An unusual case of thalassemia intermedia with inheritable complex repeats detected by single-molecule optical mapping

A 3.5-year-old female presented with mild anemia and was preliminarily diagnosed with thalassemia intermedia (TI) based on blood testing confirming microcytic, hypochromic anemia and abnormal amounts of hemoglobin (HGB, Hb) constituents. Longitudinal hematological measurements of the proband revealed HGB concentrations of 87-107 g/L (reference interval [RI]: 110-150 g/L), mean corpuscular volume of 57.8-62.1 fL (RI: 82-100 fL), mean corpuscular hemoglobin of 18.1-19.0 pg (RI: 27-34 pg), Hb A_2 of 1.6-1.9% (RI: 2.5-3.5%), and Hb F of 2.9-10.2% (RI: <2%) (Figure 1A, B). Other biochemical and blood routine parameters, including the ferritin levels, and counts of white blood cells, lymphocytes, neutrophils, monocytes and platelets were within normal ranges. Upon examination, the patient was pale, but with neither indications of thalassemic face nor hepatosplenomegaly. The patient did not receive prior transfusion therapy. The research protocol for this study was designed and implemented in accordance with the principles of the Declaration of Helsinki. This study was approved by the Ethics Committee of NanFang Hospital of Southern Medical University. Written informed consent was obtained for each participant.

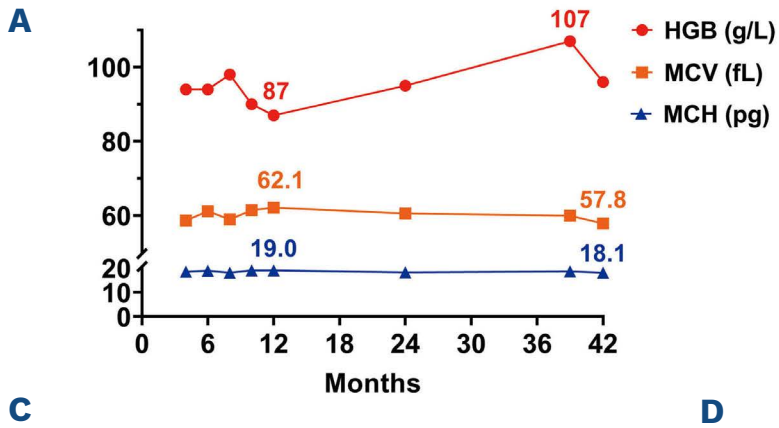
We first conducted conventional molecular diagnosis for detecting thalassemia genes including Sanger sequencing, gap-polymerase chain reaction (gap-PCR), and multiplex ligation-dependent probe amplification (MLPA).[1] The results revealed no thalassemia-causative point mutations on *HBA1/2* or *HBB* genes. Gap-PCR analysis of known α-thalassemia deletions revealed maternal inheritance of a heterozygous Southeast Asian type deletion (--^{SEA}, SEA deletion) (Figure 1B, C). However, the heterozygous SEA deletion alone is insufficient to explain the TI phenotype of the patient, prompting further investigations of other molecular defects involved in imbalanced production of globin chains. Thus, further characterization of α- and β-globin copy number variations (CNV) responsible for the TI phenotype was carried out.

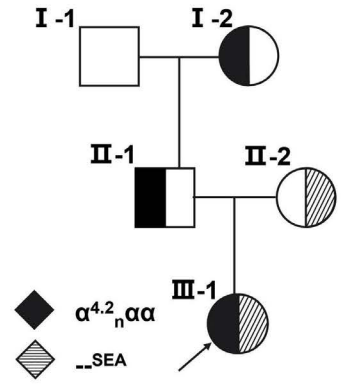
MLPA was performed using SALSA MLPA KIT P140-B2 HBA/P102 HBB (MRC-Holland, Amsterdam, Holland) according to the manufacturer's instructions. There was no copy number abnormality of the β-globin clusters in the proband. However, the results of the α-globin clusters were intriguing, as signaling of probes targeting the SEA deletion region decreased by about 50% for the proband (except for *HBA2* gene) and II-2, confirming a heterozygous SEA deletion, in agreement with the gap-PCR results. The signal intensity of the *HBA2* region was significantly increased by about 6-12-fold for the proband (III-1), father (II-1), and grandmother (I-2), suggesting the presence of highly expanded *HBA2* gene repeats, composed of multiple 4.2-kb fragments (Figure 1D). Chromosomal microarray analysis was performed on II-1 in accordance with the manufacturer's protocol by CytoScan 750K array (Affymetrix, Santa Clara, CA, USA), and no other abnormalities were detected within the probe coverage range.

Droplet digital PCR (ddPCR) is a new generation PCR technique for absolute quantification of a target DNA template. CNV of the α-globin cluster were identified by ddPCR using three specific probes targeting three non-homologous regions (UP-HBA-FAM, 4.2-HBA-FAM, and 3.7-HBA-FAM) (Figure 1E) with the QX200 ddPCR system (Bio-Rad Laboratories, Hercules, CA, USA). The copy numbers of I-2, II-1, and the proband (III-1) detected with the 4.2-HBA-FAM probe were 23.18±1.26, 24.53±1.74, and 23.05±1.05 respectively. The results of the other two probes in the family members were consistent with former genotypes. The duplications in this pedigree were inferred as 22-26 repeats of 4.2-kb (Figure 1F).

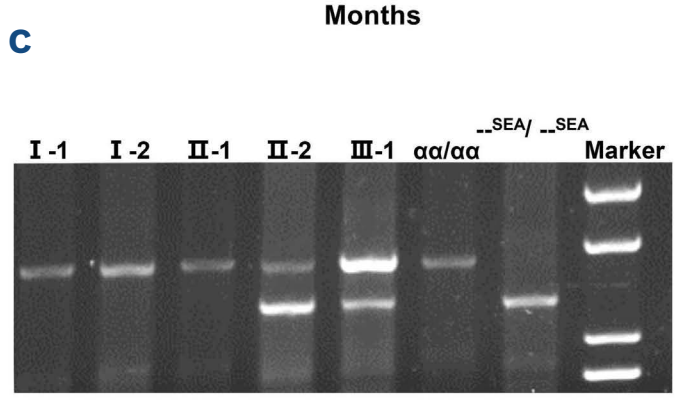
In order to elucidate the structure of the 4.2-kb repeats, SMRTbell Express Template Prep Kit 2.0 was employed to prepare a library from sample II-1, followed by continuous long read sequencing using the Sequel IIe platform (Pacific Biosciences, San Diego, CA, USA). Total bases yield was 1,210 Gb and unique molecular yield was 228 Gb, with an average sequencing depth of 76X and average read length of 50 kb. Four representative long reads covering the left flanking and five full anti-4.2 repeats (read 1), eight full anti-4.2 repeats (read 2), 12 full anti-4.2 repeats and the right flanking (read 3), as well as wild-type allele (read 4) were displayed in integrative genomics viewer plot (Figure 1G). The anti-4.2 repeats occurred head-to-tail at the α-globin locus and the left and right breakpoints were delineated to be chr16:169454-170793 (X2 box) and chr16:173711-175048 (X1 box), respectively. However, due to repeat length limitation and lack of specific SNP for gene assembly in the region of repeats, even long-read sequencing could not identify the exact number of anti-4.2 repeats.

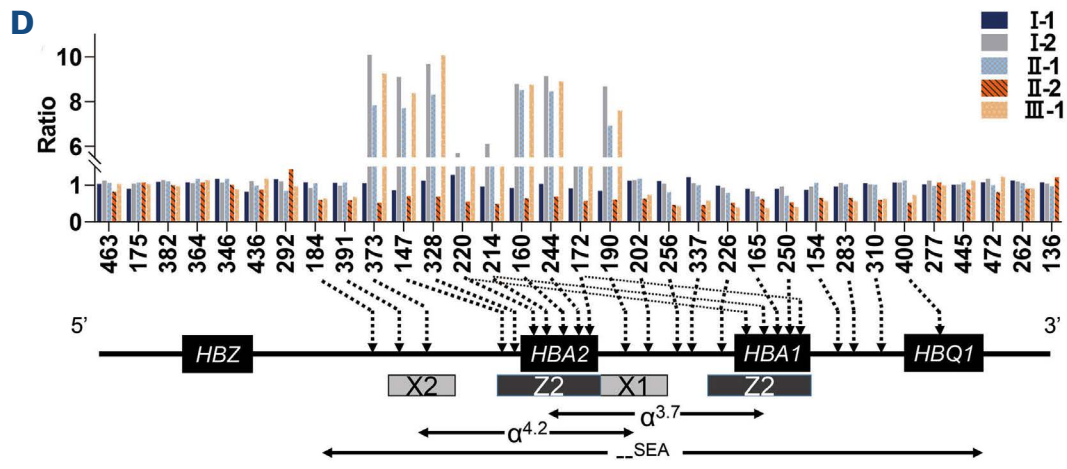
Globin analysis of red blood cells found no significant differences in α/β-globin levels among the II-1, II-2, and normal controls, while the proband had lower α/β-globin levels than SEA deletion carriers (Figure 1H). Quantita-

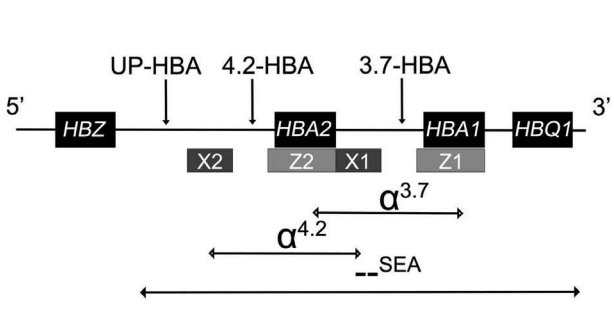


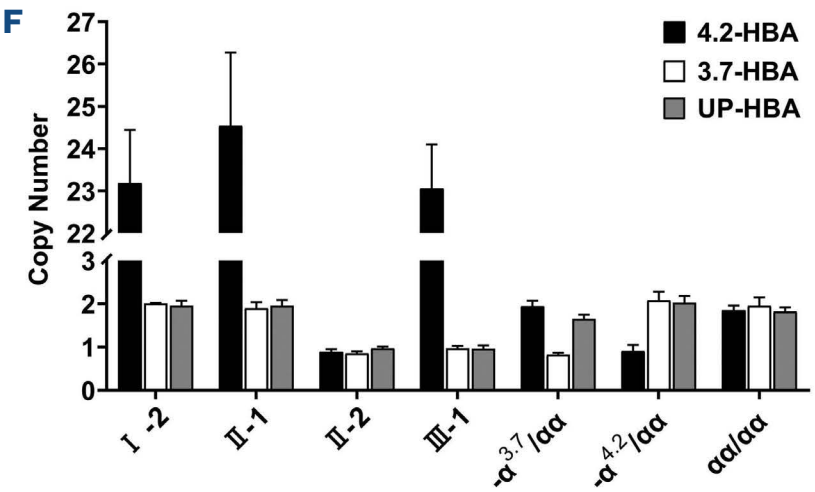


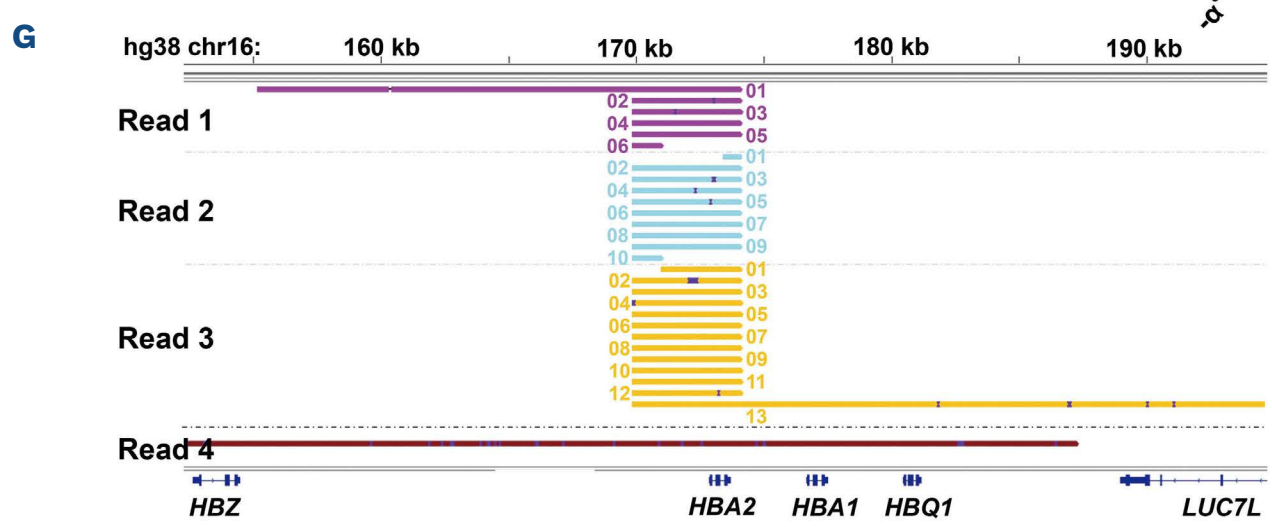
	HGB (g/L)	MCV (fL)	MCH (pg)	Hb A_2 (%)	Hb F (%)
I -1	137	102.4	33.0	2.7	0.4
I -2	130	90.0	29.5	2.5	2.2
II-1	155	89.1	29.2	2.7	1.2
II-2	115	72.4	22.3	2.0	1.1
III-1	96	57.8	18.1	1.9	2.9

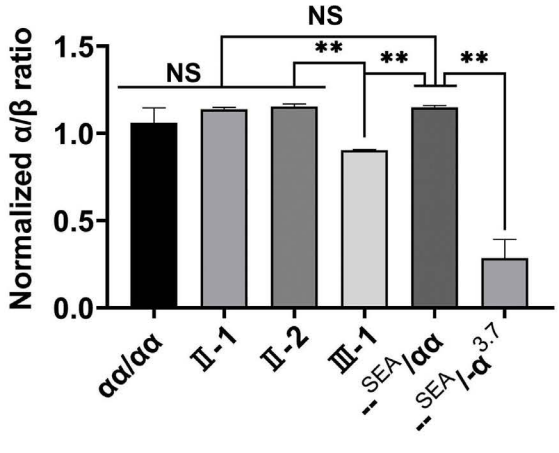


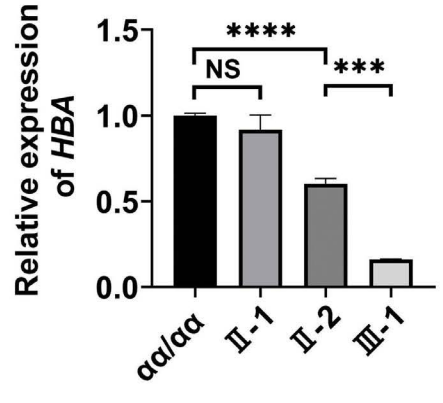












Continued on following page.

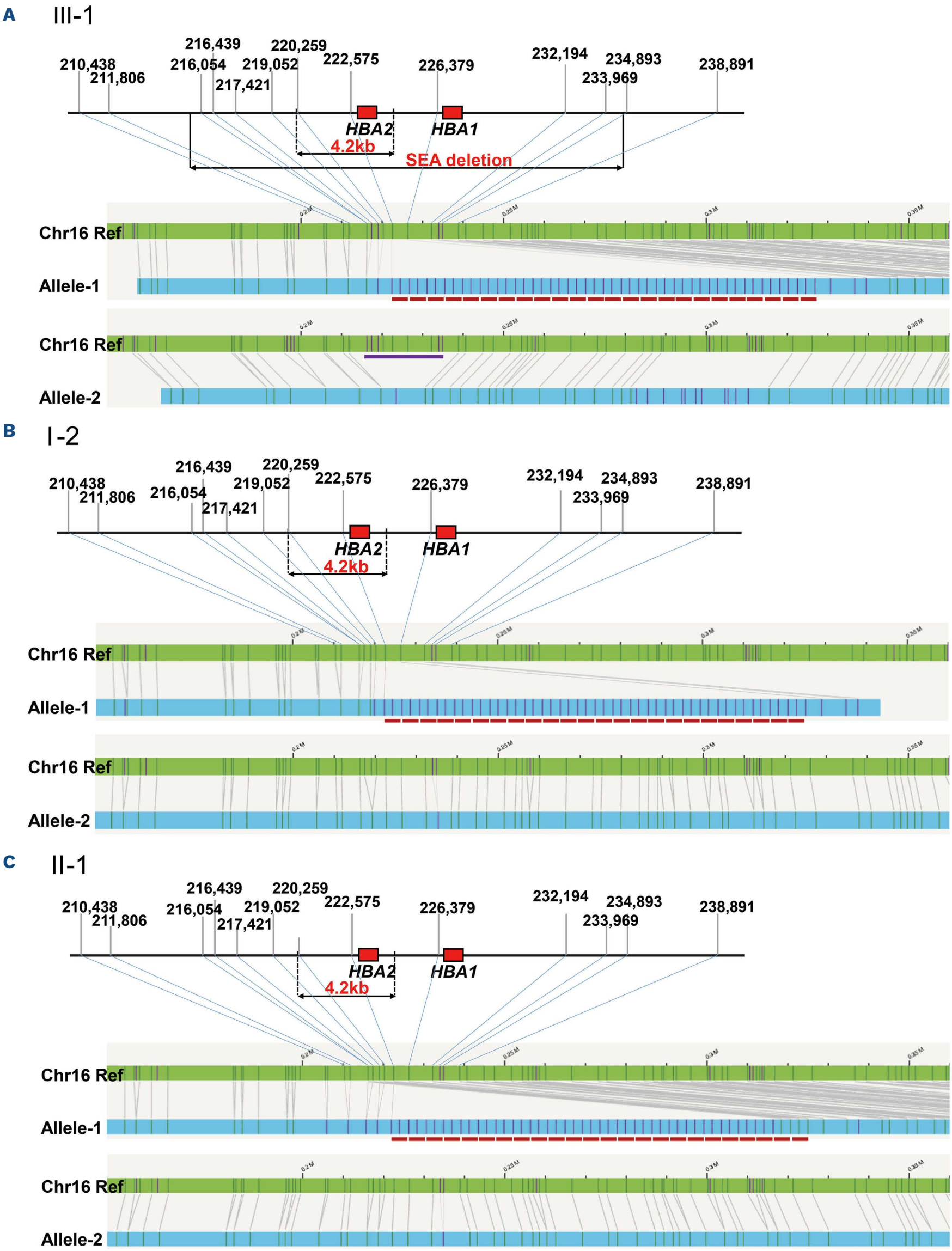
Figure 1. Hematological and molecular data of the pedigree. (A) Longitudinal hematological features of the proband. The highest and lowest values are labeled on the figure. (B) Pedigree of the investigated family with copy number variations of the α-globin cluster. The proband is indicated by the arrow. (C) The Southeast Asian type deletion (--^{SEA}, SEA deletion) detected in the proband (III-1) and mother (II-2) by agarose gel electrophoresis and gap-polymerase chain reaction (gap-PCR). (D) The multiplex ligation dependent probe amplification (MLPA) results targeting α-globin cluster among the family members. The histogram shows duplication of the 4.2-kb fragment for I-2, II-1, and III-1. II-2 and III-1 both carried the SEA deletion. The targeting regions of each specific probe are indicated below. (E) Targeting regions of the droplet digital PCR (ddPCR) probes. (F) The copy number values of the 4 family members as determined by ddPCR. Normal controls, -$\alpha^{3.7}$/αα and -$\alpha^{4.2}$/αα samples were tested using 2 biological and 2 analytical repeats. Samples from 4 family members were tested using 2 analytical repeats. (G) Representative reads from long-read sequencing of II-1. The allele with anti-4.2 repeats (read 1/2/3) and wild-type allele (read 4). Raw reads were aligned to hg38, cut at positions chr16:169454 and chr16:173711, and then displayed in integrative genomics viewer plot from left to right and from top to bottom. (H) The ratios of α- to β-globin chains were calculated by reversed-phase high performance liquid chromatograph (LC-20AT, Shimadzu Corporation, Kyoto, Japan). Normal controls, SEA deletion carriers and HbH diseases samples were tested using 2 biological and 2 analytical repeats. Samples from II-1, II-2, and III-1 were tested using 2 analytical repeats. There were no significant differences in α/β ratios among the 4 groups (normal controls, II-1, II-2, and SEA deletion carriers). The α/β ratios of III-1 were significantly lower than those of II-1, II-2 and the SEA deletion carriers, but higher than the HbH diseases samples. (I) Real-time quantitative PCR (RT-qPCR) for assessing the α-globin mRNA level. α-globin mRNA was slightly decreased in II-1 with no significance, decreased by 50% in II-2, while more decreased in III-1.

tive real time PCR (RT-qPCR) showed that compared to normal controls, α-globin mRNA was slightly decreased in II-1 with no significance, decreased by 50% in II-2, while more decreased in III-1 (Figure 1I).

Single-molecule optical mapping technology (Bionano Genomics, San Diego, CA, USA) was used to further deconstruct the sequence structure to determine the exact size of the complex repeats of the proband (III-1), I-2, and II-1. First, high molecule weight genomic DNA was isolated using the Bionano Prep Blood and Cell Culture DNA Isolation Kit. Second, 1 μg of DNA was nicked, labeled, repaired and stained with NLRS kit to fluorescently label DNA at the Nb. BssSI recognition motif CACGAG. Third, the labeled DNA was loaded into the flow cell of the Saphyr Chip. On average, 494 (range, 486-500) Gb of data per sample were collected. An average of 12.2 (range, 11.78-12.67) labels per 100 kb, and the average mapping rate was 76.9% (range, 75.0-78.8%), equating to an average effective coverage depth of 122.78X (range, 120.25-124.34X). Fourth, raw DNA molecules were filtered based on a minimum length of 150 kb and a minimum of nine labeled sites per molecule. Filtered single molecules were assembled in to consensus genome maps with Bionano Solve v3.2.1 using RefAligner module. The hg19 reference genome maps was obtained by *in silico* Nt.BssSI digestion of reference genome sequence. Finally, the structure variants were called by aligning consensus genome maps to reference genome maps. We determined the allelic configuration of the proband, which was composed of the 24 extra tandem duplications of the 4.2-kb in Allele-1 and the SEA deletion in Allele-2 (Figure 2A). This duplication variant is designated as $\alpha^{24\ repeats}\alpha\alpha$, namely, adding 24 extra copies of α–globin gene to the normal two α-globin genes in one chromosome. For the proband, the specific genotype of the α-globin gene was --^{SEA}/$\alpha^{24\ repeats}\alpha\alpha$. I-2 and II-1 harbored 24 extra α-globin genes (αα/$\alpha^{24\ repeats}\alpha\alpha$), which included a 100-kb repeat sequence adjacent to the normal 4.2-kb allele (Figure 2B, C), the same as that of the proband.

Typically, tandem duplications and increased copy numbers are frequent somatic alterations in various cancer.[2] In the present study, this unusual tandem duplication, which extended over 100 kb, was inherited over three generations of the family. α-globin gene triplication (ααα/αα) is generally common, while quadruplication (αααα/αα) or more copy numbers are quite rare. Duplication of α-globin genes is attributable to unequal crossover between misaligned homologous segments in the α-globin gene cluster during meiosis, which is always benign and exerts no obvious influence on hematological parameters when found in isolation, rendering it barely detectable in clinical practice. The cases with co-inheritance of α-globin gene duplication and one β-thalassemia mutation would exacerbate the imbalance in α and β-globin chains, leading to TI phenotype.[3,4]

We report an extremely rare, inheritable complex repeats of *HBA2* within the human α-globin cluster by single-molecule optical mapping. Due to the long repeat fragments and the existence of highly homologous sequences, MLPA, ddPCR, PacBio sequencing fail to elucidate the characteristics of CNV/structural variation (SV) in this pedigree. In single molecule ordered restriction maps, we found that the pattern of the Nb.BssSI-based labels covering *HBA2*, was repeated additional 24 times, indicating a 100-kb repeat sequence adjacent to the normal *HBA2*. Among the current molecular diagnostic techniques, single-molecule optical mapping is unrivaled for deciphering highly repetitive regions and detecting the genome structure variations. It utilizes very long reads, up to megabases, by eliminating the need to break DNA molecules into short fragments and subsequent amplification. Currently, single-molecule optical mapping is predominantly applied for genome assembly and detection of large structure variation. In the field of molecular diagnosis, it is useful to identify and characterize gene duplications, deletions and genomic rearrangements,



Continued on following page.

Figure 2. Characterization of the duplications by single-molecule optical mapping technology. Single-molecule optical mapping analysis of the proband (A), I-2 (B) and II-1 (C). Schematic representation showing the locations of α-globin genes and the region of 3 copy number variations (-$\alpha^{3.7}$, -$\alpha^{4.2}$, --^{SEA}) are present above the each panel. The α-globin cluster reference maps (green) aligned to the 2 alleles maps (blue) are indicted by horizontal boxes. The vertical lines within boxes are label sites. Matching labels between alleles from samples and reference genome maps are connected by gray lines. In Allele-1, the distance between the 2 labels, 222,575 and 226,379, was much larger than that of the reference genome, indicating sequence insertion. The pattern of label 220,259 to label 222,575 was amplified with extra 24 regular repeats compared to reference genome. The 24 extra tandem duplications of the 4.2-kb segment in Allele-1 are indicated by red dotted lines. In Allele-2 of the proband (A), the distance between the 2 labels, 211,806 and 234,893, was shorter than that of the reference genome, indicating sequence deletion. The SEA deletion region in Allele-2 is indicated by a purple line. Coordinates are based on human genome build GRCh37/hg19.

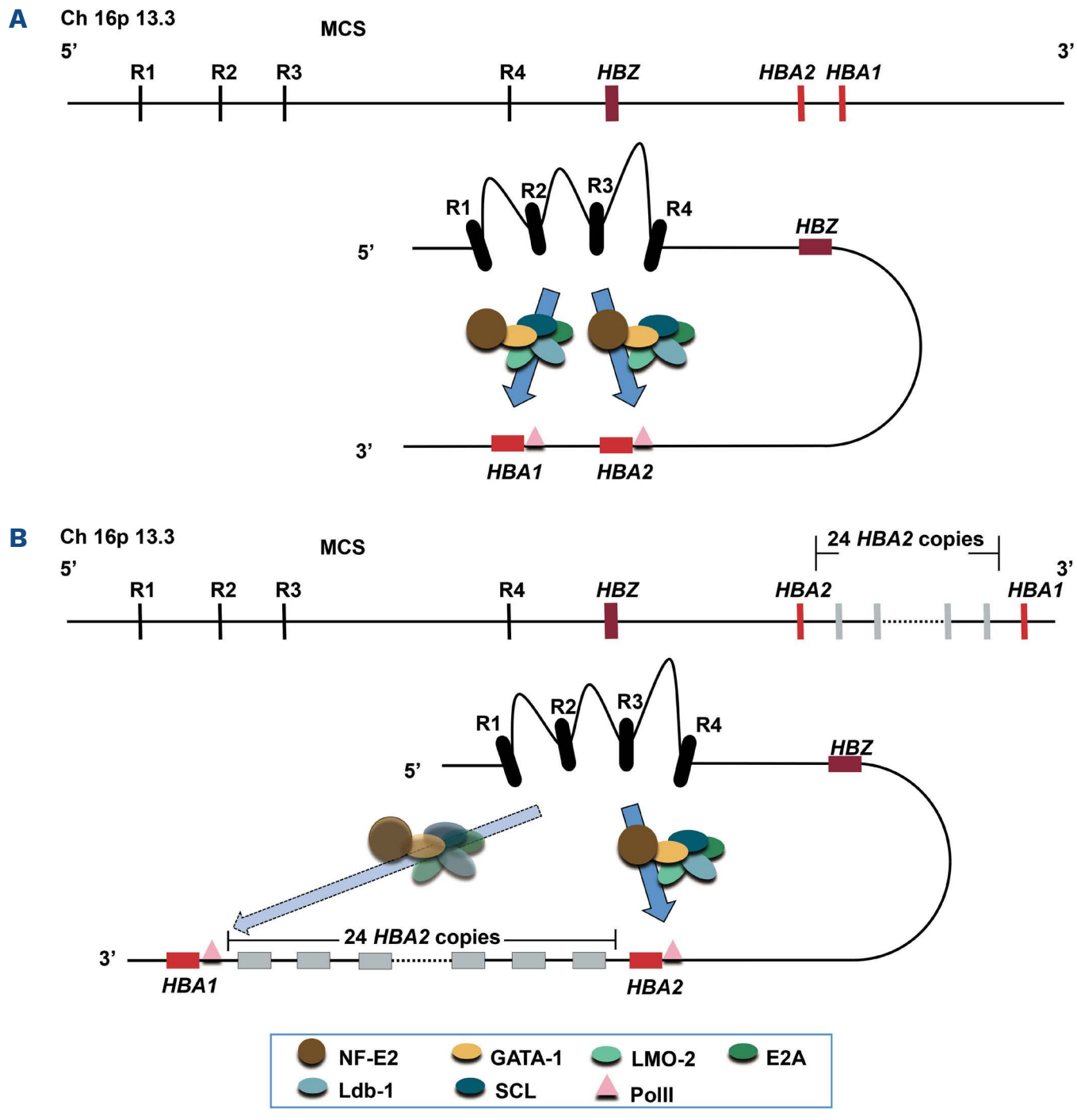


Figure 3. A hypothetical model to explain that the tandemly repeated copies of *HBA2* inhibit the normal α-globin expression. (A) The α-globin locus in human, including *HBZ*, *HBA2*, *HBA1* and 4 multispecies conserved sequences (MCS R1-4)) associated with erythroid-specific DNase1 hypersensitive sites (HS) (upper panel). Normally, HS contact α-globin gene promoters via chromatin looping, facilitated by the erythroid-specific transcription factors and proteins, activates high-level transcription of the *cis*-linked α-globin genes (bottom panel). (B) The inheritable complex repeats detected in this study, inserted 24 extra α-globin genes between *HBA2* and *HBA1* (upper panel). The repeats largely extend the distance between the HS and *HBA1* by approximately 100 kb, diminish HS/*HBA1* contacts, thus the *HBA1* expression is impaired.

ranging in size from kilobases to megabases.[5] The addition of this new molecular diagnostic technique might be beneficial to identify potentially pathogenic CNV/SV, similar to this study.

A large increase in the copy number of the α-globin gene was initially assumed to produce excess of α-globin chains, resulting in thalassemic phenotypes due to imbalanced globin chain synthesis, although this effect

was not observed in this pedigree. I-2 and II-1 had extra 24 heterozygous repeats of the α-globin genes, showing normal hematological manifestations. Analysis of red blood cells found no differences in the α/β-globin ratio of II-1, II-2, and normal controls (Figure 1H), suggesting no linear relationships of globin expression with the gene copy number and that the extra repeats of the α-globin gene were inactive. Tandem duplications often produce tandemly arrayed genes which are usually non-functional, similar to pseudogenes.[6] In this study, the SEA deletion was likely the cause of microcytic hypochromic anemia and low Hb A_2 of the proband, but her anemia burden was likely greater than that of the general SEA deletion carriers, as reflected by the much lower HGB, mean corpuscular volume and mean corpuscular hemoglobin levels. Notably, the α-globin mRNA and α/β-globin ratio of the proband were all lower than those of the SEA deletion carriers (Figure 1H, I), suggesting tandemly repeated copies of *HBA2*, in turn, inhibit the normal α-globin expression, which might be related to the long-range looping interactions of distant functional elements with the α-globin genes upon gene activation. Normally, functional elements, characterized by the presence of DNAse I hypersensitive sites (HS) located 53 to 68 kb upstream of *HBA2*, form loop structures with downstream globin genes, which are necessary for α-globin genes expression (Figure 3A).[7] In this regulatory loop, the α-globin genes expression level is determined by looping efficiency, which depends on the flexibility of chromatin fiber, the distance between HS and α-globin gene promoters. For the proband, insertion of non-functional tandem duplications before *HBA1* would largely extend the distance between the distant functional elements and *HBA1* by approximately 100 kb, which could have decreased *HBA1* transcription (Figure 3B).[8-10] There are normally four α-globin genes arranged as linked pairs (αα/αα), whereas one of the four genes is impaired, the overall α-globin production is practically unchanged, such as in I-2 and II-1. However, in the proband, both α-globin genes on the same chromosome were deleted, and *HBA1* expression on the other chromosome were decreased at the same time. This would be obvious functional change brought about by these two defects, and thereby lead to imbalanced α/β ratio and a milder TI. However, we did not detect the presence of HbH in proband's peripheral blood, which might due to its low levels below the detection limit of the capillary electrophoresis system.[11]

This case report highlights the important role of structural organization of the globin locus in gene activity, which would provide a new clue to investigate the distal enhancer and globin gene promoter communication during erythroid development. Moreover, it is important to consider the possibility of more intricate underlying mechanisms, such as methylation-induced silencing, which warrants further investigation, but it is hindered by the current limited sample availability.

Authors

Qianqian Zhang,[1,2*] Peng Lin,[1,2*] Aiping Mao,[3] Yongqiong Liu,[1] Xuan Shang,[1] Xiaofeng Wei,[1] Yuezhen Li,[3] Bin Lin[4] and Xiangmin Xu[1]

[1]Innovation Center for Diagnostics and Treatment of Thalassemia, Nanfang Hospital; Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong; [2]Dongguan Maternal and Child Health Care Hospital, Postdoctoral Innovation Practice Base of Southern Medical University, Prenatal Diagnosis Center, Dongguan Maternal and Child Health Care Hospital, Dongguan, Guangdong; [3]Department of TGS Research and Development, Berry Genomics Corporation, Beijing and [4]Guangzhou Jiexu Gene Technology Co., Ltd., Guangzhou, Guangdong, China

*QZ and PL contributed equally as first authors.

Correspondence:
X. XU - xixm@smu.edu.cn

https://doi.org/10.3324/haematol.2023.282902

Received: February 9, 2023.
Accepted: September 20, 2023.
Early view: September 28, 2023.

Disclosures

No conflicts of interest to disclose.

Contributions

QZ and YQ-L collected samples and guided diagnosis. PL, YQ-L and XW performed molecular experiments. AM and YZ-L performed optical mapping, PacBio sequencing, data analysis and interpretation. BL conducted the chromosomal microarray analysis. QZ and PL developed the figures. QZ, PL and AM wrote the paper. BL, XS and XX revised the manuscript. XW provided administrative support. XX supervised the study.

Funding

This work was supported by research funding from National Natural Science Foundation of China (to XX; grant U20A20353), National Key Research and Development Program of China (to XX; grants 2018YFA0507800 and 2018YFA0507803).

Data-sharing statement

Data are available upon reasonable request addressed to the corresponding author.

References

1. Vijian D, Wan Ab Rahman WS, et al. Molecular detection of alpha thalassemia: a review of prevalent techniques. Medeni Med J. 2021;36(3):257-269.
2. Chiba K, Shiraishi Y, Nagata Y, et al. Genomon ITDetector: a tool for somatic internal tandem duplication detection from cancer genome sequencing data. Bioinformatics. 2015;31(1):116-118.
3. Shang X, Xu XM. Update in the genetics of thalassemia: what clinicians need to know. Best Pract Res Clin Obstet Gynaecol. 2017;39:3-15.
4. Shang X, Peng ZY, Ye YH, et al. Rapid targeted next-generation sequencing platform for molecular screening and clinical genotyping in subjects with hemoglobinopathies. EBioMedicine. 2017;23:150-159.
5. Mantere T, Neveling K, Pebrel-Richard C, et al. Optical genome mapping enables constitutional chromosomal aberration detection. Am J Hum Genet. 2021;108(8):1409-1422.
6. Wagner A. The fate of duplicated genes: loss or new function? Bioessays. 1998;20(10):785-788.
7. Baù D, Sanyal A, Lajoie BR, et al. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. Nat Struct Mol Biol. 2011;18(1):107-114.
8. Shimotsuma M, Matsuzaki H, Tanabe O, et al. Linear distance from the locus control region determines epsilon-globin transcriptional activity. Mol Cell Biol. 2007;27(16):5664-5672.
9. Fraser P. Transcriptional control thrown for a loop. Curr Opin Genet Dev. 2006;16(5):490-495.
10. Voon HP, Vadolas J. Controlling alpha-globin: a review of alpha-globin expression and its impact on beta-thalassemia. Haematologica. 2008;93(12):1868-1876.
11. Greene DN, Pyle AL, Chang JS, et al. Comparison of sebia capillarys flex capillary electrophoresis with the BioRad variant II high pressure liquid chromatography in the evaluation of hemoglobinopathies. Clin Chim Acta. 2012;413(15-16):1232-1238.