Annals of Hematology (2025) 104:2529–2535
https://doi.org/10.1007/s00277-025-06223-2

CASE REPORT

# An unusual transfusion-dependent hemoglobin H disease caused by a novel complex inverted duplication involving the α-globin regulatory elements and α-thalassemia—-SEA deletion

**Fan Jiang[1] · Jieying Wu[1] · Manqiu Yang[1] · Xiaojun Chen[2] · Ru Li[1] · Qiuxia Yu[1] · Liandong Zuo[1] · Jianying Zhou[1] · Yongling Zhang[1] · Li Zhen[1] · Xuewei Tang[1] · Ren Zhang[2] · Shiping Chen[3] · Chenyu Wang[1] · Can Liao[1] · Dongzhi Li[1]**

Received: 5 November 2024 / Accepted: 20 January 2025 / Published online: 12 February 2025


**Abstract**
Four multi-species conserved sequences (MCSs) are important enhancers which affect α-globin expression. Deletions of MCS can cause α-thalassemia. So far, duplication of MCS has never been reported to account for thalassemia. In this study, an unusual transfusion-dependent case of hemoglobin H disease was identified by whole-genome sequencing, optical genome mapping and longer PCR with special primers, which was caused by a familial 96,620-bp inverted duplication (from MCS-R1 to MCS-R4), inserted between chr16:199348 and 199349 (GRCh37/hg19) within MCSs. The duplication segment included an inverted repeat sequence from chr16:102712 to176193 and one direct repeat sequence from chr16:176208 to 199348. The associated α-thalassemia trait was confirmed to result from disrupted topological chromatin domains using ATAC-seq and the dual-luciferase reporter assay system. This case presents a new mechanism of α-thalassemia, and may aid our understanding of the effects of enhancers on gene expression and the differential contribution of the four enhancer elements in the human a-globin locus.

**Keywords** Thalassemia · Inverted duplication · Regulatory elements · Multi-species conserved sequences

## Introduction

A-thalassemia is the most common single gene disorder in southern China. Hemoglobin H (HbH) disease, the most severe non-fatal type of α-thalassemia, results from the loss of three functional α-globin genes (*HBA2* and *HBA1*). The compound heterozygosity for $\alpha^0$-thalassemia (–/) and deletional $\alpha^+$-thalassemia (-α/) or non-deletional mutations ($\alpha^T\alpha$ or $\alpha\alpha^T$) can cause HbH disease [1]. Deletions of the multi-species conserved sequences (MCSs), 40 kb upstream of the α-globin gene, can cause α-thalassemia trait [2, 3]. As a result, the genotype of Hb H disease can be occasionally consequence of compound heterozygosity for $\alpha^0$-thalassemia and an MCS-R2 deletion [4, 5]. However, the duplication of MCS has never been reported.

## Case description

We present a severe case of HbH disease in a 9-year-old Chinese girl (III-1). Her prenatal course was uneventful, with a weight of 3.0 kg at full term birth. She was noticed to be pale due to anemia at the age of 4 months, and the anemia became severe (5–6 g/dL) at 6 months old, since then she required regular blood transfusions at a 3-month interval. She had mild hepatosplenomegaly, with a hemoglobin level of about 11 g/dL. On this referral, her weight and height were at 5thcentile for age and sex. She had been diagnosed as HbH disease based on the presence of HbH band (13.3%) and a plenty of HbH inclusions visualized in red cells with supravital stains. However, only an

✉ Can Liao
canliao7981@21cn.com

✉ Dongzhi Li
drlidongzhi2014@sina.com

[1] Prenatal Diagnostic Center, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Jinsui Road 9, Zhujiang New Town, Guangzhou 510623, Guangdong, China

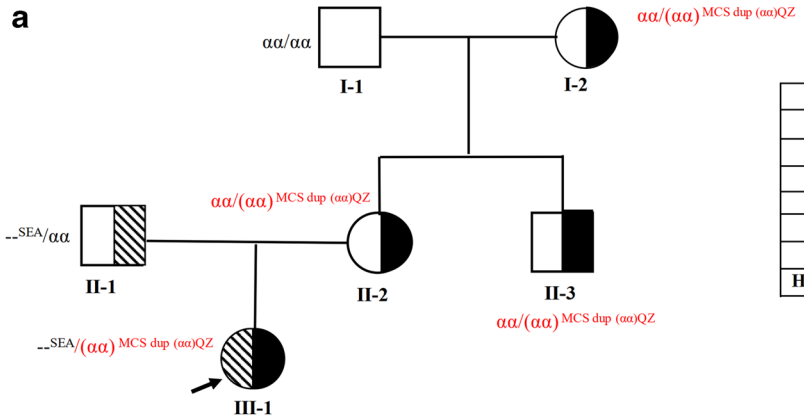
[2] Guangzhou University of Chinese Medicine, Guangzhou, China

[3] BGI Genomics, BGI-Shenzhen, Shenzhen, China

$\alpha^0$-thalassemia deletion ($-^{SEA}$), inherited from her father (II-1) was revealed by traditional genotyping methods. Sanger sequencing failed to detect pathogenic variants in *HBA1*/*HBA2*. Multiplex ligation-dependent probe amplification (MLPA) (SALSA MLPA KIT P140-B2 HBA, MRC-Holland) confirmed the heterozygous $-^{SEA}$ deletion, and, surprisingly, detected a duplication of MCS-R2 (HS-40) in the patient. This MCS duplication came from her mother (II-2), The carriers in family members all demonstrated an $\alpha^+$-thalassemia trait (Fig. 1A, B).

## Results and discussion

To investigate the genomic rearrangement, optical genome mapping (OGM) testing (Bionano Genomics, San Diego, CA, USA) was performed in the proband's mother according to its standard operating procedures and data analysis was performed using OGM specific pipelines managed via Bionano Access™ (v.1.7) software, which detected a 97-kb inversion with two breakpoints located within chr16:96953–130508 and chr16:176157–197960 in the proband's mother (GRCh37/hg19) (Fig. 1C) (Detailed materials and methods are described in the Supplementary




| Test | I-1 | I-2 | II-1 | II-2 | II-3 | III-1 |
|---|---|---|---|---|---|---|
| Hb(g/dL) | 14.3 | 12.5 | 13.8 | 11.4 | 14.6 | 7.6 |
| MCV(fL) | 87.3 | 78.4 | 68.1 | 80.4 | 74.1 | 77.4 |
| MCH(pg) | 28.4 | 26.3 | 21.5 | 26.0 | 24.4 | 19.8 |
| HbA2(%) | 2.3 | 2.1 | 2.4 | 2.7 | 2.4 | 1.2 |
| HbF(%) | 0.3 | / | / | / | / | / |
| HbH (%) | / | / | / | / | / | 13.0 |
| HbBarts (%) | / | / | / | / | / | 0.3 |

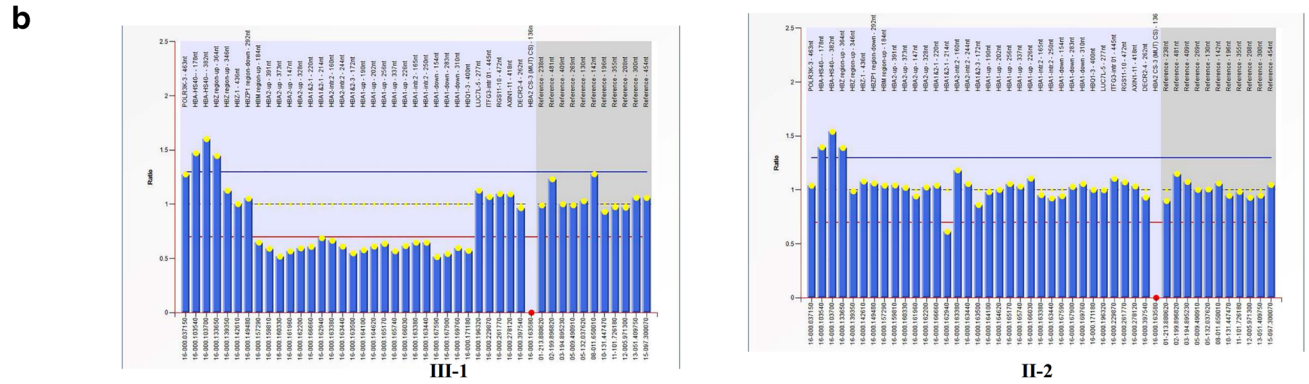



**Fig. 1** **A** Family pedigree with hematological characterization, showing that all MCS dup $(\alpha\alpha)^{QZ}$ members presented borderline MCV and MCH levels. **B** MLPA studies showing heterozygous MCS regulatory elements duplications in the patient (III-1) and her mother (II-2). The patient also presents the $-^{SEA}$ deletion in a heterozygous state. **C** Schematic representation of optical genome mapping (OGM) in II-2, showing the locations of MCS duplication. Horizontal boxes are used to indicate the α-globin cluster reference maps (green, hg38) aligned to the sample maps (blue). The vertical lines within boxes are labels for locations on chromosome 16. Matching labels between reference genome map and the sample (II-2) map at the normal (blue) chromosomal positions are connected by gray lines, and matching labels between reference genome map and the sample (II-2) map at the abnormal (orange) chromosomal positions are connected by blue lines. On the rearranged chromosome, the duplication sequences are in an inverted orientation. The schematic representation of the rearrangement in red box shows inverted orientation for duplicated fragments. The locations of the two breakpoints can be roughly targeted, with one in the range ofchr16:96953–130508, and the other in the range of chr16:176157–197960 (GRCh37/hg19). **D** Identification of the structure and the precise breakpoints of the MCS duplication. The F1 and R1 primers for the first breakpoint lead to a 2834-bp fragment and the F2 and R2 primers for the other breakpoint generate a 5506-bp fragment only in carriers and the proband. Sequencing of the two PCR products reveal the duplication is inserted between chr16:199,348 and 199,349, which includes two fragments. One is the inverted fragment from chr16:102712 to 176193 (73481 bp) and the breakpoint is joined with a 22-bp sequence. Another direct repeat sequence comes from chr16:176209 to 199348 and is inserted between 102712 and 176209 (23139 bp). The duplication is about 96,620 bp with two breakpoint junctions

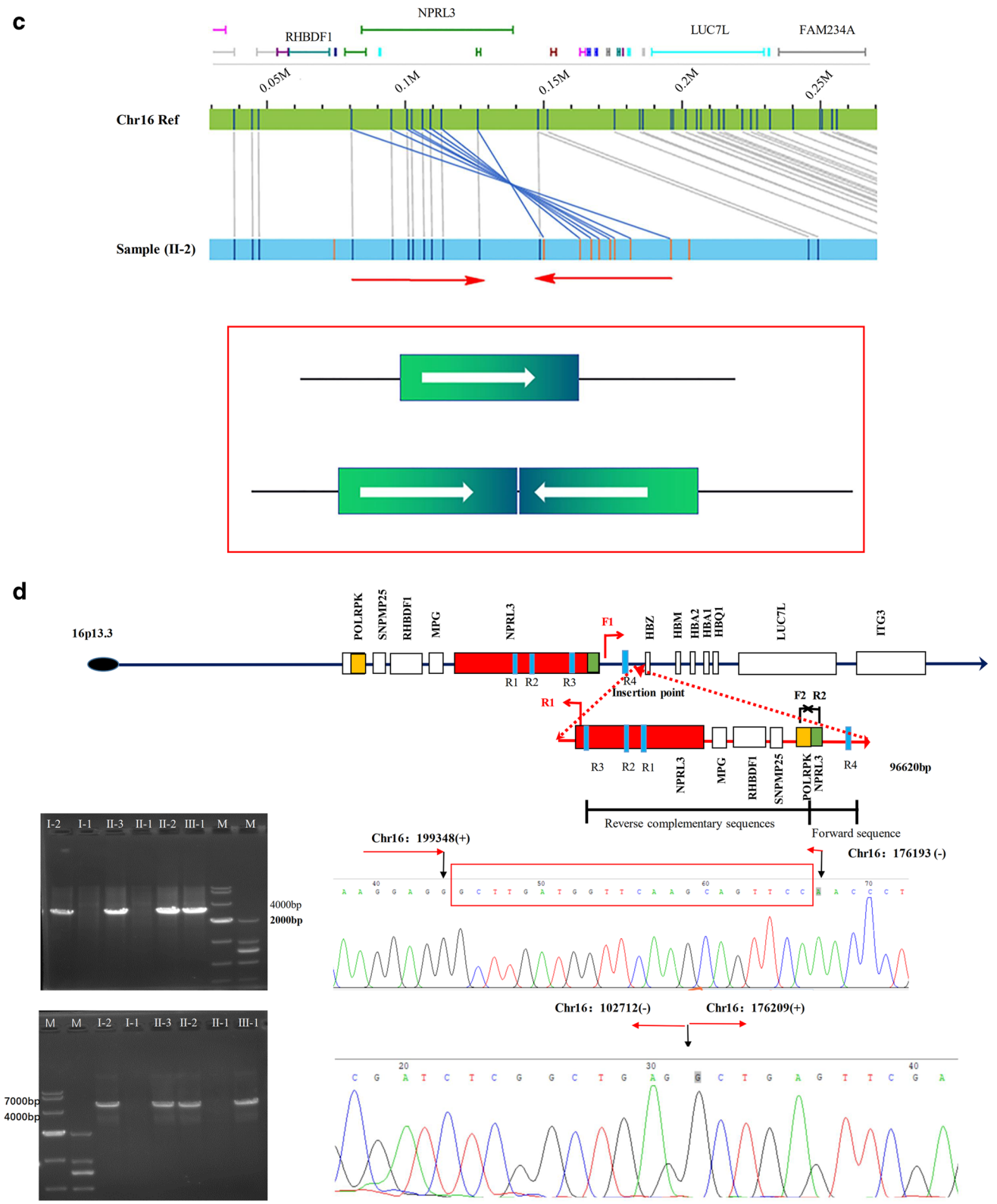



**Fig. 1** (continued)

data) [6]. Whole-genome sequencing (WGS) with NovaSeq6000(Illumina) was further used to get more sequence information about the genomic coordinates of the breakpoints, which narrowed the locations with one breakpoint happening within chr16:102711- 176193 and the other within chr16:176208–199348 (Figure S1) (Detailed materials and methods are described in the Supplementary data). With these valuable results obtained from OGM and WGS, we revealed successfully a 96,620 bp complex duplication segment inserted between chr16:199348 and 199349 bylong-range PCR amplification with two pairs of specific PCR primers (Table 1). The duplication segment has been identified to be an inverted repeat sequence from chr16:102712 to176193 followed by one direct repeat sequence from chr16:176208 to 199348. The inverted duplication segment included the reverse fragments of three MCSs: an additional MCS-R3 (HS-33) is inserted before MCS-R2 (HS-40), and an additional MCS-R1(HS-48) is inserted after MCS-R2 (HS40) (Fig. 1D). Therefore, this genomic rearrangement, named as MCS dup $(\alpha\alpha)^{QZ}$ results in three copies of MCSs in carriers, confirmed by using three specially designed MLPA probes. Another specially designed MLPA probe targeting on downstream breakpoint also provides the details of this complex structure (Figure S2). The information on MLPA probes is showed in Table 2. Normally, four multi-species conserved sequences (MCS-R1 to R4) form loop structures contacting with α globin genes and regulate their expression [7]. For the proband and the carriers, the complex inverted duplication with two breakpoint junctions may disrupt the loop structure (Figure S3), which may be caused by multiple non-homologous end joining (NHEJ) events [8].

Topologically associated domains (TADs) are the basic units of transcriptional regulation. In TADs, genes are co-located with their regulatory elements. A high degree of self-interaction especially enhancer-promoter interaction exists within the regions of the TAD [9]. 11.8% of pathogenic CNVs (Copy number variations) were believed to exert their effect through disruption of TADs [10], which have been identified to cause disease without altering gene sequences. For instance, X-linked acrogigantism is a TADopathy due to a duplication disrupting a conserved TAD border and neo-TAD formation that causes aberrant gene expression [11]. The 19p13.3 duplication has also been reported to disrupt a TAD, which breaks down the interactions between enhancers and their targeted genes [12]. Franke et al. reported another intra-TAD duplication causing sex reversal phenotype due to

**Table 1** Two special primers for detecting the locations with two breakpoints

| Name | Sequencing | Chromosome location(GRCh37/hg19) | Product length (bp) |
|---|---|---|---|
| F1 | 5'-GAGGAGCCGCTGTGAAATGT-3' | Chr16: 199303–199322 | 2834 |
| R1 | 5'- GTGCCTCATCAGTCACTCTTCAAC-3' | Chr16: 173392–173415 | |
| F2 | 5'-TGACGCTTTGACACATGATTGACGC-3' | Chr16: 104660–104684 | 5506 |
| R2 | 5'-GCCAAAGATCCACAGAAGGGAT-3' | Chr16: 175206–175227 | |

**Table 2** MLPA Probes designed for the three major regulatory elements (MCS R1-R3) and the upstream breakpoint

| Name | Sequencing | Chromosome location(GRCh37/hg19) | Product length (bp) |
|---|---|---|---|
| HS-48 | GGGTTCCCTAAGGGTTGGACATCCTGACAATGACTATGCAAAAGGGAATGG CTGGC<br>GCTTTGGCTGGGGCACCTATCTCCTGTCACAGACTCTAGATTGGATCTTGCTGG CAC | Chr16:155174–155210<br>Chr16:155136–155169 | 113 |
| HS-33 | GGGTTCCCTAAGGGTTGGAGCTTCCCCAAGAGAGTGCACCCTCCTGACCTAGAT ATTAG<br>ATATATTTTTCCTGGGTGGAGAGATATGTATGGGACTTCTCTCTCTCTAGATTG GATCTTGCTGGCAC | Chr16:170211–170655<br>Chr16:170166–170210 | 127 |
| HS-10 | GGGTTCCCTAAGGGTTGGACTCTGCTAAGACCCCACACCTCCAAGTCTCCTCAT TTTA<br>CCTTTAAATAGCTGTTTCATGACCTGCTTTTTTGACTCTAGATTGGATCTTGCT GGCAC | Chr16:193013–193051<br>Chr16:193052–193087 | 117 |
| MCR-DUAN1* | GGGTTCCCTAAGGGTTGGACGAGGAACAGACCAGCAAGGAGGATCCACGCAGT<br>GCTAGAAGGGAGTTCCTGGAAGCCTGGTGGAGTCTAGATTGGATCTTGCTG GCAC | Chr16:199326–199359<br>Chr16:199360–199351 | 108 |

*the probe is for the upstream breakpoint of the duplication

changing the frequency of interactions with the entire *SOX9* gene, causing abnormal expression of SOX9 [13]. Therefore, TAD are one of the mechanisms which are associated with abnormal gene expression and phenotypic changes. However, the pathogenesis is not easy to be clarified. Hi-C is a method to identify chromatin interactions across an entire genome without bias [14]. ATAC-seq is often used to identify chromatin structure and accessibility [15]. Those two techniques have been used for this purpose [16]. Four major regulatory elements (MCS R1-R4) and two α-globin genes lie together in a well-defined 65 kb sub-topologically associating domain (sub-TAD) [17]. A complex intra TAD duplication involving two sets of MCS element was observed in this study. The duplication with reverse insertion of MCS-R1, R2, and R3 changes the length of TAD. Unfortunately, we could not obtain enough $CD34^+$ cells from the patient or carriers for Hi-C analysis. Alternatively, more than 50,000 $CD71^+$ cells were obtained using magnetic cell sorting from 50 mL peripheral blood of the proband's mother and one normal female as reference for ATAC-seq, which was used to demonstrate enhancers' accessibility to transcription factors in vivo. ATAC libraries were generated according to the protocol and were sequenced on the Illumina NovaSeq 6000.Index of the reference genome was built using BWA v 0.7.17 and clean reads were aligned to the reference genome using BWA mem v 0.7.17. After mapping reads to the reference genome, we used the MACS14 version 1.4.2 peak finding algorithm to identify regions of enrichment over background. A p-value threshold of enrichment of 1e-5 was used for all data sets. We observed a significant reduction in the ATAC peak corresponding to R2 and R3 in the carrier (Fig. 2A). However, we did not observe any new peaks forming in regions surrounding the α-globin genes, or any compensatory increase in the peak heights of the remaining enhancers (R1 and R4). Since the relative contributions of MCS-R2 to α globin gene expression has been suggested to be 90%, this finding suggests that the complex rearrangements affect α-globin gene expression.

We also attempted to use ClinTAD to confirm the clinical significance of the MCS CNV in the context of TAD. The CNV crossing a TAD boundary and the patient's phenotypes associated with nearby genes are the two important parameters used in this tool [18]. As expected, the patient had the CNV crossing a TAD boundary and HPO phenotype matches in adjacent TADs using ClinTAD tool (Figure S4), increasing the pathogenic possibility of the MCS duplication.

Considering the main role of MCS-R2 as the enhancer for α-globin gene expression, we used the dual-luciferase reporter assay system, performed in K562 erythroid cells, to determine whether the reverse insertion of MCS-R2 had a negative impact on the regulation of α-gene expression. The fragments of HS-40 (356 bp), reverse sequence of HS-40 (356 bp) and the promoter of *HBA2* (2000 bp) were cloned into the plasmid vector containing the firefly luciferase reporter gene. We used the ratio of firefly

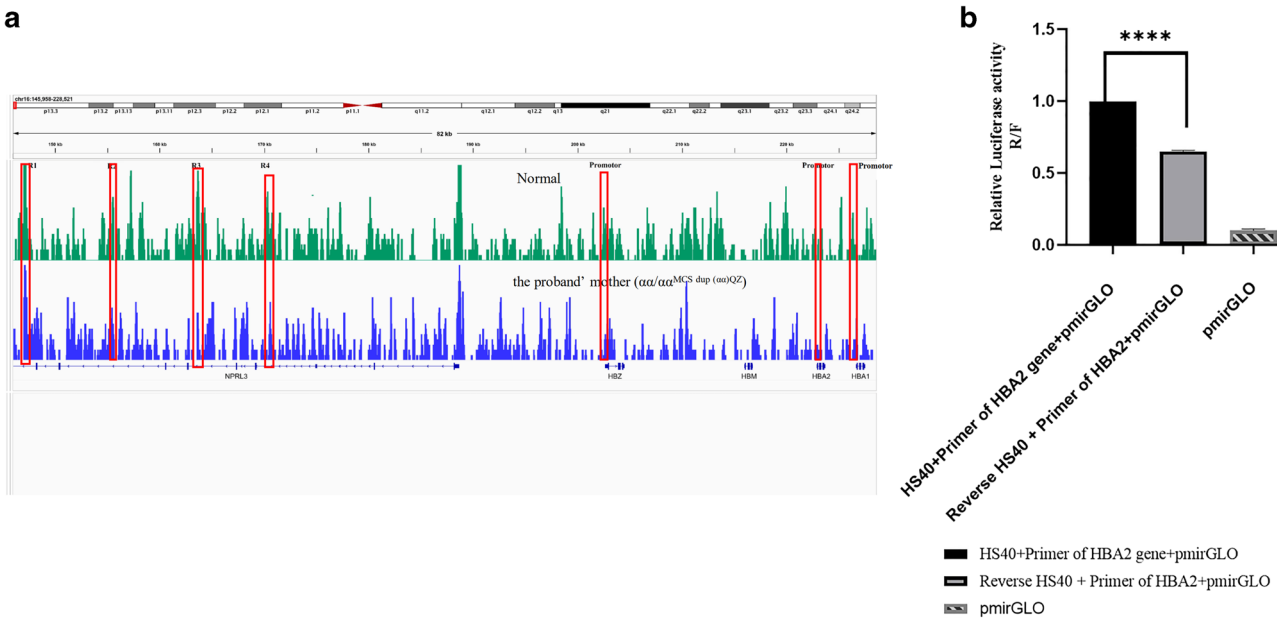



**Fig. 2** **A** The open chromatin landscape at the α-globin locus (ATAC-seq) in CD71 + cells from one female control sample (αα/αα) and the proband' mother (αα/(αα).$^{QZ}$. In the carrier's cells, no remarkable peak is observed at the R2 enhancer, indicating reduction in ATAC signal. **B** Luciferase reporter assays in erythroid K562 cell lines with two constructs including MCS-R2wt and antiMCS-R2 respectively. Histograms of average luciferase activity with standard deviation normalized to the MCS-R2wt activity. The *** indicates $p < 0.01$

luciferase activity to the corresponding Renilla luciferase activity to express the results. All assays were repeated three times (Detailed materials and methods are described in the Supplementary data). As expected, the reverse insertion of MCS-R2 showed a lower enhancer activity than that of intact MCS-R2 (Fig. 2B). MCS-R2 is the major regulatory element of the human α-globin locus, but its performance requires the coordination of MCS-R1 and MCS-R3. The effect of other regulatory elements on MCS-R2 regulation seems depending on their physical position [19]. One previous study reported similar effect on β- and γ-globin globin expression, in which one and two copies of 355 bp MCS-R2 core sequence were inserted upstream of the β- or γ-globin promoter, respectively [20]. This was consistent with ATAC seq. All those findings indicate that the tandem rearrangement of MCS may disrupt the TAD structure and impair its function as an enhancer, which reduces the level of α-globin gene expression.

## Conclusion

In summary, this case presents the first tandem duplication of enhancer elements of α-globin cluster causing $\alpha^+$-thalassemia. The duplication not only extends the distance between normal enhancer elements and promoters, but also modifies the physical locations of three regulatory elements; both alterations impair the effect of MCS on the regulation of α-globin gene expression. The evidences of HbH phenotype in the proband, co-segregation of this duplication with $\alpha^+$-thalassemia carriers in family members, and the results of ATAC-seq and ClinTAD, all support the causative property of this CNVs. This case presents a novel mechanism of α-thalassemia.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00277-025-06223-2.

**Author contributions** F.J, L.D.Z collected clinical data and wrote the initial draft of the paper. C.L. contributed the central idea. JY-W and CY-W performed Magnetic cell sorting and flow cytometry. XJ-C performed dual luciferase reporter assay. MQ-Y, QX-Y analyzed data of ATAC Seq. SP-C perfumed Whole-genome sequencing. RL revised the manuscript. JY-Z and XW-T were responsible for capillary electrophoresis. LZ performed genetic counseling. LD-Z was responsible for project management. YL-Z collected clinical data. RZ designed research.

**Funding** This work was supported by Guangdong Natural Science Foundation (grant number 2023A1515010402), Guangdong Basic and Applied Basic Research Foundation (grant number 2022A 1515220140), Guangzhou Science and Technology Plan Project (grant number 202206010100, 2023A03J0875, 202201020632).

**Data availability** No datasets were generated or analysed during the current study.

## Declarations

**Ethics statement** This study was performed in accordance with the relevant ethics committee regulation of Guangzhou Women and Children's Medical Center (2022-450B00). The informed consent was obtained from all subjects in this study.

**Informed consent** The patient and her family signed informed consent.

**Competing interests** The authors declare no competing interests.



## References

1. Piel FB, Weatherall DJ (2014) The α-thalassemias. N Engl J Med 371:1908–1916. https://doi.org/10.1056/NEJMra1404415
2. Capasso S, Cardiero G, Musollino G et al (2023) Functional analysis of three new alpha-thalassemia deletions involving MCS-R2 reveals the presence of an additional enhancer element in the 5' boundary region. PLoS Genet 19:e1010727. https://doi.org/10.1371/journal.pgen.1010727
3. Badat M, Davies JOJ, Fisher CA et al (2021) A remarkable case of HbH disease illustrates the relative contributions of the α-globin enhancers to gene expression. Blood 137:572–575. https://doi.org/10.1182/blood.2020006680
4. Huang LY, Yan JM, Zhou JY, Li J, Xie XM, Li DZ (2017) A severe case of hemoglobin H disease due to compound heterozygosity for deletion of the major α-Globin regulatory element (MCS-R2) and α0-Thalassemia. Acta Haematol 138:61–64. https://doi.org/10.1159/000477531
5. Jiang F, Xu LL, Tang XW, Li DZ (2021) A novel deletion of the major regulatory element flanking the α-globin gene cluster as a cause of α(0) -thalassemia. Int J Lab Hematol 43:O190-o192. https://doi.org/10.1111/ijlh.13505
6. Iqbal MA, Broeckel U, Levy B et al (2023) Multisite assessment of optical genome mapping for analysis of structural variants in constitutional postnatal cases. J Mol Diagn 25:175–188. https://doi.org/10.1016/j.jmoldx.2022.12.005
7. Baù D, Sanyal A, Lajoie BR et al (2011) The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. Nat Struct Mol Biol 18:107–114. https://doi.org/10.1038/nsmb.1936
8. Grochowski CM, Bengtsson JD, Du H et al (2024) Inverted triplications formed by iterative template switches generate structural variant diversity at genomic disorder loci. Cell Genom 4:100590. https://doi.org/10.1016/j.xgen.2024.100590

9. Sánchez-Gaya V, Rada-Iglesias A (2023) POSTRE: a tool to predict the pathological effects of human structural variants. Nucleic Acids Res 51:e54. https://doi.org/10.1093/nar/gkad225
10. Ibn-Salem J, Köhler S, Love MI et al (2014) Deletions of chromosomal regulatory boundaries are associated with congenital disease. Genome Biol 15:423. https://doi.org/10.1186/s13059-014-0423-1
11. Naves LA, Daly AF, Dias LA et al (2016) Aggressive tumor growth and clinical evolution in a patient with X-linked acrogigantism syndrome. Endocrine 51:236–244. https://doi.org/10.1007/s12020-015-0804-6
12. Jouret G, Egloff M, Landais E et al (2023) Clinical and genomic delineation of the new proximal 19p13.3 microduplication syndrome. Am J Med Genet A 191:52–63. https://doi.org/10.1002/ajmg.a.62983
13. Franke M, Ibrahim DM, Andrey G et al (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature 538:265–269. https://doi.org/10.1038/nature19800
14. Lieberman-Aiden E, van Berkum NL, Williams L et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293. https://doi.org/10.1126/science.1181369
15. Luo L, Gribskov M, Wang S (2022) Bibliometric review of ATAC-Seq and its application in gene expression. Brief Bioinform (23). https://doi.org/10.1093/bib/bbac061
16. Dubois F, Sidiropoulos N, Weischenfeldt J, Beroukhim R (2022) Structural variations in cancer and the 3D genome. Nat Rev Cancer 22:533–546. https://doi.org/10.1038/s41568-022-00488-9
17. King AJ, Songdej D, Downes DJ et al (2021) Reactivation of a developmentally silenced embryonic globin gene. Nat Commun 12:4439. https://doi.org/10.1038/s41467-021-24402-3
18. Spector JD, Wiita AP (2019) ClinTAD: a tool for copy number variant interpretation in the context of topologically associated domains. J Hum Genet 64:437–443. https://doi.org/10.1038/s10038-019-0573-9
19. Blayney JW, Francis H, Rampasekova A et al (2023) Super-enhancers include classical enhancers and facilitators to fully activate gene expression. Cell 186:5826-5839.e5818. https://doi.org/10.1016/j.cell.2023.11.030
20. Chen H, Lowrey CH, Stamatoyannopoulos G (1997) Analysis of enhancer function of the HS-40 core sequence of the human alpha-globin cluster. Nucleic Acids Res 25:2917–2922. https://doi.org/10.1093/nar/25.14.2917