



Ancestry of the major long-range regulatory site of the α -globin genes in the Portuguese population with the common 3.7 kb α -thalassemia deletion

Rita Pena¹ · Pedro Lopes¹ · Gisela Gaspar² · Armandina Miranda² · Paula Faustino^{1,3,4}

Received: 23 October 2023 / Accepted: 5 April 2024
© The Author(s) 2024

Abstract

Background The α -Major Regulatory Element (α -MRE), also known as HS-40, is located upstream of the α -globin gene cluster and has a crucial role in the long-range regulation of the α -globin gene expression. This enhancer is polymorphic and several haplotypes were identified in different populations, with haplotype D almost exclusively found in African populations. The purpose of this research was to identify the HS-40 haplotype associated with the 3.7 kb α -thalassemia deletion ($-\alpha 3.7\text{del}$) in the Portuguese population, and determine its ancestry and influence on patients' hematological phenotype.

Methods and results We selected 111 Portuguese individuals previously analyzed by Gap-PCR to detect the presence of the $-\alpha 3.7\text{del}$: 50 without the $-\alpha 3.7\text{del}$, 34 heterozygous and 27 homozygous for the $-\alpha 3.7\text{del}$. The HS-40 region was amplified by PCR followed by Sanger sequencing. Four HS-40 haplotypes were found (A to D). The distribution of HS-40 haplotypes and genotypes are significantly different between individuals with and without the $-\alpha 3.7\text{del}$, being haplotype D and genotype AD the most prevalent in patients with this deletion in homozygosity. Furthermore, multiple correspondence analysis revealed that individuals without the $-\alpha 3.7\text{del}$ are grouped with other European populations, while samples with the $-\alpha 3.7\text{del}$ are separated from these and found more closely related to the African population.

Conclusion This study revealed for the first time an association of the HS-40 haplotype D with the $-\alpha 3.7\text{del}$ in the Portuguese population, and its likely African ancestry. These results may have clinical importance as in vitro analysis of haplotype D showed a decrease in its enhancer activity on α -globin gene.

Keywords HS-40 haplotypes · α -Major Regulatory element · Alpha-thalassemia · 3.7 kb deletion

Introduction

Human hemoglobin is a globular tetrameric protein composed of two α -like and two β -like globin chains. These chains are encoded by two independent gene clusters located in different chromosomal loci: the α -globin gene cluster on chromosome 16 (16p13.3) and the β -globin gene cluster on chromosome 11 (11p15.5). The globin genes in each clusters are organized in the 5' to 3' direction, in the same order in which they will be expressed during the different stages of development: embryonic, fetal, and adult [1–3]. The α -globin gene cluster is composed by the embryonic ζ gene (*HBZ*); pseudogenes $\psi\zeta$ (*HBZps*) and $\psi\alpha 1$ (*HBA1ps*); two fetal/adult α genes (*HBA2* and *HBA1*); and pseudogene θ (*HBQ*) of unknown function [4, 5].

The high levels and correct expression of the α -globin genes depends on both local and remote *cis*-acting sequences,

✉ Paula Faustino
paula.faustino@insa.min-saude.pt

¹ Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Avenida Padre Cruz, Lisboa 1649-016, Portugal

² Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

³ Grupo Ecogenética e Saúde Humana, Faculdade de Medicina, Instituto de Saúde Ambiental, Universidade de Lisboa, Lisboa, Portugal

⁴ Laboratório Associado TERRA, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

such as the gene promoter sequences and the α -Upstream Regulatory Element (α -URE), respectively. The α -URE is composed by four highly conserved noncoding regulatory sequences called Multispecies Conserved Sequences (MCS-R1 to MCS-R4) [3, 6, 7]. The major sequence, MCS-R2, also known as HS-40 or α -MRE (α -Major Regulatory Element), is a 350 bp enhancer located 40 kb upstream of the ζ -globin gene Cap site [3, 6], and its main function is to activate and enhance the erythroid lineage-specific and development stage-specific expression of the α -globin genes in *cis* [3, 8–11].

The functional domain of this element is composed by several conserved nuclear binding sites, including two binding sites for the Nuclear Factor Erythroid 2 (NF-E2), three binding sites for GATA-1, and one CACC box, all of which are occupied in vivo in erythroid cells [6, 12]. These regions recruit general transcription factors, as well as the RNA Polymerase II, which binds to the promoter sequence of the α -globin genes [6]. The HS-40 sequence analysis performed by genomic footprinting has demonstrated the formation in vivo of specific nuclear factor DNA complexes at a subset of these sequence motifs in erythroid cells [12]. These transcription factor binding sites showed high conservation between human and other mammals, indicating their functional relevance [13, 14]. However, sequence heterogeneity within or in between these motifs of the human HS-40 fragment occurs between different human populations. Six polymorphic sites in human HS-40 sequence allowed to reconstruct six different combinations designed haplotypes, called A to F. Only A and B haplotypes are present in all groups analyzed. The other haplotypes are present in low frequencies and in specific populations [15, 16]. Haplotype D was primary described in African populations and is nearly absent in other populations [15, 16].

Alpha-thalassemia is an autosomal recessive disorder usually caused by the deletion of one or more α -globin gene that result in a deficiency or absence of α -globin chain synthesis. Alpha-thalassemia is characterized by a microcytic hypochromic anemia, and a clinical phenotype varying from almost asymptomatic to a lethal hemolytic anemia [17]. It is probably the most common monogenic gene disorder in the world and is especially frequent in Mediterranean countries, South-East Asia, Africa, the Middle East and in the Indian subcontinent [17]. Compound heterozygotes and some homozygotes have a moderate to severe form of α -thalassemia called HbH disease. Hb Bart's hydrops foetalis is a lethal form in which no α -globin chain is synthesized [17].

In the African and European populations, the most common form of α -thalassemia is due to the 3.7 kb deletion, which affects both α -globin genes (*HBA2* and *HBA1*), resulting in a single hybrid gene [17], and the same can be said

for the Portuguese population. A study conducted in 1996 using blood samples from 100 newborns showed that 7% of the individuals was heterozygous for the 3.7 kb deletion [18]. On the other hand, large deletions may occur removing all the globin distal regulatory elements as well as the complete α -globin gene cluster, giving rise to total absence of gene expression [17]. Other deletions were described removing only the distal regulatory elements, consequently the *in cis* α -globin genes are physically intact but functionally inactive [19–22]. Very rarely, the deletion that gives rise to α -thalassemia only affects one distal regulatory region, such as the HS-40, leaving the α -globin genes intact but partially inactivated [17, 23–27]. Some of these rare types of deletions that affect the regulatory elements, have also been found in Portuguese individuals, namely the $(\alpha\alpha)^{MM}$, $(\alpha\alpha)^{ALT}$, $(\alpha\alpha)^{TI}$ and $(\alpha\alpha)^{CSC}$ [22, 23, 28].

The HS-40 haplotypes can be used as markers for linkage analyses in addition to common molecular lesions, such as the common 3.7 kb α -thalassemia deletion. Therefore, the main purpose of this study was to characterize the haplotypes of the distal regulatory region HS-40 in individuals with and without α -thalassemia, and identify which haplotype is associated with the 3.7 kb α -thalassemia deletion in the Portuguese population, as well as determine the ancestry of this deletion in this population. Moreover, we intended to investigate if different HS-40 haplotypes are able to affect the hematological phenotype of α -thalassemia due to the homozygosity for the 3.7 kb deletion.

Materials and methods

Sample selection

We selected 111 anonymized DNA samples from Portuguese individuals who had already been investigated for the presence of the 3.7 kb α -thalassemia deletion by Gap-PCR as described elsewhere [29]. The criteria for sample selection was based on individuals' α -globin genotype: wild type, heterozygous for the 3.7 kb deletion, and homozygous for the 3.7 kb deletion. The hematological phenotype of each individual had previously been characterized by standard procedures and included the following hematological parameters: red blood cell count, hemoglobin (Hb) level, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration, hematocrit, and red cell distribution width. Of the selected individuals, 52 presented with normal hematological parameters, while 59 presented with microcytosis and/or hypochromia.

DNA extraction

Genomic DNAs were isolated from peripheral blood samples, collected in EDTA, using a nucleic acid automatic extractor, *MagNA pure LC 2.0* (Roche®, Germany). DNA quantity and quality were assessed using a *NanoDrop One* (Thermo Fisher Scientific, USA) spectrophotometer. DNAs were stored at 4°C.

HS-40 genotyping

To determine the sequence of the different HS-40 haplotypes, a DNA fragment of 400 bp containing the HS-40 region was amplified through conventional PCR, using primers described elsewhere [15]. The amplified PCR fragments were purified using *JET quick PCR Product Purification Spin Kit* (GENOMED) according to the manufacturer's instructions. Sanger sequencing was performed using the *ABI Prism BigDye® Terminator v1.1 Cycle Sequencing* commercial kit (Applied Biosystems) in an automated sequencer *3500 Genetic Analyzer* (Applied Biosystems). Sequences were analyzed using the *FinchTV v1.4.0* (Geospiza) software.

HS-40 haplotype reconstruction

Six single nucleotide polymorphic sites within the HS-40 fragment characterize the haplotypes A to F in humans. In order to identify the HS-40 haplotypes, the sequence variability observed in our 111 HS-40 fragments was compared to those described by Hartevelde and his collaborators in 2002 [15].

Statistical analysis

The distribution of HS-40 haplotypes and genotypes between the groups of samples was tested using the Test of Equal and Given Proportions. In order to determine the ancestry of the 3.7 kb deletion in the Portuguese population, a multiple correspondence analysis was performed and a specific function to draw the respective graphical representation was used.

For the comparison of the hematological parameters of the individuals with the HS-40 genotypes AA versus AD or

DD (using the dominant genetic test model), we started by testing the normality distribution using Shapiro-Wilk's test. When the normality of both populations was confirmed, the parametric T-test was used. The non-parametric test of Mann-Whitney was applied when there was a non-normal distribution.

All the statistical analysis were performed using *R* software and the statistical significance was established for a *p*-value lower than 0.05.

Results

Sample grouping according to the α -globin genotype

The selected 111 Portuguese non-related individuals were divided in three different groups according to their α -globin genotype: 50 without the 3.7 kb α -thalassemia deletion (genotype $\alpha\alpha/\alpha\alpha$; group 1), 34 with the 3.7 kb deletion in heterozygosity (genotype $-\alpha^{3.7}/\alpha\alpha$; group 2), and 27 with the 3.7 kb deletion in homozygosity (genotype $-\alpha^{3.7}/-\alpha^{3.7}$; group 3). All individuals of group 1 present normal levels of Hb, MCV, and MCH. Individuals of group 2 have a mean MCV of 82.7 ± 3.7 fL, MCH of 26.7 ± 2.1 pg, and Hb of 14.8 ± 1.2 g/dL for men and 12.8 ± 0.8 g/dL for women. When it comes to group 3, the individuals present with a mean MCV of 70.4 ± 4.2 fL, MCH of 21.9 ± 1.5 pg, and Hb of 13.9 ± 0.9 g/dL and 11.0 ± 1.0 g/dL for males and females, respectively. Therefore, in the latter group all individuals have hypochromia and microcytosis.

The corresponding DNAs were used to amplify the HS-40 region followed by Sanger sequencing analysis.

HS-40 genetic findings

The sequence of HS-40 regulatory region in our 111 samples revealed four distinct haplotypes labelled A, B, C, and D (Table 1). Haplotype A (CGCGGG) was the most common in all studied groups (Table 2), which was expected given that this is the ancestral sequence [6, 15]. In general, haplotype B (CACAGG) was the second most frequent; being that this was also the second most prevalent haplotype in individuals without α -thalassemia and in the carriers of

Table 1 Polymorphic substitutions in the HS-40 sequence and the corresponding haplotypes found in our samples

HS-40 haplotype	Position of polymorphic sites in the HS-40 sequence					
	+96	+130	+158	+199	+209	+212
A	C	G	C	G	G	G
B	C	A	C	A	G	G
C	C	A	C	A	A	G
D	C	G	T	G	G	G

Numbers indicate the position of polymorphic sites in the HS-40 sequence, as described [15]

Table 2 HS-40 haplotypes in the Portuguese population without the 3.7 kb α -thalassemia deletion ($\alpha\alpha/\alpha\alpha$), with the deletion in heterozygosity ($-\alpha^{3.7}/\alpha\alpha$) and in homozygosity ($-\alpha^{3.7}/-\alpha^{3.7}$)

Alpha-globin genotype	HS-40 haplotypes			
	A x (%)	B x (%)	C x (%)	D x (%)
$\alpha\alpha/\alpha\alpha$ (group 1; x = 100)	57 (57.0)	39 (39.0)	3 (3.0)	1 (1.0)
$-\alpha^{3.7}/\alpha\alpha$ (group 2; x = 68)	46 (67.7)	19 (27.9)	0 (0.0)	3 (4.4)
$-\alpha^{3.7}/-\alpha^{3.7}$ (group 3; x = 54)	31 (57.4)	8 (14.8)	0 (0.0)	15 (27.8)
Total (x = 222)	134 (60.4)	66 (29.7)	3 (1.3)	19 (8.6)

x = number of alleles

Table 3 HS-40 genotypes in the Portuguese population without the 3.7 kb α -thalassemia deletion ($\alpha\alpha/\alpha\alpha$), with the deletion in heterozygosity ($-\alpha^{3.7}/\alpha\alpha$) and in homozygosity ($-\alpha^{3.7}/-\alpha^{3.7}$)

Alpha-globin genotype	HS-40 genotypes						
	AA n (%)	AB n (%)	AD n (%)	BB n (%)	BC n (%)	BD n (%)	DD n (%)
$\alpha\alpha/\alpha\alpha$ (group 1; n = 50)	18 (36.0)	20 (40.0)	1 (2.0)	8 (16.0)	3 (6.0)	0 (0.0)	0 (0.0)
$-\alpha^{3.7}/\alpha\alpha$ (group 2; n = 34)	15 (44.1)	13 (38.2)	3 (8.8)	3 (8.8)	0 (0.0)	0 (0.0)	0 (0.0)
$-\alpha^{3.7}/-\alpha^{3.7}$ (group 3; n = 27)	7 (25.9)	7 (25.9)	10 (37.0)	0 (0.0)	0 (0.0)	1 (3.7)	2 (7.4)
Total (n = 111)	40 (36.0)	40 (36.0)	14 (12.6)	11 (10.0)	3 (2.7)	1 (0.9)	2 (1.8)

n = number of individuals

the 3.7 kb deletion. The very rare haplotype C (CACAAG) was only found in three individuals with the wild type α -globin genotype (group 1). On the other hand, haplotype D (CGTGGG) was found in nineteen alleles, 78.9% of them from individuals with the 3.7 kb deletion in homozygosity (group 3).

When it comes to the HS-40 genotypes, seven different combinations were found designated AA, AB, AD, BB, BC, BD, and DD (Table 3). We found 53 individuals (57.8%) homozygous for the HS-40 genotype. As they presented the expected hematological phenotype according to their α -globin genotype group, there was no evidence that any of them could be hemizygous rather than homozygous.

The AA and AB combinations were the most common in individuals without α -thalassemia and in those with the $-\alpha^{3.7}/\alpha\alpha$ genotype (group 1 and 2), while in patients with the $-\alpha^{3.7}/-\alpha^{3.7}$ genotype (group 3) the most prevalent combination was AD.

Haplotype D was found in three different genotypes: AD, BD, and DD. Genotype AD was the most prevalent, being found in 14 individuals, with 71.4% of them also having the 3.7 kb deletion in homozygosity (group 3). In addition, this group is the only one where we can find the very rare genotypes BD and DD.

The distribution of the diverse HS-40 haplotypes and genotypes is significantly different between individuals

without α -thalassemia and individuals with the 3.7 kb deletion in homozygosity (p -value < 0.001).

HS-40 genotype association study with α -thalassemia hematological parameters

In order to investigate if the HS-40 AA, AD, and DD genotypes are influencing the hematological phenotype of individuals with the 3.7 kb α -thalassemia deletion in homozygosity, a statistical comparison between their hematological parameters was performed using the dominant genetic test model (Table 4). However, no significant differences were found (p -value > 0.05) for any hematological parameters.

Ancestry of the 3.7 kb α -thalassemia deletion in the Portuguese population

After determining that the specific HS-40 haplotype D and genotypes AD, BD, and DD, are associated with the presence of the 3.7 kb α -thalassemia deletion in the Portuguese population, we aimed to investigate the ancestry of this deletion in this population. Initially, these genotypes were only reported in African people [15, 16], however more recently, they were also detected in Uruguayans [30]. In the two populations, these genotypes have been found mostly in individuals with the 3.7 kb deletion.

Table 4 Statistical comparison between the hematological parameters of Portuguese individuals with the HS-40 genotypes AA *versus* AD or DD, from the group with the 3.7 kb deletion in homozygosity

Hematological Parameters	Group 3 ($-\alpha^{3.7}/\alpha^{3.7}$)		p-value
	HS-40 Genotype		
	AA (n=7)	AD or DD (n=12)	
RBCs (x 10¹²/L)			
Female	5.28 ± 0.24	5.62 ± 0.61	0.514
Male	6.31 ± 0.07	4.97	NA
Hb (g/dL)			
Female	11.1 ± 0.6	10.6 ± 0.6	0.647
Male	14.1 ± 0.3	11.8	NA
HCT (%)	36.7 ± 2.3	36.8 ± 1.4	0.501
MCV (fL)	69.0 ± 2.4	72.1 ± 0.8	0.264
MCH (pg)	21.7 ± 0.9	22.2 ± 0.2	0.855
MCHC (g/dL)	31.4 ± 0.4	20.6 ± 0.3	0.139
RDW (%)	17.1 ± 1.1	14.9 ± 0.4	0.105

n=number of individuals; NA – Not Applicable (only one male individual with genotype AD or DD, not enough observations); RBC – Red Blood Cells; Hb – Hemoglobin; HCT – Hematocrit; MCV – Mean Corpuscular Volume; MCH – Mean Corpuscular Hemoglobin; MCHC – Mean Corpuscular Hemoglobin Concentration; RDW – Red Cell Distribution Width

Multiple correspondence analysis was performed in order to better visualize the similarities between the Portuguese population and other populations [15, 16, 30, 31]. This analysis showed that the Portuguese individuals who do not have α -thalassemia (PRT Normal) are grouped with other European populations, while samples with the 3.7 kb deletion (PRT $-\alpha^{3.7}/\alpha\alpha$ and PRT $-\alpha^{3.7}/-\alpha^{3.7}$) are isolated from these and found to be more closely related to the African population (Fig. 1).

Discussion

The Portuguese population is predominantly composed of haplotype A (60%) and haplotype B (30%), according to the 222 alleles analyzed for the HS-40 region sequence. Similarly, haplotype A was also reported as predominant in the Italian, Indonesian, Chinese, East Indian, Bantu-speaking-Africans, Brazilian Indians, and Uruguayan populations, with frequencies ranging from 56 to 87% [15, 16, 30], while haplotype B was found to be predominant exclusively in the Dutch population (57%) [15]. For the other populations indicated above, the haplotype B frequencies are lower and range between 13 and 43% [15, 16, 30]. On the other hand, haplotype D is characteristic of Bantu-speaking Africans (16%) and Pygmies from the Central African Republic (5%), being nearly absent in others populations [15]. Nonetheless, a high frequency of haplotype D was found in the Uruguayan population (6.4%) [30] and here in this study (8.6%). Furthermore, our results showed that the distribution of HS-40 haplotypes and genotypes are significantly different between individuals with and without the 3.7 kb α -thalassemia deletion and, consequently, that there is an association between the HS-40 haplotype D and the

presence of this deletion in the Portuguese population. For this conclusion, it certainly weighs a lot the presence of haplotype D, as well as the genotypes AD, BD, and DD, that were found mainly in individuals with the $-\alpha^{3.7}/-\alpha^{3.7}$ genotype. Thus, we hypothesize that the significant higher frequency of haplotype D in the sample with the $-\alpha^{3.7}$ deletion may be due to a predominant African origin of this deletion in the Portuguese population. The same was concluded for the Uruguayan population [30].

Haplotype D derived from haplotype A by a nucleotide substitution at position +158, which leads to a change in the consensus sequence for the AP-1/NF-E2 binding site, a composite binding site that is recognized by the transcription factor NF-E2 [32, 33]. Previous studies using murine erythroleukemia cells revealed that this transcription factor acts as an enhancer-binding protein for long-range regulation of globin gene expression and that, consequently, α -globin gene expression is highly dependent on NF-E2 [34–36]. Besides that, analysis of mice lacking NF-E2 showed that these mice exhibit some microcytosis, increase erythropoiesis, mild anemia, and their red cells present a slight decrease in hemoglobin content [34, 37]. Furthermore, other studies showed that mutated AP-1/NF-E2 binding sites lead to a 25% reduction in α -globin gene expression in transgenic mice [38], and in vitro experiments using constructs with the luciferase gene under the control of the different human HS-40 haplotypes revealed a noticeable reduction in luciferase expression in all haplotypes compared to A haplotype [39].

Consequently, interference in the NF-E2 binding site, as seen in haplotype D, may result in decreased α -globin gene expression in humans; even so, the presence of this HS-40 haplotype in heterozygosity is not enough to cause α -thalassemia. Moreover, the interference with this

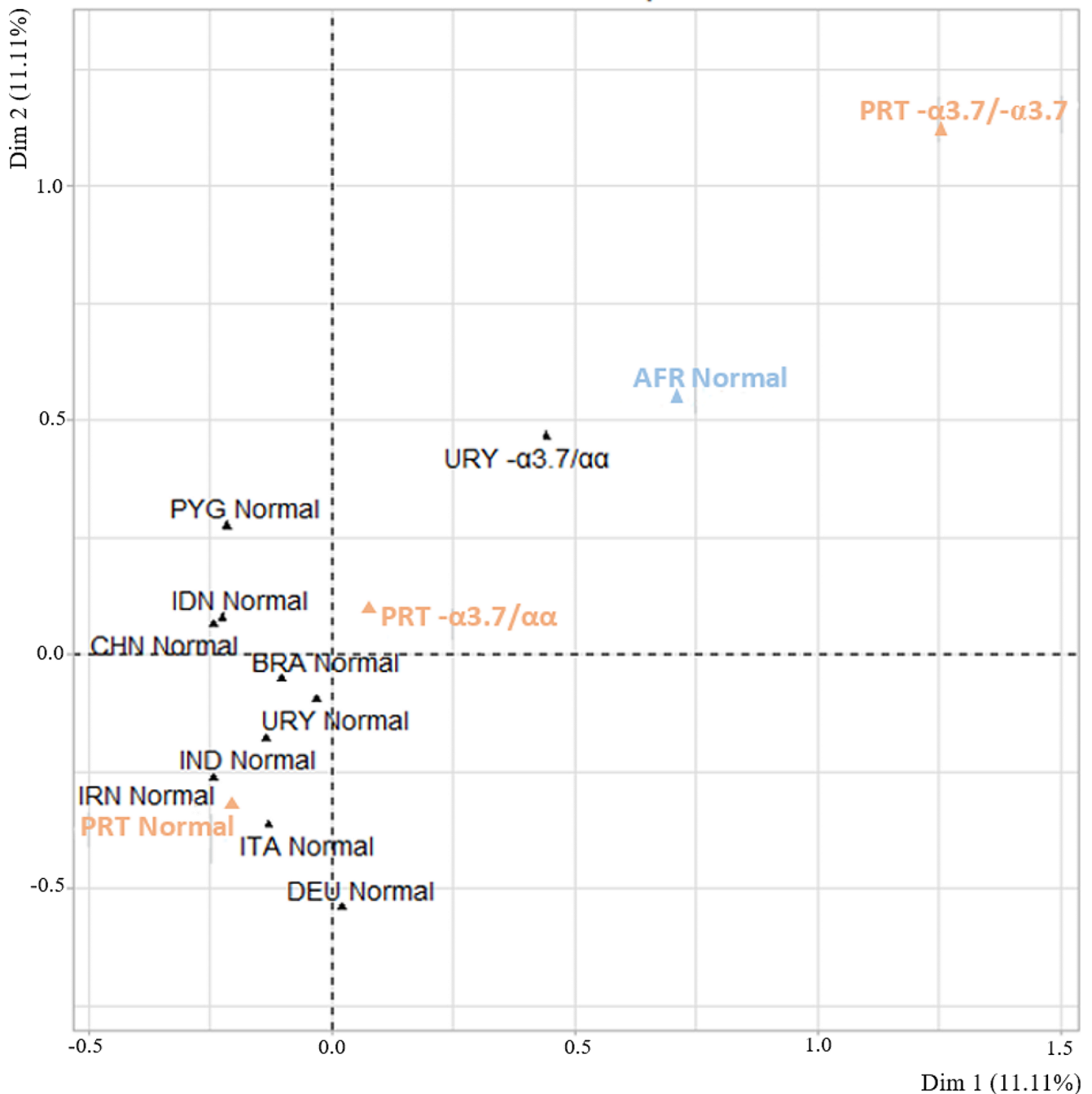


Fig. 1 Multiple correspondence analysis of the $-\alpha^{3.7}$ genotypes in multiple geographic populations. AFR: African; BRA: Brazilian; CHN: Chinese; DEU: Dutch; IDN: Indonesian; IND: Indian; IRN: Iranian; ITA: Italian; PRT: Portuguese; PYG: Pygmies; URY: Uruguayan. All

the genotypes from foreign populations were collected from [15, 16, 30, 31]. The Portuguese populations investigated in this study are marked as PRT normal, PRT $-\alpha^{3.7}/\alpha\alpha$, and PRT $-\alpha^{3.7}/-\alpha^{3.7}$

transcription factor binding site may have a greater impact in individuals that either have the genotype DD or that have a combination of haplotype D and an α -thalassemia defect, such as the 3.7 kb deletion. It was hypothesized that in individuals homozygotes for both the HS-40 haplotype D and the 3.7 kb deletion, α -globin gene expression may reduce below a critical level and result in the formation of

HbH (β_4 tetramers), due to an excess of unpaired β -globin chains [15]. However, our results did not reveal a significant difference between the hematological parameters of individuals with the HS-40 AA, AD, or DD genotypes, and with homozygosity for the 3.7 kb α -thalassemia deletion. Similar results were obtained by Hartevelde and collaborators [15]. These may be justified by many reasons, one of

them may be the sample size being too small to draw any conclusions, since patients homozygous for both HS-40 DD and $-\alpha^{3.7}/-\alpha^{3.7}$ are rare. Alternatively, the long-range regulation of α -globin gene expression in mice may differ from that in humans, as suggested by other studies [3, 23, 40], and is probably under a more complex mechanism, which may include epigenetic regulations.

Furthermore, in this study, a multiple correspondence analysis revealed that Portuguese individuals without α -thalassemia are grouped with other European populations, while samples with the 3.7 kb deletion are separated from these and more closely related to the African population, which reinforces the previous hypothesis and leads to the conclusion that there is a predominant African origin of the 3.7 kb α -thalassemia deletion in the Portuguese population.

Conclusion

In conclusion, this study revealed for the first time an association between the HS-40 haplotype D and the common 3.7 kb α -thalassemia deletion in the Portuguese population, and its likely African ancestry. This result contribute to the knowledge of the different genetic background between populations. Furthermore, this work highlights the importance of further studies to know better the consequences of genetic variability on the long-range regulation of α -globin genes in humans. The related experiments, carried out in vitro or in transgenic mice, revealed results that suggest clinical consequences, but these have not yet been validated in humans.

Acknowledgements We would like to thank all participants in the study. We also would like to thank the Technology and Innovation Unit of DGH/INSA for the Sanger sequencing analyses.

Author contributions Conceptualization: PF, and RP. Sample and data collection: PF, AM, GG. Laboratorial molecular experiments: RP, GG, PL. Interpretation of results: RP, PL, AM, GG, PF. The first draft of the manuscript was written by RP and PF, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by FCT|FCCN (b-on). The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. Open access funding provided by FCT|FCCN (b-on).

Declarations

Ethics approval and consent to participate Informed consent was obtained from all individual participants included in the study. The authors declare that the procedures were followed according to the regulations established by the Ethics Committee (Comissão de Ética para a Saúde do INSA, 2010DG720) and to the Declaration of Helsinki of the World Medical Association updated in 2013.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Forrester WC, Takegawa S, Papayannopoulou T et al (1987) Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res* 15:10159–10177. <https://doi.org/10.1093/nar/15.24.10159>
- Grosveld F, van Assendelft GB, Greaves DR, Kollias G (1987) Position-independent, high-level expression of the human β -globin gene in transgenic mice. *Cell* 51:975–985. [https://doi.org/10.1016/0092-8674\(87\)90584-8](https://doi.org/10.1016/0092-8674(87)90584-8)
- Higgs DR, Wood WG, Jarman AP et al (1990) A major positive regulatory region located far upstream of the human α -globin gene locus. *Genes Dev* 4:1588–1601. <https://doi.org/10.1101/gad.4.9.1588>
- Orkin SH (1978) The duplicated human α globin genes lie close together in cellular DNA. *Proc Natl Acad Sci U S A* 75:5950–5954. <https://doi.org/10.1073/pnas.75.12.5950>
- Lauer J, Shen CKJ, Maniatis T (1980) The chromosomal arrangement of human α -like globin genes: sequence homology and α -globin gene deletions. *Cell* 20:119–130. [https://doi.org/10.1016/0092-8674\(80\)90240-8](https://doi.org/10.1016/0092-8674(80)90240-8)
- Jarman AP, Wood WG, Sharpe JA et al (1991) Characterization of the major regulatory element upstream of the human alpha-globin gene cluster. *Mol Cell Biol* 11:4679–4689. <https://doi.org/10.1128/MCB.11.9.4679>
- Vyas P, Vickers MA, Simmons DL et al (1992) Cis-acting sequences regulating expression of the human α -globin cluster lie within constitutively open chromatin. *Cell* 69:781–793. [https://doi.org/10.1016/0092-8674\(92\)90290-S](https://doi.org/10.1016/0092-8674(92)90290-S)
- Sharpe JA, Chan-Thomas PS, Lida J et al (1992) Analysis of the human α -globin upstream regulatory element HS-40 in transgenic mice.pdf. *EMBO J* 11:4565–4572. <https://doi.org/10.1002/j.1460-2075.1992.tb05558.x>
- Bernet A, Sabatier S, Picketts D et al (1995) Targeted inactivation of the major positive regulatory element (HS-40) of the human alpha-globin gene locus. *Blood* 86:1202–1211. <https://doi.org/10.1182/blood.v86.3.1202.1202>
- Vernimmen D, Marques-Kranc F, Sharpe JA et al (2009) Chromosome looping at the human α -globin locus is mediated via the major upstream regulatory element (HS -40). *Blood* 114:4253–4260. <https://doi.org/10.1182/blood-2009-03-213439>
- Mettananda S, Gibbons RJ, Higgs DR (2015) α -globin as a molecular target in the treatment of β -thalassemia.pdf. *Blood* 125:3694–3701. <https://doi.org/10.1182/blood-2015-03-633594>

12. Strauss EC, Andrews NC, Higgs DR, Orkin SH (1992) In vivo footprinting of the human alpha-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. *Mol Cell Biol* 12:2135–2142. <https://doi.org/10.1128/mcb.12.5.2135-2142.1992>
13. Kielman MF, Smits R, Devi TS et al (1993) Homology of a 130-kb region enclosing the α -globin gene cluster, the α -locus controlling region, and two non-globin genes in human and mouse. *Mamm Genome* 4:314–323. <https://doi.org/10.1007/BF00357090>
14. Kielman MF, Smits R, Bernini LF (1994) Localization and characterization of the mouse α -Globin Locus Control Region. *Genomics* 21:431–433. <https://doi.org/10.1006/geno.1994.1289>
15. Harteveld CL, Muglia M, Passarino G et al (2002) Genetic polymorphism of the major regulatory element HS-40 upstream of the human α -globin gene cluster. *Br J Haematol* 119:848–854. <https://doi.org/10.1046/j.1365-2141.2002.03917.x>
16. Ribeiro DM, Figueiredo MS, Costa FF, Sonati MF (2003) Haplotypes of α -globin gene regulatory element in two Brazilian native populations. *Am J Phys Anthropol* 121:58–62. <https://doi.org/10.1002/ajpa.10193>
17. Harteveld CL, Higgs DR (2010) α -thalassaemia. *Orphanet J Rare Dis* 5:13. <https://doi.org/10.1186/1750-1172-5-13>
18. Peres MJ, Carreiro MH, Machado MC et al (1996) Rastreio neonatal de Hemoglobinopatias Numa população residente em Portugal. *Acta Med Port* 9:135–139. <https://doi.org/10.10400.17/633>
19. Wilkie AOM, Lamb J, Harris PC et al (1990) A truncated human chromosome 16 associated with α thalassaemia is stabilized by addition of telomeric repeat (TTAGGG)_n. *Nature* 346:868–871. <https://doi.org/10.1038/346868a0>
20. Hatton C, Wilkie A, Drysdale H et al (1990) Alpha-thalassaemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. *Blood* 76:221–227. <https://doi.org/10.1182/blood.v76.1.221.221>
21. Liebhaber SA, Griese EU, Weiss I et al (1990) Inactivation of human α -globin gene expression by a de novo deletion located upstream of the α -globin gene cluster. *Proc Natl Acad Sci U S A* 87:9431–9435. <https://doi.org/10.1073/pnas.87.23.9431>
22. Romao L, Osorio-Almeida L, Higgs DR et al (1991) α -thalassaemia resulting from deletion of regulatory sequences far upstream of the α -globin structural genes.pdf. *Blood* 78:1589–1595. <https://doi.org/10.1182/blood.V78.6.1589.1589>
23. Coelho A, Picanço I, Seuanes F et al (2010) Novel large deletions in the human α -globin gene cluster: clarifying the HS-40 long-range regulatory role in the native chromosome environment. *Blood Cells Mol Dis* 45:147–153. <https://doi.org/10.1016/j.bcmd.2010.05.010>
24. Sollaino MC, Paglietti ME, Loi D et al (2010) Homozygous deletion of the major alpha-globin regulatory element (MCS-R2) responsible for a severe case of hemoglobin H disease. *Blood* 116:2193–2194. <https://doi.org/10.1182/blood-2010-04-281345>
25. Wu M-Y, He Y, Yan J-M, Li D-Z (2016) A novel selective deletion of the major alpha-globin regulatory element MCS-R2 causing alpha-thalassaemia. *Br J Haematol* 176:984–999. <https://doi.org/10.1111/bjh.14005>
26. Luo S, Chen X, Zhong Q et al (2020) Analysis of rare thalassaemia caused by HS-40 regulatory site deletion. *Hematol (United Kingdom)* 25:286–291. <https://doi.org/10.1080/16078454.2020.1799587>
27. Capasso S, Cardiero G, Musollino G et al (2023) Functional analysis of three new alphathalassaemia deletions involving MCS-R2 reveals the presence of an additional enhancer element in the 5' boundary region. *PLoS Genet* 19:1–21. <https://doi.org/10.1371/journal.pgen.1010727>
28. Ferrão J, Silva M, Gonçalves L et al (2017) Widening the spectrum of deletions and molecular mechanisms underlying alpha-thalassaemia. *Ann Hematol* 96:1921–1929. <https://doi.org/10.1007/s00277-017-3090-y>
29. Dodé C, Krishnamoorthy R, Lamb J, Rochette J (1993) Rapid analysis of α 3.7 thalassaemia and α aaanti 3.7 triplication by enzymatic amplification analysis. *Br J Haematol* 83:105–111. <https://doi.org/10.1111/j.1365-2141.1993.tb04639.x>
30. Soler AM, Piellusch BF, Silveira L, da et al (2021) Alpha thalassaemia and alpha-MRE haplotypes in Uruguayan patients with microcytosis and hypochromia without anemia. *Genet Mol Biol* 44:1–6. <https://doi.org/10.1590/1678-4685-gmb-2020-0399>
31. Alimohammadi-Bidhendi S, Azadmehr S, Razipour M et al (2021) Regulatory Mutation Study in cases with Unsolved Hypochromic Microcytic Anemia and α -Major Regulatory Element Haplotype Analysis in Iran. *Hemoglobin* 45:37–40. <https://doi.org/10.1080/03630269.2021.1882482>
32. Andrews NC, Erdjument-Bromage H, Davidson MB et al (1993) Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* 362:722–728. <https://doi.org/10.1038/362722a0>
33. Mignotte V, Wall L, Deboer E et al (1989) Two tissue-specific factors bind the erythroid promoter of the human porphobilinogen deaminase gene. *Nucleic Acids Res* 17:37–54. <https://doi.org/10.1093/nar/17.1.37>
34. Shivdasani RA, Orkin SH (1995) Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc Natl Acad Sci* 92:8690–8694. <https://doi.org/10.1073/pnas.92.19.8690>
35. Lu SJ, Rowan S, Bani MR, Ben-David Y (1994) Retroviral integration within the Fli-2 locus results in inactivation of the erythroid transcription factor NF-E2 in friend erythroleukemias: evidence that NF-E2 is essential for globin expression. *Proc Natl Acad Sci U S A* 91:8398–8402. <https://doi.org/10.1073/pnas.91.18.8398>
36. Kotkow KJ, Orkin SH (1995) Dependence of globin gene expression in mouse erythroleukemia cells on the NF-E2 heterodimer. *Mol Cell Biol* 15:4640–4647. <https://doi.org/10.1128/mcb.15.8.4640>
37. Shivdasani RA, Rosenblatt MF, Zucker-Franklin D et al (1995) Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell* 81:695–704. [https://doi.org/10.1016/0092-8674\(95\)90531-6](https://doi.org/10.1016/0092-8674(95)90531-6)
38. Loyd MR, Okamoto Y, Randall MS, Ney PA (2003) Role of AP1/NFE2 binding sites in endogenous α -globin gene transcription. *Blood* 102:4223–4228. <https://doi.org/10.1182/blood-2003-02-0574>
39. Ribeiro DM, Zaccariotto TR, Santos MNN et al (2009) Influence of the polymorphisms of the α -major regulatory element HS-40 on in vitro gene expression. *Brazilian J Med Biol Res* 42:783–786. <https://doi.org/10.1590/S0100-879X2009005000014>
40. Anguita E, Sharpe JA, Sloane-Stanley JA et al (2002) Deletion of the mouse α -globin regulatory element (HS-26) has an unexpectedly mild phenotype. *Blood* 100:3450–3456. <https://doi.org/10.1182/blood-2002-05-1409>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.