deletion breakpoints. *Hum Genet*. 1997;99(4):528-534. https://doi.org/10.1007/s004390050401

20. Kutlar F, Gonzalez-Redondo J, Kutlar A, et al. The levels of zeta, gamma, and delta chains in patients with Hb H disease. *Hum Genet*. 1989;82(May):179-186.

21. Harteveld C, Losekoot M, Heister A, Van der Wielen MJ, Giordano PC, Bernini L. alpha-Thalassemia in The Netherlands: a heteroge-neous spectrum of both deletions and point mutations. *Hum Genet*. 1997;100:465-471.

22. Fortina P, Dianzani I, Serra A, et al. A newly-characterized alpha-thalassaemia-1 deletion removes the entire alpha-like globin gene cluster in an Italian family. *Br J Haematol*. 1991;78(4):529-534. https://doi.org/10.1111/j.1365-2141.1991.tb04483.x

23. Vandenplas S, Higgs DR, Nicholls RD, Bester AJ, Mathew CG. Characterization of a new α° thalassaemia defect in the South African population. *Br J Haematol*. 1987;66:539-542.

24. Phylipsen M, Chaibunruang A, Vogelaar IP, et al. Fine-tiling array CGH to improve diagnostics for α- and β-thalassemia rearrange-ments. *Hum Mutat*. 2012;33(1):272-280. https://doi.org/10.1002/humu.21612

25. Phylipsen M. *Development of New Technological Applications for Post- and Prenatal Diagnosis of the Hemoglobinopathies*. CPI - Wöhrmann Print Service, Zutphen; 2013.

26. Vanin EF, Henthom PS, Kioussis D, Grosveld F, Smithies O. Unexpected relationships between four large deletions in the human β-globin gene cluster. *Cell*. 1983;35(3 PART 2):701-709. https://doi.org/10.1016/0092-8674(83)90103-4